



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Cognition in two languages: bilingualism and second language competency as determinants of cognitive performance
Author(s)	Arndt, Sophia
Publication Date	2019-07-15
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/15258

Downloaded 2024-05-02T21:05:41Z

Some rights reserved. For more information, please see the item record link above.





National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh

Cognition in two Languages:
Bilingualism and Second Language Competency as
Determinants of Cognitive Performance.

Thesis submitted for the Degree of Doctor of Philosophy

Sophia Arndt, BSc (Psychology)

Research Supervisor: Dr Mark A. Elliott
School of Psychology
College of Arts, Social Sciences, and Celtic Studies,
National University of Ireland,
Galway, Ireland

Submission Date: April 2019

Table of Contents

Declaration	v
Abstract	vii
Acknowledgements.....	ix
List of Abbreviations	xi
List of Figures	xiii
List of Tables	xv
0. Foreword: Language and Culture	1
1. Introduction.....	9
1.1. Defining bilingualism.....	9
1.1.1. Proficiency.....	10
1.1.2. Age of acquisition	10
1.2. Difficulties with research practices in bilingualism.....	11
1.2.1. Socioeconomic status (SES) and culture.....	11
1.2.2. The context of second language learning and usage.....	12
1.2.3. Language combination dependency	13
1.3. Challenges of defining bilingualism.....	13
1.4. Cognitive changes stemming from bilingualism	13
1.4.1. Bilingual performance in verbal tasks	14
1.4.2. Influence on executive functions.....	16
1.4.3. Questioning bilingual advantage on low-level cognitive processes	19
1.5. High-level cognition and bilingualism	22
1.5.1. Defining high-level cognition	22
1.5.2. Critical thinking and bilingualism.....	23
1.5.3. Problem solving and bilingualism.....	24
1.5.3.1. Monolingual and bilingual performance in problem solving	25
1.5.3.2. Problem solving performance across languages.....	25
1.5.4. Verbal arithmetic skills and bilingualism.....	26
1.5.5. High-level cognition in language skills	27
1.5.6. Bilingualism throughout the lifespan	28
1.6. Critical Thinking	30
1.6.1. Hypothesis Testing (HT)	31
1.6.2. Verbal Reasoning (VR).....	31
1.6.3. Judging Likelihood and Probability (JL).....	32
1.6.4. Argumentation Analysis (AA)	32
1.6.5. Problem Solving (PS)	32
1.6.6. Summary	33

2.	The Assessment of Critical Thinking	35
2.1.	Hypothesis Testing	38
2.2.	Verbal Reasoning	44
2.3.	Judging Likelihood and Probability	45
2.4.	Argumentation Analysis	49
2.5.	Problem Solving	50
3.	Comparison across Languages	55
3.1.	Methods.....	55
3.1.1.	General method.....	55
3.1.1.1.	Participants.....	55
3.1.1.2.	Design and Procedure	64
3.1.1.3.	Data Collection and Grouping.....	65
3.1.2.	English-Irish.....	68
3.1.2.1.	Participants.....	68
3.1.2.2.	Translation process	69
3.1.2.3.	Data collection	69
3.1.3.	German-English.....	70
3.1.3.1.	Participants.....	70
3.1.3.2.	Translation process	70
3.1.3.3.	Data collection	70
3.1.4.	Italian-English	71
3.1.4.1.	Participants.....	71
3.1.4.2.	Translation process	71
3.1.4.3.	Data collection	72
3.1.5.	Russian-English.....	72
3.1.5.1.	Participants.....	72
3.1.5.2.	Translation process	73
3.1.5.3.	Data collection	73
3.1.6.	Japanese-English	74
3.1.6.1.	Participants.....	74
3.1.6.2.	Translation process	74
3.1.6.3.	Data collection	75
3.1.7.	Chinese-English/ -Japanese	75
3.1.7.1.	Participants.....	75
3.1.7.2.	Translation process	75
3.1.7.3.	Data collection	76
3.1.8.	Chinese-Japanese.....	76
3.1.9.	Turkish-English /-German	76
3.1.9.1.	Participants.....	76

3.1.9.2.	Translation process	76
3.1.9.3.	Data collection	76
3.1.10.	Turkish-German	77
3.2.	Materials	77
3.2.1.	Socioeconomic Status	77
3.2.2.	Language Experience and Proficiency Questionnaire	77
3.2.3.	Mixed Measures	78
3.2.4.	Raven's Standard Progressive Matrices	79
3.3.	Research questions	80
3.4.	Design	80
3.5.	Statistical measures.....	81
3.6.	Results	83
3.6.1.	General results	83
3.6.1.1.	Comparison mono- versus bilinguals (between comparison)	90
3.6.1.2.	Comparison L1 versus L2 (within comparison)	92
3.6.2.	L1 English- L2 Irish.....	95
3.6.2.1.	Comparison mono- versus bilinguals (between comparison)	96
3.6.2.2.	Comparison L1 versus L2 (within comparison)	96
3.6.3.	L1 German- L2 English.....	97
3.6.3.1.	Comparison mono- versus bilinguals (between comparison)	97
3.6.3.2.	Comparison L1 versus L2 (within comparison)	98
3.6.4.	L1 Italian- L2 English	98
3.6.4.1.	Comparison mono- versus bilinguals (between comparison)	99
3.6.4.2.	Comparison L1 versus L2 (within comparison)	99
3.6.5.	L1 Russian- L2 English.....	100
3.6.5.1.	Comparison mono- versus bilinguals (between comparison)	100
3.6.5.2.	Comparison L1 versus L2 (within comparison)	100
3.6.6.	L1 Japanese- L2 English	101
3.6.6.1.	Comparison mono- versus bilinguals (between comparison)	101
3.6.6.2.	Comparison L1 versus L2 (within comparison)	103
3.6.7.	L1 Chinese- L2 English.....	103
3.6.7.1.	Comparison mono- versus bilinguals (between comparison)	104
3.6.7.2.	Comparison L1 versus L2 (within comparison)	104
3.6.8.	L1 Chinese- L2 Japanese.....	105
3.6.8.1.	Comparison mono- versus bilinguals (between comparison)	105
3.6.8.2.	Comparison L1 versus L2 (within comparison)	106
3.6.9.	L1 Turkish- L2 English.....	106
3.6.9.1.	Comparison mono- versus bilinguals (between comparison)	107
3.6.9.2.	Comparison L1 versus L2 (within comparison)	107

3.6.10.	L1 Turkish- L2 German	108
3.6.10.1.	Comparison mono- versus bilinguals (between comparison)	108
3.6.10.2.	Comparison L1 versus L2 (within comparison)	109
3.7.	Summary of the results	109
3.7.1.	L1 English- L2 Irish	111
3.7.2.	L1 German- L2 English	111
3.7.3.	L1 Italian- L2 English	112
3.7.4.	L1 Russian- L2 English	112
3.7.5.	L1 Japanese- L2 English	113
3.7.6.	L1 Chinese- L2 English	113
3.7.7.	L1 Chinese- L2 Japanese	114
3.7.8.	L1 Turkish- L2 English	114
3.7.9.	L1 Turkish- L2 German	115
4.	General Discussion	117
4.1.	Conclusion	131
4.2.	Implications	132
4.3.	Directions for future research	133
4.4.	Concluding remarks	134
	References	137
	Appendices	151
Appendix 1:	Website Used for Data Collection	151
Home Page		151
English/Irish Native Speakers		152
Appendix 2:	SES Background Questionnaire	153
Appendix 3:	Language Questionnaire (LEAP-Q)	154
Appendix 4:	Assessment <i>Mixed Measures</i>	157
English		157
Irish		169
German		181
Italian		193
Russian		203
Japanese		215
Chinese		227
Turkish		237
Appendix 5:	Example of Raven's Standard Progressive Matrices	247

Declaration

I, the candidate, assert that this thesis is my own work, and that I have not obtained a degree in this university or elsewhere on the basis of any of the thesis content.

Sophia Arndt

Abstract

Introduction. This project emphasises the importance of two aspects in research on bilingualism: (i) the inclusion of multiple language sets to reach generalisable results, which ensures that differences originate in changes due to bilingualism rather than the specific effect of certain language sets implemented, and (ii) the comparison of a bilingual individual's performance in first and second language for a deeper insight into the influence of bilingualism on cognition. This is best achieved through the implementation of verbally-based measures focussing on higher cognition, such as tasks assessing critical thinking.

Methods. Participants of nine different language sets were presented with the critical thinking assessment *Mixed Measures*, a compilation of published measures covering five subcategories of critical thinking: hypothesis testing, verbal reasoning, judging likelihood and probability, argumentation analysis, and problem solving. Socioeconomic status and language experience and proficiency were measured, as well as a number of Raven's Standard Progressive Matrices which were implemented to control for fluid intelligence. Outcomes of mono- and bilingual individuals were compared, as well as the performance of bilingual individuals in first and second language.

Results. Comparative calculations of monolingual and bilingual individuals were not found to be consistent across language groups. Differences were discovered within four of the nine language groups, with an advantage of bilingual participants in three of these four groups. Due to the lack of consistency throughout all groups, these findings were not conclusive. When comparing bilinguals' first and second language, two findings stood out: an advantage of solving verbal reasoning scenarios in the first language, whereas questions assessing judging likelihood and probability were solved better in the second language.

Conclusions. Higher exposure to the first language could explain advantages within verbal reasoning, whereas a certain emotional detachment might explain advantages when solving judging likelihood and probability measures in the second language.

Acknowledgements

I would like to sincerely thank and show my full gratitude and respect to my supervisor, Dr. Mark A. Elliott. Conducting a project of this scope would not have been possible without such relentless support. Thank you, Mark, for showing us, your supervisees, what great supervision involves, treating us as equal researchers, and respecting our opinions and input. Not only did you teach me how to approach each of the steps of a successful research project and how to develop self-confidence with it, you taught me how to be a researcher on an international level, cultivating collaborations with fellow researchers across the world.

Moreover, I would like to thank all the incredible individuals that were involved in the project on a translation-, pilot testing-, and data collection-level. I would like to extend a special thank you to Prof. Nakajima and the researchers in Fukuoka, especially Kaori Kojima and Shimeng Liu (Lucy), not only for making my time in Japan possible and as unforgettable as it was, but also for their patience with the translation of the material into Japanese and Chinese, and their effort with data collection. I would like to thank Dr. Yuko Yamashita for reimbursement of participants; Zimin Bao and Yuan Chen (Cindy), for their help from afar and the exceptional cooperation. My gratitude goes out to Peadar Mac Fhlannchadha and Cnag in Galway for sponsoring translation of the material into Irish. Thanks to Ethan Keeney for helping with data collection, and to Dr. Stanislava Antonijevic for advice on the project. Moreover, the scope of this study would not have been possible to reach without Sarah Volkmer, Chiara Ferrari, and Olga Bolbocean, who provided translations to German, Italian, and Russian. I cannot thank you enough for your time and effort! Moreover, I would like to share my appreciation for Dr. Arzu Özkan Ceylan in Hacettepe University, Ankara, Turkey, and Assist. Prof. Evrim Gülbetekin in Akdeniz University Antalya, Turkey, for both, translating the material as well as being of such valuable help collecting data. Focussing on the past few weeks I would like to express my gratitude to my diligent proof readers, Amanda, Diana, Jess, Kevin, Kristin, Naomi, and Sarah. Each of the individuals mentioned above, plus everybody else that provided me with important feedback on the project, helped to advertise my study, tolerated my endless posts on social media, and was still happy to share them, as well as all those hundreds of participants around the world, are those people, without whom this research project would not have been possible. I would like to thank the School of Psychology for providing me with space and equipment to conduct my research, technical support of all kinds, thanks to never-resting, tech-virtuoso Declan Coogan, and Olive O'Grady for helping to solve any administrative matters. Beyond that I'm thankful for professional as well as personal support at all stages provided by my fellow PhD-students, all those wonderful individuals who without a doubt added to this PhD being a great experience and such an influential time of my life. Last but not least, thanks to my family for always placing such trust and confidence in me; thanks to Maura and Aidan for providing me with a home away from home, to my number One (You, Kev!), for always putting my feet back on the ground whenever I got carried away and for providing me with the best puzzles, and to Ludovico Einaudi for composing the piece that turned into the soundtrack of my Ph.D., so beautiful I have no words, *Nuvole Bianche, Una Mattina*.

List of Abbreviations

Bi	=	Bilingual
Mono	=	Monolingual
L1	=	First Language
L2	=	Second Language
EF	=	Executive Functioning
CT	=	Critical Thinking
HT	=	Hypothesis Testing
VR	=	Verbal Reasoning
JL	=	Judging Likelihood and Probability
AA	=	Argumentation Analysis
PS	=	Problem Solving
AoA	=	Age of Acquisition
SES	=	Socioeconomic Status
LEAP	=	Language Experience and Proficiency Questionnaire
RSPM	=	Raven's Standard Progressive Matrices
MANOVA	=	Multivariate analysis of variance
reMANOVA	=	Repeated measures MANOVA

List of Figures

Figure 0.1 Overview of Language Families, Main Language Branches, Sub-Branches and Individual Languages included in this Project.....	4
Figure 3.1. Scatterplot representing correlation between L1total and Raven Scores... 84	84
Figure 3.2. Scatterplot representing correlation between L2total and Raven Scores... 84	84
Figure 3.3. Histogram showing normality of data for total scores of L1.	88
Figure 3.4. QQ-plot showing normality of data for total scores of L1.	88
Figure 3.5. Histogram showing normality of data for total scores of L2.	89
Figure 3.6. QQ-plot showing normality of data for total scores of L2.	89
Figure 4.1. Overview of Scores comparing Mono- and Bilinguals, divided into Sub-Categories.	119
Figure 4.2. Overview of Scores of all Language Groups comparing Mono- and Bilinguals, divided into Sub-Categories.	121
Figure 4.3. Overview of Scores comparing L1 and L2, divided into Sub-Categories.	124
Figure 4.4. Overview of Scores of all Language Groups comparing L1 and L2, divided into Sub-Categories.	127
Figure 4.5. Overview of Overall Scores of CT in L1, divided by Language Group.	129
Figure 4.6. Overview of Overall Scores of CT in L2, including all Bilinguals, divided by Language Group.....	129

List of Tables

Table 2.1. Overview of Sources for each Question implemented in Mixed Measures divided by Sub-category	36
Table 3.1. Number of Participants, Age, and Gender divided by Language Group and Mono- and Bilinguals	56
Table 3.2. Percentage of Participants having visited a State School and Years of Education divided by Language Group and Mono- and Bilinguals.....	57
Table 3.3. Percentage of Participants having obtained a University Degree; divided by Language Group and Mono- and Bilinguals	58
Table 3.4. Overview of Participants' Parents' Education; divided by Language Group and Mono- and Bilinguals	60
Table 3.5. Overview of Participants' Income; divided by Language Group and Mono- and Bilinguals	61
Table 3.6. Overview of Participants' First Language Exposure; divided by Language Group	67
Table 3.7. Overview of Participants' Second Language Exposure; divided by Language Group	68
Table 3.8. Cronbach's Alpha Values divided by Language Group and Mono- and Bilinguals	85
Table 3.9. L1: Influence of Age, Gender and Education on L1total divided by Language Group	86
Table 3.10. L2: Influence of Age, Gender and Education on L2total divided by Language Group	87
Table 3.11. Between-Group Comparisons: MANOVA comparing Monolinguals and Bilinguals divided by Language Group.....	90
Table 3.12. F-Values and Means comparing Monolinguals and Bilinguals divided by Sub-Category and Language Group.....	91
Table 3.13. Within Group Comparison: reMANOVA comparing L1 and L2 divided by Language Group.....	92
Table 3.14. F-Values and Means comparing L1 and L2 divided by Sub-Category and Language Group.....	93
Table 3.15. Overview of average Scores for each Subcategory, divided by Language Group and Mono- and Bilinguals.....	95
Table 3.16. Overview of average scores including all Japanese Participants divided by Gender for each Subcategory	102

0. Foreword: Language and Culture

Language and culture are two interwoven aspects defining human interaction (de Montes, Semin, & Valencia, 2003). Culture is a complex concept, reaching across a multitude of dimensions within an individual, such as thoughts, beliefs, and actions. Research mostly focusses on culture in terms of geographical circumstance and acculturation, which describes the slight adaptation of international families to new centres of life abroad, including a different culture (Yang, Yang, & Lust, 2011). It is impossible to separate language from culture and vice versa, yet it is not fully clear what effect and influence they have upon one another (Jiang, 2000), and it is important to keep in mind that bilingualism, the use of more than one language on a daily basis (Grosjean, 2010) does not equal biculturalism, which describes the participation in behaviours and traditions of two or more cultures (Grosjean & Byers-Heinlein, 2018). Language as a human's tool implants cognition into communication (Semin, 2000), and is hugely contextualised by culture. Connotation and meaning of words are dependent on where they are used and in which context they are implemented (Oyserman & Lee, 2008). For instance, words describing meals such as breakfast, lunch and dinner elicit completely different images within someone from the western world compared to someone of Asian origin (Jiang, 2000). Language and choice of words change depending upon the relationship to the receiver and are culturally dependent. Western cultures tend to lean more towards abstract language, for instance shown in the preferred use of adjectives when describing an individual, whereas Asian languages (for example, Japanese) predominantly include concrete language, using mostly verbs when describing an individual (Maass, Karasawa, Politi, & Suga, 2006). This does not only affect an individual's behaviour for instance in terms of individualism or collectivism, it also influences cognition and thought (Oyserman & Lee, 2008). Findings in this area of research should be treated with some caution, correlations of differences in international comparisons do not indicate causation and with culture being a multi-layered concept it is difficult to pinpoint the determining factor, which in turn influences language and cognition (Oyserman & Lee, 2008). Moreover, research studies focusing on only one nationality or country deliver inconclusive outcomes as to possible generalizations towards whole culture streams such as those emphasizing collectivism or individualism (Oyserman & Lee, 2008). Overall, findings in the research area are sparse but clearly indicate an interaction between language and culture.

The influence of culture on language and the interplay of both, language and culture, can be highlighted through the introduction of another field of research. This

can be achieved through focussing on language use and its implementation in connection to thought. Language is considered a part of cognition in the Western world, with the assumption of talking being connected to thinking (Kim, 2002) especially as mentioned in early research by Watson (1924). In this particular study, language was used as a premise of thought and any mental process (Kim, 2002). In comparison, within East Asian cultures, talking itself has been given a subordinate role and a connection between thought and language is not presumed. Instead, the importance of silence and introspection are emphasised to reach a thoughtful mental state. Culture appears to affect the subconscious use of language in connection to thinking and influences engagement with thought, originating in varying levels of reliance on language and thought (Kim, 2002). Exposure to a novel, second culture (e.g. through student exchange) is associated with a multitude of positive outcomes for thinking and cognition. These improvements are culture-specific but are also found on a general level (Lee, Therriault, & Linderholm, 2012). However, the effects of culture, language and possible second language learning in those circumstances are hugely interwoven and difficult to disentangle. Therefore, the origin of cognitive changes remains unclear and should be regarded as an interaction of all these aforementioned aspects (Yang, Yang, & Lust, 2011).

Language

Current estimates state that close to 7000 different languages are currently spoken worldwide. Each language differs in the number of speakers and varies in its reach. Historically, linguists are noted to have explored the origins of language. Part of this explorative process led to establishing the concept of language-families. The Concise Oxford Dictionary of Linguistics (Matthews, 2003) defines a language family as a group of languages, which can be traced back to the same language origin. Indo-European and Austronesian are examples of language families. Within each family, there are subfamilies, which are described as branches. In each branch, languages seem to be slightly more closely related than other languages in the family (Marian, 2018). Examples of language branches are the Germanic branch, the Celtic branch, and the Latin branch. It is important to mention that there is not one universal name or description for each branch, or even one generally accepted way of grouping those branches. Rather, these appear to be a variety of categorisations within the literature. Differences are small within bigger branches of languages but vary widely within the grouping of smaller languages, and their belonging to one or another language group. In the following, grouping criteria and family/branch names of the Concise Oxford Dictionary of Linguistics (Matthews, 2003) are employed.

Different crucial criteria to establish which family a language belongs to can be employed (Campbell & Poser, 2008). There are three main sources of importance: the vocabulary of a language, grammar, and sound correspondences (Campbell & Poser, 2008). This may seem self-explanatory but taking into account the research field of linguistics and trends of clustering languages up until the end of the twentieth century, these are critical grouping criteria.

It is important to further question the origins of similarities in languages when clustering them into families (Campbell & Poser, 2008). Basing groups only on superficial similarities is insufficient. Similarities between languages can be led back to a number of different explanations, and only some of them provide a reason for grouping languages together. Origins of similarities between languages might be due to coincidence. This can be traced back to chance, or accident, rather than a shared origin (Campbell & Poser, 2008). Moreover, borrowing words and phrases between languages is a common practice, especially with the growing interconnectivity through the internet and the growing influence of English at a global level. Languages which show similarities due to borrowing through language connection cannot always reliably be traced back to the same language origin. Another way of creating new words in a language is onomatopoeia (sound symbolism), which is a way of creating words dependent on the sound of what they describe. Different languages might use similar words for sounds or events if they perceive the according sound accompanying the event as similar. This does not describe shared origins and so sound symbolism should be excluded as a criterion when examining language origins. Lastly, universal rules or grammar should also be taken into account. These universal concepts can be traced back to nearly all languages and as such, have been found to describe the roots of language rules. Considering them when deciding on language families would distort the outcome. This leads to the only true reason any language should be considered part of a bigger group, a branch or family: Genetic (phylogenetic, genealogical) relationships which can be traced back to a common ancestor (Campbell & Poser, 2008). Genetic relationships describe inheritance of vocabulary, grammar or sounds. All other factors should be considered first and should be excluded, before drawing the conclusion of genetic relationships between languages (Campbell and Poser, 2008).

With this in mind, it is interesting to consider different reasons as to why languages change over time. Researchers in the area of linguistics have investigated this throughout the years and developed numerous theories for diversification of languages. Prominent examples include wars and migration, but also trade, geographical seclusion, and technological advantages, are noted to stimulate change in language. A

multitude of possibilities exist to support this phenomenon including adaptation to fellow trade partners or to linguistically signify a defined group identity.

The following aims to explore the origins of language, while shedding light on some language families in particular. Special attention will be paid to certain branches to investigate the roots of the languages employed in the subsequent research project, which will illustrate the breadth and depth of this project. Figure 0.1 below depicts an overview of language families and branches that are further explored.

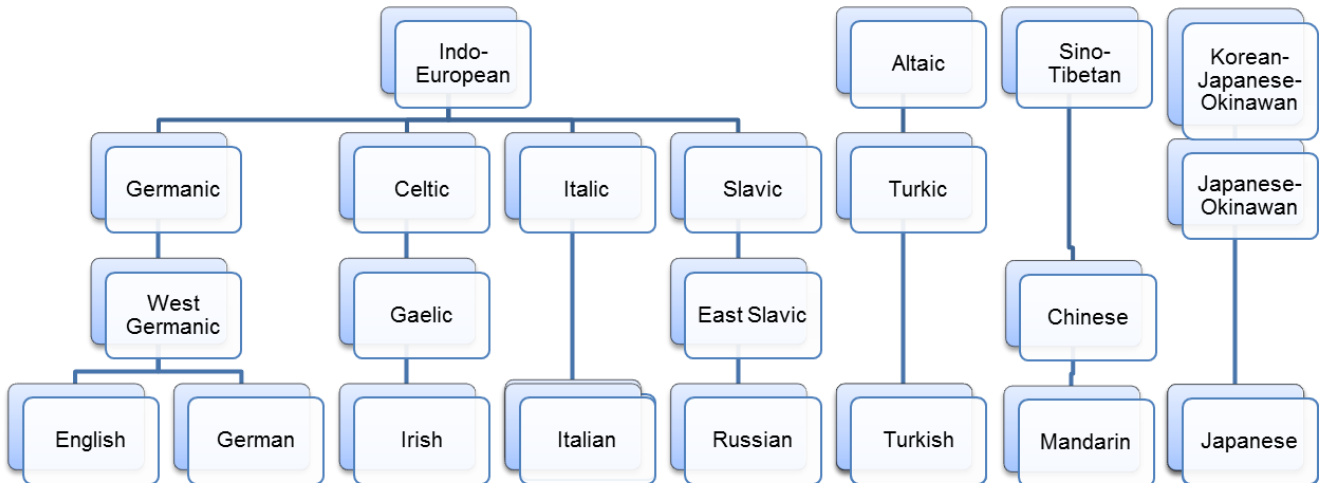


Figure 0.1 Overview of Language Families (first row), Main Language Branches (2nd row), Sub-Branched (3rd row) and Individual Languages (4th row) included in this Project.

Indo-European is a language family most prevalent and spoken in Europe, which describes the western limit of its outreach (Matthews, 2014). It reaches as far as the Indian continent, where its languages are represented, together with languages of other language families. According to the Concise Oxford Dictionary of Linguistics (Matthews, 2003), the Indo-European language family splits into eleven main branches including some extinct ones (Anatolian, Tocharian). Greek, Indo-Iranian, Armenian, Baltic, and Albanian are examples of members of the Indo-European language family which will not be evaluated further.

Germanic as a branch of the Indo-European language family includes different sub-branches, one of which is referred to as West Germanic. West Germanic languages are English, German, and Dutch, whereas Finnish, Swedish, Norwegian, among others, can be labelled as North Germanic, also referred to as Scandinavian languages; with East Germanic languages, also known as Gothic languages.

English, as it is spoken today, is thought to have descended from Anglo-Saxon, Old English, which evolved into Middle English before it was heavily influenced by French, especially in terms of vocabulary (Matthews, 2014). The standard form of English spoken today began to develop at the end of the Middle Ages. Through colonization, dominion or economic influence by the British Empire as well as due to the post-war economic power of the United States of America, English is currently prevalent all over the world (Matthews, 2014). It is spoken as a first language in North America, the British Isles, Australia, New Zealand, and South Africa among others. English is also an official second national language in multiple countries including India and Nigeria (Matthews, 2014). This gives English the status of an international language, especially in its form of American English. Due to the wide reach of English, many regional varieties were formed, e.g. Indian-English, Singapore English. Core rules of grammar and vocabulary are the same for these languages, but are adjusted to and by the use of groups of people as mostly second-language speakers.

German as a language is mainly spoken in Germany, Austria and parts of Switzerland, Italy, and Luxembourg. It can be divided into two main dialects, low German, spoken in the North of Germany, and high German, more prevalent in the South. Both descend from the original form called Old High German (Matthews, 2014).

The Celtic branch of Indo-European can be divided into two parts, of which one, Continental Celtic, is now extinct. Continental Celtic was formerly spoken in the north of Italy (Lepontic), France (Gaulish) and Spain (Celtiberian) before it was eliminated by Latin through the growing influence of the Roman Empire. Insular Celtic can be split into Brittonic which includes Welsh, Breton, Cornish, and Gaelic. One of the two surrogates of Gaelic is Irish Gaelic, which describes one of the native languages in the Republic of Ireland (apart from spoken English). Nowadays it is more prevalent in terms of teaching and speaking in the western part of the country. Scottish Gaelic was originally carried from Ireland to Scotland by emigrants in the Dark Ages and is now spoken in parts of north-west Scotland (Matthews, 2014).

The Italic language branch represents the Romance languages in the current study. Dating back to Latin, Italic languages were originally spoken in continental Europe, with languages like French, Italian, Portuguese, Spanish, and Romanian among others, including dialects like Catalan and Sardinian. Historical developments such as colonialization resulted in the Italic branch to spread to the American continent. Focussing on one of the members of this language branch, Italian has its origins in Latin, specifically from a Tuscan dialect of Florence, and can be traced back as far as the middle ages. Nowadays many dialects of the Italian language can be found, mostly within the borders of Italy, which date back to regional varieties (Matthews, 2014).

A fourth language branch within the Indo-European language family represented in this project is Slavic. Slavic spreads from Central Europe towards the east and divides into three branches, East Slavic, West Slavic, and South Slavic. Russian and Ukrainian can be clustered in the East Slavic sub-branch, Polish, Czech, Slovak and Sorbian make up West Slavic, and Serbian (and derivative languages, e.g. Croatian and Macedonian), Bulgarian and Slovenian are representatives of South Slavic. Within the Indo-European language branch, there seems to be a closer similarity of the Slavic language branch with the Baltic branch compared to others, which is why these two branches are sometimes aggregated and referred to as the Balto-Slavic language branch (Matthews, 2014).

Russian is the most widespread language of Eurasia and the largest native language in Europe with roughly 144 million native speakers. Russian is noted to have diverged from Ukrainian in late 11th century and is currently written in Cyrillic. Cyrillic is implemented by other Slavic languages also, mainly going back to the dominance of the Orthodox Church and is based on modern forms of the Greek alphabet (Matthews, 2014).

A second language family represented within this research project is Altaic. Altaic describes the family of languages which includes the language branch Turkic. Geographically, Turkic reaches from the Balkans across Central Asia into Siberia. Languages categorised under Turkic are for instance Uzbek, Azerbaijani, Turkmen, and Turkish, with the latter accepted as the largest representative. Turkish is the official language of Turkey, and spoken by minorities in the Balkans and in Northern Cyprus. More recently, immigration has led to a spread of the Turkish language in South Eastern and more recently Western Europe. Turkish was originally written in the Arabic alphabet but since the early 20th-century, it is now communicated using adjusted Roman letters. Strong influences of Arabic and Persian can be found, dating back to the Middle Ages.

Japanese is an example of a language, which has barely any association with any other language or language family with the exception of a few dialects closely related to it. It is sometimes grouped under the language family Korean-Japanese-Okinawan, within the language branch Japanese-Okinawan, even though this connection is very weak (Lyovin, 1997). It is assumed to have expanded throughout Japan from the southern island Kyushu, where it was brought through agriculture from the Korean Peninsula (Bellwood & Renfrew; 2002). Early recordings from the 8th century show the Japanese language written using Chinese characters (Kanji), which is still the case nowadays, combined with two forms of Japanese characters: Katakana and Hiragana.

The last example of a language family is Sino-Tibetan. This family consists of two distinctly differentiable branches, Chinese and Tibeto-Burman. Chinese can be split into certain sub-languages, or dialects, the most prevalent of which is Mandarin. Mandarin describes the official language in China, Taiwan, and Singapore beyond others. Other dialects include Hokkien (Southern Min) and Cantonese (Yue), which are the most widely spoken dialects after Mandarin due to immigration. All Chinese dialects can be traced back to “old Chinese” which goes back as far as the fifth-century BCE and was spoken until roughly the third-century AD, to then develop into “Classical Chinese”. The writing system is estimated to originate around the second half of the second millennium BCE.

Languages implemented within this research project reach across the globe and represent a wide mixture of cultures and communities. This is a unique approach to explore effects of bilingualism on cognition while using the recommended methods to discover reliable differences between monolinguals and bilinguals. Finally, the research also seeks to explore possible effects within the first and second language of a bilingual individual.

1. Introduction

Researchers estimate that more than 50% of today's world population is bilingual. The numbers vary from continent to continent and from country to country, with some countries' rates as high as 99% (e.g. Luxembourg).

These figures explain how the concept of bilingualism and its possible influence on humans caught the attention of researchers. Research began in the early twentieth century with only a few scientists (Saer, 1923; Sanchez, 1934), and massively expanded within the past thirty or forty years (Bialystok, 2007; Gathercole, 2014; Grosjean, 2001; Marian, 2018; Paap, 2016). Opinions about bilingualism and its effect on the brain shifted noticeably throughout the century. Early researchers considered it as a clear disadvantage, with the bilingual speaker being "[...] much more confused in this respect than monoglots [...]" (Saer, 1923, p.25), and that "[...] there can be no doubt that the child reared in a bilingual environment is handicapped in his language growth." (Thompson, 1962, p.385; as cited by Hakuta, 1986). Today views and findings are much more positive. Even though opinions still vary, the main perception of speaking two or more languages on a daily basis is perceived as an advantage in most regards.

1.1. Defining bilingualism

Bilingualism as a lay definition solely focusses on speakers with two mother tongues who master both languages perfectly. This depicts bilingualism as the "native-like control of two languages" (Bloomfield, 1933, p.56), implying a high degree of mastery of the individual's second language. Rather than this commonly held idea which refers to competency in the second language, bilingualism is also definable in terms of second-language (L2) usage, labelling individuals who use more than one language on a daily basis as bilingual (Grosjean, 2010). This definition depicts bilingualism as the capability of communicating through two or more languages without an explicitly defined frequency of use of the second language.

When defining bilingualism, three aspects of importance are (i) Proficiency of/ fluency in the second language, (ii) age of acquisition of said language and (iii) exposure to both languages on an everyday life basis (Grosjean & Byers-Heinlein, 2018). The ratio and interaction of those three factors (i-iii) underlie individual variations and varies within research studies, determined by the respective researcher.

1.1.1. Proficiency

An important factor in bilingualism is proficiency in an individual's L2. L2 proficiency provides a crucial factor for differentiating monolingual from bilingual speakers. It can be measured with language tests or through self-rating scales.

Proficiency tests give a broad picture of a participant's language skills. They aim to standardize measurement and objectify procedures. Tests have been developed in most major languages with a focus on various topics adjusted to the target group, e.g. in a playful way for children or a more serious manner when targeted at adults. However, it is questionable how comparable the results of tests across languages are. Moreover, assessments are very time intensive and expensive to implement.

Self-rating scales are more economical, but they have the disadvantage of subjectivity. Furthermore, researchers rely on participants' capability to estimate their own language competency. Proficiency tests can easily be adjusted and translated into different languages, but outcomes might differ depending on cultural and personal backgrounds, as well as variability in traits such as self-confidence.

To date, there is no ideal measure to identify language proficiency. Language appears to be a fluent construct on a continuum rather than categories. Despite the clear difficulties in accurately measuring language proficiency, a measure of proficiency is essential for research on bilingualism. Supporting the validity of both proficiency tests and self-rating scales, research shows a clear correlation between the two measures when matching participants' proficiency levels (Bialystok, 2001; Marian, Blumenfeld, & Kaushanskaya, 2007), which implies a certain interchange-ability and reliability.

1.1.2. Age of acquisition

Another important factor in defining bilingualism is the age of L2 acquisition. A common way of differentiating between participants' age of acquisition is the categorization early versus late bilinguals, where the cut-off point is usually around the age of ten years. A different way of categorizing speakers is comparing simultaneous bilingualism versus sequential bilingualism. Simultaneous bilingual speakers are exposed to both languages from birth on, whereas sequential bilinguals acquire one language first and the other language later.

The concept of the age of acquisition of L2 is difficult for a few reasons: It has to be decided which age is taken as the comparison, the age a person starts learning a new language, or once fluency in the language is achieved. Fluency is subjective and difficult to determine, as discussed above. Nonetheless, the learning curve and

intensity of language study of two individuals who start acquiring a language at the same time can be crucially different and this influences L2 competency.

One of the reasons why the age of acquisition is considered essential when assessing the influence of bilingualism on an individual is a theory hypothesizing a critical period in language learning, implying that the process of language learning would be facilitated if the learner was of, or under a certain age. Language learning after this critical period, in adolescence and adulthood, is considered more effortful. There is a vast amount of support for this theory, but more recently criticism has been added: Bialystok (2001) summarizes that more efficient language learning during childhood can be traced back to a range of different reasons. Children are assigned more time to learn a language. This time intensity of learning a new language is rarely compatible with the hectic life of an adult to the same extent as a child. Teaching and learning rarely happen with the same level of formal instruction and sympathetic methods and demands of adult learners are increased rapidly with content that is more difficult. These points do not suggest the existence or non-existence of a critical period, but they are crucial aspects to keep in mind when researching the area, possibly confounding, and definitely correlated with, findings on the concept of a critical period (Bialystok, 2001).

To summarise, L2 proficiency and age of acquisition are essential aspects when working with bilingual individuals and assessing possible effects of bilingualism on L2 speakers. To date, there is no generally agreed way of defining and grouping bilinguals in research, which leads to a wide range of variety and makes it difficult to compare outcomes and studies of researchers.

1.2. Difficulties with research practices in bilingualism

1.2.1. Socioeconomic status (SES) and culture

Bilingualism does not originate from a deliberate decision in the majority of individuals (Bialystok, Craik, Green, & Gollan, 2009). This suggests that bilingualism is not the outcome of increased cognitive functions if those are present, but a possible source of changes within a bilingual's brain. Environment and circumstances in which one person grows up in determine the likelihood of growing up bilingual. Not only the reasons for learning a second language are mostly predestined, but this also applies to SES, which influences language development and skills as well as cognition and must thus be considered when assessing bilingualism (Gathercole, 2014). Bilinguals are part of every societal group, and SES varies between below average, for example within refugees, up to above average, for instance in countries where learning and regularly

speaking a second language is considered elite. This leads to the question of a possible influence of culture (Tran, Arredondo, & Yoshida, 2018) and cultural diversity on cognition. Balanced bilinguals often grow up in immigrated families or live in a country other than their country of origin. This changes their cultural environment compared to their monolingual peers and possibly adds the influence of a second culture. Growing up with more than one culture affects not only language development and the status of the languages as perceived by the bilingual individual but also the development of executive functions, mostly influencing verbal response inhibition (Tran, Arredondo, & Yoshida, 2018). While SES is commonly controlled for in most published studies, culture is a factor that is often overlooked in research. One way to control for the effects of culture on cognitive changes within studies assessing bilingualism is through the inclusion of multiple language sets covering diverse cultures implementing the same measurements, followed by a comparison of the outcomes. If outcomes prove to be stable and consistent throughout all languages, differences unrelated to culture or specific language sets are implied. This does not only apply to cultural backgrounds and their effect on language speaking, but also to specific language sets and their effect on the brain. It is questionable if data, which is collected covering only two or three languages - as commonly done in projects investigating bilingualism -, is generalizable. Outcomes might be leading back to differences between certain languages rather than differences due to cognitive changes caused by bilingualism. At present, research mainly focusses on the combination of two languages, or one native language combined with various second languages, rather than implementing big samples covering several language sets. This is often due to the feasibility of the research procedures and accessibility of bilingual participants. Bialystok and colleagues (Bialystok, Craik, Klein, & Viswanathan, 2004) assessed this matter briefly and came to the conclusion that at least concerning non-linguistic tasks (like the Simon task implementing coloured squares), the bilingual advantage does not depend on certain combinations of languages which are spoken by the bilingual. Beyond non-linguistic tasks, there is a lack of research and cross-language findings.

1.2.2. The context of second language learning and usage

SES and culture are linked directly to another set of inevitable questions, which refers to the context of language learning and language use. In which context were both languages learned and in which context are they used? Are both languages used equally or is one language used more? Are certain contexts connected only to one specific language? A bilingual person might be fully fluent in one language concerning one aspect of their life, for instance, work, but barely know any of the vocabularies in

this language for another aspect of their life, in which they merely use the other language.

1.2.3. Language combination dependency

Research studies focus mostly on the combination of two languages, or one native language mixed with different second languages, rather than big samples covering several languages. As mentioned, this is done due to the feasibility of the research procedures and accessibility of bilingual participants. It is questionable if data, which is collected on a restricted amount of language sets, is generalizable. Effects might be leading back to differences of certain languages rather than cognitive differences due to changes caused by bilingualism: Bialystok and colleagues (Bialystok, Craik, & Viswanathan, 2004) assessed this matter briefly and came to the conclusion that at least concerning non-linguistic tasks, the bilingual advantage does not depend on certain combinations of languages which are spoken by the bilingual. There is a lack of research focussing on the broad variety of tasks implemented and possible cross-language effects, which might affect current findings in a restrictive matter.

1.3. Challenges of defining bilingualism

In summary, it is obvious that this current uni-dimensional practice in the research area of bilingualism has to be overcome, and a wider focus must be used, one which looks beyond only proficiency or age of acquisition (Hamers & Blanc, 2000). As stated above and by Ardila (1998), variability in research subjects in the field of second language learning is huge and no bilingual speaker of any set of languages is the same. There are numerous factors influencing the proficiency and understanding of bilinguals (Bialystok et al., 2009; Hoff, 2013). Points like the age and sequence of acquisition, proficiency of spoken and written language, preferences towards one language, also dependent on cultural identification, frequency and context of use and the method and context of acquisition are inevitable to consider (Ardila, 1998). All these aspects are crucial to define the degree of bilingualism in an individual (Hoff, 2013).

1.4. Cognitive changes stemming from bilingualism

Speaking an L2 is sometimes compared to a form of brain training and to activities like juggling, or playing video games (Bialystok, Craik, & Luk, 2012). This reference indicates two perspectives on bilingual cognition: Not only is bilingualism identified with brain activity, but the skills accompanying the use of two languages are associated with those low-level cognitive mechanisms typically targeted by brain training (Bialystok, 2007). Supporting this idea is a set of studies, which have been focused towards investigating differences in performance of bilinguals relative to monolinguals. These

contain tasks, that specifically index the performance of lower-level cognitive processes, for instance non-verbal executive functioning tasks (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2009; Bialystok et al., 2004; Bialystok, Craik, & Luk, 2012; Carlson & Meltzoff, 2008; Gold, Kim, Johnson, Kryscio, & Smith, 2013; Hilchey & Klein, 2011). Considering their superior performance in basic executive functions, bilinguals have been argued to benefit from what is referred to as a “bilingual advantage”. Evidence for the bilingual advantage will be explored in more detail in 1.4.2. First, a slightly negative effect of bilingualism on human cognition will be introduced, the effect of bilingualism on verbal tasks.

1.4.1. Bilingual performance in verbal tasks

It has been found that when tested on verbal tasks, bilinguals show a slight disadvantage relative to monolinguals (Bialystok, Craik, & Luk, 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Kaushanskaya & Marian, 2007; Lehtonen, Soveri, Laine, Järvenpää, de Bruin, & Antfolk, 2018; Roberts, Garcia, Desrochers, & Hernandez, 2002). This small negative influence of bilingualism on a person’s verbal skills, which applies to both the first language (L1) and L2 (Bialystok, Craik, & Luk, 2012; Ivanova & Costa, 2008) has been examined using multiple tasks. Tasks including word naming, where bilinguals show to be slightly slower at naming random words of certain categories: For instance, words with specific first letters in a given time (Roberts et al., 2002). Moreover, lexical retrieval tasks, where bilingual individuals take longer to retrieve names for images they are presented with (Bialystok, Craik, & Luk, 2008). Also included were verbal fluency tasks in which bilinguals delivered a less fluent (Gollan et al., 2005), and less accurate (Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007) description of scenes or scenarios on pictures compared to monolinguals, even when responding in their dominant language (Ivanova & Costa, 2008). Finally, bilinguals appear to possess a smaller receptive vocabulary size than monolinguals (Bialystok & Luk, 2012; Bialystok, Luk, Peets, & Yang, 2010; Anderson, Mak, Chahi, & Bialystok, 2018; Trenkic & Warmington, 2018).

These results have led to the following explanation: When speaking language A, both languages A and B become simultaneously activated (Bialystok, 2007; Green, 1998; Kroll & Bialystok, 2013; Preston & Lambert, 1969) introducing a requirement for executive control in the form of inhibitory mechanisms. These mechanisms suppress the unnecessary activation of the dormant language B so that the fluent and correct use of language A can be maintained (Kroll, Bobb, Misra, & Guo, 2008). The parallel activation of languages A and B is supported by between and within language comparisons. Using eye tracking, Marian and Spivey (2003) found an increase in the

number of incidental glances to objects with names that were phonologically similar across the two languages in bilingual speakers indicating parallel activation of both lexicons via phonologically similar words. Thus, representation of a lexical item itself activates all phonologically relating items, irrespective to the language currently in use (Beauvillain & Grainger, 1987; Marian & Spivey, 2003). The co-activation of two lexicons has also been revealed using event-related potentials (ERPs) in the electroencephalogram (EEG): Thierry & Wu (2007) showed that Chinese-English bilinguals had implicit access to their L1 (Chinese) lexicon when asked to judge semantic relatedness between two words written in L2 (English). The study showed that the N400 ERP component, sensitive to semantic priming, was reduced when two semantically unrelated English words shared a Chinese character. (For an overview of the prominent N400 ERP component see Kutas and Federmeier (2011).) Importantly, participants were performing the task exclusively through English and were not instructed to refer to Chinese translations or rate similarities due to characters. This study confirms that when presented with an isolated word, brain activation can be language-non-selective and word candidates from both languages become active (Dijkstra, 2005). While automated activation of the unused language B, when presented with verbal information in language A, cannot be fully suppressed, a variation of the decision criteria helps to optimize correct activation by adjusting the threshold of required input in language A or B dependent to the situation (Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & van Heuven, 2002). The requirement to suppress language B by engagement of inhibitory mechanisms is argued to reduce available cognitive capacity (Bialystok, 2009; Green, 1998): In monolinguals, there is no second language to be suppressed, and it is argued that this cognitive capacity can be deployed to other (in this instance) language-related cognitive processes. However, the suppression of language B induces additional cognitive load in bilinguals and therefore negatively impacts the processing of language A, resulting in poorer performance for bilinguals when compared to monolinguals in tasks such as lexical retrieval (Bialystok, Craik, & Luk, 2008), and verbal fluency (Gollan et al., 2005). Green (1998) referred to this constant need for focused inhibition within bilinguals in terms of an "inhibitory control model". Effects are recorded across the lifespan and persist with age (Gollan et al., 2007). Moreover, performance differences might root in balanced bilinguals' restricted exposure to and thus lower experience with each language (Lehtonen et al., 2018). This also indicates that potential slower language learning in childhood or low language proficiency of a person just in the process of acquiring a new language are likely to be the superficial presentations of a performance difference

between monolinguals and bilinguals, which has its origins in resource allocation within low-level, capacity-limited cognitive mechanisms.

One area of verbal skill in which bilinguals show an advantage compared to monolinguals is in new language learning including acquiring new vocabulary (Kaushanskaya & Marian, 2009) and specific grammatical rules. Bilingual speakers seem to have a better grasp of the concept of a language, the universal language concepts (Jessner, 1999). This enables the learning of a foreign language by helping to comprehend and internalize rules and exceptions. Having learned a second language seems to act not only as the learning of the actual language itself but also as the learning of a metalinguistic “language concept” (Bialystok, 1986).

1.4.2. Influence on executive functions

As stated above, bilingualism does not only affect language-related cognition: There are differences in the performance of mono- and bilinguals across a range of non-verbal cognitive tasks, particularly basic executive-function tasks, describing tasks concerned with attentional control and cognitive inhibition (Adesope et al., 2010; Bialystok, 2018; Diamond, 2013). Differences in executive-function performance between mono- and bilinguals also seem to vary as a property of the language skills of bilinguals, with balanced bilinguals generally showing the best performance on non-verbal cognitive tasks such as the Flanker Task (Thomas-Sunesson, Hakuta, & Bialystok, 2018). Typically, executive functions of this sort are assessed using experimental tasks such as the Stroop, Simon or the Flanker tasks:

In the Flanker task, “flankers” are placed to the left and right of a target arrow, which is itself presented at the centre of a computer monitor screen indicating to the left or right. Flankers are either congruent with the target, which means they are also arrows pointing in the same direction as the target, or they are incongruent, meaning they point in the opposite direction to the target. Flankers may also be neutral, without any directional orientation. Participants are asked to point out the direction of the target as quickly as possible and the reaction time is taken as a measure of response inhibition based upon selective attention (Eriksen & Eriksen, 1974).

The Stroop task assesses the ability to focalize attention and thereby suppress unnecessary information, including semantic information carried by lexical stimuli (Williams, Mathews, & MacLeod, 1996). In the Stroop task, participants are presented with congruent and incongruent conditions, a congruent condition consisting of a colour word printed in black ink or the correlating colour. By contrast, an incongruent condition consists of a colour word printed in a different colour. Experimental participants are required to report the colour in which the word is printed and in order to do this, they

must ignore the colour denoted by the word. Conversely, participants may be asked to ignore the printed colour of the word and report the denoted colour. Report time is faster, and errors are generally lower for congruent compared to incongruent trials, which is due to a disruption in processing and a consequent delay in response when the presented colour and the written colour word are different (MacLeod, 1991).

The Simon task similarly requires participants to ignore irrelevant information. The position of a target needs to be ignored, presented to the left or right of a computer monitor screen, and attention must be focussed on other, more task-relevant information. For example, a green target may indicate the participant is required to press the button to the left, while a red target indicates that the right button has to be pressed (Bialystok et al, 2004). Reaction times are generally faster when the target is presented closer to the side on the screen that is congruent with the side of the correct button to be pressed. In the Simon Task, the dimension is not lexical, but is instead spatial, measuring participants' ability to ignore this irrelevant spatial information and to attend to the task-relevant non-spatial stimulus. As was the case with the other measurements, reaction times in the Simon task tend to be faster for congruent relative to incongruent targets (Bialystok et al., 2004) showing that a decrease in disruptive information available, in this case, spatial information increases focus upon task-relevant information.

Bilinguals show a general advantage over monolinguals in each of these tasks: Bilinguals complete the Flanker task significantly faster than monolinguals (Bogulski, Rakoczy, Goodman, & Bialystok, 2015), with early bilinguals who acquired L2 before the age of ten outperforming late bilinguals (Hartanto & Yang, 2019). They are also faster than monolinguals to respond correctly in Stroop tasks. This is attributed to advanced suppression skills for unnecessary information as well as a superior selection process of relevant information compared to irrelevant input due to practice through language selection (Bialystok, Craik, & Luk, 2012). The response behaviour of bilinguals is also different to that of monolinguals in the Stroop task, in which overall longer response-initiation times combine with overall faster manual responses towards a selected target (colour) option (Incera & McLennan, 2016). Damian and colleagues (Damian, Ye, Oh, & Yang, 2018) summarize this initial delay within bilinguals, followed by a quicker, more direct reply compared to monolinguals, as a more "efficient" response pattern, as observed through mouse tracking within the Flanker, Simon and Stroop Task. This evidence shows the bilingual advantage extends beyond information processing, with an effect in the patterning of the manual response during selective-attention tasks. The superior performance of bilinguals on the Simon task has been

shown across the lifespan from young children to older adults (Bialystok, 2006; Bialystok et al., 2004; Lu & Proctor, 1995; Martin-Rhee & Bialystok, 2008).

These outcomes confirm the existence of a bilingual advantage in tasks measuring the control of relatively low-level cognitive processing. Given the nature of the tasks concerned, this advantage is likely to originate in an enhanced inhibition, or the ability to selectively prioritize and in so doing deselect irrelevant input for processing (Peal & Lambert, 1962). The effects of inhibition result from varying attentional or executive performance in different task environments. Thus, bilingual performance is interpreted in terms of a generalized enhanced inhibitory capability in bilinguals. It is related to the earlier mentioned requirement to suppress the unnecessary activation of the dormant language in order to maintain the fluent and correct use of the language currently in use (Bialystok et al., 2004).

Bilingual proficiency, including that of full language-immersed, 'balanced' bilinguals, develops over time. The question arises of at which point in time, or at which level of language proficiency, does bilingualism have an effect on cognitive performance? To answer this question Luk, De Sa and Bialystok (2011) investigated whether the bilingual advantage in cognitive performance varies depending upon the age at which a person became bilingual. Reporting data collected from samples of young adults they found a negative correlation between the AoA of L2 and the successful completion of the Flanker task, indicating that early bilinguals performed significantly better than late bilinguals – who in fact scored similarly to monolinguals. Luk and colleagues conclude that this was due to the early development of language skills in both languages, and is less attributable to proficiency in the languages (Luk, De Sa, & Bialystok, 2011). Supporting these findings, Sörman and colleagues (Sörman, Hansson, & Ljungberg, 2019) failed to find differences between monolinguals and late bilinguals, who learned their L2 after the age of six, when assessed on Flanker, Stroop and Simon Task. It seems likely that an early age of L2 acquisition may better facilitate low-level cognitive performance than later acquisition. Young adult bilinguals would have acquired superior inhibitory control during childhood at which point brain development is as yet not fully matured, while the types of cognitive processes studied are closer to brain function, than processes of reasoning or problem solving.

In addition to an effect of age on L2 performance, Pelham and Abrams (2014) have shown that both the levels of language proficiency, as well as the regular usage of the two languages have a significant effect on executive function as tested with the Attention Network Task, which combines a Flanker task with a cued reaction-time task. As mentioned above, factors such as AoA of L2, language proficiency and usage of

both languages, are crucial aspects when assessing the influence of bilingualism on cognitive performance (Kuhl, Stevenson, Corrigan, van den Bosch, Can, & Richards, 2016). Nevertheless, the influence of these factors may differ depending upon the exact type of cognitive performance under examination and these differences have not yet been sufficiently explored.

To summarize the evidence discussed thus far, there is an indication that bilinguals outperform monolinguals in tasks designed to measure low-level, executive functions concerned with the inhibition of one response (attentional or behavioural) relative to another. It has been conjectured that this bilingual advantage might originate in the capacity for a bilingual speaker to selectively process one of two possibly competing languages, whilst suppressing the other (Bialystok, 2009; Bialystok et al., 2004). Considering the vast amount of variables affecting cognitive performance and the wide variety of covariates within monolingual and bilingual individuals, current research has introduced a novel approach to explore the effects of bilingualism in more detail, entailing comparisons of bilinguals with varying degrees of proficiency and AoA.

On the basis of this, as well as evidence for the simultaneous activation of L1 and L2 lexicons, it can be concluded that everyday language use involves a continuous process of lexical-item activation and selection, during which simultaneously activated lexical items from the dormant or unused language require suppression. In this context, and perhaps given the frequency with which this process of suppression is required, a generalized improvement in selection and inhibitory control develops, which is superior to selection and inhibitory control in monolinguals. Indeed, according to Bialystok (1999), this improvement transfers from low-level cognition to mental suppression as well as to non-language related tasks (Adesope et al., 2010; Bialystok, 2007; Hilchey & Klein, 2011).

1.4.3. Questioning bilingual advantage on low-level cognitive processes

Reservations have nevertheless been expressed concerning the reliability of differences reported between monolinguals and bilinguals (Valian, 2015). For instance, Paap and Greenberg (2013) find that when implementing several assessments with the same participants, covering multiple measures such as Flanker and Simon task, which are commonly generalized as measures of inhibitory control, outcomes of the individual tests do not correlate (Paap & Greenberg, 2013). The lack of cross-task correlations makes a domain-general interpretation of the results difficult (Paap & Greenberg, 2013). This lack of correlation is not restricted to research on bilingualism but also applies to the broader field of research on executive functions (Lehtonen et al., 2018). Nevertheless, the breadth of research measures implemented ensures external validity

and should not be threatened by the homogenization of measurements, which would lead to an increase in internal validity at the expense of external validity. Draheim, Hicks, and Engle (2016) suggested a more applicable solution to this problem, stating the origin of non-correlations might lie in the variation of variables collected and compared, with either reaction time or accuracy scores being considered. A combination of scores of both, speed and accuracy, into one single metric might represent a more accurate base for the comparison of groups (Draheim, Hicks, & Engle, 2016).

Additionally, participants across different studies vary in pairs of languages, language skill levels and AoA of L2 as well as participant numbers. Although the explanation that inhibition of language B during language processing could lead to better executive control is plausible, studies that are more systematic are needed to confirm the relationship between bilingualism and changes in basic executive function. In summary, Paap (2015) queries the construct of the bilingual advantage, questioning the consistency in experimental methods and the consensus of definitions of bilingualism as a whole, and of grouping criteria of bilingual participants implemented by researchers between different studies leading to inconsistent results. Paap then calls for common guidelines setting minimum sample sizes to help justify sample sizes in future research and specification of language history of participants as well as cultural background when describing subjects in research studies (Paap, 2015; Surrain & Luk, 2017). A recent study by Hartano and Yang (2019) showed that significant differences within accuracy scores on the Flanker task between early and late bilinguals disappeared after matching participants on confounding variables such as demographic details, but differences in response time increased. Hartano and Yang's (2019) findings highlight the necessity of matching participants on non-linguistic covariates to strengthen results and enable the observation of true differences or non-differences.

The lack of a clear definition of bilingualism raises a problem (Bialystok, 2018; Kaushanskaya & Prior, 2015; Marian, 2018; Paap, 2015; Surrain & Luk, 2017): presently, the specific definition 'bilingual' is a matter of interpretation on the part of researchers. As a result, the basic set of constructs used to define bilingualism, as a variable, is not standardized across studies. This naturally leads to different inclusion criteria and means that bilingual performance, when considered across studies, may be influenced by variations in proficiency, AoA, and use of L2 through differences in factors that include participants' education, socioeconomic status, and geographical location, alongside the socio-cultural factors that may have influenced their bilingualism. Any confounding variable which has not been explicitly measured,

consequently describes a potential influence on outcomes of a research study (Marian, 2018), making it impossible to clearly generalize results from individual studies and adding the threat of significant findings originating in sampling errors. On a broader level, this problem compounds with an absence of specification in some studies of the language sets, as well as the different proficiencies or exposure of participants to both languages.

Issues with consistency in relation to the above-mentioned relevant factors might explain why some studies have failed to confirm a bilingual advantage and report either no differences or a disadvantage for bilingual participants (Paap & Greenberg, 2013). De Bruin, Treccani and Della Sala (2015) take this argument a step further and see the origin of the currently widely assumed bilingual advantage as a form of publication bias: The publication of negative or null findings in the area is very rare. Studies with positive findings, which tend to support the bilingual advantage are much more likely to be accepted by journals, followed by studies with mixed findings, whereas negative findings towards the bilingual advantage, or those showing a disadvantage, are up to twice as likely not to be published (De Bruin, Treccani, & Della Sala, 2015). This trend seems to be slightly reversing since 2015 and the number of publications challenging the bilingual advantage is increasing (Sanchez-Azanza, López-Penadés, Buil-Legaz, Aguilar-Mediavilla, & Adrover-Roig, 2017). Lehtonen and colleagues (Lehtonen et al., 2018) conducted a meta-analysis including corrections for publication-bias and summarized that when correcting observed differences of published and unpublished studies, any evidence for a bilingual advantage in inhibition, shifting and working memory, a core aspect of executive control (Thomas-Sunesson, Hakuta, & Bialystok, 2018), disappears.

Additionally to the publication bias, studies with small sample sizes seem to comprise of stronger positive outcomes than studies with big sample sizes, which in turn show weaker effects overall (Lehtonen et al., 2018), posing an effect on meta-analyses in the research domain.

Based on these critical evaluations, Lehtonen and colleagues (Lehtonen et al., 2018) suggest the pre-registration of research studies exploring a bilingual advantage. Another solution to the problem might be a shift of standards and a novel approach for future directions of the research area, pointing towards comparisons of bilinguals of different proficiencies rather than comparisons of monolinguals versus bilinguals (Bialystok, 2017; Lehtonen et al., 2018; Mishra, 2015). The comparison of bilinguals can help to eliminate covariates such as culture and SES through facilitated matching procedures due to similar target groups and might lead to important insight into underlying cognitive changes evoked by certain AoA or L2 proficiency.

In summary, low-level cognition may not say everything about bilingual cognitive performance because it is limited to a subset of cognitive operations. The research field exhibits a lack of consistency in terms of the characteristics of participants and, in particular, the language ability of participants to the extent that it might be difficult to conclusively define an overarching explanatory framework due to difficulty in generalizing findings across studies. Adopting a different approach is necessary to successfully explore the question of how speaking multiple languages on a daily basis influences human thinking. This approach must be taking into account all of the above-mentioned limitations (Valian, 2015) whilst addressing an additional lack of clarity in the definition of bilingualism (Paap, 2016). A focus on more complex cognition which consolidates both, lower executive functioning itself and the language it is based on, might lead to more consistent results as will be shown in the following.

1.5. High-level cognition and bilingualism

1.5.1. Defining high-level cognition

High-level cognition is characterized by processes that allow a level of cognitive control, including the assimilation and contextualization of information before making a decision or coming to a conclusion, and metacognition such as varying levels of confidence about the correctness of a judgment that may be adjusted in relation to the evidence available (Evans & Stanovich, 2013). High-level cognition, such as reasoning and problem solving, can be based on an interplay of various basic executive functions and thus entail the support and reflect the constraints of lower-level cognitive processing (Anderson, Matessa, & Lebiere, 1997; Blanchette & Richards, 2010). Examples for this interaction include working memory in problem-solving or decision-making, both representing higher cognition (Anderson, Matessa, & Lebiere, 1997). High-level cognition is often consciously available, volitional in that deliberate mental effort is applied, and contains a certain level of complexity. Instances of high-level cognition, all of which can involve a series of steps throughout the thinking process (Evans, 2008; Frith & Dolan, 1996), are the interpretation of ambiguous scenarios, the estimation of the likelihood of different outcomes, decision making, including choosing between several options presented, problem solving and reasoning (Blanchette & Richards, 2010; Evans, 2008; Sternberg, 2004).

In the majority of processes in high-level cognition the individual has immediate access to a functional structure. In this respect, high-level cognition differs from the type of cognitive performance that has contributed to the concept of 'bilingual advantage'. High-level cognitive processes are embedded either in language directly or

in heuristics, which are similar to language in that content or statements may be constrained by rules and operators, and thus by systems that are functionally similar to a syntax. On this basis, high-level cognitive performance may influence or be influenced by L2 ability (Frith & Dolan, 1996). This basis of higher-order cognition in language or heuristics enables researchers who are investigating possible effects of bilingualism on higher-order cognitive functions to extend the reach of their outcomes and not only set up comparisons between monolinguals and bilinguals but also to measure differences within bilinguals' first and second language. This provides researchers with valuable information aiding to examine the influence of L2 acquisition and usage on cognition. Whilst we return to the relationship between syntax processing in natural languages and bilingualism in a later section, in the following section we examine the current research linking high-level cognitive performance with bilingualism.

1.5.2. Critical thinking and bilingualism

A relevant superordinate, describing the everyday use of high-level cognition is a cluster of cognitive operations referred to generically in terms of critical thinking (CT), the ability of purposefully evaluating any problem in a goal-orientated manner (Facione, 1990; Halpern, 2002). CT will be defined in more detail in section 1.6. Consisting of complex processes (Halpern, 1998) CT is critical in situations demanding problem-solving, decision making or the consideration of likelihoods of certain outcomes of a situation, ideally leading to a desirable result (Halpern, 2002). Consequently, CT is a set of tools aiding judgment and consisting of different steps: Interpretation of a scene or scenario, its analysis and evaluation, followed by an inference based upon well-considered judgment that is led by reason and evidence considering all options and avoiding any subjective bias (Facione, 1990; Paul & Elder, 2005). CT can be broken down into five subcategories or sub skills (Halpern, 2010): These are verbal reasoning skills, argument analysis, skill in hypothesis testing, likelihood and uncertainty analysis, and decision making and problem-solving skills.

To date, there is very little research investigating the influence of bilingualism on critical thinking, depicting a lack of tangible evidence of a possible effect of bilingualism on higher cognition. One of the few published research studies reveals a relationship between bilingualism and critical-thinking disposition (Albert, Albert, & Radsma, 2002), with dispositions describing the aptitude of a person for effective critical thinking (Facione, 1990). Albert, Albert and Radsma (2002) assessed participants performance with the California Critical thinking Skills Test and Disposition Inventory, a measure asking participants to agree or disagree with statements about opinions related to CT. Test takers underwent assessment in the language of their

choice, either L1 or L2, preventing comparison between languages. This study lacks a specific definition of CT, which makes it difficult to differentiate performance relative to particular CT subcategories. Results are kept very general only stating the significant relationship between bilingualism and CT disposition without further specifying this relationship or the specific CT disposition affected. Bataineh and Zghoul (2006) examined high-proficiency bilinguals in the process of completing a teaching degree using the Cornell Critical Thinking Test, presented to them in their second language. The Cornell Critical Thinking Test consists of short passages of text that are followed by questions related to statements in the text, to be answered via multiple choice answer options. Whilst the subsection 'judging conclusions' scored highest, overall scores were lower compared to normed scores for the assessment, indicating a poorer performance in L2, but the lack of a control group make statements about the outcome problematic (Bataineh & Zghoul, 2006). However, reservation must be exercised in the case of both, Albert, Albert, and Radsma (2002) and Bataineh and Zghoul (2006) because neither study includes a within-subjects comparison. Consequently, the evidence from these studies remains inconclusive concerning both the cognitive performance of bilinguals and any difference in performance between L1 and L2 within bilinguals, and the need for comprehensive research remains.

CT is an umbrella term covering a variety of different forms of high-level cognition. This raises the possibility that performance differences could be specific to particular types of CT, a possibility raised by Bataineh and Zghoul's (2006) finding that scores on 'judging conclusions' were higher than the other forms of CT tested. To examine this possibility, and to deepen the insight into a possible effect of CT on bilingualism, research on some of the subcategories of CT, including problem-solving and verbal skills, as introduced above will be reviewed in the following sections, in which differences between L1 and L2 performance will also be examined. Differences between monolinguals and bilinguals or differences in performance in L1 and L2 within a bilingual individual in different subcategories of CT would indicate an influence of bilingualism on CT, and in this respect an influence of bilingualism on high-level cognition.

1.5.3. Problem solving and bilingualism

Problem solving entails finding a solution to a possibly complicated situation and can be considered a central subcategory of critical thinking (Halpern, 2002). Problem solving defines two kinds of problems: Insight and non-insight problems. Examples of insight problems are spatial object-move problems of a creative kind, like the triangle of coins. Participants are presented with ten coins laid out in a triangular shape and are

asked to move three coins for the triangle to point to the opposite direction (Cushen & Wiley, 2011). Non-insight problems are more math based, for instance, “Solve for Y” problems, where participants are presented with a number of equations with letters and numbers (e.g. $3x+y=7$) and are asked to solve the equation.

1.5.3.1. Monolingual and bilingual performance in problem solving

Overall, bilingualism positively affects an individual’s problem-solving skills, especially concerning tasks, which are based on executive control like cognitive flexibility and abstract thinking (Adesope et al., 2010; Kharkhurin, 2017). Early bilinguals who have acquired both languages before the age of six, show an advantage over monolinguals in the creative insight problem task “Triangle of Coins” (described above, see Cushen & Wiley, 2011), whereas monolinguals seem to show an advantage over bilinguals in earlier mentioned “Solve for Y” math-based problems (Cushen & Wiley, 2011). This has been interpreted as early bilinguals having grown up speaking two languages benefitting from a higher level of cognitive flexibility (Adesope et al., 2010; Kharkhurin, 2017) and an ability to adjust to changes in creative tasks more efficiently (Cushen & Wiley, 2011). The advantage of bilingualism for creative problem solving is believed to originate in the broader cultural experiences of bilingual participants enabled by their L2 knowledge (Cushen & Wiley, 2011).

1.5.3.2. Problem solving performance across languages

When presenting bilinguals with problems in their L1 versus their L2, they tend to make decisions in their second language less intuitively and more deliberately (Costa & Sebastián-Gallés, 2014; Costa, Vives, & Corey, 2017; Keysar, Hayakawa, & An, 2012). This so-called ‘foreign-language effect’ includes a decrease in risk aversion in L2. This decrease in risk aversion in L2 might result (for example) in: i) a reduced sensitivity to the presentation of options in terms of gains or losses; ii) a reduction in illusory fallacy, which might lead to expectation of a continued positive outcome predicated on a number of positive outcomes, albeit across unrelated events; and iii) a more outcome-based decision-making process concerning moral judgment, as in the footbridge dilemma, in which participants have to sacrifice the life of one person to save five others (Costa & Sebastián-Gallés, 2014; Costa, Vives, & Corey, 2017; Keysar, Hayakawa, & An, 2012). Different susceptibilities to cognitive bias dependent upon the language in which the bias-inducing material is presented are suggestive of different processing paths, involving interplays of intuition, deliberation, and creative thinking that might be activated by native compared to foreign language contexts. More specifically, if the same person is presented with a scenario in their L2 rather than L1, that person might take a more deliberate approach at solving it, including a more

neutral stand to the phrasing of the options, either emphasizing gains or losses, being aware of fallacies and having an increased focus on the outcomes in moral decisions.

Outcome bias and representativeness heuristics, on the other hand, seem to be used to the same extent in both languages, possibly indicating the boundaries of the foreign-language effect (Vives, Aparici, & Costa, 2018). In the foreign-language-effect literature, L2 is typically referred to as a 'foreign language', but definition and specifications of the proficiency and usage of the second language are often not given. High language proficiency (Eilola & Havelka, 2011) and early age of L2 acquisition (Ferré, García, Fraga, Sánchez-Casas, & Molero, 2010) seem to lead to an absence of the foreign-language effect. This means the effect might not be as straightforward as originally assumed and underlines the importance of assessment of both, proficiency and AoA of L2. Language competency might remain an issue in all bilinguals except in the case of balanced bilinguals, who have typically been raised from infancy in an environment including two languages. This may also refer to the effects of low-level cognition on other psychological dimensions: For instance, fully proficient bilinguals score similarly to monolinguals and show the same pattern in differences on the emotional and taboo Stroop task based upon an evaluation of response times to neutral as compared against those to emotional and taboo words (Eilola & Havelka, 2011). In addition, when assessed on memorizing positive, negative and neutral words no difference in outcomes was found between L1 and L2 in early and late bilinguals of high proficiency (Ferré et al., 2010). A recent study exploring the foreign-language effect found no differences comparing L1 and L2 in terms of outcome bias, the use of representativeness heuristics (which describe judgment under uncertain conditions), and emotion-based decision-making (Vives, Aparici, & Costa, 2018; also Eilola & Havelka, 2011). Further studies are needed to specify the effect of proficiency, especially high proficiency, on the foreign-language effect. With this in mind, it is worth noting that in the study of Vives, Aparici, and Costa (2018) participants were included if they had acquired their L2 in a classroom setting while participants who had previously lived in a country where their L2 was spoken were excluded.

1.5.4. Verbal arithmetic skills and bilingualism

Comparisons of bilinguals' verbal arithmetic skills between their L1 and L2 find improved performance when solving problems in their L1 (Bernardo, 2002; Bernardo & Calleja, 2005; Van Rinsveld, Schiltz, Brunner, Landerl, & Ugen, 2016). As the task was increased in difficulty and/or complexity, there was an increased difference between L1 and L2 performance, with faster and more accurate L1 performance on arithmetic tasks that require a verbal solution. This indicates reliance upon verbal processes embedded

in L1 for complex mathematical tasks. Similar results have been obtained with children in primary school who were brought up through L1 and educated through L2 (Van Rinsveld et al., 2016). However, Van Rinsveld and colleagues did not measure language proficiency, and this may have had an influence on performance. Interestingly, when prompted by a preceding verbal judgment task in L2, which primed a language context, L2 performance was improved (Van Rinsveld et al., 2016). This is consistent with the 'language mode hypothesis' (Grosjean, 2001) in which bilingualism can be described as varying activation of both languages on a performance continuum. This means that an individual can be situated closer to the monolingual mode, with one language mostly suppressed, or closer to the bilingual mode, with both languages equally active, dependent on the momentary linguistic demand and language context (Grosjean, 2001; Grosjean & Byers-Heinlein, 2018). The current location of a bilingual on the performance continuum is flexible and capable of changing frequently, making a bilingual's processing system flexible and dynamic in its nature (Grosjean & Byers-Heinlein, 2018). The crucial factors here are the environment and usage of both languages, which also indicate the importance of measuring contextual variables when defining bilingualism.

1.5.5. High-level cognition in language skills

Grammatical knowledge is at the heart of language competency and syntactic priming can be considered one of the most basic indicators that high-level cognition is influenced by bilingualism. Lexical processing, consisting of word recognition and lexical access, is considered the base of language processing, representing low-level cognition. Syntactic information processing, on the other hand, tends to rely upon low-level as well as high-level cognitive processing (Pulvermüller & Shtyrov, 2006). Sentence processing can be enhanced if an individual is presented with a prime sentence comprising similar syntactic structures (Bock, 1986; Jackson & Ruf, 2017). In verbal speech production (Hartsuiker, Pickering, & Veltkamp, 2004) as well as in written speech production (Kantola & van Gompel, 2011) bilingual individuals are inclined to copy the grammatical structure of a previously presented sentence structure, even across languages. There is evidence, that this type of cross-linguistic grammatical priming influences both the production and comprehension of sentences (Hartsuiker, Pickering, & Veltkamp, 2004) with short-term priming associated with a facilitating effect on L2 learning (Jackson & Ruf, 2017). Language priming effects suggest that the different syntactic structures of both languages are stored together, and are language independent, at least as long as grammatical constructs between languages are comparable (Hartsuiker, Pickering, & Veltkamp, 2004). Nevertheless,

syntactic priming that represents high-level cognition tends to be weaker than low-level lexical priming (Schoonbaert, Hartsuiker, & Pickering, 2007).

In a more general context, and applying verbal executive functioning findings to higher level cognition within a bilinguals' everyday life, it has been found that there is a significantly lower understanding of, and ability to summarize written content presented in L2 when compared to monolingual speakers (Trenkic & Warrington, 2018). Overall, lower exposure to L2 and more experience in L1 is thought to result in more automated L1 processing and the origin of lower performance in L2 (Trenkic, Mirkovic, & Altmann, 2014). This might explain why bilinguals listening to speech in L2 missed around 30% of the content presented compared to monolingual controls (Charles & Trenkic, 2015). Moreover, the process of suppressing L1 when producing content in L2 is more effortful compared to suppressing L2 when speaking in L1 and this could lead to poorer outcomes in cognitively demanding verbal tasks in L2 (Austin, Pongpaioj, & Trenkic, 2015; Marian & Spivey, 2003). Reasons for a disadvantageous L2 processing might be the gap between the L1 and L2 lexicons. This applies both in terms of their size and organisation, which favours cognitive operations in L1 (Skehan, 2009), structural competition in the form of interference from L1 during L2 processing (Trenkic, Mirkovic, & Altmann, 2014) and the erroneous input of lexical elements stemming from L1 grammar systems (Luk & Shirai, 2009). Nonetheless, a within-subject comparison of performance in L1 and L2 would add unbiased data avoiding results due to matching-errors, which cannot be excluded based on research conducted to date. Presented findings alongside the fact that speech production is fundamentally based upon lexical-syntactic systems, which contain structures for both languages, lead more directly than any other set of findings, to the question 'What effect does bilingualism have on higher-order cognition?' Addressing this question would serve to extend knowledge on the influence of bilingualism on cognitive performance beyond studies of executive function. It seems likely to provide results of more direct relevance to ability with and the use of L2.

1.5.6. Bilingualism throughout the lifespan

Current research on bilingualism has focused mainly on the performance of children and the elderly. In these groups, the effects of executive control on language are more often reported compared to in young adults (Bialystok, Craik, & Luk, 2012; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009). The bilingual advantage seems to be less distinct within participants around the age of twenty years (Gold et al., 2013). This could be attributable to the fact that the cognitive ability is at peak efficiency at this age (Craik & Bialystok, 2006) and bilingualism offers no further improvement. In other

words, the high cognitive ability in young adults, as a baseline, makes a distinction between monolinguals and bilinguals more difficult to establish at that age.

Supporting this assumption are outcomes within young adults, that prove that only tasks including high monitoring conditions and complex, difficult tasks show differences (Bialystok, 2006) whereas comparably easy settings that show differences in children or older adults remain very small when conducted with young adults compared to monolingual scores (Bialystok, Craik, & Luk, 2012).

Whether the degree of advantage and disadvantage of bilingualism varies depending upon the age at which a person became bilingual was investigated by Luk, de Sa, and Bialystok (2011), who sought to find a relation between the onset age of bilingualism and its positive effects on cognitive control. In contrast to previous studies, onset age is defined by the age that the two languages are actively being used as opposed to the first contact with the second language. Participants in this study were 100 young adults around the age of twenty, divided into three groups, monolinguals, early bilinguals, and late bilinguals. Division followed a self-report questionnaire of language skills. The researchers aimed to assess executive control, which they conducted by implementing the Flanker Task (Eriksen & Eriksen, 1974), which uses irrelevant conflicting flankers for distraction. Outcomes revealed a positive relationship between the age of onset of bilingualism and the successful completion of the task, indicating that early bilinguals performed significantly better than late bilinguals, who scored similarly to monolinguals. Luk and colleagues conclude that this was due to the early development of language skills in both languages and is less attributable to proficiency in the languages.

Pelham and Abrams (2014) employed an oral assessment. Monolingual English-speaking students, early bilingual students with English Spanish as their two languages, in which they were fluent no later than with seven years of age, and late bilingual students, who acquired their second language after the age of thirteen, were recruited. Categorisation took place following the assessment of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) and oral language production in Spanish.

Two assessments were undertaken, a picture naming task for lexical access and an attentional network task (similar to the Flanker task) for executive functioning.

Contrary to Luk, de Sa and Bialystok (2011), Pelham and Abram's (2014) findings show that the level of language proficiency, as well as the regular usage of both languages, have a significantly larger effect on executive functioning than the age of second-language-acquisition. Early and late bilinguals did not differ in outcomes,

meaning that the crucial factor would be language proficiency. This can be achieved at any stage in life.

Influence of bilingualism on older adulthood has been the focus of attention of many researchers in recent years. Studies have proven that bilingualism has a protective effect on the age-related cognitive decline (Bak, Nissan, Allerhand, & Deary, 2014; Bialystok, Craik, & Luk, 2008). A positive influence on the age of onset of symptoms of dementia was shown. Symptoms within bilinguals usually appeared around three to four years later than for monolinguals (Bialystok, Craik, & Luk, 2012), due to a delay of age-related deteriorations in neural efficiency (Gold et al., 2013).

1.6. Critical Thinking

“The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit.” (Facione, 1990; page 6).

The abundance of information reaching individuals through a variety of different sources leads to the increasing necessity of filtering reliably from poor sources, useful from useless input, and important from redundant information. Thinking critically is a form of higher-order thinking (Facione, 1990; Marin & Halpern, 2011), which includes weighing up options in a purposeful and goal orientated manner when faced with any problem (Facione, 1990; Halpern, 2002). The expert panel of the Delphi Report on Critical thinking (Facione, 1990) divided the concept of CT into two categories, critical thinking dispositions and CT skills (Halpern, 2006). CT dispositions describe a person’s consistent internal motivation to implement CT, characterising each individual through distinguishable features (Facione, 2000; Halpern, 2006). CT skills are exhibited in an individual’s performance, they are pervasive and purposeful (Facione, 2000). CT skills describe the ability to think critically, which is not a predisposition but can be improved through teaching (Halpern, 2006). The growing importance of CT skills in education and the workplace, with links to intelligence, combined with the possibility to measure them, explain the focus of research on CT skills over dispositions (Halpern, 2006). The research area of CT established multiple definitions of the construct with varying levels of detail, generally agreeing on similar aspects to be of importance (Halpern, 2002). As defined by Diane Halpern (1998; 2010), a successful process of critical thinking can be broken down into five subcategories or sub-skills (Halpern, 1998, 2010): Verbal reasoning skills, argument analysis skills, hypothesis testing skills, likelihood and

uncertainty skills and decision making and problem-solving skills. These subcategories, which will be described in more detail below, are overlapping and interconnected and do not act independently. They cannot be isolated but work together in any scenario. Any given task requiring critical thinking entails a different combination of the five sub-skills (Halpern, 1998, 2010).

Overall, CT is a self-regulatory tool for judgment (Facione, 1990; Paul & Elder, 2006), implemented by individuals to increase the chances of a “desirable outcome” (Halpern, 2002, page 6). It is a high-level cognitive skill consisting of complex sub-processes (Halpern, 1998) which are crucial in situations demanding problem-solving, decision making or the consideration of likelihoods of certain outcomes of a situation (Halpern, 2002). Every individual holds a different set of strengths and weaknesses within these sub-skills, and it is not a prerequisite to excel in all those skills to be a good critical thinker. As mentioned above, the improvement of thinking skills is possible through critical and self-aware engagement with thinking processes, the aim for objective and logic decision making, and by broadening one’s own area of expertise (Facione, 1990; Halpern, 2006; Marin & Halpern, 2011).

1.6.1. Hypothesis Testing (HT)

Hypothesis testing describes the skill to use information to confirm or disconfirm a given hypothesis. In successful hypothesis testing, rushed generalizations are avoided by highlighting the importance of appropriate sample size or the implementation of control conditions. HT is an important factor to consider in any everyday situation as well as in scientific research. Examples of poor hypothesis testing are a new colleague being in a bad mood leading to the conclusion that this colleague must be habitually irritable. Here, a single time point is taken as a sample size rather than gathering information about that colleague throughout a few days or weeks to create a repeated measure sample. Another instance for poor hypothesis testing is concluding, one gets a cold by not wearing a hat after a single prior situation without a hat leading to a cold. The absence of a control condition in this situation weakens the conclusion.

1.6.2. Verbal Reasoning (VR)

Verbal reasoning addresses the ability to analyse language and the awareness of how language influences thought. Applied verbal reasoning skills help to recognize the use of pervasive or misleading language, to perceive verbally delivered messages without being swayed by possible persuasion embedded in language. Prior knowledge about related issues is being used to draw conclusions about the truth or correctness of the fact being assessed and helps humans to figure out relationships between different constructs. Reasoning by analogies is applied daily in decision-making (Sternberg,

1977), for instance when purchasing a good of a certain brand because the last purchase of this brand was positively annotated or following the advice of a person after another piece of advice of the same person was helpful.

1.6.3. Judging Likelihood and Probability (JL)

Judging likelihood and probability skills refer to unknown situations in everyday life. They describe the realistic estimation of the chance of an event, applying relevant principles of probability. Basic examples are throwing dice or flipping a coin. These judgments underlie general misinterpretations, as some of the rules are slightly counterintuitive, with the first answer which springs to mind usually being incorrect, and further evaluation being needed. Concepts covering likelihood and probability are dependent on a learning process. They are taught as the regression to the mean, the fact that extreme events are likely to be followed by less extreme events (Lehman, Lempert, & Nisbett, 1988), gambler's fallacy, describing that every flip of a fair coin has the same probabilities, even after three flips had the same outcome, the fourth flip is not more likely to land on the other side (Lee & Smith, 2002), and the law of large numbers, describing the growing chance of an expected outcome the higher the sample size (Toplak, West, & Stanovich; 2011).

1.6.4. Argumentation Analysis (AA)

Argumentation analysis describes the ability to judge the quality of an argument, consisting of statements, reasons, and conclusions. It includes judgment of the individual speaker, the reasons and conclusions drawn as well as the ability to filter out irrelevant information. Halpern (2002) defines three parts of argumentation analysis, acceptability and consistency of the argument, the relationship between argument and conclusion, and neglected aspects within the argument. Being able to differentiate between a conclusion and an assumption and making judgments based on reasoned thinking rather than uninformed opinion show good thinking and skilled argumentation analysis.

1.6.5. Problem Solving (PS)

Problem solving, parallel to its colloquial definition, entails finding a solution to a, possibly complicated, situation. Good problem-solving skills help an individual to identify the core of a problem including different problem statements, leading to the generation of feasible alternatives and the selection of a solution amongst the alternatives. It partly contains the above-mentioned subcategories of CT and can be considered an overarching skill for critical thinking, due to the fact that all other subcategories of CT are involved in the problem solving process (Halpern, 2010).

1.6.6. Summary

In summary, critical thinking describes the evaluation, analysis, and assessment of a situation, and its critical contemplation in regards to the outcome (Paul & Elder, 2005). Researchers have linked CT to certain aspects of intelligence, dependent on the definition applied (Halpern, 2002). This is especially applicable to intelligence defined as a fluid concept, improving through study, even though somewhat restricted by genetic disposition, and leading to individuals excelling in everyday situations. Intelligence entails a significant overlap with the concept of CT in terms of its definition, including emotional intelligence, necessary for interpersonal situations (Halpern, 2002; Stanovich & Stanovich, 2010). IQ, as measured in standardised tests, on the other hand, builds on intelligence defined as a fixed disposition. Correlations between CT and IQ defined as this fixed disposition are often not observable, which is why the connection between IQ and CT is often kept vague (Halpern, 2002), but a correlation between the two concepts going beyond current measures and including more fluid aspects of intelligence, can be assumed, based on similarities of skills and abilities included in both terms as per definition (O'Hare & McGuinness, 2009; Stanovich & Stanovich, 2010).

Critical thinking assessments, such as Halpern's Critical Thinking Assessment (HCTA, Halpern, 2010) or the California Critical Thinking Assessment (CCTA, Facione, 1990) have been validated and correlate with real-world critical thinking outcomes (Butler, 2012) providing the valuable evidence that the assessment of CT as currently defined is possible and reliable. CT can be assessed in a multiple-choice format (California CT Assessment, HCTA test form S2), with open-ended questions, or in a mixture of both (HCTA test form S1). Each measurement method includes advantages and limitations (Ennis, 1993). Multiple choice (MC) format represents an efficient method of assessing participants in terms of test duration and scoring, through selection of the most accurate alternatives, however not assessing participants' own ideas but focussing on the evaluation of options presented. Open-ended questions assess spontaneous, creative CT, restricted due to the necessity to phrase thoughts, which favours better writers, as well as an extensive scoring process. A mixture of both MC and open-ended questions, cumulates advantages in one assessment, but still entails limitations such as a high duration of assessment and an extensive scoring process (Ennis, 1993; Halpern, 2010). Availability of published CT assessments in languages other than English is restricted to a few specific languages (e.g. Spanish). This makes a bilingual assessment difficult and confines research to available

languages. An offer for cooperation with the authors of HCTA in the form of professional translations of the existing assessment was rejected which led to the necessity of creating a new critical thinking measure. This measure consists of multiple published tests widely used in the literature around critical thinking and is based on Halpern's definition of critical thinking, including the five subcategories HT, VR, AA, JL, PS. Due to its nature, the accumulation of published assessments, it will be referred to as *Mixed Measures*. It will be introduced in detail in the next section.

2. The Assessment of Critical Thinking

A total of 21 short tests with a published scientific literature base, were added together into what will be referred to as *Mixed Measures*. This represents a collection of published assessments with different measures, amounting to a generalized assessment of critical thinking that will be described in detail below. All questions follow a multiple-choice (MC) answer format, due to feasibility (Ennis, 1993) and the large amount of different languages implemented, which limits the effectiveness of scoring of the assessment using an open-ended question format. The MC format also ensures scoring objectivity, which can be assumed due to scoring being automated, predetermined and coded before data collection started. Each participant was presented with the same instructions independent of the researcher, ensuring administrative objectivity. The creation of the assessment followed one of the main definitions and sub-categorization in the research field of critical thinking by Halpern (1998, 2002), including the five subcategories hypothesis testing (HT), verbal reasoning (VR), argumentation analysis (AA), judging likelihood and probability (JL), and problem solving (PS). Measures were selected based on a literature search, and available tests were grouped dependent on requirements for each category, and were matched for length and difficulty. The difficulty was determined through the percentage of correct solutions reported in papers which previously implemented each particular test (for details see the description of each subcategory) and by comparison to other available tests in the literature. High difficulty levels were employed to avoid ceiling effects and to spread out and emphasise possible differences observed within and between groups. Difficulty levels vary slightly depending on the subsection and the tests assessed. Outcomes were aimed to be analysed in a comparative manner by focussing on between- and within comparisons respective to mono- and bilinguals, and including L1 and L2 performance within participants. Performance comparisons between the five subcategories are not meaningful, because difficulty levels across sections are not standardised and raw scores are used for the calculations, which vary in range and data type between sections. As such, each of the five subsections was aimed to require a similar amount of time regarding time required for reading and for solving each scenario. The overall duration of the assessment was kept at a tolerable level of one hour. Moreover, participants were given the possibility to pause and continue the assessment later and are exposed to low stress levels due to no temporal restrictions which would lead to unnecessary pressure. Similarly to the HCTA (Halpern, 2010), *Mixed Measures* is assumed to be resistant to bias due to the fact that the most

socially appropriate response is not necessarily the correct response. Finally, the fairness of the assessment is especially important to consider in a multilingual, multicultural data collection process. Therefore, cultural adaptations dependent on the language set were undertaken based on the input of native speakers who were consulted during the translation process and pilot testing, which specifically aimed at the correct comprehension of all scenarios.

Table 2.1. Overview of Sources for each Question implemented in Mixed Measures divided by Sub-category

Category	Strategy	Question	Source
HT	Statistical heuristics	College Scenario	Fong, Krantz, & Nisbett, 1986; Lehman, Lempert, & Nisbett, 1988; Nisbett, Krantz, Jepson, & Kunda, 1983; Stanovich & West, 1998; Toplak, Liu, MacPherson, Toneatto, & Stanovich, 2007
	Methodological reasoning	Police Chief Dilemma	Lehman, Lempert, & Nisbett, 1988
	Inference problems	Unknown Car Task	Mynatt, Doherty, & Dragan, 1993
	Gathering information	Party Guests Scenario	Beyth-Marom & Fischhoff, 1983; Stanovich & West, 1998; Toplak et al., 2007
	Conjunction fallacies	Linda Problem	Kahneman & Tversky, 1982; Reyna, 1991; Toplak, West, & Stanovich, 2011; Tversky & Kahneman, 1983
	Weighting options	New Car Scenario	Fong, Krantz, & Nisbett, 1986; Nisbett, Borgida, Crandall, & Reed, 1976; as cited in Nisbett & Ross, 1980; Stanovich & West, 1998; Toplak, West, & Stanovich, 2011
VR	Verbal reasoning	Analogies	Dermott, 2002; Sternberg, 1977; Matsuoka & Lepage, 2014
JL	Sample size	Squash Problem	Kahneman & Tversky, 1982; Toplak, West, & Stanovich, 2011
	Regression to the mean	Baseball Scenario	Lehman, Lempert, & Nisbett, 1988; Toplak et al., 2007; Toplak, West, & Stanovich, 2011
	Gambler's fallacy	Slot Machine	Toplak et al., 2007; Toplak, West, & Stanovich, 2011

CHAPTER TWO – MIXED MEASURES

		Tossing a Coin	Toplak et al., 2007; Toplak, West, & Stanovich, 2011
	Probabilistic reasoning	Marble Game	Kirkpatrick & Epstein, 1992; Toplak et al., 2007
AA	Syllogistic reasoning	Syllogisms	Gilinsky & Judd, 1994; Klauer, Musch, & Naumer, 2000; Newstead, Pollard, Evans, & Allen, 1992; Toplak, West, & Stanovich, 2011
PS	Impulsive answers	Cognitive Reflection Test	Frederick, 2005; Gilhooly & Murphy, 2005; Metcalf & Wiebe, 1987; Nagin & Pogarsky, 2003; Sternberg, 1986; Sternberg & Davidson, 1983
	Insight problems	Matching Socks	Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005; Metcalf & Wiebe, 1987; Sternberg, 1986
		Trains and Bird	Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005; Posner, 1973
		Horse-Trade	Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005; Maier & Burke, 1967; Maier & Solem, 1952; Metcalf, 1986; Webb, Little, & Cropper, 2017
	Non-insight problem	Crime	Schooler, Ohlsson, & Brooks, 1993
	Conditional reasoning	Card Sorting Test	Wason, 1968

Note. HT=hypothesis testing; VR=verbal reasoning; JL=judging likelihood and probability;
AA=argument analysis; PS=problem solving

The original version of the assessments, Version A, was slightly changed to develop an alternative version, Version B, as close to the original assessment as possible referring to wording and answer options but with changed subjects and objects in each question. For instance, if Version A is about Mrs W buying X, Version B would contain Mr Y. selling Z. Both versions of the assessment were translated into seven different languages (English (original), Irish, German, Italian, Russian, Turkish, Chinese, and Japanese) by native speakers, using back and forth translations and

controlled by multiple reviewers. This process was similar to published studies in the research area of bilingualism (e.g. Eilola, Havelka, & Sharma, 2007). For details about the translation process please see sub-section 3.1.2.2, 3.1.3.2, etc., each introducing a specific language set. An overview including all measures can be found in Table 2.1. In the following section 2.1-2.5, each of the five subcategories will be introduced including the array of tests to assess it. The entire assessment including all questions can be found in Appendix 4.

2.1. Hypothesis Testing

The aim of Hypothesis Testing (HT) is to assess of the ability to judge the validity of a hypothesis. Factors assisting the judgment of this validity include the law of large numbers (describing the growing chance of an expected outcome the higher the sample size, Toplak, West, & Stanovich; 2011), as well as the awareness of the need for a control condition in decision making. Moreover, successful hypothesis testing includes gathering of information for each alternative and the ability to judge the quality of a source of information, which in turn might lead to an adjustment of one's own opinion of this piece of information. *Mixed Measures* assesses hypothesis testing with a range of six questions covering different strategies: statistical reasoning (Nisbett, Krantz, Jepson, & Kunda, 1983), methodological reasoning (Lehman, Lempert, & Nisbett, 1988; Mynatt, Doherty, & Dragan, 1993), inference problems (Beyth-Marom & Fischhoff, 1983) and conjunction fallacy (Tversky & Kahneman, 1983; Nisbett & Ross, 1980).

The College Scenario (Lehman, Lempert, & Nisbett, 1988; Nisbett et al., 1983; Stanovich & West, 1998) tests the use of statistical heuristics in everyday inductive reasoning. Nisbett and colleagues define heuristics as “judgmental tools that are rough intuitive equivalents of statistical principles” (p. 339), in this case implemented to test participants' recognition of the relevance of the law of large numbers in their reasoning about events, especially concerning personal preferences. The scenario describes a student having to make a choice between two colleges with two sets of experience reports, one for each college. He briefly visited one of the colleges himself, while the other one is described to him by several of his friends attending the college. In essence, the participants are presented with opinions from two sources: one provides a one-time only, brief and superficial interaction with the object - the student's own experience - while the other describes deeper, more extensive contact and a larger sample size - several friends' opinions. Test-takers have to recognize the relevance of the sample size and put less emphasis on the student's personal experience.

A high school student had to choose between two colleges A and B. The student had several friends, who were similar to himself in values and abilities, at each college. All of his friends at college A liked it on both educational and social grounds; all of them at college B had problems on both grounds. The student visited both colleges for a day, and his impressions were the opposite of their reports.

Which college should the high school student choose?

- College A
- College B

When employing the scenario, Nisbett and colleagues presented half of their 157 participants to a scenario similar to the one implemented in *Mixed Measures*, and the other half to the same scenario with additional help: emphasis on the possibilities for error in the scenario. 74% of students chose the wrong answer of the group excluding the extra help, whereas error rates dropped to 56% for the scenario including additional help. These findings demonstrate that participants realize the importance of sample size once prompted to do so, but are otherwise easily drawn to follow their own, faulty judgment (Nisbett et al., 1983). *Mixed Measures* implements the scenario without additional aid. This creates an adequate level of difficulty and tests participants on their own realization for the need for a sufficient sample size. Fong, Krantz and Nisbett (1986) implemented a similar scenario on the topic of purchasing a car rather than choosing a college. Similar to the example implemented in *Mixed Measures*, subjects had to come to a conclusion about a decision based on objective input rather than subjective experience. Fong and colleagues investigated the effect of instruction on reasoning skills of their participants. This included formal training and guided induction, such as the teaching of a rule by means of examples. They tested an overall of 540 subjects in two studies and found both forms of training to have a positive effect on performance by improving test-takers' statistical reasoning. More recently Toplak, Liu, MacPherson, Toneatto, and Stanovich (2007) used the scenario in a study of 107 male participants. A probabilistic thinking composite of ten tasks was constructed, including the above-mentioned car-purchase-scenario. All tasks focussed on heuristics and biases literature and aimed to test cognitive errors that might lead to persistent gambling. *Mixed Measures* implements the *College Scenario* to represent and assess participants' ability to gather information in everyday life situations, and weight the quality of different sources as well as to test the awareness of the importance of sample sizes.

Lehman, Lempert and Nisbett (1988) published the *Police Chief Dilemma*, used as an example of methodological reasoning in everyday life. *The Police Chief Dilemma* is one of the everyday-life content methodological problems introduced by the authors, and it focuses on the realization of the need for a control group in methodological problems. The scenario depicts a situation describing a hypothesis. The work quality of

a police chief is evaluated by the reduction of crime rates, without the necessary consideration of a control group - in this case, represented by reductions of crime rates in neighbouring cities. Participants have to detect the weakness of the scenario by choosing one of four possible critical points within the hypothesis.

A city called Kingston has had an unpopular police chief for a year and a half. He is politically active and a colleague of the mayor, and he had little previous experience in police administration when he was appointed. The mayor has recently defended the chief in public, announcing that in the time since he took office, crime rates decreased by 12%. Which of the following pieces of evidence would most weaken the mayor's claim that his chief is competent?

- *The crime rates of the two cities closest to Kingston in location and size have decreased by 18% in the same period.*
- *An independent survey of the citizens of Kingston shows that 40% more crime is reported by respondents in the survey than is reported in police records.*
- *Common sense indicates that there is little a police chief can do to lower crime rates. These are for the most part due to social and economic conditions beyond the control of officials.*
- *The police chief has been discovered to have personal contacts with people who are known to be involved in organized crime.*

Five-hundred and fifty-three students were tested overall, of which two-hundred and six took part in a longitudinal study, a two-year follow up. Researchers investigated the effect of graduate training in different domains, law, medicine, psychology and chemistry, on reasoning skills of students and their ability to reason about problems using particular rule systems, and discovered a slightly bigger improvement in scores in psychology and medicine courses over law and chemistry (Lehman, Lempert, & Nisbett; 1988). Graduate education of a specific sort, especially the kind implemented in psychological and medical courses, seems to have a training effect on methodological reasoning skills and helps to improve those significantly within the duration of the degree (Lehman, Lempert, & Nisbett, 1988). *Mixed Measures* uses the *Police Chief Dilemma* to assess participants' realization of the need for a control condition, which is an important aspect within hypothesis testing skills.

Mynatt's *Unknown Car-Task* tests another important factor: inference problems (Mynatt, Doherty, & Dragan, 1993). It assesses hypothesis testing with so-called inference tasks. Inferences describe the varying likelihood of a hypothesis to be true within real-world problems, especially problems where evidence supports an alternative thesis. Participants have to test a hypothesis by determining its relative likelihood. In this particular example, test-takers are presented with two pieces of information about two car brands, with two cars of one brand each, and have to decide which additional piece of information out of four options would be the most useful to them to determine which of the cars belongs to which brand.

Consider a decision about whether a car is of the brand X or of the brand Y. Assume that two types of information are potentially available about each alternative—the percentage of car X and car Y that have a consumption of more than 10 l/100km (liters per 100 km) and the percentage of car X and car Y that have had no major mechanical problems in the first two years of ownership.

Assume that an unknown car (either of the brand X or Y) does over 10 l/100 km and that it has had no major mechanical problems in the first two years of ownership. Also assume that 65% of car brand X consume more than 10 l/100 km.

If you could find out one of the three following types of information, which one would help you to decide which one the unknown car is, car X or car Y?

- Percentage of cars of brand Y that get over 10 l/100 km
- Percentage of cars of brand X that have had no major mechanical problems in the first two years
- Percentage of cars of brand Y that have had no major mechanical problems in the first two years

When testing hypotheses, individuals prefer to gather information supporting the alternative they already gathered information about rather than the one they have less information about (Doherty, Mynatt, Tweney, & Schiavo, 1979 as cited in Mynatt, Doherty, & Dragan, 1993). This leads to individuals usually testing the hypothesis they believe to be true rather than the alternative hypothesis. Overall, participants only seem to select information based on one alternative. Mynatt, Doherty, and Dragan (1993) tested the performance of one-hundred and seven participants, of which half were presented with the scenario implemented in *Mixed Measures*, representing an “information selection task”, or inference task. The authors acknowledge the reason for the restricted gathering of information in the capacity of working memory, and argue that the maximum number of maintained and computed hypotheses is one. *Mixed Measures* implements the *Unknown Car-Task* to assess participants’ awareness of the importance of gathering sufficient information about all alternatives for an informed decision about a hypothesis.

Beyth-Marom and Fischhoff’s (1983) example of *Party Guests* extends the assessment of the gathering of relevant information to test a hypothesis. In this assessment, subjects are asked to intuitively evaluate a set of questions in terms of their relevance to a hypothesis. Participants are given a piece of information about a person and an array of relevant questions to help them determine that person’s occupation. Out of the array of four possible questions, the one question irrelevant to answering the occupation has to be identified by the test-takers.

Mr. Maxwell attended a party to which only university professors and business executives were invited (whereas guests would be either or, not both of the positions combined). The only thing you know about Mr. Maxwell is that he is a member of the Bear's Club. You are asked to judge the probability that Mr. Maxwell is a university professor by asking 1, 2, 3, or 4 of the questions given below. You are asked to check their relevancy. A relevant question is one the answer to which will help you in your judgement. Please consider each of the questions separately and mark the question that is most relevant for your task.

- What percentage of the people at the party are university professors?
- What percentage of the Bear's Club members are at the party?
- What percentage of the university professors at the party are members of the Bear's Club?
- What percentage of the business executives at the party are members of the Bear's Club?

Two-hundred and seventy-one students were presented with three adjusted versions of the selection and probability assessment task (as utilised by *Mixed Measures*) in three studies. Overall results of the studies show that participants were much more efficient at just using relevant information rather than explaining the necessity of a piece of information for the decision making process, or at seeking it out, as asked to do in this scenario (Beyth-Marom & Fischhoff, 1983). The same adjusted version was implemented by Stanovich and West (1998), who tested nine-hundred subjects overall and their performances on deductive and inductive reasoning. They showed that participants who were successful at gathering relevant information scored higher in assessments of cognitive ability and were significantly better at reasoning, emphasizing the importance of the inclusion of this task in *Mixed Measures*. A previously mentioned study linking cognitive failure with gambling habits (Toplak et al., 2007) reused another adapted version of the *Party Guests Scenario*, focussing more on pure statistical probability and assessing participants' understanding of it. The authors conclude that the domain of probabilistic reasoning is remediable through teaching, if erroneous in gamblers' behaviour. When re-implementing the scenario, Toplak, West, and Stanovich (2011) tested three-hundred and forty-six students on the task and found that only 22.3% of participants came to the right conclusion, which proves its suitability for the assessment due to sufficient difficulty levels. An essential part of hypothesis testing consists of gathering sufficient, high-quality information about options, which is why this scenario was added to *Mixed Measures*.

Conjunction fallacies are covered by Kahneman and Tversky's (1982) *Linda-problem*, introducing the representativeness heuristic and the conjunction effect, which aim to explore the nature of error in decision-making (Kahneman and Tversky, 1982). Scenarios make use of the conjunction rule, which describes the fact that the probability of a joint conjunction of two events cannot exceed the likelihood of either of the two events individually (Kahneman & Tversky, 1982; Tversky & Kahneman, 1983; Reyna, 1991). To induce this effect, participants are presented with a personality

description hinting towards certain character traits and hobbies of a fictional character. Participants are asked to decide which hypothesis is the more likely case about the character when guessing: one involving a job title of that character which seems quite unrepresentative for the description, the other one listing that job title plus a hobby which appears representative for the description of the person (Tversky & Kahneman, 1983).

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Which of two alternatives is more probable:

- *Linda is a bank teller*
- *Linda is a bank teller and is active in the feminist movement*

Kahneman and Tversky (1982) report that in a large sample of undergraduate students, 86% were misled and ignored the conjunction rule, whereas only 50% of graduate students made this mistake. This clearly shows the influence of statistical training and college education on successful hypothesis testing (Kahneman & Tversky, 1982). In 1996, Kahneman and Tversky (1996) adjusted the Linda problem slightly and tested 69 participants on the alterations. The outcomes show that the conjunction rule is applied in direct comparisons and that test-takers use representativeness to estimate probability. Toplak, West, and Stanovich (2011) reused the scenario and only 19.1% of 346 student participants reached the correct conclusion, which proves the task suitable for the compiled assessment, *Mixed Measures*, entailing a sufficient difficulty level. The inclusion of conjunction fallacies aids the assessment of participants' knowledge of this principle of probability theory.

Hypothesis testing is round up with the *New-Car Scenario* (Nisbett, Borgida, Crandall, & Reed, 1976, as cited in Nisbett & Ross, 1980; Stanovich & West, 1998) where participants have to weigh options and choose statistical knowledge over personal experience. Test-takers have to decide whether to put more emphasis on a statistical report or on personal experience in terms of the purchase of a new car. Nisbett and colleagues assess misapplication of schemas and intuition, which may lead to an incorrect choice.

You wish to buy a new car. Today you must choose between two alternatives: to purchase either a car from company A or B. You use only one criterion for that choice, the car's life expectancy. You have information from Consumer Reports that in a sample of several hundred cars the car A has the better record. Yesterday a neighbour told you that his new car of brand A broke down. Which car do you buy?

- *Car A*
- *Car B*

Critical and logical thinking can prevent participants from making inferential errors by ignoring statistical considerations, mainly sample size and sample bias (Nisbett & Ross, 1980). The authors label those errors "cognitive failings" rather than motivational

errors. The origin of the misjudgement is described to be cognitive and subconscious, rather than conscious (Nisbett & Ross, 1980). Fong, Krantz, and Nisbett (1986) used the *New Car Scenario* as one of their examples of teaching their subjects about the law of large numbers. Instruction increases both frequency and quality of statistical reasoning in various day-to-day scenarios and the researchers state their belief of an “abstract inferential rule system” which intuitively helps people to apply the law of large numbers correctly (Fong, Krantz, & Nisbett, 1986). Toplak, West and Stanovich (2011) resumed research on the *New Car Scenario*. They implemented it together with a range of other tasks in previously mentioned studies, assessing over nine-hundred participants in total. The *New Car Scenario* represents one of the heuristics and biases tasks, which contains a special focus on the causal base rate. The base rate principle describes statistical reasoning while taking into account provided knowledge; wrongly supporting implicit beliefs of people (Cao & Banaji, 2016). *Mixed Measures* implements the *New Car Scenario* to assess participants’ judgment of information sources as well as the law of large numbers.

2.2. Verbal Reasoning

Mixed Measures implements verbal reasoning using a selection of eighteen of Dermott’s (2002) 501-word analogy questions. They depict similar but modernized word pairs to Sternberg’s (1977) verbal analogies. Sternberg (1977) explored analogical reasoning in terms of information processes underlying cognition. Test-takers are asked to identify relationships between two sets of two words each. One of the four words is erased, and participants are presented with a range of four alternatives to choose the correct word to complete the sets (Dermott, 2002). An example of this would be “puppy : dog :: kitten : _____”, read as “puppy relates to dog as kitten relates to blank.” The answer in this instance is “cat” which would be presented underneath the analogy together with three related but incorrect words.

_____ : horse :: board : train

- stable
- shoe
- ride
- mount

A clear understanding of the relationship between the pairs of words is necessary to select the correct choice to complete the analogy, which describes the core of verbal reasoning as defined by Halpern (1998, 2002). Deduction and drawing of logical conclusions aid the analysis of the relationship and help participants to detect the missing word to correctly complete the analogy. Relationships found in analogies can be grouped into several subcategories. Instances are relations of parts of an object to

the whole object (e.g. spoke : wheel; with a spoke being part of a wheel), type and category (e.g. orange : citrus; with an orange being a type of citrus) or degree of intensity (shower : monsoon; with a shower being a lighter version of rain and a monsoon being heavy rain). Analogies rely heavily on language comprehension and precision, which lead to certain difficulties in the translation process. Generalizability was aimed to be achieved throughout the selection process of the 18 analogies implemented in *Mixed Measures*, but certain samples had to be replaced in some languages, due to missing concepts or translations in those languages (e.g. the absence of a specific word in German for kitten which does not include the word cat, which led to the replacement of the word pair cat-kitten to dog-puppy). Word pairs falling under one of the categories, relationships in symbols and representations, were avoided in order to keep cultural backgrounds and concepts to a minimum. Factual content questions were also excluded, to avoid specific background knowledge to be required, making it a covariate. Word pairs were carefully chosen for each version, A and B, with equal levels of difficulty for both, and varying difficulty levels throughout the nine pairs. Matsuoka and Lepage (2014) implemented 400 of Dermott's (2002) word analogies, following similar rules as used for *Mixed Measures* and excluding word pairs requiring domain-specific knowledge. The analogies were used to test a computational model measuring similarities between word pairs. Word analogies are presently and have been deployed widely for the assessment of verbal skills, for instance in the American university SAT entrance exam (Turney, 2008) and the Air Force Officer Qualifying Test (AFOQT; Berger, Gupta, & Berger, 1990). Measuring verbal ability, the appropriate use of vocabulary for certain situations as well as inferring relationships between words are the central aspect of word analogies (Berger, Gupta, & Berger, 1990). Consequently, analogies are considered suitable for *Mixed Measures* in the subcategory verbal reasoning.

2.3. Judging Likelihood and Probability

Assessing the ability to judge the likelihood and probability of certain outcomes is the overall goal of JL as a subcategory. Implemented in *Mixed Measures* are five different questions per version A and B, focussing on the assessment of participants' judging-likelihood skills. Following Halpern's definition of JL (1998; 2002), several concepts are of importance. These include the understanding of a sufficient sample size (Kahneman & Tversky, 1982; Toplak, West, & Stanovich, 2011); test-takers' awareness of the regression to the mean (Lehman, Lempert, & Nisbett, 1988), and describing the tendency that an extreme measurement or score will be followed by a score closer to the average at the second time of testing; the avoidance of the gambler's fallacy

(Toplak, West, & Stanovich, 2007; Toplak et al., 2007) - the cognitive error that the frequent occurrence of something will make it less likely to happen in the future or vice versa - and probabilistic reasoning (Kirkpatrick & Epstein, 1992; Toplak et al., 2007) - the estimation of the probabilities of outcomes.

Kahneman and Tversky (1982) designed the *Squash Problem*, focussing on the understanding of sample sizes and estimations relating to the real mean. The scenario describes two squash players of different skill levels, and questions whether the better player will be more likely to win in a shorter or a longer game. This question requires the comprehension of the effect of sample size and sampling errors.

A game of squash can be played to either 9 or 15 points. If A is a better player than B, which scoring scheme would give player A a better chance of winning?

- 9 point game
- 15 point game

Hasty decision-making often leads participants to disregard the length of the game as an example of sample size, as Kahneman and Tversky (1982) show. Even though most of their subjects reported an understanding of statistics, the researchers had to prompt participants after a wrong answer before they realized the correct solution. Toplak, West, and Stanovich (2011) revisited the *Squash Problem* and implemented it in a study on heuristics and biases, assessing whether people follow baseline rules of rational thought. A total of 346 students took the test and results showed a high difficulty level with only 15.6% of participants answering the test correctly. The authors explain the assessment to be testing the understanding of large sample sizes and the increasing likelihood of an expected outcome the higher the sample size; in this case, the points played (Toplak, West, & Stanovich; 2011) makes the scenario valid for *Mixed Measures*.

Lehman and colleagues' *Baseball Scenario* (Lehman, Lempert, & Nisbett, 1988) explores participants realization of the regression to the mean explained in the following. The scenario assesses statistical reasoning in everyday life situations. Easily confused with the law of averages/large numbers (as described in 1.6.3), regression to the mean implements the statistical rule that if a good team performs well above the average, it will not continue to perform that highly. It will regress to the mean (even though not fully) as the law of large numbers would predict, but remain slightly above average, based on the skills and talent which made the team perform better in the first place (Lee & Smith, 2002). Extreme values are less likely to be followed by another extreme score when the same sample is retested (Lehman, Lempert, & Nisbett, 1988). The scenario describes a baseball high score board. The question asked is why the board has a much higher average for players after two weeks of the season than at the

end of the season. Participants are expected to recognize the statistical rule and prove their awareness of the regression to the mean to answer the scenario correctly.

After the first 2 weeks of the major league baseball season, newspapers begin to print the top 10 batting averages. Typically, after 2 weeks, the leading batter often has an average of about .450. However, no batter in major league history has ever averaged .450 at the end of the season. Why do you think this is? Tick one:

- *When a batter is known to be hitting for a high average, pitchers bear down more when they pitch to him.*
- *Pitchers tend to get better over the course of a season, as they get more in shape. As pitchers improve, they are more likely to strike out batters, so batters' averages go down.*
- *A player's high average at the beginning of the season may be just luck. The longer season provides a more realistic test of a batter's skill.*
- *A batter who has such a hot streak at the beginning of the season is under a lot of stress to maintain his performance record. Such stress negatively affects his playing.*
- *When a batter is known to be hitting for a high average, he stops getting good pitches to hit. Instead, pitchers "play the corners" of the plate because they don't mind walking him.*

Toplak et al. (2007) implemented the scenario as part of a taxonomy categorizing cognitive errors possibly resulting in gambling behaviour. Moreover, Toplak, West and Stanovich (2011) applied the same scenario to assess test-takers sensitivity to the regression to the mean, which depicts an important aspect of judging likelihood skills as by Halpern's definition (2002) implemented by *Mixed Measures*.

The gambler's fallacy is considered another important aspect of judging likelihood and is represented with two questions in the assessment. It describes the expectation that several negative outcomes will be less likely to be followed by another negative one, or vice versa (Ayton & Fischer, 2004). Negative recency creates a psychologically perceived falsification despite the same statistical odds for either outcome (Ayton & Fischer, 2004). Believing in a negative correlation of a random sequence (Croson & Sundali, 2005) - for instance when a coin landed on heads for a few times - the gambler's fallacy would predict a tail to be "due" for the next throw, even though the likelihood is 50% for either of the two outcomes (Croson & Sundali, 2005). The two fallacies used in *Mixed Measures* are *Slot Machines* and *Tossing a Coin*, both prior implemented by Toplak and colleagues (Toplak, West, & Stanovich, 2011; Toplak et al., 2007). *Slot Machines* describes a scenario of three wins at a slot machine with the winning likelihood of one in ten, and participants have to estimate the likelihood of the next play (Toplak, West, & Stanovich, 2011; Toplak et al., 2007).

*When playing slot machines, people win something about 1 in every 10 times. Julie, however, has just won on her first three plays. What are her chances of winning the next time she plays?
_____ out of 10*

Tossing a Coin consists of an experiment where test-takers have to guess the likelihood of the next flip of a coin after five throws resulted in heads to be on top (Toplak, West, & Stanovich, 2011; Toplak et al., 2007).

Imagine that we are tossing a fair coin (a coin that has a 50/50 chance of coming up heads or tails) and it has just come up heads 5 times in a row. For the 6th toss do you think that:

- *It is more likely that tails will come up than heads.*
- *It is more likely that heads will come up than tails.*
- *Heads and tails are equally probable on the sixth toss.*

Both scenarios test participants' awareness of the gambler's fallacy and their ability to judge the likelihood of unknown situations, demanding the correct implementation of principles of probability to reach the solution.

The last question of the subcategory Judging Likelihood and Uncertainty which covers probabilistic reasoning is the *Marble Game* (Kirkpatrick & Epstein, 1992; Toplak et al., 2007). The scenario describes a situation in which participants are presented with two trays with black and white marbles, each with a different ratio to one another. Participants are promised a prize if they blindly draw a marble of one specific colour, and have to decide from which tray to pick for a higher chance of winning. Even when participants are aware of the higher percentage of winning marbles on one tray, they often decide irrationally against this tray. They choose the tray with a higher actual number of winning marbles, which is the tray with a lower percentage of winning marbles overall, thinking their chances might increase (Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992).

Assume that you are presented with two trays of black and white marbles: a large tray that contains 100 marbles and a small tray that contains 10 marbles. You must draw out one marble from either tray without looking. If you draw a black marble, you win \$2. Consider a condition in which the small tray contains 1 black marble and 9 white marbles, and the large tray contains 8 black marbles and 92 white marbles. From which tray would you prefer to select a marble in a real situation?

- *Large tray*
- *Small tray*

Kirkpatrick and Epstein (1992) conducted three studies with an overall sample of 1331 participants implementing the marble game in order to examine irrational decision making in humans. They concluded that subjective probability influences suspiciousness about coincidences (Kirkpatrick & Epstein, 1992). As a possible explanation for irrational decision-making, the authors mention the limited capacity of humans to process information, which leads to the use of cognitive shortcuts to solve everyday problems (so-called heuristics). Decision-making studies based in the laboratory correlate with heuristic errors and gambling behaviour in real life (Denes-Raj & Epstein, 1994), which makes them a good indicator for heuristic processing and justifies their implementation in *Mixed Measures*.

2.4. Argumentation Analysis

Mixed Measures utilizes syllogistic reasoning to assess argumentation analysis ability. Assessed are response bias, belief bias, and the acceptance of both response and belief bias. Klauer, Musch, and Naumer (2000) published an array of valid and invalid syllogisms with believable, neutral, and unbelievable conclusions, based on reputable templates of syllogisms (Newstead, Pollard, Evans, & Allen, 1992; Gilinsky & Judd, 1994). Klauer, Musch, and Naumer (2000) constructed statements with varying content for each category, believable-valid, believable-invalid, unbelievable-valid and unbelievable-invalid. Believable statements seemed correct and plausible without further reflection, whereas unbelievable statements seemed farfetched and unlikely at first glance (Andrews, 2010; Gilhooly, Logie, & Wynn, 1999; Gilhooly & Murphy, 2005; Stanovich & West, 1998). This does not determine the correctness of a statement. Deeper thought and insight into the problem as such is necessary to encode the problems' validity correctly, which is essential in the implemented definition of the subsection argumentation analysis. It has been shown that test-takers are more likely to accept congruent conclusions (believable-valid, unbelievable-invalid) which describe the core of the belief bias (Klauer, Musch, & Naumer, 2000), with incongruent samples testing participants' argumentation analysis skills. All statements follow the same construct, starting with two premises followed by a conclusion, built with three terms A, B, C. The middle term is present in both statements: "*All A are B. All B are C. Therefore, all A are C.*" Qualifiers vary between *all, no, some* and *some..., not* (Klauer, Musch, & Naumer, 2000).

An example of an unbelievable, valid syllogism is the following:

- a) No fish swim in the river.
- b) Some trout swim in the river.
- c) Therefore, some trout are not fish.

Statement a) and b) are to be taken as a valid baseline for the scenario, and participants are asked to indicate whether statement c) is a valid or invalid conclusion under those circumstances. Statements do not always follow basic instincts and general knowledge, and assess a test-taker's conclusion-acceptance in different biases (Klauer, Musch, & Naumer, 2000; Reyna, 1991).

- *No cigarettes are inexpensive.*
- *Some addictive things are inexpensive.*
- *Therefore, some addictive things are not cigarettes.*

Valid – Invalid statement

Toplak and colleagues (Toplak, West, & Stanovich, 2011) applied analogies to assess rational thinking within a syllogistic reasoning task. Participants' ability to judge the quality of an argument, and their ability to differentiate between a statement and a

conclusion, are the core skills of argumentation analysis represented in *Mixed Measures*. Nine syllogisms are implemented per version, including a mixture of valid believable, invalid believable, valid unbelievable, and invalid unbelievable examples.

2.5. Problem Solving

Problem Solving is, by definition, a relatively broad section of critical thinking and is mainly distinguished by the ability to select the right option from a range of choices (see 1.6.5 for details). Scenarios implemented in *Mixed Measures* include several published measures. Included are the three questions of the Cognitive Reflection Test (CRT; Frederick, 2005), for which the initial response tendency is usually wrong, demanding further reflection to arrive at the correct solution. Moreover, *Mixed Measures* includes three insight problems requiring selective encoding due to the solution not being clear at the first glance (Maier & Solem, 1952; Posner, 1973; Sternberg, 1986). Implemented are also a measurement assessing general mathematic problem solving without the necessity for any calculation (Posner, 1973); a non-insight problem where logical thinking leads to the correct solution (Schooler, Ohlsson, & Brooks, 1993); and a conditional reasoning task, Wason's Card Sorting Test (Wason, 1968), where statements have to be tested by determining the correct exclusion criteria.

Frederick's (2005) *Cognitive Reflection Test* including all three questions marks the first set. Questions are designed in a way, in which following one's intuition will not lead to a correct response to the question, "impulsive" answers are wrong. Reflection upon the problem and its solution are necessary to arrive at the correct response. This applies to the *Bat and Ball Question*, which focusses on a monetary problem-solving task. Participants are told the cumulative price for a bat and have to figure out individual prices.

A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

- 5 cents
- 10 cents
- 55 cents
- 100 cents

It also applies to the *Five Machines Question*; implementing an easy version of the rule of three by describing the time it takes a certain amount of machines, to produce a number of widgets, and participants are asked to figure out how long it would take if the number of machines would change.

If it takes 5 machines 5 minutes to make 5 plastic parts, how long would it take 100 machines to make 100 plastic parts?

- 1 minute
- 5 minutes
- 25 minutes
- 100 minutes

Lastly, it applies to the *Lake with Lily Pads*, which is best solved not with calculations but with logical thinking, describing a scenario where a lake is gradually being covered by lily pads and test-takers have to figure out the amount of time it takes for half of the lake to be covered (Sternberg, 1986; Metcalfe & Wiebe, 1987).

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 24 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

- 2 days
- 4 days
- 12 days
- 23 days

Frederick (2005) assessed 3428 participants in 35 separate studies on the CRT and aimed to test cognitive reflection - defined as the ability to not only resist the first response that comes to mind, but to undergo further thoughts which lead to the correct answer (Frederick, 2005). Out of the three questions, the average amount of questions answered correctly was 1.24, which showcases the difficulty level of the set. Toplak, West, and Stanovich (2011) describe the CRT as similar to the prior problem-solving scenarios implemented by Gilhooly and Murphy (2005) but with an additional level of difficulty: an extra step in the problem-solving process, triggered by the realization of the incorrect initial response. Correctly responding to the CRT seems to reinforce the “value of reflection” (Paxton et al., 2012). The adjustment of intuitive responses and the realisation that initial choices might be wrong is considered essential for good problem-solving skills. Prior literature, for example Nagin and Pogarsky (2003), implemented the *Bat and Ball Question* as a measure for problem-solving ability. Within *Mixed Measures*, version B of the *Lily Pad Lake Question* is Sternberg and Davidsons’ (1983) *Murples Problem*, which describes the same scenario with different subjects: a container being filled with rocks in a certain amount of days (Gilhooly & Murphy, 2005).

Insight problems are problems that require a certain mental restructuring of a scenario before being able to reach the solution (Gilhooly & Murphy, 2005). The solution is not obvious at a first glance, but the problem solver has to approach different solutions before the scenario is solved successfully (Knoblich, Ohlsson, Haider, & Rhenius, 1999). The difficulty level of insight problems stems from the activation of incorrect paths during the problem-solving process (Knoblich et al., 1999). Cognitive flexibility, as for instance tested through finding alternative uses for an object, is an important skill for solving those insight problems in comparison to non-insight

problems, and strategic switching and inhibition skills can predict insight-problem task performance (Gilhooly & Murphy, 2005). The *Matching Socks* question, taken from Sternberg (1986), is an insight problem, which requires selective encoding. Pictured in this scenario is a drawer with two kinds of loose socks, and participants are asked to report how many socks must be drawn in a blindfolded manner before a matching pair is complete. Test-takers should realize the irrelevance of the ratio of different socks in the drawer for the completion of the task (Sternberg, 1986), and again, rather than by calculations, reach the correct solution through a logical problem-solving process.

There are black and brown loose socks in a drawer mixed in a ratio of 4 to 5. How many separate socks would you have to take out without looking to be sure of getting a pair of the same colour?

- 1 sock
- 2 socks
- 3 socks
- 4 socks
- 5 socks

Gilhooly and Murphy (2005), Gilhooly and Fioratou (2009), Metcalfe & Wiebe (1987) and more implemented the task to assess participants' ability to solve problems correctly. The *Matching Socks* question is considered a classic example within insight problems in the literature (Webb, Little, Cropper, & Roze, 2017).

Posner's (1973) *Trains and Bird Problem* assesses general mathematic understanding of speed measurement, which is necessary to solve the scenario. Nevertheless, despite the first impression, no mathematical calculations are needed. The scenario describes two trains driving towards each other for one hour at a certain speed from a certain distance, including a bird, which keeps flying back and forth between them at a reported speed. Participants are asked to state which distance the bird will have flown in total, once the two trains meet (Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005; Posner, 1973).

Two trains, 50 km apart start towards each other at 25 km/h (km per hour) each. As the trains start a bird flies from the front of one train towards the second. On reaching the second train the bird turns round and flies back to the first train and so on until the trains meet. If the bird flies at 60 km/h how many km will the bird have flown before the trains meet?

- 15 km
- 30 km
- 45 km
- 60 km

When attempting to solve the problem, facts like the speed of the trains driving towards each other and the bird turning around regularly have to be ignored and emphasis must only be put on the duration of the bird flying and the speed at which it flies. This creative problem-solving scenario falls, similar to the *Matching Socks* scenario described above, under the category of insight problems (Lin & Lien, 2013).

The last insight problem implemented in *Mixed Measures* is the *Horse Trading Problem* (Maier & Burke, 1967; Metcalfe, 1986; Maier & Solem, 1952; Gilhooly & Murphy, 2005; Gilhooly & Fioratou, 2009; Webb, Little, & Cropper, 2017). The scenario describes a trading process of three sales at different prices, and the test taker is asked to report the overall profit made by the trading person. The solution process includes simple calculations but minimizes necessary creative input to come to the correct solution (Maier & Burke, 1967). An obvious but incorrect solution raises the difficulty level of this scenario and leads to a high number of participants failing to answer correctly (Maier & Burke, 1967).

A man buys a horse for £60, sells it for £70, buys it back for £80 and sells it finally for £90. How much has he made?

- £10
- £20
- £30
- £40

Participants were also presented with a non-insight problem called *Crime* (Schooler et al., 1993). Non-insight problems can be solved in a logical manner; they are straightforward and include no distractors in formulation and solution process. (Schooler, Ohlsson, & Brooks, 1993). The scenario describes four contradicting witness statements, and participants have to decide which one of the statements is correct, implementing logical problem-solving processes.

The police were convinced that either A, B, C, or D had committed a crime. Each of the suspects, in turn, made a statement, but only one of the four statements was true.

- A said, "I didn't do it."
- B said, "A is lying."
- C said, "B is lying."
- D said, "B did it."

Who is telling the truth? Who committed the crime?

Truth: A – B – C – D

Crime: A – B – C – D

Non-insight problems put a heavier load on working memory during the solving process (Gilhooly & Murphy, 2005). The solution strategy can be found relatively quickly, and is utilized in an incremental path until the actual solution has been reached (Gilhooly & Murphy, 2005).

The last task in the range of problem-solving scenarios is Wason's (1968) *Card Sorting Test*, a conditional reasoning task (Wason, 1968). This test asks participants to evaluate and use exclusion criteria to determine the correctness of a statement. Test-takers are presented with four cards of different prints on each side and have to verify or falsify a statement about the cards. To do so, participants can decide to turn over cards to test the statement but are asked to do so in a limited manner. As few cards as possible should be turned, and only those which are necessary to turn for

verification or falsification of the statement. Formal logic will help participants to avoid turning cards which don't add important knowledge for judging if the given rule is violated (Cosmides, 1989).

You are shown a set of four cards placed on a table, each of which has a number on one side and a colour on the other side. The visible faces of the cards show 3, 8, red and brown. Which card(s) do you have to turn over in order to test the truth of the proposition that if a card shows an even number on one face, then its opposite face is red? Try to work in the most efficient way by turning around the least amount of cards necessary.

Mark as many as apply: 3 – 8 – red – brown

Sperber and colleagues (Sperber, Cara, & Girotto, 1995) examined tasks of the same logic, called selection tasks, and showcase subjects' incomplete processing and drawing of restricted inferences. This is based on the accessibility of inferences and participants' interpretation of the scenario, which leads to an average error rate of up to 90% for tasks like the Four Card Selection Task (Sperber, Cara, & Girotto, 1995) and makes it a suitable assessment for problem-solving skills within *Mixed Measures*.

3. Comparison across Languages

3.1. Methods

3.1.1. General method

3.1.1.1. Participants

A total of 737 participants fully completed the assessment. These were divided into nine language sets: English-Irish, German-English, Italian-English, Russian-English, Japanese-English, Turkish-English, Turkish-German, Chinese-English, and Chinese-Japanese, with the first language mentioned representing the native language. Three-hundred and forty-five (46.8%) of those participants were grouped as monolinguals and 392 (52.1%) as bilingual. An a priori power analysis using G*Power (Erdfelder, Faul, & Buchner, 1996) indicated that to conduct a MANOVA, 24 subjects are required in each of the two groups, monolinguals and bilinguals, with 80 percent power for detecting a large sized effect ($d=.8$), when employing the traditional .05 criterion of statistical significance. This also covers minimum required sample size of ten for within-group comparisons via reMANOVA at the same power, for the assessment of first and second language within participants.

On average, participants were 27 years old at the time of testing ($M=26.99$; $SE=.37$), with a slight discrepancy between monolinguals ($M=24.04$; $SE=.42$ years) and bilinguals ($M=29.57$; $SE=.57$ years), who were on average five and a half years older. Within each individual language set, differences between mono- and bilingual participants are less prominent, with differences observable across languages (with age variations between language sets). The difference within the merged data set stems from varying participation numbers within each language set (see Table 3.1 for details) and thus does not affect calculations within each language. The majority of test takers were female; 71.5% ($n=528$), with 75.4% within monolinguals and 68.1% for bilinguals. 28.3% ($n=209$) were male. One person did not identify with either gender. A detailed overview divided by the nine language sets is presented in Table 3.1.

Table 3.1. Number of Participants, Age, and Gender divided by Language Group and Mono- and Bilinguals

Language	N		Age: M (SE)		Female (N; %)		Male (N; %)	
	Mono	Bi	Mono	Bi	Mono	Bi	Mono	Bi
English-Irish	55	59	26.6 (09.7)	26.5 (.89)	39; 70.9%	36; 61.0%	16; 29.1%	22; 37.3%
German-English	20	97	27.9 (08.2)	31.2 (1.20)	18; 90.0%	70; 72.2%	02; 10.0%	27; 27.8%
Italian-English	23	63	25.0 (07.3)	27.3 (.94)	15; 65.2%	42; 66.7%	08; 34.8%	21; 33.3%
Russian-English	09	48	39.4 (16.0)	36.14 (1.59)	08; 88.9%	35; 72.9%	01; 11.1%	13; 27.1%
Chinese-English	34	30	19.9 (04.4)	21.08 (.64)	27; 79.4%	23; 76.7%	07; 20.6%	07; 23.3%
Chinese-Japanese	32	33	19.1 (01.2)	21.31 (.41)	28; 87.5%	18; 54.5%	04; 12.5%	15; 45.5%
Japanese-English	37	21	30.3 (08.1)	32.69 (1.35)	15; 40.5%	16; 76.2%	22; 59.5%	05; 23.8%
Turkish-English	68	20	21.7 (03.3)	22.60 (.46)	55; 80.9%	11; 55.0%	13; 19.1%	09; 45.0%
Turkish-German	67	21	21.8 (03.3)	25.05 (.90)	55; 82.1%	16; 76.2%	12; 17.9%	05; 23.8%
All merged	345	392	24.0 (.42)	29.6 (.57)	260; 75.4%	267; 68.1%	85; 24.6%	124; 31.6%

Note. Mono=Monolinguals; Bi=Bilinguals. Participants were given the option to refrain from providing any background information.

Most participants reported not to suffer from any impairment which might have influenced the performance of the assessment. Within the reported impairments, the most common one cited was affected vision. No participants reported suffering from a hearing impairment or a learning disability. All visual impairments were reported to be corrected to normal.

Seventy-nine percent of all participants underwent education through state-school secondary education, with a higher percentage reported within bilinguals, 86.2% and a lower amount of 71.9% within monolinguals. 7.6% visited a private school and the remaining 12.9% reported to either have visited a mixture of public and private school, were home-schooled, did not remember, or preferred not to specify their education. On average participants spent a total amount of 15 years ($M=15.15$; $SE=.13$) in education, bilinguals on average half a year more ($M=15.53$; $SE=.19$) than monolinguals ($M=14.72$; $SE=.17$). An overview of educational background divided by language set is presented in Table 3.2.

Table 3.2. Percentage of Participants having visited a State School and Years of Education divided by Language Group and Mono- and Bilinguals

Language	State School			Edu (years)		
	All	Mono	Bi	All (M; SE)	Mono (M; SE)	Bi (M; SE)
English-Irish	91.2%	94.5%	88.1%	16.15 (.37)	16.58 (.50)	15.75 (.55)
German-English	95.7%	100%	94.8%	13.58 (.29)	13.20 (.63)	13.66 (.32)
Italian-English	90.7%	87.0%	92.1%	15.64 (.38)	14.26 (.70)	16.14 (.43)
Russian-English	94.7%	100%	93.8%	17.40 (.53)	16.11 (.63)	17.65 (.62)
Chinese-English	96.9%	97.1%	96.7%	14.25 (.32)	13.44 (.28)	15.17 (.56)
Chinese-Japanese	90.8%	96.9%	84.8%	14.74 (.27)	13.56 (.31)	15.87 (.33)
Japanese-English	72.4%	78.4%	61.9%	17.72 (.37)	17.81 (.42)	17.57 (.71)
Turkish-English	93.3%	94.1%	90.0%	14.50 (.33)	13.92 (.31)	16.30 (.91)
Turkish-German	95.5%	95.5%	95.2%	13.90 (.28)	13.93 (.31)	14.00 (.62)

Note. Mono=Monolinguals; Bi=Bilinguals; Edu=Education.

Forty-eight percent ($n=355$) of participants were enrolled in a Bachelor programme at the time of testing. When comparing monolinguals and bilinguals, 67.5% of monolinguals were still in the process of obtaining their Bachelor's degree, whereas only 31.1% of bilinguals reported being at that stage in their higher education. Overall, 17.8% ($n=131$) have received a Bachelor's degree in the past, with a higher percentage within bilinguals (23.0%) compared to monolinguals (11.9%); 16.4% ($n=121$) obtained a Master's degree, again with a noticeably higher amount within bilinguals (23.5%) compared to monolinguals (11.9%); 4.9% ($n=36$) have a Ph.D., with numbers of monolinguals and bilinguals being similar (4.1%; 5.6% respectively); and 8.5% ($n=63$) reported to have obtained some other form of higher education (monolinguals 5.8%; bilinguals 11%). 4.3% ($n=32$) have not undergone higher education, with a lower percentage of 2.6% within monolinguals and a slightly higher number of 5.9% in bilinguals. In summary, participants within the monolingual group are mainly still enrolled in their Bachelor and find themselves on average at an earlier stage of higher education in comparison to bilinguals, who have overall obtained higher degrees of higher education. An overview of the higher education levels divided by language set is presented in Table 3.3. Differences vary, similarly to the age of participants, across language groups. Within each language set there is less variation, in comparison to the merged data set including all participants, which shows a bigger difference overall.

Table 3.3. Percentage of Participants having obtained a University Degree; divided by Language Group and Mono- and Bilinguals

	Undergraduate (N; %)	Undergraduate Mono	Undergraduate Bi	B degree (N; %)	B Mono	B Bi	M degree (N; %)	M Mono
English Irish	53; 46.5%	30; 54.5%	23; 39%	25; 21.9%	09; 16.4%	16; 27.1%	20; 17.5%	7; 12.7%
German English	32; 27.4%	05; 25.0%	27; 27.8%	14; 12.0%	04; 20.0%	10; 10.3%	15; 12.8%	0
Italian English	27; 31.4%	10; 43.5%	17; 27.0%	23; 26.7%	06; 26.1%	17; 27.0%	15; 17.4%	0
Russian English	07; 12.3%	0	07; 14.6%	14; 24.6%	02; 22.2%	12; 25.0%	23; 40.4%	03; 33.3%
Chinese English	48; 75%	31; 91.2%	17; 56.7%	08; 12.5%	02; 05.9%	06; 20.0%	07; 10.9%	01; 02.9%
Chinese Japanese	40; 61.5%	30; 93.8%	10; 30.3%	16; 24.6%	02; 06.3%	14; 42.4%	08; 12.3%	0
Japanese English	08; 13.8%	04; 10.8%	04; 19.0%	21; 36.2%	14; 37.8%	07; 33.3%	14; 24.1%	09; 24.3%
Turkish English	74; 83.1%	62; 91.2%	12; 60.0%	04; 04.5%	01; 01.5%	03; 15.0%	09; 10.1%	04; 05.9%
Turkish German	66; 75.0%	61; 91.0%	05; 23.8%	06; 06.8%	01; 01.5%	05; 23.8%	10; 11.4%	04; 06.0%

M Bi	PhD (N; %)	PhD Mono	PhD Bi	other (N; %)	other Mono	other Bi	
13; 22.0%	07; 06.1%	05; 09.1%	02; 03.4%	09; 07.9%	04; 07.3%	05; 08.5%	English Irish
15; 15.5%	07; 06.0%	0	07; 7.2%	24; 20.5%	06; 30.0%	18; 18.6%	German English
15; 23.8%	01; 01.2%	0	01; 01.6%	20; 23.3%	07; 30.4%	13; 20.6%	Italian English
20; 41.7%	05; 08.8%	0	05; 10.4%	04; 07.0%	02; 22.2%	02; 04.2%	Russian English
06; 20.0%	01; 01.6%	0	01; 03.3%	0	0	0	Chinese English
08; 24.2%	0	0	0	01; 01.5%	0	01; 03.0%	Chinese Japanese
05; 23.8%	12; 20.7%	09; 24.3%	03; 14.3%	03; 05.2%	01; 02.7%	02; 09.5%	Japanese English
04; 20.0%	01; 01.1%	0	01; 05.0%	0	0	0	Turkish English
06; 28.6%	02; 02.3%	0	02; 09.5%	02; 02.3%	0	02; 09.5%	Turkish German

Note. Mono=Monolinguals; Bi=Bilinguals; B=Bachelor; M=Master; other=other higher education.

40.9% ($n=302$) of participants reported that their parents underwent higher education, and 54.9% ($n=405$) stated they obtained no degree. The remaining participants either did not want to specify or did not know the answer.

Table 3.4. Overview of Participants' Parents' Education; divided by Language Group and Mono- and Bilinguals

	Parents higher edu (N; %)	Parents no higher edu (N; %)	no comment (N; %)	Parents edu Mono (N; %)	Parents no edu Mono (N; %)	Parents n.c. Mono (N; %)	Parents edu Bi (N; %)	Parents no edu Bi (N; %)	Parents n.c. Bi (N; %)
English Irish	55; 48.2%	57; 50.0%	02; 01.8%	27; 49.1%	27; 49.1%	01; 01.8%	28; 47.5%	30; 50.8%	01; 01.7%
German English	40; 34.2%	74; 63.2%	03; 2.6%	05; 25.0%	15; 75.0%	01; 01.8%	35; 36.1%	59; 60.8%	01; 01.7%
Italian English	32; 37.2%	53; 61.6%	01; 1.2%	05; 21.7%	17; 73.9%	01; 04.3%	27; 42.9%	36; 57.1%	0
Russian English	45; 78.9%	12; 21.1%	0	07; 77.8%	02; 22.2%	0	38; 79.2%	10; 20.8%	0
Chinese English	16; 25.0%	45; 70.3%	03; 04.7%	09; 26.5%	24; 70.6%	01; 02.9%	07; 23.3%	21; 70.0%	02; 06.7%
Chinese Japanese	29; 44.6%	35; 53.8%	01; 01.5%	08; 25.0%	23; 71.9%	01; 03.1%	21; 63.6%	12; 36.4%	0
Japanese English	37; 67.2%	18; 31.0%	01; 01.7%	24; 64.9%	13; 35.1%	0	15; 71.4%	05; 23.8%	01; 04.8%
Turkish English	28; 31.5%	60; 67.4%	01; 01.1%	17; 25.0%	50; 73.5%	01; 01.5%	11; 55.0%	09; 45.0%	0
Turkish German	18; 20.5%	69; 78.4%	01; 01.1%	16; 23.9%	50; 74.6%	01; 01.5%	02; 09.5%	19; 90.5%	0

Note. Mono=Monolinguals; Bi=Bilinguals; higher edu=obtained a degree in higher education; no higher edu=did not undergo higher education; n.c.=participant preferred not to reply or did not know the answer.

When comparing monolinguals and bilinguals on these scores, the number of parents who obtained a university degree of some sort was higher within bilinguals (monolinguals' parents: degree 34.2%; no degree: 60.2%; bilinguals' parents: degree: 46.9%; no degree: 50%). Parental education is one of the indicators for socioeconomic status. An overview of the parental higher education divided by language set is presented in Table 3.4.

Another important factor to determine an individual's social economic status is income. To assess income, participants were asked to report which out of four categories they would fall into, adjusted to local currencies in exchange rate and value: no income, low income (which includes an average yearly wage of up to €5000), medium income (participants who reported to earn between €5000 and €20000), and high income (income higher than €20000). The amounts were chosen based on student wages, and participants were given the option to refrain from providing any information regarding their income.

33.5% of all participants indicated that they fall into the low-income category; earning less than €5000 a year (monolinguals 44.4%; bilinguals 23.5%). 25.2% reported to be earning between €5000 and €20000 a year (monolinguals 21.2%; bilinguals 29.6%) and 18% indicated to be earning more than €20000 in 12 months (monolinguals 10%; bilinguals 24.7%). The remaining participants either had no income (12.9% overall; monolinguals 15.8%; bilinguals 10.2%) or preferred not to comment on their income. An overview of participants' income divided by language set is presented in Table 3.5.

Table 3.5. Overview of Participants' Income; by Language Group and Mono-and Bilinguals

	Low income	Medium
English Irish	21; 18.4%	43; 37.7%
German English	24; 20.5%	39; 33.3%
Italian English	13; 15.1%	36; 41.9%
Russian English	07; 12.3%	15; 26.3%
Chinese English	15; 23.4%	06; 09.4%
Chinese Japanese	23; 35.4%	08; 12.4%
Japanese English	05; 08.6%	26; 44.9%
Turkish English	75; 84.3%	05; 05.6%
Turkish German	64; 72.7%	08; 09.1%

	High income	No income	no comment	Mono Low	Mono Medium	Mono High	Mono No	Mono n.c.
English Irish	30; 26.3%	06; 05.3%	14; 12.3%	12; 21.8%	24; 36.4%	12; 21.8%	02; 03.6%	09; 16.4%
German English	38; 32.5%	07; 06.0%	09; 07.7%	01; 05.0%	12; 60.0%	04; 20.0%	02; 10.0%	01; 05.0%
Italian English	20; 23.3%	07; 08.1%	10; 11.6%	03; 13.0%	10; 43.5%	05; 21.7%	02; 08.7%	03; 13.0%
Russian English	09; 15.8%	04; 07.0%	22; 38.6%	02; 22.2%	02; 22.2%	0	02; 22.2%	03; 33.3%
Chinese English	01; 01.6%	36; 56.3%	06; 09.4%	08; 23.5%	02; 05.9%	0	21; 61.8%	03; 08.8%
Chinese Japanese	01; 01.5%	30; 46.2%	03; 04.6%	06; 18.8%	01; 03.1%	0	22; 68.8%	03; 09.4%
Japanese English	20; 34.5%	01; 01.7%	06; 10.3%	04; 10.8%	17; 45.9%	14; 37.8%	0	02; 05.4%
Turkish English	04; 04.5%	02; 02.2%	03; 03.4%	60; 88.2%	03; 04.4%	0	02; 02.9%	03; 04.4%
Turkish German	10; 11.4%	02; 02.3%	04; 04.5%	59; 88.1%	03; 04.5%	0	02; 03.3%	03; 04.5%

	Bi Low	Bi Medium	Bi High	Bi No	Bi n.c.	
	09; 15.3%	23; 39.0%	18; 30.5%	04; 06.8%	05; 08.5%	English Irish
	23; 23.7%	27; 27.8%	34; 35.1%	05; 05.2%	08; 08.2%	German English
	10; 15.9%	26; 41.3%	15; 23.8%	05; 07.9%	07; 11.1%	Italian English
	05; 10.4%	13; 27.1%	09; 18.8%	02; 04.2%	19; 39.6%	Russian English
	07; 23.3%	04; 13.4%	01; 03.3%	15; 50.0%	03; 10.0%	Chinese English
	17; 51.5%	07; 21.2%	01; 03.0%	08; 24.2%	0	Chinese Japanese
	01; 04.8%	09; 42.8%	06; 28.6%	01; 04.8%	04; 19.0%	Japanese English
	15; 75.0%	02; 10.0%	03; 15.0%	0	0	Turkish English
	05; 23.8%	05; 23.8%	10; 47.6%	0	01; 04.8%	Turkish German

Note. Mono=Monolinguals; Bi=Bilinguals; Low=low income (<€5000); Medium=medium income (€5000–€20000); High=high income (>€20000); No=no income; n.c.=no comment.

Participants in this study came from all over the world, with their origins spread across the globe: Ireland, the US, the UK, Canada, Germany, Austria, Italy, Russia, Ukraine, China, Japan, Turkey, Malaysia, Singapore, Lithuania, Rumania, Romania, Albania, Bangladesh, Latvia, Belarus, Kazakhstan, Uzbekistan, Kirgizstan, Azerbaijan, Brazil, and Macedonia. When asked about their country of residency at the time of participation, a few countries were added to the list, including Australia, Luxembourg, Vietnam, New Zealand, Greece, UAE, South Korea, Netherlands, Sweden, Brazil, and Finland.

3.1.1.2. *Design and Procedure*

The five subcategories of the assessment, following Halpern's concept of critical thinking (2010), can be divided into two sections. Two of the five subcategories, Verbal Reasoning, and Argumentation Analysis, can be grouped as "verbal subcategories", while the other three can be labelled "non-verbal subcategories": Hypothesis Testing, Likelihood and Uncertainty, and Problem Solving, indicating less emphasis on verbal structures in these subcategories (Halpern, 2010). This categorization suggests the importance of two different paths to solving each of the underlying questions. Those grouped as verbal subcategories place high emphasis on language and its understanding, whether written or spoken, and knowledge of vocabulary and grammar are a prerequisite for successful responding. Non-verbal subcategories, on the other hand, place emphasis on deeper structured problems and can be phrased in multiple ways without losing meaning or accuracy. Those questions can be led back to an understanding of concepts or rules not directly related to language. With this in mind, it is evident that different levels of precision have to be met when translating *Mixed Measures* from its original version in English to other languages. It proposes a very strict translation, with as little variation as possible, for verbal subcategories, not to influence underlying structures. Where direct translations are not possible due to missing words in certain languages, substitute questions with similar difficulty but avoiding those language barriers have to be utilized. Non-verbal subcategories, on the other hand, seem to contain more flexibility in phrasing and can be adjusted to individual language understanding and cultural differences. Details of changes undertaken in wording and phrasing during the translation process can be found in the specific sections of each language (3.1.2- 3.1.10).

Repetition Recognition

Before full completion of the assessment was reached, participants were asked if they would like to report anything to the researchers. This was done through an open text box and refers to the similarity of version A and B within bilingual testing. Possible recognition of the repeating questions in both languages was noted, as well as prior knowledge of any of the scenarios, which are all published measures and might be familiar to participants. Test takers could provide feedback regarding the assessment and share any thoughts or ideas and input with the researchers. Exclusion of participants recognising the repetition, or the implementation of separate calculations for this group, were considered, which applied to around 40% of the bilingual participants. Different aspects fed into the debate of how to proceed: all bilingual participants were presented with two versions of the same scenario, one version in L1 and the other version in L2, but only some reported to have noticed this. Recognition of

repeating scenarios indicates high levels of attention of this individual, as well as better memory and higher metacognitive awareness of the task compared to those bilinguals not noticing similarities between versions. Exclusion of all individuals recognising the repetition would have led to a distorted distribution within bilinguals in comparison to monolinguals, considering that all individuals with naturally higher attention levels and memory within bilinguals would have been excluded. Additionally, it is important to note that bilingual individuals were not provided with correct responses after finalizing the first version of the assessment, which in turn would have improved their performance in the second round of the assessment and would have led to practice effects. When solving the second version of any question, participants still relied on their cognitive performance and the ability to arrive at the correct result. Another reason for recognition of scenarios might have been prior knowledge of the question as mentioned above. The likelihood for a participant to recognize questions based on prior knowledge was estimated to be equal in both groups, monolinguals and bilinguals, and thus balanced in the overall effect added by prior knowledge. Taking all these points into consideration it seems reasonable to include individuals who recognized repetition in the overall calculations, and not to exclude them, taking into account the bias in data this would cause and the minor effect inclusion might entail. Moreover, comparative analyses of both groups, including all individuals and excluding those who recognised the repetition, showed the same outcomes, which confirms the validity of inclusion of all participants.

3.1.1.3. Data Collection and Grouping

Data Collection

The data collection process was spread out over the duration of 20 months from January 2017 until August 2018. Data for different language sets was collected in a staggered manner, dependent on finalisation of translations. Data collection took place online via the survey platform surveygizmo.com, which was utilized to host all assessments. Links to all language groups were distributed with the help of a purpose-built website (<https://bilingualthinking.weebly.com/> ; see Appendix 1), through collaborating researchers, email and social media. Word-of-mouth advertisement was encouraged within test takers and both personal and professional networks of the researchers were utilized to reach potential participants. Once a participant received the link to the study, participation was flexible concerning testing time and duration. Test takers were able to decide when to start the assessment and could take breaks from it if needed. The overall duration of the assessment was around half an hour for monolingual participants, who were only presented with the social economic status

questions as well as *Mixed Measures* in their native language, and around one hour for bilingual participants, which included SES, LEAP and *Mixed Measures* in both their languages. Participation was incentivised through course credit (English-Irish language set only), raffles to win Amazon vouchers, and payment (only for the Chinese-English language group due to additional funding). Details for each language set on incentives will be provided in the subsection for each language (e.g. 3.1.2.3, 3.1.3.3, 3.1.4.3. etc.) If interested in the results, participants could contact the researchers after completing the assessment successfully and receive the correct answers to the assessment, which in turn could help them to improve their critical thinking.

Grouping

Participants with no second language knowledge or individuals, who rated their second language knowledge as less than adequate on the LEAP-Questionnaire as introduced in section 3.2.2 (up to a self-rating score of 4 on a scale till 10) were labelled as *monolinguals* (0 none, 1 very low, 2 low, 3 fair, 4 slightly less than adequate, 5 adequate, 6 slightly more than adequate, 7 good, 8 very good, 9 excellent, and 10 perfect). Participants with a score of five or higher on overall second language proficiency and regular use of both languages were grouped as *bilinguals*, which applied to around 50% of participants (see Table 3.1).

Additionally, another grouping method was utilized to control for the correct understanding of the task, focussing on participants' second language fluency. All those who reported fluency of eight or above (on earlier mentioned scale to 10) were grouped as highly fluent. Those reporting fluency between five and seven were grouped as sufficiently fluent (0 none, 1 very low, 2 low, 3 fair, 4 slightly less than adequate, 5 adequate, 6 slightly more than adequate, 7 good, 8 very good, 9 excellent, 10 perfect). Moreover, monolingual participants can be separated into two groups, those who report never to have been in contact with any L2, and those who rate their L2 proficiency as very low. A mixture of calculations including both grouping criteria, broad group comparisons of all monolinguals and bilinguals, and comparisons where individuals fall into one of four groups, monolinguals with no L2 experience, monolinguals with little L2 experience, bilinguals with low L2 proficiency and highly fluent bilinguals, can add valuable insight into the nature of observed differences (Anderson et al., 2018).

Table 3.6. Overview of Participants' First Language Exposure; divided by Language Group

	LEAP N	L1 country	L1 family	L1 school/work	L1 friends
English Irish	80	100% > 17 y	96.5% > 13 y	90.2% > 10 y	96.5% > 10 y
German English	110	95.6% > 17 y	98.2% > 14 y	93.7% > 10 y	96.4% > 10 y
Italian English	69	95.7% > 17 y	98.8% > 11 y	98.8% > 13 y	100% > 10 y
Russian English	55	98.2% > 10 y	100% > 16 y	16.4% < 8y; 83.6% > 10y	96.4% > 10 y
Chinese English	48	100% > 12 y	100% > 12 y	95.8% > 12 y	87.5% > 10 y
Chinese Japanese	30	98% > 10 y	97% > 10 y	98.5% > 12 y	89.8% > 10 y
Japanese English	37	94.6% > 13 y	100% > 17 y	97.3% > 10 y	100% > 10 y
Turkish English	31	100% > 14 y	98.9% > 18 y	100% > 10 y	100% > 13 y
Turkish German	30	10% (N=3) never; 6.7% (N=2) 1 y; 73.3% > 1y	100% > 19 y	16.7% (N=5) never; 6.7% (N=2) 1y; 76.7% > 1y	100% > 8 y

Note. L1=first language; country/family/school/work/friends=time spent in L1 environment; y=years.

Table 3.6 provides an overview of participants' exposure and proficiency of their first language, whereas Table 3.7 focusses on participants' second language environment. Most participants have spent at least ten years in a country where their L1 was spoken and spoke this language with their family and friends as well as in school. This indicates high exposure to and proficiency in this language and describes the basis of this research.

Table 3.7. Overview of Participants' Second Language Exposure; divided by Language Group

	L2 >5 (N)	L2 acqu age	L2 fluency age	L2 country	L2 family	L2 school/work	L2 friends
English Irish	59	96.6% < 10 y; 3.4% > 10 y	42.4% < 10 y; 57.6% > 10 y	80.5 > 17 y	50% never	1.7% never; 3.4% < 1 y; 94.9% > 1 y	22% never; 5.1% < 1 y; 72.9% > 1 y
German English	97	45.5% < 10 y; 54.5% > 10 y	12.4% < 10 y; 87.6% > 10 y	44.3% never; 12.4% < 1 y; 43.3% > 1 y	55.7% never; 8.2% < 1 y; 36.1% > 1 y	33% never; 10.3% < 1 y; 56.7% > 1 y	36.1% never; 15.4% < 1 y; 48.5% > 1 y
Italian English	63	66.7% < 10 y; 33.3% > 10 y	3.2% < 10 y; 57.1% > 10 y	20.6% never; 31.8% < 1 y; 47.6% > 1 y	49.2% never; 27% < 1 y; 23.8% > 1 y	12.7% never; 33.4% < 1 y; 53.9% > 1 y	19% never; 23% < 1 y; 58% > 1 y
Russian English	48	68.8% < 10 y; 31.2% > 10 y	39.6% < 10 y; 60.4% > 10 y	6.3% never; 16.7% < 1 y; 77% > 1 y	66.7% never; 8.3% < 1 y; 25% > 1 y	8.3% never; 12.5% < 1 y; 79.2% > 1 y	12.5% never; 20.8% < 1 y; 66.7% > 1 y
Chinese English	30	86.7% < 10 y; 13.3% > 10 y	46.7% < 10 y; 53.3% > 10 y	63.3% never; 13.3% < 1 y; 23.3% > 1 y	3.3% never; 6.7% < 1 y; 90% > 1 y	10% never; 10% < 1 y; 80% > 1 y	30% never; 20% < 1 y; 50% > 1 y
Chinese Japanese	36	5.6% < 10 y; 94.4% > 10 y		19.4% never; 25% < 1 y; 56.6% > 1 y	5.6% never; 25% < 1 y; 69.4% > 1 y	11.1% never; 25% < 1 y; 63.9% > 1 y	27.8% never; 38.9% < 1 y; 33.3% > 1 y
Japanese English	21	42.9% < 10 y; 57.1% > 10 y		19% never; 4.8% < 1 y; 76.2% > 1 y	57.1% never; 14.3% < 1 y; 28.6% > 1 y	9.5% never; 14.3% < 1 y; 76.2% > 1 y	9.5% never; 14.3% < 1 y; 76.2% > 1 y
Turkish English	20	81% < 10 y; 19.1% > 10 y		61.9% never; 4.8% < 1 y; 33.3% > 1 y	90.5% never; 4.8% < 2y; 4.7% < 4y	33.3% never; 4.8% < 1 y; 61.9% > 1 y	42.9% never; 9.5% < 1 y; 47.6% > 1 y
Turkish German	21	90.5% < 10 y; 9.5% > 10 y		4.8% never; 4.8% < 3 y; 90.5% > 16 y	47.6% never; 4.8% < 1 y; 47.6% > 10 y	4.8% < 6 y; 95.2% > 13 y	4.8% 6y; 95.2% > 13y

Note. L2=second language; L2>5= fluency in L2 rated higher than 5 on LEAP-Q; acqu age= age at language acquisition; fluency age= age of fluency in language; country/family/school/work/friends=time spent in L2 environment; y=years.

3.1.2. English-Irish

3.1.2.1. Participants

Seventy-seven percent ($n=88$) of participants were originally from Ireland, with a significantly higher number (93.2%) bilinguals relative to (58.2%) monolinguals. Fifteen percent ($n=17$) were from the United States of America (monolinguals 27.3%, bilinguals 3.4%), 5.3% ($n=6$) from the United Kingdom, and one person respectively from Canada, Malaysia, and Singapore. Similarly, 80.7% ($n=92$) of participants reported their parents being of Irish origin, 63.6% of monolinguals and 94.9% of bilinguals; 13.2% ($n=15$) of these were American (monolingual 25.5%; bilingual 1.7%), and 3.5% ($n=4$) from the UK. One participant per country reported having Canadian, Indian, and

Malaysian parents. When asked about their country of residency, 76.3% ($n=87$) reported currently living in Ireland, 63.6% of monolinguals and 89.8% of bilinguals; 12.3% ($n=14$) in the US (21.8% of monolinguals and 3.4% of bilinguals), 5.3% ($n=6$) in the UK, 1.8% ($n=2$) in Australia, and one participant respectively in Canada, New Zealand, Japan, Luxembourg, and Vietnam.

3.1.2.2. Translation process

The original form of *Mixed Measures* in English was presented to all participants, both monolingual and bilingual. Random assignment carried out with the help of coding on the survey platform SurveyGizmo led to the grouping of version A and B, with around 50% of participants in each group. Bilingual participants were additionally presented with the opposite version of *Mixed Measures* in Irish. Translations of the assessment from English to Irish were enabled by Peadar Mac Fhlannchadha, Advocacy Manager and Deputy General Secretary of Projects of Conradh na Gaeilge. Conradh na Gaeilge is a democratic forum which promotes the Irish language and its use, represented by roughly 180 branches throughout the country of Ireland. Peadar Mac Fhlannchadha arranged and hired a professionally trained translator for the translation process to ensure the highest quality of work. Back-and-forth translations, proofreading and pilot testing were undertaken partly by the translator, partly by independent native Irish speakers who were not presented the material before that. This ensured full understanding only through the translated document. Minor corrections had to be undertaken to avoid obvious answer options within analogies, whereas the changes overall were kept to a minimum.

3.1.2.3. Data collection

Recruitment of participants was achieved in various ways. The Sona system, a university-wide system to enable and coordinate research participation, was used to recruit with participants reimbursed for their time with research participation credits. The involvement of a national Irish association, Conradh na Gaeilge (Cnag), was used to advertise the study to their members via posters, online and at an Irish language conference. Campus-wide advertisement of the study at the National University of Ireland, Galway, was implemented by posters, especially targeting the Irish department. Posters were printed in English and in Irish. The study was posted on international online platforms such as callforparticipants.com and in Irish speaking groups on social media (Facebook and LinkedIn). Word-of-mouth advertisement of participants who took the assessment was encouraged. Emails were directly sent to all university members of staff at every Irish department in the country, who were asked to pass on the information to students.

3.1.3. German-English

3.1.3.1. *Participants*

Eighty-eight percent of test takers ($n=103$) reported being of German origin, with a similar number of mono- and bilinguals (Mono: 90%; Bi: 87.6%). Eight percent were from Austria, and one individual respectively from Ireland, Lithuania, Rumania, Turkey, and the US. Participants of non-German speaking origin were assessed for language skills and have German as a first language with respect to usage and proficiency. Parents' origin was reported to be Germany in 84.6% of cases ($n=99$), 90% for monolinguals, of which the remaining 10% originated from Austria, and 81% within bilinguals, with the remainder being from Austria, the US, Afghanistan, Belgium, Ireland, Italy, Lithuania, Rumania, or Sri Lanka. As to the current country of residency, 53.8% ($n=63$) reported being situated within Germany, 17.1% in the US, 9.7% in Austria, 4.3% in the UK, 6.8% in New Zealand, and less than 5% in Australia, Ireland, Canada, Greece, and Italy.

3.1.3.2. *Translation process*

A German native-speaking research assistant living in Ireland, Sarah Volkmer, translated the assessment from English to German. The first version was proofread and back-and-forth translated by the primary investigator, Sophia Arndt, who is a German native speaker trained in language and translation through secondary education. Minor changes were implemented due to slight differences in the two versions. This refers to word analogies; for instance, “kitten” and “cats” were replaced by “puppies” and “dogs”, seeing as there are two different words for dogs and their young but not for cats, in German. Three German native speakers proofread the document and after collecting pilot data of ten participants, answers were pre-analysed to check for possible difficulties, which were shown to be absent. Random assignment led to the grouping of version A and B, with around 50% of participants in each group. Bilingual participants were additionally presented with the opposite version of *Mixed Measures* in English; this original form of *Mixed Measures* in English was presented only to bilingual participants.

3.1.3.3. *Data collection*

Data collection for the German-English data set was mainly pursued via Facebook. Bilingual participants were addressed on pages of immigrants in several English-speaking countries (America, Australia, Canada, the UK, etc.). Monolinguals were targeted through multiple sites of college courses and societies within Germany. Recruitment of bilingual participants was approached through English posters and

posts, whereas monolingual recruitment occurred in German. In total, nine Amazon vouchers with an overall value of €190 were given away (shared between participants of language sets Italian-English and Russian-English) in a raffle to support recruitment and reward participants. Apart from this, participation was voluntary with no other incentives except a potential personal interest in the topic. Additionally to posts in groups on Facebook, private messages were sent within the researcher's networks, and individuals who did participate were encouraged to share the study with colleagues and friends. LinkedIn and email communication were used as means to reach possible participants. Moreover, the study was posted on international online platforms like callforparticipants.com. A number of Goethe Institutes, a non-profit German cultural association advocating the study of the German language, were contacted in order to advertise the study through their websites and platforms, specifically those located in English speaking countries. Word-of-mouth advertisement by participants who took the assessment was encouraged.

3.1.4. Italian-English

3.1.4.1. Participants

Overall, 82 participants (95.4%) reported originating from Italy. This included all monolingual participants of this language group and 93.6% of bilinguals. The remaining four participants of the Italian-English language set originated from Albania, Bangladesh, the US, and Romania. Most of the participants' parents were originally from Italy (94.2%), with one participant, in each case, reporting their parents' origin to be from Albania, Australia, Bangladesh, Romania, and Spain. All of these fell under the bilingual grouping, with the exception of the Spanish parents. 59.3% of all test takers were currently living in Italy, 24.4% in the UK, 5.8% reported Ireland to be their country of residence, and less than 5% mentioned Australia, New Zealand, Canada, the UAE, and the US.

3.1.4.2. Translation process

An Italian research assistant fluent in English, Chiara Ferrari, implemented translations of *Mixed Measures* from English to Italian. Two Italian native speakers individually proofread the first draft of translations and corrections were considered during a discussion and back and forth translation with the primary investigator to ensure proximity of translations to the original document. Corrections entailed of, parallel to the other language sets, mainly changes within the verbal reasoning subsection "Analogies". Five of the analogies were slightly changed to keep the difficulty level steady while achieving better understanding. Changes were minor; an example would

be the change from “rat” to “mouse” to avoid word repetition and from “frame” to “scene”, seeing as “frame” is not usually used for individual shots in moving pictures as described in this scenario. Additional to that, a pilot testing sample of about ten Italian individuals completed the assessment and provided the researchers with input about wording before concluding the final version. Random assignment led to the grouping of version A and B, with around 50% of participants in each group. Bilingual participants were additionally presented with the opposite version of *Mixed Measures* in English, this original form of *Mixed Measures* in English was presented only to bilingual participants.

3.1.4.3. Data collection

Data collection for the Italian-English data set was mainly pursued via Facebook. Bilingual participants were addressed on pages of immigrants in several English-speaking countries (America, Australia, Canada, the UK, etc.). Monolinguals were targeted through multiple pages of college courses and societies within Italy. Recruitment of bilingual participants was approached through English posters and posts, whereas monolingual recruitment was advertised in Italian. In total, nine Amazon vouchers with an overall value of €190 were given away (shared between participants of language sets German-English and Russian-English) in a raffle to support recruitment and motivate participants. Participation was voluntary, without any other incentives but a personal interest in the topic. In addition to advertising posts in groups on Facebook, private messages were sent within the researcher’s networks. Individuals who participated were encouraged to share the study with colleagues and friends. LinkedIn and email communication were used as a sub-method to reach potential participants. Moreover, the study was posted on international online platforms like callforparticipants.com. Word of mouth advertisement of participants who took the assessment was encouraged.

3.1.5. Russian-English

3.1.5.1. Participants

The majority of participants of the Russian-English language set reported to have their origins in Russia ($n=37$). Nine individuals were from Ukraine, whereas four originated from Latvia, three from Belarus, two from Kazakhstan and the UK, and one each respectively, from the United States, Bulgaria, Uzbekistan, Lithuania, Kirgizstan, and Azerbaijan. The origins of participants’ parents were distributed similarly, with 43 test takers reported their parents to come from Russia, eight from Ukraine, four from Latvia and Belarus, two from the UK and Kazakhstan, and one from Bulgaria, Uzbekistan, and

Kyrgyzstan. Participants were situated in the UK ($n=16$), in the US ($n=12$), Russia ($n=8$), Australia ($n=6$), Canada ($n=5$), Germany ($n=4$), Ireland ($n=3$), the Ukraine ($n=3$), Belarus ($n=2$), and one each in Hong Kong, the UAE and South Korea ($n=1$).

3.1.5.2. Translation process

Translations into Russian were conducted by research assistant Olga Bolbocean, a Russian native speaker living in Ireland. Prior to the translations, the main investigator and Olga Bolbocean discussed the main points of the assessment to ensure understanding and correct translations. Mariia Naumovets undertook proofreading, a Russian native speaker working for the Technical University in Kaiserslautern, Germany. A pilot sample of 80 participants was collected, and input regarding translations was received before the final version excluded any errors that arose. The reason for this relatively high number of pilot participants was the difficulty of translating some of the subsections, with CT assessments seemingly being relatively uncommon in the Russian language. Participants were encouraged to provide feedback about any occurring difficulties and unclear sections, and pilot data collection was on going until feedback shifted from suggestions for change to no noticeable difficulties.

3.1.5.3. Data collection

Data collection for the Russian-English data set was mainly pursued via Facebook. Bilingual participants were addressed on pages of immigrants in several English-speaking countries (America, Australia, Canada, the UK, etc.). Monolinguals were targeted through multiple sites of college courses and societies within Russia. Recruitment of bilingual participants was approached through English posters and posts, whereas monolingual recruitment was advertised in Russian. In total, nine Amazon vouchers with an overall value of €190 were given away (shared between participants of language sets German-English and Italian-English) in a raffle to support recruitment and motivate participants. Participation was voluntary, with no incentives other than personal interest in the topic. In addition to advertising posts in groups on Facebook, private messages were sent within the researcher's and research assistant's networks. Individuals who participated were encouraged to share the study with colleagues and friends. LinkedIn and email communication were used as a sub-method to reach potential participants. Moreover, the study was posted on international online platforms like callforparticipants.com. Word of mouth advertisement of participants who took the assessment was encouraged.

3.1.6. Japanese-English

3.1.6.1. *Participants*

All participants of the Japanese English language set originate from Japan, with the exception of eight individuals coming from China, and one from Brazil and the UK. All parents were originally from Japan bar seven, who were from China. Concerning the residency of participants, 51 reported living in Japan, six in the UK, two in Ireland, China, and Canada, and one in Vietnam, Australia, Brazil, Malaysia, the US, and Finland.

3.1.6.2. *Translation process*

Professor Yoshitaka Nakajima and Kaori Kojima conducted translations of the assessment from English to Japanese. The main investigator, Sophia Arndt, was present to answer any occurring questions regarding concepts or understanding of the questions. The interactive translation process helped to eliminate possible misunderstandings and facilitated the transfer of the scenarios from English to Japanese. Once finished, the translated document was proofread by four independent Japanese native speakers and suggestions were made as to adaptations of phrases and wording. Finally, pilot data of ten participants was collected to ensure correct understanding and structure. Within the translation process, certain sections had to be adapted due to cultural backgrounds and language construct barriers. Changes were tried to be kept to a minimum but were implemented where inevitably necessary. Changes for the non-verbal sections Hypothesis Testing, Likelihood and Uncertainty, and Problem Solving were used to aid understanding and increase natural-sounding language. These had no influence on the solution process and are not worthy of being mentioned in detail; two examples are changes from “business executives” to “businessmen” and from “high school” to “college”. Changes within verbal subcategories, on the other hand, had to be treated with care and will be listed in the following to clarify the procedure. The analogies used in Verbal Reasoning were based on relationships of words, and reasoning needed for the solution process is based on those interconnections. Four of the eighteen questions (nine per version; A and B) had to be changed due to close similarity of words or missing translations for certain words. Replacement analogies were taken from the original source (Dermott, 2002), matched for difficulty and adequacy. Another seven analogies were adapted to enable translation into Japanese. This entailed changing individual words or adapting translations to the original. For instance, the word “porch” was replaced with “entrance” due to a missing translation, or “bungalow” was replaced with “cottage”. These changes were conducted carefully and with the overall outcome in mind, with

awareness to the sensitivity of the question to change. Details changed for the second verbal subcategory, “Argumentation Analysis” were limited to the answer options, which were changed from the original “valid-invalid” buttons to “logically correct - logically incorrect”, since the word ‘valid’ has a slightly different, in this context not suitable, connotation in Japanese.

3.1.6.3. Data collection

Japanese-English data collection was enabled through the help of cooperating researchers at Fukuoka University (Professor Nakajima and his laboratory members), as well as at Shibaura Institute of Technology in Tokyo (Dr Yuko Yamashita). Participants were approached directly by collaborating researchers in person or via email, and the option was given to receive the correct answers of the assessment after taking part, which could help to improve participants’ awareness of critical thinking as an incentive for participation. Additionally, posts in Facebook groups of immigrants in English speaking countries were implemented. Five Amazon vouchers with an overall value of €250 were raffled as a reward system between all participants that opted for entering the draw.

3.1.7. Chinese-English/ -Japanese

3.1.7.1. Participants

All participants from the two language groups, Chinese-English and Chinese-Japanese, reported being from China, with the exception of three Malaysians. Two test takers had Japanese parents and two had parents from Malaysia, all other participants’ parents were of Chinese descent. As countries of residence, 58 individuals noted China, 39 were living in Japan, three in Germany, Malaysia, and the UK, and one in the Netherlands, Ireland, and Sweden.

3.1.7.2. Translation process

Zhimin Bao, a Chinese native speaker fluent in English, provided translations of the assessment into Chinese. Shimeng Liu helped to proofread, discuss and back-and-forth translate the first draft of translations, and questions were adjusted accordingly. Slight adjustments included certain words with no equivalent individual symbol in Chinese; for instance, “rat” was replaced with “rabbit” and “icing” was replaced with “cream”. It was ensured that none of the changes influenced the structure of the assessment and that difficulty levels were kept at an even level.

3.1.7.3. Data collection

The Chinese assessment for native speakers was used in combination with two languages as a second language: English and Japanese. Research assistant Yuan Chen (Cindy) recruited Chinese-English bilinguals and Chinese monolingual speakers, whereas Chinese-Japanese bilinguals were recruited with the help of research assistant Shimeng Liu (Lucy), based in Fukuoka, Japan. Participants recruited through Shimeng Liu were reimbursed with a voucher that was sponsored and provided by Dr Yamashita at Shibaura Institute of Technology in Tokyo.

3.1.8. Chinese-Japanese

See section 3.1.7 for further detail.

3.1.9. Turkish-English /-German

3.1.9.1. Participants

Turkish participants all reported originating from Turkey ($n=127$) with the exception of eight individuals, who said they were from Germany and one who reported origins from Macedonia and Azerbaijan. Most of the participants' parents were originally from Turkey ($n=116$), with three originating from Bulgaria and one from Armenia, Azerbaijan, and Macedonia. Most test takers were living in Turkey, bar 14 individuals who reported living in Germany, four in the US, and one in the UK, the Netherlands, and Cyprus.

3.1.9.2. Translation process

Dr Arzu Özkan Ceylan in Hacettepe University, Ankara, Turkey, provided all translations of the assessment from English to Turkish. These were proofread, back and forth translated, and changes were discussed with the main investigator. Pilot data collection involved the testing of ten participants who were encouraged to bring forward any concerns or uncertainties concerning phrases and translations within the assessment.

3.1.9.3. Data collection

Turkish data collection was enabled by Dr Arzu Özkan Ceylan in Hacettepe University, Ankara, Turkey, and by Assist. Prof. Evrim Gülbetekin at Akdeniz University, Antalya, Turkey. The Turkish assessment, similar to the Chinese assessment, was combined with two languages as a second language; English and German. Further data collection took place through social media (Facebook) with the incentive of five €50 Amazon vouchers to win in a raffle. Additionally, Turkish German participants were contacted through a local mosque (Meulana Mosque, Fuerstenfeldbruck, Germany) with the help of poster advertisement, as well as through academics and societies

throughout Germany and Turkey which were contacted by the researcher and asked to advertise the study.

3.1.10. Turkish-German

See section 3.1.9 for further detail.

3.2. Materials

3.2.1. Socioeconomic Status

Questions assessing participants' socio-economic status cover personal background information. Participants are asked about their schooling background (e.g., the type of school and years of formal education they underwent), their living situation (e.g. house owner, sole tenant, or sharing), average income, and cultural background. For details see Appendix 2. Socioeconomic status can hugely vary within samples of bilingual populations and for that reason is important to be taken into account. Moreover, research has shown that socioeconomic background influences cognition (Hackman, Farah, & Meaney, 2010; Noble, Houston, Kan, & Sowell, 2012) and intelligence (Fischbein, 1980) including critical thinking and problem-solving skills, mostly focussed on factors like higher education.

3.2.2. Language Experience and Proficiency Questionnaire

The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) was implemented to assess language proficiency and language use. The LEAP-Q is a validated, reliable assessment with established internal and criterion-based validity (Marian, Blumenfeld, & Kaushanskaya, 2007). Marian, Blumenfeld, and Kaushanskaya (2007) measured internal validity through factor analysis and multiple regression analyses, to ensure each question assessing the same construct.

Self-reported measures are commonly used in language assessment, with a special focus on research on bilingualism (for instance Eilola, Havelka, & Sharma, 2007; Palmer, van Hooff, & Havelka, 2010). They were compared to standardized, objective language assessments to assess external reliability in first and second language and were shown to be strongly correlated, proving the LEAP-Q to be a reliable instrument in the assessment of bilinguals (Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAP-Q is available in electronic format and can be filled in independently. It is freely accessible in a wide range of languages, for instance, English, German, Italian Japanese, Mandarin Chinese, Malay, Russian and Turkish (<https://bilingualism.northwestern.edu/leapq/>), representing all languages implemented

in this project. LEAP-Q takes approximately 15 minutes to complete for bilingual speakers.

The questionnaire includes the assessment of various details about participants' language history and background: language knowledge, language dominance, the age of acquisition, percentages of time speaking and reading first or second language, language preferences, as well as time spent in particular language environments (Anderson et al., 2018; Luk & Bialystok, 2013). For details please see Appendix 3. LEAP-Q is a published measure which is widely used in the field of research on bilingualism, especially useful in those cases with high participation numbers where individual assessment would not be feasible. It has been implemented in various languages (Barbeau et al., 2017; Krizman, Marian, Shook, Skoe, & Kraus, 2012; Libben & Titone, 2009; Luk & Bialystok, 2013; Marian, Lam, Hayakawa, & Dhar, 2018; Pelham & Abrams, 2014) and proves to be an efficient, reliable way of assessing participants' language history and background.

3.2.3. Mixed Measures

Critical thinking (CT) scores of every participant were gathered through an assembly of scenarios, covering the five critical thinking subcategories as defined by Halpern (Halpern, 2010). A total of 21 short, published tests were added together into *Mixed Measures*, a comprehensive assessment of CT described in chapter 2. For details of the assessment please see Appendix 4. All questions follow a multiple-choice answer format, due to feasibility and the number of different languages implemented. The original version of the assessments, version A, was slightly changed to develop an alternative version, version B, as close to the original assessment as possible in terms of wording and answer options but with alternative subjects and objects in question. For instance, if version A is about Mrs A buying C, version B would contain Mr B selling D. Possible differences in difficulty due to slight changes between the versions were balanced by randomisation of participants. 50 per cent of participants were presented with version A in their first language (L1) and version B in their second language (L2), and the other 50 per cent conversely with version B in L1 and version A in L2. Both versions of the assessment were professionally translated into seven different languages (English (original), Irish, German, Italian, Russian, Turkish, Chinese, and Japanese) by native speakers, all of whom professionals, using back and forth translations, controlled by multiple reviewers, similarly to published studies in the research area of bilingualism (e.g. Bernardo & Calleja, 2005; Eilola, Havelka, & Sharma, 2007).

Each participant was presented with *Mixed Measures* in their first language and if bilingual in their second language. Full randomization and counterbalancing, automatically executed through coding within the survey tool *SurveyGizmo*, was implemented throughout the assessment: Participants were randomly presented with group assignment A or B for each language, moreover the order of languages L1 or L2 alternated across participants, as well as the presentation of subcategories, questions within the subcategories and answer options were randomized for each individual. Randomization was used to control for order effects and to avoid confounding variables, helping to reduce biases and to counterbalance negative effects on the assessment.

3.2.4. Raven's Standard Progressive Matrices

Raven's Standard Progressive Matrices (RSPM) were deployed as a nonverbal control measure for fluid intelligence (Diamond, 2013; Mani, Mullainathan, Shafir, & Zhao, 2013; Prabhakaran et al., 1997; Raven, 2000) as defined by Cattell (1963). Fluid intelligence is a factor of general intelligence, based on novel scenarios to be solved by inductive and deductive reasoning. It is independent of prior knowledge, in comparison to crystallized intelligence, which is based on knowledge and experience (Cattell & Horn, 1978; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). Ensuring the validity of the implemented critical thinking (CT) measure, participants' scores on RSPM are expected to correlate with scores of *Mixed Measures*, outlining the correlation between intelligence and critical thinking (Halpern, 2006; Marin & Halpern, 2011). RSPM is a robust and validated assessment of test takers' mental ability and abstract reasoning skills (Bors & Stokes, 1989; Raven, 2000), with its first version standardized in 1938 on 1407 children in the UK (unpublished thesis; Raven, 1936; as cited in Bilker, Hansen, Brensing, Richard, Gur, & Gur, 2012; Raven, 2000). Reliability studies focussing on internal consistency and test-retest reliability as well as validity studies have been successfully conducted: the RSPM has been tested internationally and has shown to be consistent in different cultural and economic settings (Raven, 2000). Strengths lie in the adaptability to all age groups and its usefulness for comparative studies (Raven, 2000). RSPM was developed for the use with participants of average intelligence, whereas it is less reliable on both extreme ends of the scale (Raven, 2000). It consists of 60 tasks. In each task, participants are presented with a pattern, which contains a blank space, a cut-out within the image. The task is to complete this pattern by choosing the one of the six to eight pieces presented below the pattern, which fills the blank space correctly. Only one of the pieces below the pattern takes all aspects of the pattern into consideration, such as shape and fill. See Appendix 5 for an example.

Bilker et al. (2012) developed two non-overlapping short forms of this measure, using a Poisson predictive model. Model fitting and validation have been ensured via split sample testing and cross-validation to verify results (Bilker et al., 2012). The short form consists of nine items per version, achieving correlations of .98 between both versions. The short form of the SPM proves to be an appropriate measure in terms of test duration and scope for this assessment, as a well-validated tool covering basic cognitive functioning within various cultural, ethnic and socioeconomic pools of participants from all around the world. Raven Matrices are language independent and the only written words within the measure are instructions explaining the task. The exact wording is secondary with the primary importance being the comprehension of the task. Translation of the sentence of instruction of the RSPM was adjusted to cultural understanding and undertaken by experienced native speakers.

The implementation of the order of the individual images of the RSPM followed a randomisation process and varied for every participant. Participants were presented with each image of the RSPM randomly throughout the assessment. This led to the presentation of written scenarios of *Mixed Measures* alternating with images, which in turn aided concentration and focus of participants, requiring different problem-solving techniques, and counteracted the monotony of written content.

3.3. Research questions

The study was conducted to answer two primary research questions:

- (I) Do bilinguals score differently compared to monolinguals in measures of critical thinking (CT)?
- (II) Do bilinguals differ in CT scores when assessed in their dominant language (L1) compared to their second language (L2)?

3.4. Design

To answer these questions, data of nine different language sets was collected and a cross-sectional, mixed within- and between-subjects measure was carried out. The analysis included two groups with two layers each, state of bilingualism: monolingual or bilingual; as well as language mode: first language and second language. Between subject calculations were implemented comparing the two states of bilingualism, monolingual individuals versus bilingual individuals, whereas a within measures design was applied for the comparison of bilingual subjects' first and second language scores. Demographic questions, as well as questions about the social economic status, were presented to all test-takers. These were followed by a question asking for any second language skills. If any knowledge of a second language was indicated, participants were presented with the LEAP-Questionnaire, as introduced

above, assessing their language background. Subsequently, all participants were presented with the critical thinking test battery *Mixed Measures*, monolinguals only in their native language, either version A or B, bilinguals in both languages, version A in one language, version B in the other language. This process followed full randomisation concerning the order of presentation of versions and languages, subcategories, questions within those subcategories as well as answer options.

The following exclusion criteria were applied: a participant's data was excluded if the first language, either by acquisition or by proficiency, was indicated to be a language not covered by the relevant *Mixed Measures* assessment. For instance, the participant was filling in the English (L1) –Irish (L2) questionnaire, but indicated to have learned Spanish as a first language, or to be most proficient in Spanish rather than English or Irish. The second exclusion criterion was if the assigned first language L1 of *Mixed Measures*, English in the above example, was marked by the participant as the third proficient language or the third language the participant reported to have learned. The third exclusion criterion concerned the exposure an individual had to an unrelated language, referring to the example above, e.g. Spanish. If the exposure to this unrelated language, not covered by *Mixed Measures*, exceeded 40% as indicated by the participant it led to exclusion of the data. This was done to ensure the reported languages to be applicable to test-takers and participants to fall under the category of bilingualism in comparison to multilingualism, both of which are shown to have differing effects on an individual (Anderson et al., 2018).

3.5. Statistical measures

Examination of the data was undertaken using the Statistical Package for Social Sciences (IBM SPSS statistics 23.0, 2015). Additionally to assessing individual outcomes within each language set, an overall comparison was undertaken including all language sets. Through this, a unique combination of individual languages covering multiple linguistic families, outcomes could be attributed to fundamental cognitive performance rather than differences between specific languages or cultural contexts.

To assess the validity of *Mixed Measures*, a Pearson product-moment correlation was run to determine the relationship between the overall score on *Mixed Measures* measuring critical thinking (CT) and Raven's Standard Progressive Matrices (RSPM), measuring fluid intelligence.

Between-group comparisons were undertaken with a multivariate analysis of variance (MANOVA), which made the comparison of multiple means of sub scores possible. Relationships of dependent and independent variables could be observed,

and trends and differences detected. Dependent variables were averages of participants raw scores on each of the five dimensions of CT, hypothesis testing (HT), verbal reasoning (VR), argumentation analysis (AA), judging likelihood and probability (JL), and problem solving (PS). Raw scores were considered appropriate as no transformation (e.g. square root, log n, arcsine) helped to increase normality levels. To assess normality within the test results Shapiro Wilk or Kolmogorov-Smirnov were not suitable considering the high numbers of participants. Normality was assessed through QQ plots of approximately normally distributed data and observation of scores on a histogram. Multiple regression was calculated to predict the effects of three factors, age, gender, and years of education, on total scores. An item analysis of all cases was implemented to assess the reliability of the scale as well as correlation and covariance within the five subcategories within the merged data set. Bonferroni correction was applied to all pairwise comparisons. Box's Test of Equality was implemented to control for homogeneity of covariance and Levene's Test helped to check whether the assumption of homogeneity of variance was met. A Pearson product-moment correlation was run to determine the relationship between participants' overall scores of *Mixed Measures* and outcomes of the Raven Matrices as a form of validation of *Mixed Measures*, with scores of the Raven representing fluid intelligence, which were expected to correlate with critical thinking.

An additional calculation was implemented to assess possible differences in detail, specifically focussed on language proficiency. This was carried out parallel to the first set of calculations. A MANOVA was conducted to control for differences within the two groups, monolinguals, and bilinguals, but only participants who reported to never have been in touch with a second language (monolingual) or those who rated their second language proficiency as highly proficient with a minimum rating of eight on a scale to ten (bilingual) were included. To deepen the insight into possible differences within the five subcategories, these two groups - monolinguals who have never been in contact with a second language, and highly proficient bilinguals - were compared to two groups, one consisting of monolinguals who had some knowledge of a second language, even though proficiency would be very low, and bilinguals with a mediocre proficiency, varying between five and seven on a scale to ten. This last calculation was only undertaken with the merged data set including participants of all language sets due to sufficient numbers in each of these four groups within the merged set, whereas all other calculations were implemented with all language sets individually as well as the merged set.

In one language set, Japanese English, supplementary calculations were added, trying to shed light on observed differences within the language set: A

MANOVA with gender as the independent variable and the five subcategories as dependent variables helped to explore the origin of differences within the group, followed by the same MANOVA but splitting participants by language setting, monolingual or bilingual. Finally, all participants that reported to live outside of Japan were excluded and the original MANOVA, with language setting, mono-or bilingual, as the independent variable, focussing on the five subcategories as a dependent variable, was undertaken.

Within-group comparisons were implemented using repeated measures MANOVA (reMANOVA). Less emphasis was put on covariates considering the nature of the repeated measurement of single individuals and the comparison of differences within each individual. Bonferroni correction was applied to all pairwise comparisons. An additional reMANOVA was carried out only including highly proficient participants, who rated their own L2 proficiency eight or higher on a scale to ten. By only focussing on highly proficient bilinguals, important details of differences within the two languages could be observed and the exclusion of comprehension difficulties or misunderstanding of scenarios as a possible covariate and source of error was ensured. Another reMANOVA was undertaken to assess a possible influence of age of acquisition of L2. All bilingual participants were divided dependent on the age of acquisition of their second language, with early bilinguals having acquired L2 before the age of ten.

A final set of calculations was undertaken to compare all language sets on overall scores, as well as to compare scores of each subcategory across languages. A MANOVA for each language setting, L1 and L2, was carried out. Language groups represented the independent variable, of which the effect on L1/2total, as well as HT, VR, JL, AA, and PS, was measured. When assessing bilinguals' L2, an additional calculation was implemented, only including those bilinguals who rated themselves as highly proficient, with a minimum score of eight on a scale to ten.

3.6. Results

3.6.1. General results

A Pearson product-moment correlation was implemented to determine the relationship between critical thinking scores and fluid intelligence of participants. There was a positive correlation between *Mixed Measures* and Raven's Standard Progressive Matrices (RSPM), which was statistically significant both in scores of L1 ($r = .41$, $N = 738$, $p < .001$) and L2 ($r = .62$, $N = 392$, $p < .001$), also visible on scatterplots created for both language settings (L1: see Figure 3.1; L2: see Figure 3.2).

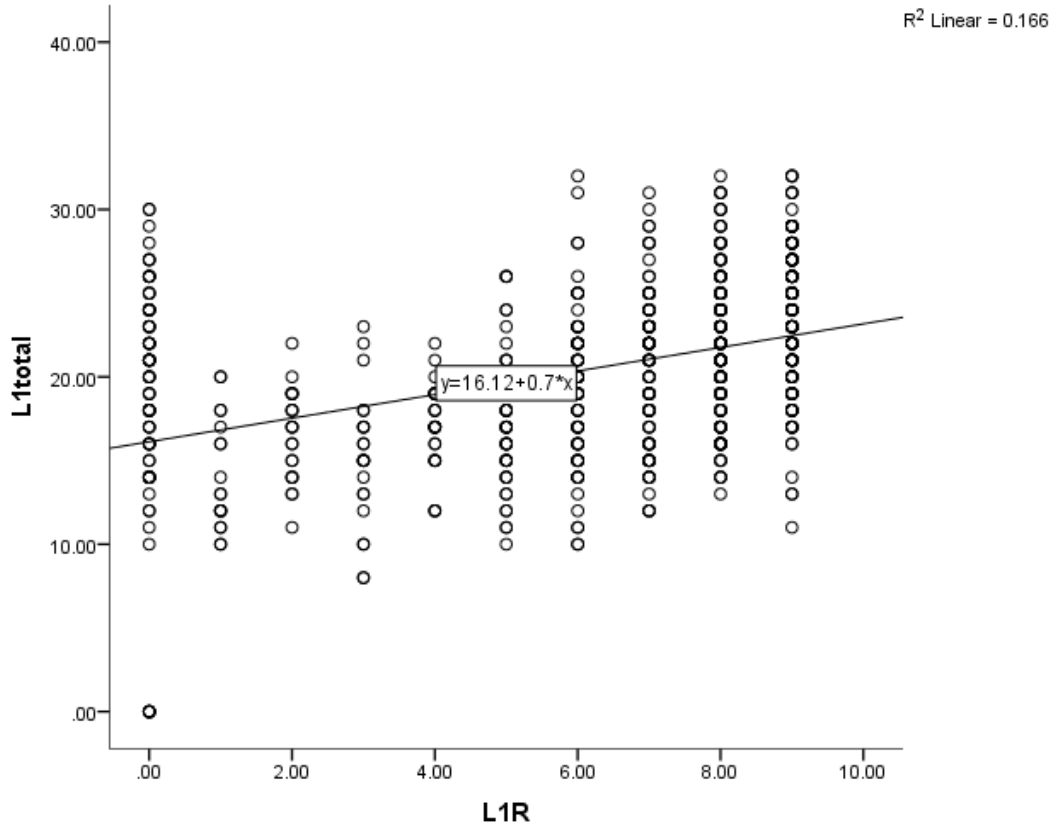


Figure 3.1. Scatterplot representing correlation between L1total and Raven Scores.

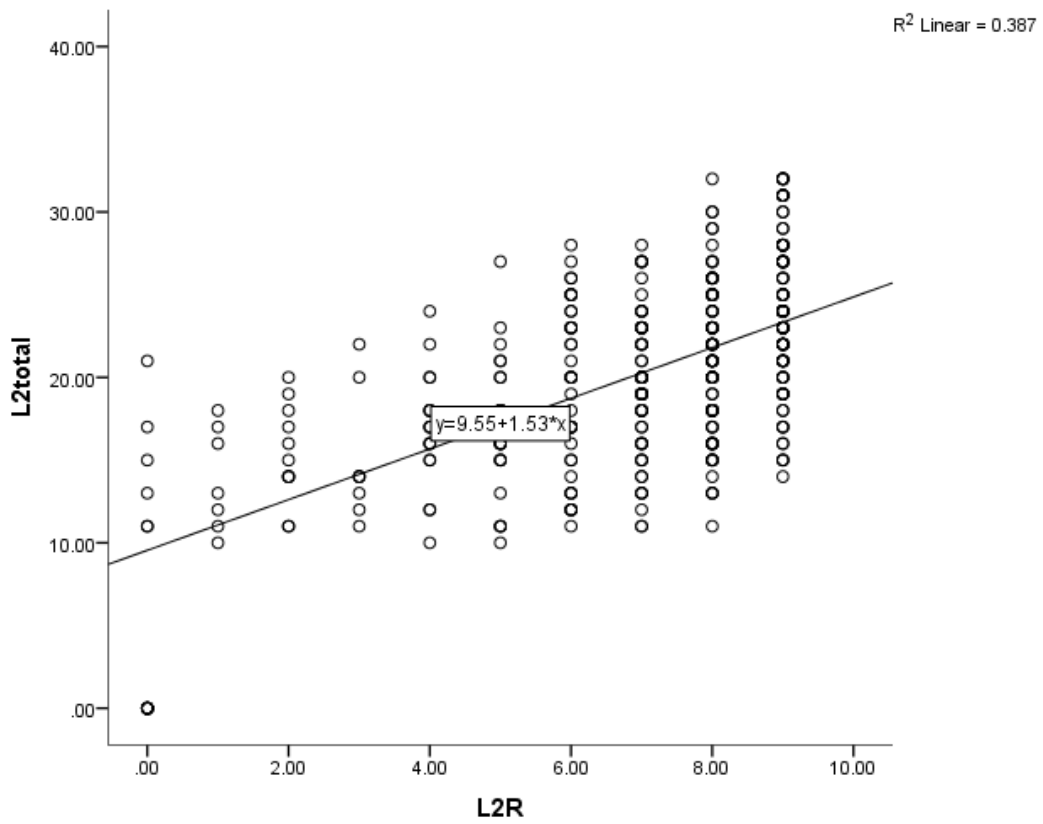


Figure 3.2. Scatterplot representing correlation between L2total and Raven Scores.

Item analysis of all cases was utilised to assess the reliability of the scale as well as correlation and covariance within the five subcategories. Cronbach's alpha assesses the range of inter-individual variation, which on the other hand is aiming to be assessed within Mixed Measures. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1total reaches a value of $\alpha(L1total)=.52$. Comparable analyses with outcomes for L2total show a value for Cronbach's alpha of $\alpha(L2total)=.64$. The inter-item correlation matrix and the inter-item covariance matrix show positive values within all subcategories. Table 3.8 includes an overview of the item analysis for all language groups.

Table 3.8. Cronbach's Alpha Values divided by Language Group and Mono- and Bilinguals

Language	N (Mono/Bi)	L1	L2	> $\alpha(L1)$	> $\alpha(L2)$
		α / α_{Stand}	α / α_{Stand}	with exclusion	with exclusion
English-Irish	114 (55/59)	.51 / .43	.57 / .54	AA (.66)	AA (.66)
German-English	117 (20/97)	.74 / .75	.77 / .80	no	no
Italian-English	086 (23/63)	.32 / .33	.53 / .52	AA (.36)	AA (.61)
Russian-English	057 (09/48)	.42 / .36	.45 / .48	JL (.45); AA (.53)	AA (.57)
Chinese-English	064 (34/30)	.42 / .37	.68 / .73	HT (.44)	no
Chinese-Japanese	065 (32/33)	.41 / .32	.65 / .68	HT (.46)	no
Japanese-English	058 (37/21)	.52 / .47	.55 / .60	AA (.58)	VR (.70)
Turkish-English	089 (69/20)	.51 / .48	.49 / .55	JL (.55)	AA (.74)
Turkish-German	088 (67/21)	.28 / .26	.44 / .44	JL (.33); AA (.32)	VR (.52)
All merged	738 (345/392)	.52 / .51	.64 / .67	AA (.55)	no

Note. Mono=Monolinguals; Bi=Bilinguals; L1=bilinguals' first language; L2=bilinguals' second language; α =Cronbach's alpha/ α_{Stand} = α based on standardized items; $\alpha>$ with exclusion= alpha score increases with exclusion of sub-category.

Multiple regression was calculated to predict L1total based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,734)=25.02$; $p<.001$; $R^2=.09$), explaining 9% of the overall variance in the model. Participants' predicted L1total score is equal to $12.32 +.03 (AGE) +1.07 (GENDER) +.39 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. Gender and years of education were found to be significant predictors of L1total ($p_{Gender}<.01$; $p_{Edu}<.001$), with participants' overall score of L1total increasing by 1.01 points dependent on gender and .39 points per additional year of

education. Males scored higher than females in two of the five subcategories (ANOVA: JL: $F(1,735)=9.60$; $p<.005$; PS: $F(7,735)=14.01$; $p<.001$) affecting the overall score significantly ($F(1,735)=7.04$; $p<.01$). It is important to mention that numbers of male and female participants are not equal and dispersed in all language groups, with a significantly higher number of females taking part in the assessment.

Table 3.9 illustrates the effects of the three factors age, gender and years of education on the individual language groups.

Table 3.9. L1: Influence of Age, Gender and Education on L1total divided by Language Group

Language	Multiple Regression	Overall	AGE	GENDER	EDU
English-Irish	$F(3,110)=7.33$; $p<.001$; $R^2=.17$	09.74;	.08; n.sig.	1.43; n.sig.	.39; $p=.001$
German-English	$F(3,113)=1.06$; n.sig.; $R^2=.03$	16.95	-.01; n.sig.	-.53; n.sig.	.31; n.sig.
Italian-English	$F(3,82)=.94$; n.sig.; $R^2=.03$	17.34	-.02; n.sig.	1.15; n.sig.	.16; n.sig.
Russian-English	$F(3,53)=5.17$; $p<.005$; $R^2=.23$	18.04	-.02; n.sig.	4.67; $p<.001$.02; n.sig.
Chinese-English	$F(3,60)=3.23$; $p<.05$; $R^2=.14$	14.11	-.14; n.sig.	-.42; n.sig.	.71; $p<.005$
Chinese-Japanese	$F(3,61)=.66$; n.sig.; $R^2=.03$	18.40	-.11; n.sig.	-.90; n.sig.	.44; n.sig.
Japanese-English	$F(3,54)=2.41$; n.sig.; $R^2=.12$	19.89	-.08; n.sig.	2.12; n.sig.	.17; n.sig.
Turkish-English	$F(3,85)=6.26$; $p<.005$; $R^2=.18$	06.14	.21; n.sig.	.97; n.sig.	.46; $p<.05$
Turkish-German	$F(3,84)=3.12$; $p<.05$; $R^2=.10$	09.15	.08; n.sig.	.40; n.sig.	.44; $p<.05$
All merged	$F(3,734)=25.02$; $p<.001$; $R^2=.09$	12.32	.03; n.sig.	1.07; $p<.01$.39; $p<.001$

Note. Edu= Education.

Multiple regression was calculated to predict L2total based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model to predict L2 including gender was non-significant, but age and education were significant ($F(3,388)=7.45$; $p<.001$; $R^2=.05$), explaining 5% of the overall variance in the model. Participants' predicted L2total score is equal to $12.89 + .06$ (AGE) + $.97$ (GENDER) + $.26$ (EDU), where age and education are measured in years and gender is coded 1=female, 2=male. Age and years of education were found to be significant predictors of L2total ($p_{Age}<.05$; $p_{Edu}<.005$), with participants' overall score of L2total increasing by $.26$ points per additional year of age and $.06$ points per additional year of education. Table 3.10 illustrates the effects of the three factors age, gender and years of education on the individual language groups.

Table 3.10. L2: Influence of Age, Gender and Education on L2total divided by Language Group

Language	Multiple Regression	Overall	AGE	GENDER	EDU
English-Irish	F(3,55)=5.59; p<.005; R ² =.23	05.84	.13; n. sig.	1.09; n.sig.	.44; p<.005
German-English	F(3,93)=1.12; n.sig.; R ² =.04	13.40	.06; n.sig.	.53; n.sig.	.27; n.sig.
Italian-English	F(3,59)=1.30; n.sig.; R ² =.06	14.98	-.03; n.sig.	1.64; n.sig.	.27; n.sig.
Russian-English	F(3,44)=1.37; n.sig.; R ² =.09	21.83	-.05; n.sig.	2.70; n.sig.	-.05; n.sig.
Chinese-English	F(3,23)=1.62; n.sig.; R ² =.16	04.80	.25; n.sig.	-.08; n.sig.	.47; n.sig.
Chinese-Japanese	F(3,29)=.67; n.sig.; R ² =.06	27.26	-.04; n.sig.	-2.60; n.sig.	-.15; n.sig.
Japanese-English	F(3,17)=2.44; n.sig.; R ² =.30	24.36	-.10; n.sig.	4.69; n.sig.	-.29; n.sig.
Turkish-English	F(3,16)=2.91; n.sig.; R ² =.35	08.01	.89; p<.05	1.39; n.sig.	-.66; n.sig.
Turkish-German	F(3,17)=1.79; n.sig.; R ² =.24	13.97	.04; n.sig.	4.41; p<.05	-.01; n.sig.
All merged	F(3,388)=7.45; p<.001; R ² =.05	12.89	.06; p<.05	.97; n.sig.	.26; p<.005

Note. Edu=Education .

Assessing normality of test results using Shapiro Wilk or Kolmogorov-Smirnov was not suitable considering the high numbers of participants. Normality of data was assured by inspection of histograms (L1 see Figure 3.3; L2 see Figure 3.5) and QQ plots (L1 see Figure 3.4; L2 see Figure 3.6). All subcategories of both languages, as well as overall scores of L1 and L2, met the requirements of a normal distribution. Transformation of scores was considered but due to a decrease in normality within all scores with square root transformations, log as well as ln transformation, original scores were used for all subsequent calculations.

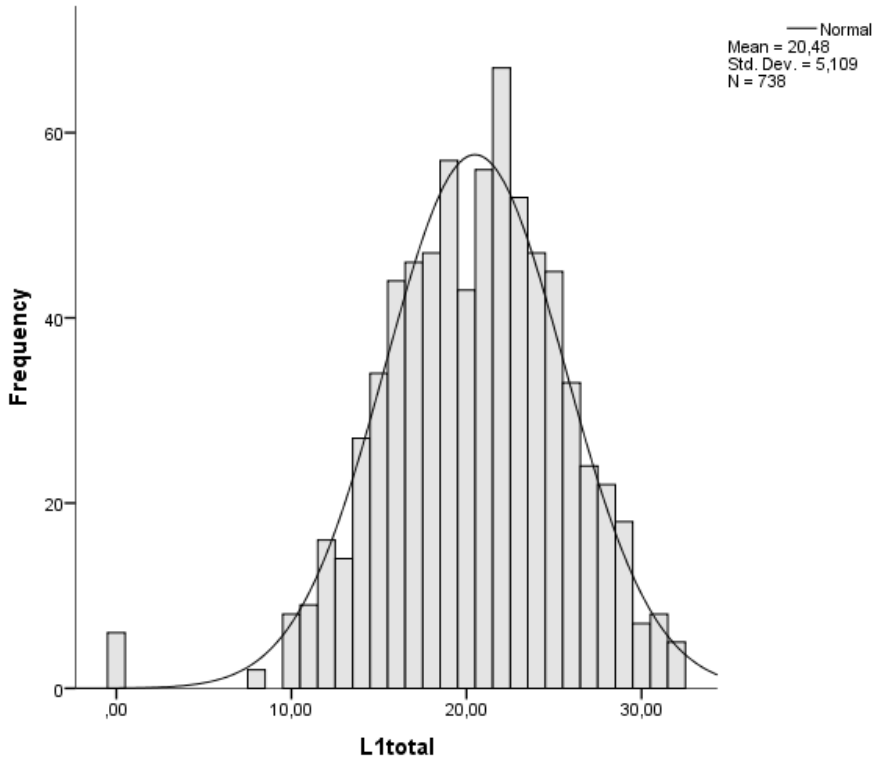


Figure 3.3. Histogram showing normality of data for total scores of L1.

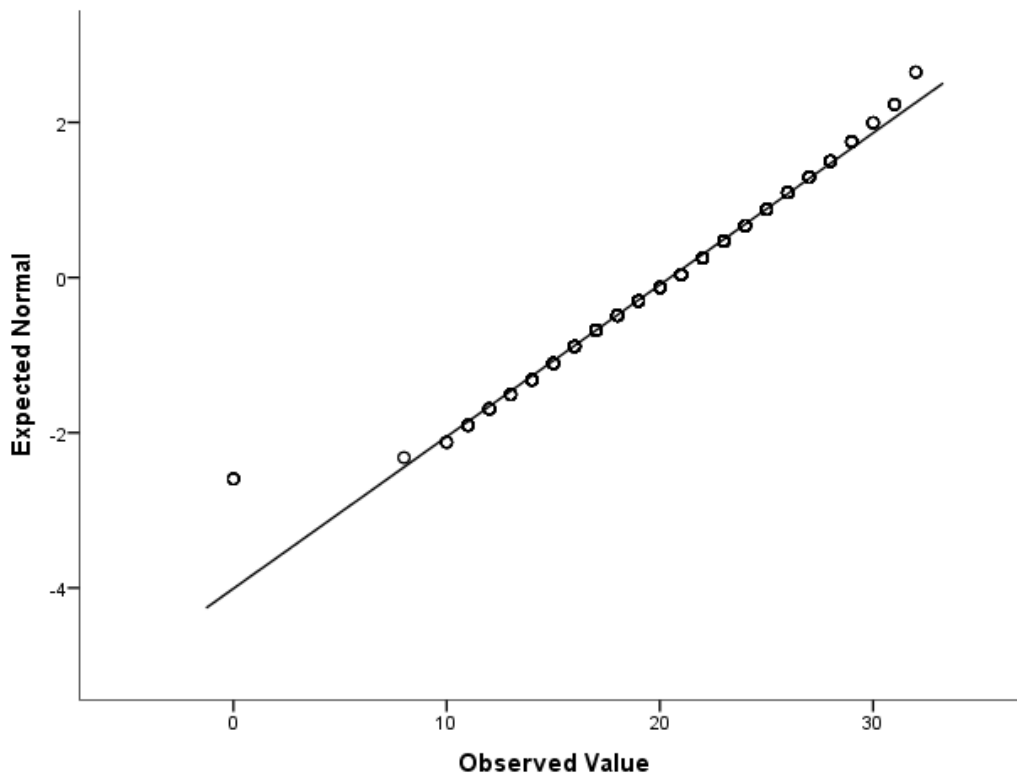


Figure 3.4. QQ-plot showing normality of data for total scores of L1.

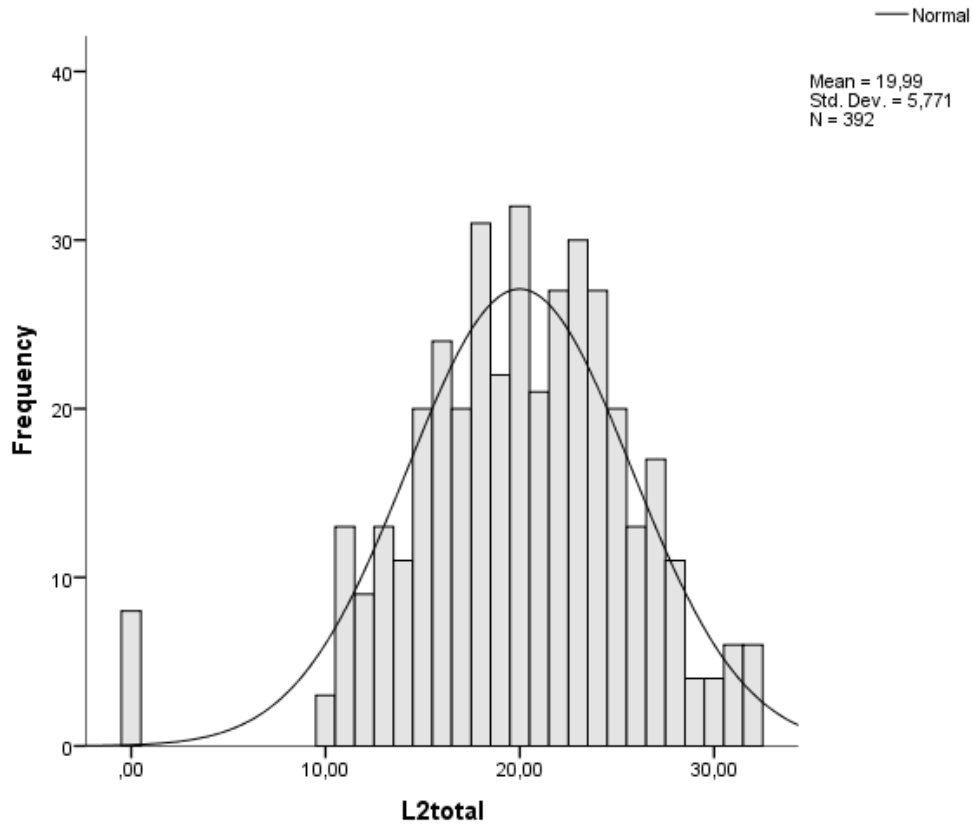


Figure 3.5. Histogram showing normality of data for total scores of L2.

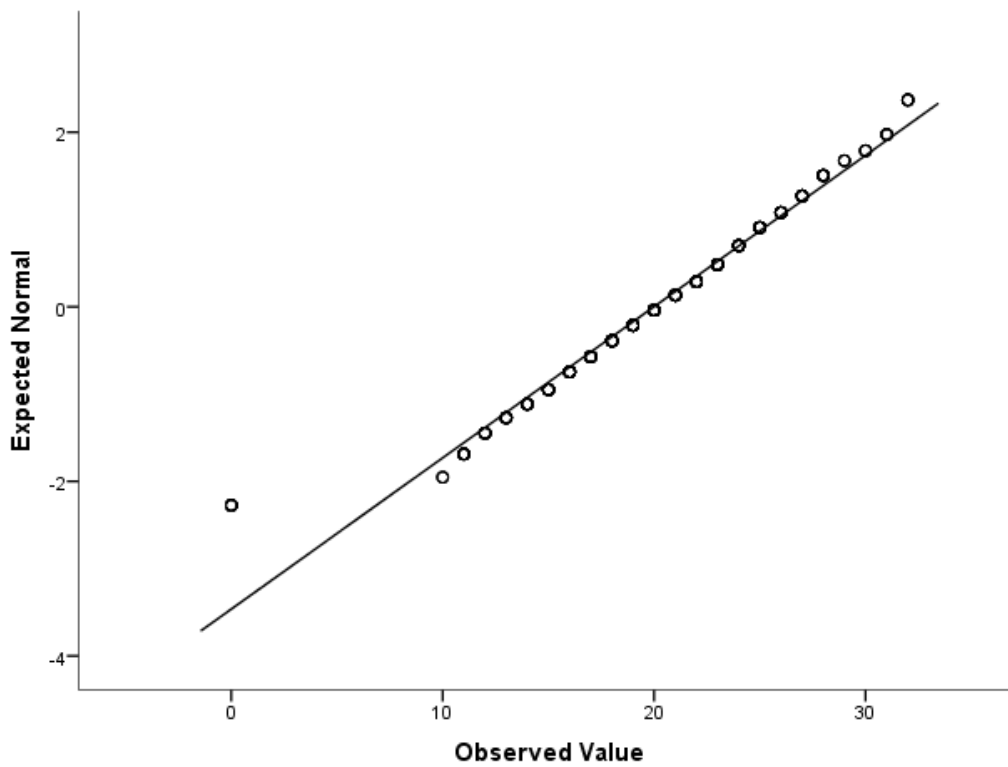


Figure 3.6. QQ-plot showing normality of data for total scores of L2.

3.6.1.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($N=345$) and bilingual ($N=392$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p>.05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances within three of the five subcategories, HT, AA, and PS. For a detailed overview divided by language group see Table 3.11.

Table 3.11. Between-Group Comparisons: MANOVA comparing Monolinguals and Bilinguals divided by Language Group

Language Group	F(df); p	Wilk's Λ	Partial η^2	Power
English-Irish	F(5,108)=1.00; n.sig.	.96	.04	.34
German-English	F(5,111)=1.90; n.sig.	.92	.08	.63
Italian-English	F(5,80)=.73; n.sig.	.96	.04	.25
Chinese-English	F(5,58)=1.18; n.sig.	.91	.09	.39
Chinese-Japanese	F(5,59)=3.48; $p<.01$.77	.23	.89
Japanese-English	F(5,52)=4.03; $p<.005$.72	.28	.93
Turkish-English	F(5,82)=3.81; $p<.005$.81	.19	.92
Turkish-German	F(5,82)=7.21; $p<.001$.70	.31	1.00
All merged	F(5,731)=16.75; $p<.001$.90	.10	1.00

Note. Russian-English excluded as a result of low numbers within the monolingual group ($N_{Mono}(Russian)=9$).

A statistically significant difference was discovered in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,731)=16.73$, $p<.001$; Wilk's $\Lambda<.001$, *partial* $\eta^2=.10$). An overview of results divided by all language groups can be seen in Table 3.11. Due to the significant result, further follow-up tests are reported (see Table 3.12). A significant difference in adjusted means between both independent groups, mono- and bilinguals, and the levels of categorical predictor variables on the outcome can be observed within three of the five subcategories, namely HT ($F=3.97$; $p<.05$; *partial* $\eta^2=.01$; *power*=.51), VR ($F=56.62$; $p<.001$; *partial* $\eta^2=.07$; *power*=1.00) and JL ($F=11.10$; $p>.005$; *partial* $\eta^2=.02$; *power*=.91). For all significantly differing subcategories, bilinguals achieved higher scores compared to monolinguals (HT_{mono}:

$M=2.65$; $SE=.07$; HT_{biling} : $M=2.84$; $SE=.07$; VR_{mono} : $M=5.26$; $SE=.11$; VR_{biling} : $M=6.38$; $SE=.10$; JL_{mono} : $M=2.11$; $SE=.06$; JL_{biling} : $M=2.36$; $SE=.05$). Bonferroni corrections for multiple testing led to the implementation of alpha levels of $\alpha_{\text{corrected}}=(.05/5)=.001$, which will be used henceforth. Consequently, scores of HT and JL did not reach significance after corrected for multiple testing.

Table 3.12. F-Values and Means (if significant) comparing Monolinguals and Bilinguals divided by Sub-Category and Language Group (Only including Groups with a significant Overall Difference)

Language	F_{HT} (M_{mono} ; M_{bi})	F_{VR} (M_{mono} ; M_{bi})	F_{JL} (M_{mono} ; M_{bi})	F_{AA} (M_{mono} ; M_{bi})	F_{PS} (M_{mono} ; M_{bi})
Chinese-Japanese	$F=.09$; n.sig.	$F=.04$; n.sig.	$F=17.83$; $p<.001$ (1.66; 2.55)	$F=.85$; n.sig.	$F=2.45$; n.sig.
Japanese-English	$F=9.54$; $p<.005$ (3.76; 2.67)	$F=2.44$; n.sig.	$F=2.62$; n.sig.	$F=5.87$; n.sig.	$F=.98$; n.sig.
Turkish-English	$F=3.94$; n.sig.	$F=11.51$; $p<.005$ (4.12; 5.70)	$F=2.80$; n.sig.	$F=.56$; n.sig.	$F=9.34$; $p<.005$ (4.31; 6.05)
Turkish-German	$F=.22$; n.sig.	$F=24.08$; $p<.001$ (4.07; 6.33)	$F=.14$; n.sig.	$F=.09$; n.sig.	$F=3.61$; n.sig.
All merged	$F=3.97$; n.sig.	$F=56.62$; $p<.001$ (5.26; 6.38)	$F=11.08$; $p<.005$ (2.11; 2.36)	$F=2.86$; n.sig.	$F=.04$; n.sig.

Note. Mono=Monolingual; Bi=Bilingual.

An additional analysis only including monolinguals without any L2 knowledge ($n=266$) and highly proficient bilinguals ($n=204$) showed similar results. The two groups differed significantly ($F(5; 464)=21.57$; $p<.001$) with scores of VR showing variations between monolinguals and bilinguals ($F=85.78$; $p<.001$). Bilinguals scored better than monolinguals (VR : $M_{\text{Mono}}=5.11$; $SE_{\text{Mono}}=.13$; $M_{\text{Bi}}=6.76$; $SE_{\text{Bi}}=.12$).

For a further, more detailed analysis of differences, participants were split into four groups dependent on fluency. The first group consisted of participants with no experience in any L2 ($n=266$), the second group were monolingual with a restricted amount of second language experience ($n=79$), group three consisted of bilinguals with a mediocre L2 proficiency ($n=188$) and group four represented highly proficient bilinguals ($n=204$). A MANOVA was implemented to observe possible differences within the groups in all sub-categories. A significant difference was found within two subcategories, VR ($F=27.07$; $p<.001$) and JL ($F=4.02$; $p<.01$). When examining results, VR, stands out, with a constant improvement of scores with increasing L2 proficiency, from the first group with the lowest scores constantly improving up to the highly proficient group scoring highest on VR.

3.6.1.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($N=392$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT as can be seen in detail in Table 3.13. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories HT, VR, AA, JL, and PS. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,387)=46.78$; $p < .001$; *partial* $\eta^2 = .38$; *power* = 1.00), suggesting further analysis of each dependent variable.

Table 3.13. Within Group Comparison: reMANOVA comparing L1 and L2 divided by Language Group

Language	Wilk's Λ	Greenhouse-Geisser	<i>partial</i> η^2	<i>power</i>
English-Irish	.53	$F(5,54)=9.45$; $p < .001$.47	1.00
German-English	.68	$F(5,92)=8.82$; $p < .001$.32	1.00
Italian-English	.48	$F(5,58)=12.61$; $p < .001$.52	1.00
Russian-English	.40	$F(5,43)=13.18$; $p < .001$.61	1.00
Chinese-English	.48	$F(5,25)=5.35$; $p < .005$.52	.97
Chinese-Japanese	.41	$F(5,28)=8.01$; $p < .001$.59	1.00
Japanese-English	.56	$F(5,16)=2.56$; n.sig.	.45	.64
Turkish-English	.42	$F(5,15)=4.13$; $p < .05$.58	.85
Turkish-German	.36	$F(5,16)=5.80$; $p < .005$.64	.96
All merged	.62	$F(5,387)=46.78$; $p < .001$.38	1.00

Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations of this kind due to its conservative nature. Values for two of the five subcategories, HT and PS, did not reach significance ($p > .01$), whereas scores for VR, JL, and AA show to be significantly different in scores of L1 and L2. VR ($F=128.71$; $p < .001$; *partial* $\eta^2 = .25$; *power* = 1.00) was absolved significantly better in L1 ($M=6.38$; $SE=.10$) compared to L2 ($M=5.14$; $SE=.11$); as well as AA ($F=7.35$; $p < .01$; *partial* $\eta^2 = .02$; *power* = .77) with L1 ($M=4.78$; $SE=.07$) being significantly higher than L2 ($M=4.54$; $SE=.08$). Scores of JL ($F=103.95$; $p < .001$; *partial* $\eta^2 = .21$; *power* = 1.00) show the opposite trend, with participants solving tasks better in L2 ($M=2.94$; $SE=.06$) than in L1 ($M=2.36$; $SE=.05$). Details divided by language group can be seen in Table 3.14.

Table 3.14. F-Values and Means (if significant) comparing L1 and L2 divided by Sub-Category and Language Group

Language	F _{HT} (M _{L1} ; M _{L2})	F _{VR} (M _{L1} ; M _{L2})	F _{JL} (M _{L1} ; M _{L2})	F _{AA} (M _{L1} ; M _{L2})	F _{PS} (M _{L1} ; M _{L2})
English-Irish	F=.23; n.sig.	F=38.93; p<.001 (6.02; 4.34)	F=4.49; n.sig.	F=4.79; n.sig.	F=.55; n.sig.
German-English	F=1.08; n.sig.	F=28.23; p<.001 (6.54; 5.41)	F=29.47; p<.001 (2.29; 2.85)	F=.77; n.sig.	F=.01; n.sig.
Italian-English	F=.00; n.sig.	F=11.47; p<.005 (6.40; 5.44)	F=54.84; p<.001 (2.25; 3.10)	F=.28; n.sig.	F=.52; n.sig.
Russian-English	F=1.15; n.sig.	F=59.05; p<.001 (7.71; 5.96)	F=7.74; p<.01 (2.56; 3.00)	F=2.96; n.sig.	F=.00; n.sig.
Chinese-English	F=7.44; n.sig.	F=9.67; p<.005 (5.67; 4.13)	F=1.24; n.sig.	F=4.36; n.sig.	F=5.83; n.sig.
Chinese-Japanese	F=.31; n.sig.	F=18.67; p<.001 (5.79; 4.39)	F=15.13; p<.001 (2.55; 3.30)	F=3.25; n.sig.	F=.97; n.sig.
Turkish-English	F=.25; n.sig.	F=1.45; n.sig.	F=22.30; p<.001 (2.60; 3.50)	F=.53; n.sig.	F=4.13; n.sig.
Turkish-German	F=2.48; n.sig.	F=.12; n.sig.	F=16.60; p<.005 (2.10; 3.05)	F=.74; n.sig.	F=.88; n.sig.
All merged	F=.83; n.sig.	F=128.71; p<.001 (6.38; 5.14)	F=103.95; p<.001 (2.36; 2.94)	F=7.35; p<.01 (4.78; 4.54)	F=3.95; n.sig.

Note. Japanese-English does not show an overall significant difference (see Table 3.13).

Above mentioned calculations are based on the inclusion of all bilingual participants ($N=392$). When only including highly proficient participants ($n=204$) a highly significant difference between first and second language performance was still present ($F(5, 199)=28.76; p<.001$). Measures of VR were robustly higher ($F=42.97; p<.001$) in L1 ($M=6.76; SE=.12$) than L2 ($M=5.79; SE=.14$). JL ($F=106.41; p<.001$) showed the opposite trend with better scores within L2 ($M=3.07; SE=.08$) compared to L1 ($M=2.33; SE=.07$). Differences within AA were non-significant ($p>.01$).

When splitting all bilingual participants dependent on their age of L2 acquisition, similar tendencies could be observed between early bilinguals who learned their L2 before the age of ten ($n=68$) and late bilinguals, who acquired L2 after the age of ten

($n=324$). Significant differences could be observed within the early bilingual group ($F(5;63)=11.15$; $p<.001$) and the late bilingual group ($F(5;319)=36.75$; $p<.001$). Parallel to above-mentioned results, early bilinguals showed a significant difference in VR ($F=21.97$; $p<.001$) with a better performance in L1 ($M=6.03$; $SE=.26$) than in L2 ($M=4.68$; $SE=.27$) but the opposite trend within the sub-category JL with better scores in L2 ($M=3.03$; $SE=$) than L1 ($M=2.19$; $SE=.12$). None of the other three groups, HT, AA, and PS, differed significantly. Late bilinguals showed the same range of differences with two categories entailing significances, VR ($F=106.72$; $p<.001$) and JL ($F=69.45$; $p<.001$) comparable to early bilinguals. L1 scores were higher compared to L2 scores in VR ($M_{L1}=6.45$; $SE_{L1}=.10$; $M_{L2}=5.23$; $SE_{L2}=.12$), whereas within JL, L2 scores ($M=2.93$; $SE=.07$) were found to be better than L1 ($M=2.40$; $SE=.06$).

The final calculation was implemented to compare overall scores, as well as scores for each subcategory of CT, of each language group. For an overview of outcomes see Table 3.15. There was a statistically significant difference in both scores, L1 and L2, based on participants' language set (L1: $F(40,3163)=6.87$, $p < .001$; *Wilk's* $\Lambda = .70$, *partial* $\eta^2 = .07$; L2: $F(40,1655)=3.06$, $p < .001$; *Wilk's* $\Lambda = .73$, *partial* $\eta^2 = .06$). Differences of scores between language groups within L1 could be observed both overall, with scores on L1total varying significantly ($F=10.80$; $p<.001$), and in the subcategories, HT ($F=9.64$; $p<.001$), VR ($F=64.94$; $p<.001$), and PS ($F=9.76$; $p<.001$). Similarly, differences of scores between language groups within L2 could be observed overall, with scores on L2total varying significantly ($F=4.83$; $p<.001$), and in the subcategories VR ($F=4.36$; $p<.001$) and PS ($F=4.83$; $p<.001$). Table 3.15 contains scores divided by language set, for each subcategory, presented as achieved by monolinguals, bilinguals in their L1 and bilinguals in L2. When assessing bilinguals' L2, an additional calculation was executed, only including highly proficient bilinguals. The overall MANOVA was still significant ($F(40,835)=2.13$; $p<.001$), but a difference in overall scores of L2total was not present ($F=1.71$; $p>.05$). Two subcategories showed significant differences depending on the language group, VR ($F=2.85$; $p<.01$) and PS ($F=2.88$; $p<.01$).

Table 3.15. Overview of average Scores (M) for each Subcategory, divided by Language Group and Mono- and Bilinguals (First and Second Language)

		English-Irish	German-English	Italian-English	Russian-English	Japanese-English	Chinese-English	Chinese-Japanese	Turkish-English	Turkish-German	ALL MERGED
HT	Mono	2.93	2.45	2.78	2.44	3.76	2.85	2.81	2.18	2.13	2.65
	Bi L1	2.90	2.87	2.95	3.19	2.67	2.50	2.73	2.80	2.29	2.84
	Bi L2	2.80	2.73	2.95	3.02	2.91	2.03	2.58	2.95	2.91	2.78
VR	Mono	5.84	5.90	6.48	5.67	6.70	5.68	5.69	4.12	4.08	5.26
	Bi L1	6.03	6.54	6.40	7.71	6.14	5.67	5.79	5.70	6.33	6.38
	Bi L2	4.34	5.41	5.44	5.96	5.00	4.13	4.39	5.00	6.14	5.14
JL	Mono	2.26	2.05	2.00	2.22	2.43	1.71	1.66	2.19	2.19	2.11
	Bi L1	2.24	2.29	2.25	2.56	2.86	2.20	2.55	2.60	2.10	2.36
	Bi L2	2.64	2.85	3.10	3.00	3.14	2.47	3.30	3.50	3.05	2.94
AA	Mono	5.20	4.95	4.91	5.00	4.95	5.03	5.09	4.82	4.82	4.96
	Bi L1	4.76	4.44	5.06	5.19	4.14	5.00	4.76	5.10	4.71	4.78
	Bi L2	4.25	4.29	4.95	4.79	4.95	4.23	4.12	4.75	5.10	4.54
PS	Mono	4.20	4.15	4.70	4.33	6.57	5.35	5.31	4.31	4.27	4.74
	Bi L1	3.81	4.33	4.18	5.63	5.95	5.77	6.27	6.05	3.24	4.78
	Bi L2	3.66	4.32	4.35	5.63	5.52	4.57	5.85	5.30	3.43	4.60

Note. Mono=Monolinguals; Bi=Bilinguals; HT=hypothesis testing; VR=verbal reasoning; JL=judging likelihood and probability; AA=argumentation analysis; PS=problem solving.

3.6.2. L1 English- L2 Irish

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (English) reached a value of $\alpha(L1)=.51$. Comparable analyses with outcomes for L2 (Irish) showed a value for Cronbach's alpha of $\alpha(L2)=.57$.

Multiple regression was calculated to predict L1total (English) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1total was significant ($F(3,110)=7.33$; $p<.001$; $R^2=.17$), explaining 17% of the overall variance in the model. Participants' predicted L1total score was equal to $9.72 + .082 (AGE) + 1.43 (GENDER) + .39 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. Years of education were found to be a significant predictor of L1 ($p<.001$), with participants' overall score of L1 increasing by .38 points per additional year of education. Multiple regression was calculated to predict L2total based on age, gender, and education of participants using the Enter method. Using

regression analysis, the overall model to predict L2 was non-significant with age, gender, and education not affecting outcomes within the variable.

3.6.2.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=55$) and bilingual ($n=59$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p>.05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances.

There was no statistically significant difference in CT sub scores based on participants' language setup, monolingual or bilingual ($F(5,108)=1.00$, $p>.01$; *Wilk's* $\Lambda=.96$, *partial* $\eta^2=.04$). Significant differences were also absent when only including monolinguals that have never been in touch with an L2 ($n=45$) and highly proficient bilinguals ($n=32$). Due to the non-significant result, no further follow-up tests are reported.

3.6.2.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=59$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda<.05$ ($F(5,54)=9.45$; $p<.001$; *partial* $\eta^2=.47$; *power*=1.00), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations of this kind due to its conservative nature. Values for four of the five subcategories, HT, JL, AA, and PS, did not reach significance ($p>.01$), whereas scores for VR showed to be significantly different in scores of L1 and L2 ($F=38.93$; $p<.001$; *partial* $\eta^2=.40$) with significantly better performance in L1 ($M=6.03$; $SE=.27$) compared to L2 ($M=4.34$; $SE=.29$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=59$). When only including highly proficient participants ($n=32$) a highly significant difference between first and second language performance could still be found ($F(5, 40)=5.50$; $p<.005$). Measures of VR were solved highly significantly better

($F=17.71$; $p<.001$) in L1 ($M=6.56$; $SE=.32$) than L2 ($M=5.06$; $SE=.42$). Measures of JL in comparison were solved significantly better ($F=12.24$; $p<.005$) in L2 ($M=2.94$; $SE=.21$) compared to L1 ($M=2.16$; $SE=.17$)

3.6.3. L1 German- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (German) reached a value of $\alpha(L1)=.74$. Comparable analyses with outcomes for L2 (English) showed a value for Cronbach's alpha of $\alpha(L2)=.77$.

Multiple regression was undertaken to predict L1total (German) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,113)=1.03$; $R^2=.03$), explaining 3% of the overall variance in the model. Participants predicted L1total score was equal to $16.95 - .01 (AGE) - .53 (GENDER) + .31 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. No significant value was present. Multiple regression calculations including L2total (English) similarly led to no significant results with age, gender, and education not affecting outcomes within the variable.

3.6.3.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=20$) and bilingual ($n=97$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices is non-significant on a $p<.001$ level, confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances.

There was no statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,111)=1.90$, $p>.05$; *Wilk's* $\Lambda=.92$, *partial* $\eta^2=.08$). Due to the non-significant result, no further follow-up tests were reported. A restricted number of monolingual participants who were never in contact with any foreign language ($n=8$) within this language set made a comparison of those individuals and highly proficient bilinguals for further exploration of differences unreasonable.

3.6.3.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=97$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable described whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,92)=8.82$; $p < .001$; $partial \eta^2 = .32$; $power = 1.00$), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations of this kind due to its conservative nature. Values for three of the five subcategories, HT, AA, and PS, did not reach significance ($p > .01$), whereas scores for VR and JL showed to be significantly different in scores of L1 and L2. VR ($F=28.23$; $p < .001$; $partial \eta^2 = .28$) with significantly better performance in L1 ($M=6.54$; $SE=2.23$) compared to L2 ($M=5.41$; $SE=2.48$). Scores of JL ($F=29.47$; $p < .001$; $partial \eta^2 = .24$) showed the opposite trend, with participants solving tasks better in L2 ($M=2.85$; $SE=1.33$) than in L1 ($M=2.29$; $SE=1.15$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=97$). When only including highly proficient participants ($n=58$) a highly significant difference between first and second language performance could still be observed ($F(5, 53)=4.11$; $p < .005$), but only within results of JL ($F=19.46$; $p < .001$). Participants score higher within L2 ($M=2.97$; $SE=.16$) compared to L1 ($M=2.36$; $SE=.14$).

3.6.4. L1 Italian- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Italian) reached a value of $\alpha(L1) = .31$. Comparable analyses with outcomes for L2 (English) showed a value for Cronbach's alpha of $\alpha(L2) = .53$. Multiple regression was calculated to predict L1total (Italian) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was non-significant ($F(3,82) = .94$; $p < .05$; $R^2 = .03$), explaining 3% of the overall variance in the model. Participants predicted L1total score was equal to $17.34 - .02 (AGE) + 1.15 (GENDER) + .16 (EDU)$, where age and education were measured in years and gender was coded

1=female, 2=male. No covariate was found to be a significant predictor of L1total. Similarly, multiple regression calculations including L2total (English) led to no significant results with age, gender, and education not affecting outcomes within the variable.

3.6.4.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=23$) and bilingual ($n=63$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p>.05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances.

There was no statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,80)=.73$, $p>.01$; *Wilks's Λ* =.96, *partial η^2* =.04). Significant differences were also absent when only including monolinguals that have never been in touch with an L2 ($n=16$) and highly proficient bilinguals ($n=38$). Due to the non-significant result, no further follow-up tests are reported.

3.6.4.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=63$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance as below the critical value of $\Lambda<.05$ ($F(5,58)=12.61$; $p<.001$; *partial η^2* =.52; *power*= 1.00), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were due to its conservative nature considered as appropriate for calculations of this kind. Values for three of the five subcategories, HT, AA, and PS, did not reach significance ($p>.01$), whereas scores for VR and JL ($F=54.84$; $p<.001$; *partial η^2* =.47) whereas scores for VR and JL showed to be significantly different in L1 and L2. VR ($F=11.47$; $p<.005$; *partial η^2* =.16) with significantly better performance in L1 ($M=6.40$; $SE=1.33$) compared to L2 ($M=5.44$; $SE=1.86$). Scores of JL ($F=54.84$; $p<.001$; *partial η^2* =.47) showed the

opposite trend, with participants solving tasks better in L2 ($M=3.10$; $SE=1.00$) than in L1 ($M=2.25$; $SE=.92$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=63$). When only including highly proficient participants ($n=38$) a highly significant difference between first and second language performance could still be found ($F(5, 33)=9.51$; $p<.001$). Scores for JL ($F=39.80$; $p<.001$) were significantly higher within L2 ($M=3.12$; $SE=.18$) compared to L1 ($M=2.26$; $SE=.13$).

3.6.5. L1 Russian- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Russian) reached a value of $\alpha(L1)=.42$. Comparable analyses with outcomes for L2 (English) showed a value for Cronbach's alpha of $\alpha(L2)=.45$.

Multiple regression was calculated to predict L1total (Russian) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,53)=5.17$; $p<.005$) with $R^2=.23$, explaining 23% of the overall variance in the model. Participants' predicted L1total score was equal to $18.04 - .02 (AGE) + 4.67 (GENDER) + .02 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. Gender was found to be a significant predictor of L1total ($p<.001$), with participants' overall score of L1total increasing by 4.67 points dependent on gender. Males scored significantly higher than females. It is important to mention that within the Russian English language set the monolingual group had an insufficient group size of $N=9$. Multiple regression calculations including L2total (English) led to no significant results with age, gender and education not affecting outcomes within the variable.

3.6.5.1. Comparison mono- versus bilinguals (between comparison)

A comparison of monolinguals and bilinguals was considered inappropriate due to insufficient participation numbers within the monolingual group ($n=9$).

3.6.5.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=48$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in

their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,43)=13.18$; $p < .001$; *partial* $\eta^2=.61$; *power*=1.00), suggesting further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were due to its conservative nature considered as appropriate for calculations of this kind. Values for three of the five subcategories, HT, AA, and PS, did not reach significance ($p > .01$), whereas scores for VR and JL showed to differ significantly in scores of L1 and L2. VR ($F=59.05$; $p < .001$; *partial* $\eta^2=.56$) with significantly better performance in L1 ($M=7.71$; $SE=1.22$) compared to L2 ($M=5.96$; $SE=1.57$). Scores of JL ($F=7.74$; $p < .01$; *partial* $\eta^2=.14$) showed the opposite trend, with participants solving tasks better in L2 ($M=3.00$; $SE=.99$) than in L1 ($M=2.56$; $SE=1.17$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=48$). When only including highly proficient participants ($n=33$) a highly significant difference between first and second language performance could still be found ($F(5, 28)=8.53$; $p < .001$). Performance in VR was highly significantly better ($F=38.11$; $p < .001$) in L1 ($M=7.76$; $SE=.22$) compared to L2 ($M=6.21$; $SE=.26$).

3.6.6. L1 Japanese- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Japanese) reached a value of $\alpha(L1)=.52$. Comparable analyses with outcomes for L2 (English) showed a value for Cronbach's alpha of $\alpha(L2)=.55$.

Multiple regression was calculated to predict L1total (Japanese) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was non-significant ($F(3,54)=2.41$; $p > .05$) with $R^2=.12$). Participants' predicted L1total score was equal to $19.89 - .08 (AGE) + 2.12 (GENDER) + .17 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. None of the covariates was found to be a significant predictor of L1total. Similarly, multiple regression calculations including L2total (English) led to no significant results with age, gender, and education not affecting outcomes within the variable.

3.6.6.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=37$) and bilingual ($n=21$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language

background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p > .05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p > .05$) meeting the assumption of homogeneity of variances.

There was a statistically significant difference in CT sub scores based on participants' language setup, monolingual or bilingual ($F(5,52)=4.03$, $p < .005$; *Wilk's* $\Lambda = .72$, *partial* $\eta^2 = .28$). Significant values within the subcategories could be observed within scores of HT ($F=9.54$; $p < .005$) with monolinguals scoring significantly higher ($M=3.76$; $SE=.21$) than bilinguals ($M=2.67$; $SE=.28$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of those individuals and monolingual individuals who were never in touch with another L2 for further exploration of differences unreasonable.

An additional MANOVA was undertaken with gender as a fixed factor, comparing all male ($n=27$) participants with female participants ($n=31$) on the five subcategories of CT. Differences were non-significant ($F(5; 52)=1.27$; $p > .05$) whereas tendencies showed a better performance of male test-takers as can be seen in Table 3.16. Even though the overall score showed a non-significant result, one of the five sub-categories, PS, entailed a significant result ($F=5.61$; $p < .05$) with males reaching an average score of $M=7.07$ ($SE=.42$) and females ($M=5.71$; $SE=.39$).

Table 3.16. Overview of average scores (M) including all Japanese Participants ($n=58$) divided by Gender ($n_{female}=31$; $n_{male}=27$) for each Subcategory

Sub-category	Gender	M	SE
HT	female	3.16	.25
	male	3.59	.27
VR	female	6.45	.24
	male	6.56	.26
JL	female	2.42	.17
	male	2.78	.19
AA	female	4.55	.23
	male	4.78	.24
PS	female	5.71	.39
	male	7.07	.42

Note. HT=hypothesis testing; VR=verbal reasoning; JL=judging likelihood and probability; AA=argumentation analysis; PS=problem solving.

When dividing all participants by language setup, monolingual ($n_{female}=15$; $n_{male}=22$) and bilingual ($n_{female}=16$; $n_{male}=05$), no significant scores could be observed ($F_{Mono}(5; 31)=.92$; $p>.05$; $F_{Bi}(5; 15)=1.14$; $p>.05$).

Exclusion of all participants currently living outside of Japan led to significant differences between monolinguals ($n=34$) and bilinguals ($n=13$; $F(5;41)=2.96$; $p<.05$), within HT ($F=8.84$; $p<.01$) with monolinguals ($M=3.85$; $SE=.22$) outperforming bilinguals ($M=2.62$; $SE=.35$). None of the other four subcategories showed significant differences.

3.6.6.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=21$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2); dependent scores were the five subcategories. The ratio of unexplained variance to total variance was above the critical value of $\Lambda<.05$ ($F(5,16)=2.56$; $p=.069$.; $partial \eta^2=.45$; $power=.64$). To investigate the Japanese-English data set more thoroughly, scores of sub-categories are going to be reported: None of the sub-categories showed significant differences between first and second language ($p>.01$), but two of the five sub-categories, VR and AA, show tendencies towards significance. Scores in VR ($F=5.57$; $p=.03$; $partial \eta^2=.22$) are better in L1 ($M=6.14$; $SE=.28$) compared to L2 ($M=5.00$; $SE=.42$). Scores in AA ($F=7.39$; $p=.01$; $partial \eta^2=.27$) on the opposite are better in L2 ($M=4.95$; $SE=.25$) compared to L1 ($M=4.14$; $SE=.22$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=21$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of L1 and L2 of those individuals for further exploration of differences unreasonable.

3.6.7. L1 Chinese- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Chinese) reached a value of $\alpha(L1)=.42$. Comparable analyses with outcomes for L2 (English) show a value for Cronbach's alpha of $\alpha(L2)=.68$.

Multiple regression was calculated to predict L1total (Chinese) based on age, gender, and education of participants using the Enter method. Using regression

analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,60)=3.23$; $p<.05$) with $R^2=.14$), explaining 14% of the overall variance in the model. Participants' predicted L1total score was equal to $14.11 - .14$ (AGE) - $.42$ (GENDER) + $.71$ (EDU), where age and education were measured in years and gender was coded 1=female, 2=male. Years of education were found to be a significant predictor of L1total ($p<.005$), with participants' overall score of L1total increasing by $.71$ points per additional year of education. Multiple regression calculations including L2total (English) led to no significant results with age, gender, and education, not affecting outcomes within the variable.

3.6.7.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=34$) and bilingual participants ($n=30$) regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p>.05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances.

There was no statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,58)=1.18$; $p>.05$; *Wilk's* $\Lambda=.91$; *partial* $\eta^2=.09$; *power*=.39). Due to the non-significant result, no further follow-up tests are reported. A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of those individuals and monolingual individuals who were never in touch with another L2 for further exploration of differences unreasonable.

3.6.7.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=30$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda<.05$ ($F(5,25)=5.35$; $p<.005$; *partial* $\eta^2=.52$; *power*=.97), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations of this kind due to its conservative nature. Values for

four of the five subcategories, HT, JL, AA, and PS, did not reach significance ($p > .01$), whereas outcomes for VR ($F=35.27$; $p < .01$) with significantly better performance in L1 ($M=5.67$; $SE=.30$) compared to L2 ($M=4.13$; $SE=.46$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=30$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of L1 and L2 of those individuals for further exploration of differences unreasonable.

3.6.8. L1 Chinese- L2 Japanese

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Chinese) reached a value of $\alpha(L1)=.41$. Comparable analyses with outcomes for L2 (Japanese) showed a value for Cronbach's alpha of $\alpha(L2)=.65$.

Multiple regression was calculated to predict L1total (Chinese) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was non-significant ($F(3,61)=.66$; $p > .05$; $R^2=.03$). Participants' predicted L1total score was equal to $18.40-.11 (AGE) - .90 (GENDER) + .44 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. None of the covariates was found to be a significant predictor of L1total. Multiple regression calculations including L2total (Japanese) led to no significant results with age, gender, and education, not affecting outcomes within the variable.

3.6.8.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=32$) and bilingual ($n=33$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p > .05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p > .05$) meeting the assumption of homogeneity of variances.

There was a statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,59)=3.48$; $p < .01$; *Wilk's* $\Lambda=.77$; *partial* $\eta^2=.23$; *power*=.89). Within the subcategories, HT ($F=.09$; $p > .01$), VR ($F=.04$; $p > .01$), AA ($F=.85$; $p > .01$) and PS ($F=2.45$; $p > .01$) did not differ significantly, whereas within scores for JL ($F=17.83$; $p < .001$) a difference between monolingual and

bilingual participants could be observed, with bilinguals ($M=2.55$; $SE=.15$) outperforming monolingual individuals ($M=1.66$; $SE=.15$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of those individuals and monolingual individuals who were never in touch with another L2 for further exploration of differences unreasonable.

3.6.8.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=33$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,28)=8.01$; $p < .001$; $partial \eta^2 = .59$; $power = 1.00$), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations of this kind due to its conservative nature. Values for three of the five subcategories, HT, AA, and PS, did not reach significance ($p > .01$), whereas scores for VR ($F=18.67$; $p < .001$) and JL ($F=15.13$; $p < .001$) with significantly different performance in L1 and L2. VR was solved better in L1 ($M=5.79$; $SE=.35$) compared to L2 ($M=4.39$; $SE=.33$), JL was solved better in L2 ($M=3.30$; $SE=.20$) compared to L1 ($M=2.55$; $SE=.15$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=33$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of L1 and L2 of those individuals for further exploration of differences unreasonable.

3.6.9. L1 Turkish- L2 English

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Turkish) reached a value of $\alpha(L1) = .51$. Comparable analyses with outcomes for L2 (English) showed a value for Cronbach's alpha of $\alpha(L2) = .49$.

Multiple regression was calculated to predict L1total (Turkish) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,85)=6.26$; $p < .005$; $R^2 = .18$), explaining 18% of the overall variance in the model. Participants predicted L1total score was equal to $6.14 + .21 (AGE) + .97$

(GENDER) + .46 (EDU), where age and education were measured in years and gender was coded 1=female, 2=male. Years of education were found to be a significant predictor of L1total ($p < .05$), with participants' overall score of L1total increasing by .38 points per additional year of education. Multiple regression calculations including L2total (English) led to no significant results with age, gender, and education not affecting outcomes within the variable.

3.6.9.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=68$) and bilingual ($n=20$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p > .05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p > .05$) meeting the assumption of homogeneity of variances.

There is a statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,82)=3.81$; $p < .005$; *Wilk's* $\Lambda = .81$; *partial* $\eta^2 = .19$; *power* = .92). A difference in scores could be observed in two of the five subcategories, VR ($F=11.51$; $p < .005$) and PS ($F=9.34$; $p < .005$), with bilinguals (VR: $M=5.70$; $SE=.41$; PS: $M=6.05$; $SE=.50$) outperforming monolinguals (VR: $M=4.12$; $SE=.22$; PS: $M=4.31$; $SE=.27$) in both sections. A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of those individuals and monolingual individuals who were never in touch with another L2 for further exploration of differences unreasonable.

3.6.9.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=20$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable was whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,15)=4.13$; $p < .05$; *partial* $\eta^2 = .58$; *power* = .85), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations due to its conservative nature. Values for four of the five subcategories, HT, VR, AA, and PS, did not reach significance ($p > .01$), whereas

scores for JL showed to be significantly different depending on the language L1 and L2. JL ($F=22.30$; $p<.001$; partial $\eta^2=.54$) with significantly better performance in L2 ($M=3.50$; $SE=.69$) compared to L1 ($M=2.60$; $SE=.60$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=20$). A restricted number of highly proficient bilingual participants ($n=6$) within this language set made a comparison of L1 and L2 of those individuals for further exploration of differences unreasonable.

3.6.10. L1 Turkish- L2 German

An item analysis of all cases was implemented to assess scale reliability. When including all cases on all categories, Hypothesis Testing (HT), Verbal Reasoning (VR), Judging Likelihood (JL), Argumentation Analysis (AA), and Problem Solving (PS), Cronbach's alpha of L1 (Turkish) reached a value of $\alpha(L1)=.28$. Comparable analyses with outcomes for L2 (German) showed a value for Cronbach's alpha of $\alpha(L2)=.44$.

Multiple regression was calculated to predict L1total (Turkish) based on age, gender, and education of participants using the Enter method. Using regression analysis, the overall model including age, gender, and education to predict L1 was significant ($F(3,84)=3.12$; $p<.05$; $R^2=.10$), explaining 10% of the overall variance in the model. Participants' predicted L1total score was equal to $9.15 + .08 (AGE) + .40 (GENDER) + .44 (EDU)$, where age and education were measured in years and gender was coded 1=female, 2=male. Years of education were found to be a significant predictor of L1total ($p<.05$), with participants' overall score of L1total increasing by .44 points per additional year of education. Multiple regression calculations including L2total (German) led to no significant results with age, gender, and education.

3.6.10.1. Comparison mono- versus bilinguals (between comparison)

A one-way MANOVA was conducted to assess a possible difference between outcomes of monolingual ($n=67$) and bilingual ($n=21$) participants regarding the five sub scores of CT: HT, VR, JL, AA, and PS. Independent variable was the language background of a participant, being either monolingual or bilingual, dependent scores were the five subcategories. Box's Test of Equality of Covariance Matrices was non-significant ($p>.05$) confirming the assumption of homogeneity of covariances. Levene's Test of Equality of Error Variances was non-significant ($p>.05$) meeting the assumption of homogeneity of variances.

There was a statistically significant difference in CT sub scores based on a participants language setup, monolingual or bilingual ($F(5,82)=7.21$; $p<.001$; *Wilk's* $\Lambda=.70$; partial $\eta^2=.31$; *power*=1.00). A difference in scores could be observed in one of the five subcategories, VR ($F=24.08$; $p<.001$), with bilinguals ($M=6.33$; $SE=.40$)

outperforming monolinguals ($M=4.07$; $SE=.23$). Sixty of the monolingual participants were never in touch with a L2, and twenty of twenty-one individuals rated their L2 proficiency to be high (with a score of eight or above on a scale to ten), which made a further analysis only including monolinguals without L2 experience and highly proficient bilinguals unnecessary.

3.6.10.2. Comparison L1 versus L2 (within comparison)

For the calculations of first versus second language scores, all monolingual participants' responses were excluded and only data of bilingual participants ($n=21$) was utilized. A repeated measures MANOVA was used to explore possible differences within first and second language scores on the five subcategories of CT. The independent variable described whether participants were presented with the assessment in their first or second language (L1/L2), dependent scores were the five subcategories. The ratio of unexplained variance to total variance was below the critical value of $\Lambda < .05$ ($F(5,16)=5.80$; $p < .005$; $partial \eta^2 = .64$; $power = .96$), suggesting a further analysis of each dependent variable. Outcomes of the Greenhouse-Geisser measure were considered as appropriate for calculations due to its conservative nature. Values for four of the five subcategories, HT, VR, AA, and PS, did not reach significance ($p > .05$), whereas scores for JL showed to be significantly different depending on the language L1 and L2. JL ($F=16.60$; $p < .005$; $partial \eta^2 = .45$) with significantly better performance in L2 ($M=3.05$; $SE=1.02$) compared to L1 ($M=2.10$; $SE=1.14$).

Above-mentioned calculations were based on the inclusion of all bilingual participants ($n=21$). Twenty of those twenty-one individuals rated their L2 proficiency to be high (with a score of eight or above on a scale to ten), which made a further analysis only including highly proficient bilinguals unnecessary.

3.7. Summary of the results

For each language group, a mixture of between- and within- group comparison was undertaken to assess possible differences in critical thinking (CT) scores between monolinguals and bilinguals, as well as within bilinguals' first (L1) and second language (L2). CT was divided into five subcategories, hypothesis testing (HT), verbal reasoning (VR), judging likelihood and probability (JL), argumentation analysis (AA), and problem-solving (PS), following the definition of Halpern (1998, 2002). Observed differences between groups might indicate an effect of bilingualism on cognitive structures. Item analysis was implemented to assess scale reliability. An increase of the reliability levels through exclusion of subcategories (e.g. AA) was present within some language sets (e.g. L1 English- L2 Irish; see Table 3.8). Taking into account that the construct of

the assessment followed the definition of CT according to Halpern (2010) - a widely accepted sub-categorization of CT into five sections, HT, VR, JL, AA, PS - exclusion of any of the sub-sections was ruled out. The informative details each subsection delivered supported this decision.

Multiple regression was employed to assess the influence of external variables on the overall score of CT. Included in the calculations were participants' age, gender, and years of education. A positive effect on participants' overall score of years of education is supported by findings in the literature, proving an increase in CT through education, showcasing that CT can be taught and improved through training. Concerning scores of critical thinking (CT) on *Mixed Measures*, expectations based on prior findings show an advantage for bilinguals in non-verbal subcategories focussed on executive functioning and a slight disadvantage for bilinguals concerning verbally based tests. However, the age group targeted in the current study were all young adults, and the advantages and disadvantages highlighted in previous studies rarely applied to this age category.

When comparing bilinguals' first and second language, findings in the literature suggest better performance in L1 in the subcategory VR. Deep understanding of language and word structure was a prerequisite to resolving questions within this section successfully and bilingual individuals are expected to benefit from the depth of knowledge of L1 and slightly lack verbal reasoning skills in L2 when solving given questions. This is not surprising considering the fact that on average, participants have higher proficiency in and more exposure to L1 compared to L2.

A second category, judging likelihood and probability (JL), was expected to be solved significantly better in L2 than in L1. To explore the depth of possible differences, additional calculations excluding participants rated as bilinguals with a lower than high proficiency in L2, and only including those individuals who reported to have a very good grasp of L2 (with a self-reported L2 proficiency of eight or higher on a scale to ten) were undertaken. This would prove a robust effect depicting a difference between performances. JL described the sub-category of CT, which was mostly based on mathematical and statistical knowledge, such as the awareness of the regression to the mean. Solving assessments within this category was grounded on prior learning of the rules, of which some were slightly counterintuitive. This might be where the difference between L1 and L2 might lie. Bilingual individuals seem to disregard their intuition and follow logical thinking more within their second language (Costa, Vives, & Corey, 2017; Keysar, Hayakawa, & An, 2012).

3.7.1. L1 English- L2 Irish

Item-analysis indicated mid-range scores of scale reliability, with a higher score within participants' second language than their first language. The exclusion of one subcategory, AA, increased the reliability, both in English and Irish (see 3.6 for details).

Multiple regression showed a positive effect of years of education on participants' overall score in L1.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences were not present, even when only analysing monolinguals who had no prior experience with an L2 and only bilinguals who were highly proficient.

When comparing bilinguals' first and second language, questions in sub-category verbal reasoning (VR) were solved significantly better in L1. Even when excluding participants rated as bilinguals with a lower than high proficiency in L2, and only including those individuals who reported to have a very good grasp of L2 (with a self-reported L2 proficiency of eight or higher on a scale to ten) the significant difference in results between L1 and L2 within VR remained. This will be evaluated further in section 4., General Discussion.

3.7.2. L1 German- L2 English

Item analysis showed the scale to be reliable, with a slightly higher score within participants' second language than their first language.

Multiple regression showed no significant effect of any of the variables on the overall score of CT.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences between them were not present.

When comparing bilinguals' first and second language, questions in sub-category verbal reasoning (VR) were solved significantly better in L1. Even when excluding participants rated as bilingual with a lower than high proficiency in L2, and only including those individuals who reported to have a very good grasp of L2 (with a self-reported L2 proficiency of eight or above on a scale to ten) the significant difference in results between L1 and L2 within VR remained.

This was also the case for a second category, judging likelihood and probability (JL). Mean scores in L2 were significantly higher than in L1, when only including highly proficient bilinguals in the calculations but also including all individuals ranked as bilingual in the calculations. These results will be evaluated further in section 4, General Discussion.

3.7.3. L1 Italian- L2 English

Item-analysis indicated mid-range scores of scale reliability, with a higher score within participants' second language than their first language. The exclusion of one subcategory, argumentation analysis (AA), increased the reliability slightly, in both Italian and English (see 3.6 for details).

Multiple regression showed no significant effect of any of the variables on the overall score of CT.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences were not present, even when analysed using only monolinguals who had no prior experience with an L2 and only bilinguals who were highly proficient bilinguals.

When comparing bilinguals' first and second language, questions in subcategory VR were solved significantly better in L1. Even when excluding participants rated as bilingual with a lower than high proficiency in L2, and only including those individuals who reported to have a very good grasp of L2 (with a self-reported L2 proficiency of eight or higher on a scale to ten) the significant difference in results between L1 and L2 within VR remained.

This was also the case for judging likelihood and probability (JL). Mean scores in L2 were significantly higher than in L1, when only including highly proficient bilinguals in the calculations but also including all individuals ranked as bilingual in the calculations. These results will be evaluated further in section 4., General Discussion.

3.7.4. L1 Russian- L2 English

Item-analysis indicated mid-range scores of scale reliability, with a slightly higher score within participants' second language than their first language. Calculations indicated that the exclusion of one subcategory, argumentation analysis (AA), increased the reliability slightly, both in Russian and in English (see 3.6 for details).

Multiple regression showed a positive effect of gender on participants' overall score in L1. This effect was not present in L2. Possible implications of the effect of gender will be discussed in section 4., General Discussion.

A comparison between monolinguals and bilinguals was considered inappropriate due to insufficient participation numbers within the monolingual group.

When comparing bilinguals' first and second language, questions in subcategory VR were solved significantly better in L1. Even when excluding participants rated as bilingual with a lower than high proficiency in L2, and only including those individuals who report to have a very good grasp of L2 (with a self-

reported L2 proficiency of eight or higher on a scale to ten) the significant difference in results between L1 and L2 within VR remained.

This is also the case for a second category, judging likelihood and probability (JL). Mean scores in L2 were significantly higher than in L1, when only including highly proficient bilinguals in the calculations but also including all individuals ranked as bilingual in the calculations. These results will be evaluated further in section 4., General Discussion.

3.7.5. L1 Japanese- L2 English

Item analysis indicated mid-range scores of scale reliability, with a higher score within participants' second language than their first language. Calculations indicated that the exclusion of one subcategory, argumentation analysis (AA), slightly increased the reliability within L1, and the exclusion of verbal reasoning (VR) increased the reliability in L2 (see 3.6 for details). Multiple regression showed no significant effect of any of the variables on the overall score.

Both groups, monolinguals and bilinguals, showed significant differences in results. Japanese monolinguals scored better than Japanese English bilinguals did in hypothesis testing (HT). This trend was not reflected in any of the other language groups and will be evaluated in more detail in section 4., General Discussion. The result met the expectations based on prior findings showing a slight disadvantage for bilinguals concerning verbally based tests.

When comparing bilinguals' first and second language, questions were solved similarly with no significant difference. Scores in two sub-categories, VR and AA, showed tendencies, participants solved VR better in their first language whereas they were better at solving AA in their second language. This goes in line with the results of the other language groups.

3.7.6. L1 Chinese- L2 English

Item-analysis indicated mid-range scores of scale reliability, with a higher score within participants' second language than their first language. Calculations indicated that the exclusion of one subcategory, hypothesis testing (HT), increased the reliability of the Chinese part of the scale (see 3.6 for details).

Multiple regression showed a positive effect of years of education on participants' overall score in L1.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences were not present.

When comparing bilinguals' first and second language including participants of all fluency levels in the calculations, questions in sub-category verbal reasoning (VR)

were solved significantly better in L1. When excluding participants rated as bilingual with a lower than high proficiency in L2, and only including those individuals who reported to have a very good grasp of L2 (with a self-reported L2 proficiency of eight or higher on a scale to ten) significant differences within VR disappeared. This indicates that variations were going back to language difficulties of bilinguals with low proficiency, having difficulties with L2 comprehension. This will be evaluated further in section 4., General Discussion.

3.7.7. L1 Chinese- L2 Japanese

Item-analysis indicated mid-range scores of scale reliability, with a higher score within participants' second language than their first language. The exclusion of one subcategory, hypothesis testing (HT), increased the reliability of the Chinese part of the assessment.

Multiple regression showed no significant effect of any of the variables on the overall score of CT.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences were not present in four of five subcategories. Within one subcategory, judging likelihood and probability (JL) bilinguals scored significantly higher than monolinguals. This finding goes in line with Cushen and Wiley (2011) who reported bilinguals to be better in creative insight problem tasks as well as cognitive flexibility. Differences are going to be discussed in more detail in section 4., General Discussion. Due to a restricted amount of highly proficient bilinguals (N=6) a comparison of those bilinguals and monolinguals without prior L2 experience to explore the nature of differences observed was not possible in this language set.

When comparing bilinguals' first and second language, questions of subcategory verbal reasoning (VR) were solved significantly better in L1.

This was also the case for a second category, JL, with a significantly better performance in L2 compared to L1. These results will be evaluated further in section 4., General Discussion.

3.7.8. L1 Turkish- L2 English

Item analysis indicated mid-range scores of scale reliability concerning overall scores for both languages. The exclusion of one subcategory, judging likelihood and probability (JL), slightly increased the reliability of the Turkish assessment and argumentation analysis (AA) that of the English assessment.

Multiple regression showed a positive effect of years of education on participants' overall score in L1.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment in three subcategories, whereas significant differences could be observed in two sections, verbal reasoning (VR) and problem-solving (PS). In both cases, bilingual participants outperformed monolinguals and scored significantly higher. Differences will to be discussed in more detail in section 4., General Discussion.

When comparing bilinguals' first and second language, questions in subcategory JL were solved significantly better in L2 compared to L1, when only including highly proficient bilinguals in the calculations, but also including all individuals ranked as bilingual in the calculations. This will be evaluated further in section 4., General Discussion.

3.7.9. L1 Turkish- L2 German

Item analysis indicated poor scores of scale reliability concerning overall scores in Turkish, and mid-range scores within the German assessment. Calculations indicated that the exclusion of two subcategories, judging likelihood and probability (JL) and argumentation analysis (AA), increased the reliability within L1, and exclusion of verbal reasoning (VR) increased reliability within L2.

Multiple regression showed a positive effect of years of education on participants' overall score in L1.

Both groups, monolinguals and bilinguals, performed similarly on the CT assessment and significant differences were not present for four of the five groups. In subcategory VR, participants showed a significant advantage when bilingual. The difference remains significant even when only including monolinguals that have never been in touch with any L2 and highly proficient bilinguals. The differences are going to be discussed in more detail in section 4., General Discussion.

When comparing bilinguals' first and second language, questions were solved significantly better in L2 in the subcategory JL, when only including highly proficient bilinguals in the calculations but also including all individuals ranked as bilingual in the calculations. This was proof for a robust effect depicting a difference between performances and will be evaluated further in section 4., General Discussion.

4. General Discussion

The following paragraphs contain both, the discussion of results of the merged data set as well a general discussion of outcomes of each language set as presented in section 3.6. Correlations between scores on *Mixed Measures* and Raven's Standard Progressive Matrices (RSPM) were undertaken to assess the validity of *Mixed Measures*, based on the relationship between critical thinking (CT) and intelligence (Halpern, 2006; Marin & Halpern, 2011). A positive correlation between scores of *Mixed Measures* and RSPM was found in both language settings, participants' first language (L1) and their second language (L2), which supported the validity of *Mixed Measures*, in line with the association between the two constructs, CT and intelligence (Halpern, 2006; Marin & Halpern, 2011). With increasing scores of fluid intelligence, participants' overall score of CT increased.

As mentioned above, additionally to analyses with the specific language groups, a mixture of between- and within- group comparison calculations was undertaken with the merged data set including all participants, to assess possible differences of CT scores between monolinguals and bilinguals, as well as within bilinguals' L1 and L2. Observed differences between groups might indicate an effect of bilingualism on cognitive structures.

Item analysis was implemented indicating mid-range scores of scale reliability with all language sets merged, with a higher score within participants' second language than their first language. The assessment of individual language sets showed that the reliability of one language group, German English, exceeded all other scores with good reliability levels, making it reasonable to include all subcategories. Moreover, taking into account that the construct of the assessment followed the definition of CT according to Halpern (2010) - a widely accepted sub-categorization of CT into five sections, hypothesis testing (HT), verbal reasoning (VR), judging likelihood and probability (JL), argumentation analysis (AA), and problem solving (PS) - exclusion of any of the sub-sections was ruled out. The informative details each subsection might deliver support this decision. Concerning the variation across language groups on alpha scores, including particularly low scores in groups Italian-English and Turkish-German, differences might be rooted in cultural differences or language related critical thinking strategies. Considering the aim of testing inter-individual variations across languages this might bear additional meaningful information.

Multiple regression was employed to assess the influence of external variables on the overall score of CT. Included in the calculations were participants' age, their gender and their years of education. Years of education had a positive effect on

participants' overall score in both languages, L1 and L2, proving an increase of CT through education, highlighting that CT can be taught and improved through training. Facione (1990) stated that education helps to improve CT. Differences in CT originating in higher education are observable from as little as hours and days spent studying (Terenyini, Springer, Pascarella, & Nora, 1995), with the observed influence of years of education on CT making it reasonable to implement all questions. Participants' gender showed an effect only in L1, with male participants performing better compared to females. An effect of gender within CT was not represented in prior studies on CT (Butler, 2012; Butler et al., 2012). Taking into consideration the unequal number of male and female participants with a significantly lower percentage of male subjects, it is questionable if this result reflects a real difference. Gender differences within CT have been explored in a very restricted manner and, if present, were mostly found to be non-significant (Halpern, 2010; Walsh & Hardy, 1999). The significant influence of participants' gender disappeared within L2, with participants being distributed slightly more equal in this group. Examining the individual language groups, an effect of gender did not appear in eight out of nine groups. On the contrary, the average difference between both genders in the merged data set was relatively high, with male participants scoring higher than female participants. Test takers' age had a small effect on total CT scores within L2. Bilinguals within this participant sample, from whom L2 scores were derived, varied quite significantly in age compared to monolinguals. Monolinguals on average were younger and varied less in terms of their age. This variation combined with a higher average age within bilinguals showed that participants within the bilingual group were at different stages of their lives; assumingly at the early stage of higher education for those who were younger compared to older participants who had already undergone a few years of higher education. This was reflected in the influence years of education on CT but could also be presumed from the covariable *Age*.

Considering the nature of the analyses, the influence of any external variable only plays a role in between subject comparisons, comparing monolinguals with bilinguals. In within comparisons, comparing L1 and L2 of a participant, external variables did not affect results, seeing as two scores of each bilingual participant were used for calculations.

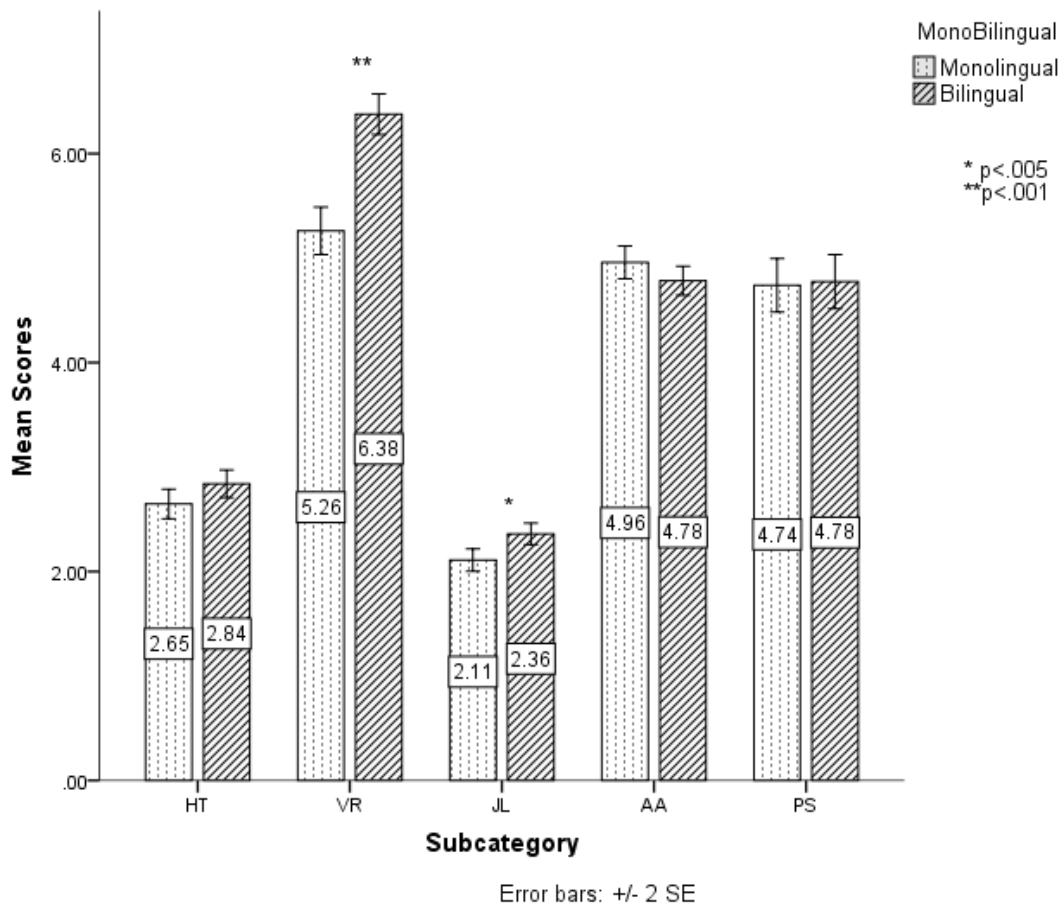


Figure 4.1. Overview of Scores (All Language Groups Merged) comparing Mono- and Bilinguals, divided into Sub-Categories.

As presented in Figure 4.1, the comparison of monolinguals and bilinguals of the merged data set revealed significant differences in two subcategories: VR, and JL. Within both sub-categories, bilinguals scored higher than monolinguals. An advantage on tasks falling under the category of non-verbal sub-sections, which applies to JL, was predicted on the basis of Bialystok (2009; also Bialystok et al., 2004; Bialystok, Craik, & Luk, 2012; Carlson & Meltzoff, 2008; Gold et al., 2013; Hilchey & Klein, 2011). The bilingual advantage was confirmed in low-level cognitive processing tasks, originating in enhanced inhibition and activation (Peal & Lambert, 1962). Low-level cognition is one of the building blocks for any higher-level executive function, making it an important aspect of CT. The ability to inhibit irrelevant information was shown to enhance an individuals' capacity to develop original and useful ideas (Kharkhurin, 2011), which might benefit bilingual individuals when presented with scenarios where CT is essential. Moreover, bilinguals develop increased mental flexibility and greater creative thinking (Cushen & Wiley, 2011; Lee & Kim, 2011), which aids in solving given

scenarios. The more proficient a bilingual speaker is in L2, the more creative this individual seems to be when it comes to problem-solving (Kharkhurin, 2011). The broader cultural experience supports this advantage of a second language in creativity and problem solving (Cushen & Wiley, 2011). Considering this it is surprising that the subcategory PS showed no significant difference between the two groups.

Verbal subcategories, VR and AA, were expected to be solved with a slight disadvantage for bilinguals. Bilinguals outperformed monolinguals in solving VR questions. This might develop from a deeper understanding of word structure and a higher awareness of relationships of words (Ben-Zeev, 1977), assisting bilinguals in completing analogies correctly.

When excluding monolinguals with any L2 experience and bilinguals that only report mediocre L2 proficiency (and only assessing monolinguals without any prior L2 experience and highly proficient bilinguals) to explore differences between the groups further (Anderson et al., 2018), the bilingual group outperformed monolinguals only on VR, and differences in JL disappeared.

Balanced bilinguals, in particular, show an advantage over monolinguals on non-verbal executive functioning tasks (Thomas-Sunesson, Hakuta, & Bialystok, 2018). This can be considered the basis of CT. Supporting this idea are results of an additional explorative analysis: individuals were grouped into four fluency categories dependent on their self-rated proficiency. Group one had no prior experience with a second language (monolingual). Group two reported L2 proficiency to be very low (between one and four on a scale of one to ten, with ten being highly proficient; monolingual with some L2 experience). Group three was at a medium level of language fluency (rated between five and seven; bilingual with mediocre L2 proficiency). Group four included highly proficient individuals who reported their fluency to be above eight (bilingual with high proficiency). When comparing individuals' performance on the five subsections, results of VR as a sub-category stood out. The more fluent individuals reported to be, the higher they scored on VR, suggesting that increasing proficiency of L2 led to increased performance. This difference was highly significant, and furthermore, no other sub-category showed a trend as unambiguous as VR, with the least proficient group performing worst and the most proficient group performing best. Bilingualism aids a better understanding of the concept of a language (Jessner, 1999). This superior metalinguistic "language concept", originating in the learning process of a second language (Bialystok, 1986) is considered to be the reason for such an outcome, with participants' grasp of the concept of language increasing with proficiency in L2. VR, as implemented in *Mixed Measures*, was based on a clear understanding and awareness of the construct of language and the relationship of words (Halpern, 2010).

When speaking a second language differences and similarities between words stand out, especially when differing within languages. This creates an increased receptiveness for such details within language.

When assessing differences between mono-and bilinguals in more detail, it is important to mention that significant differences refer to the overall comparison of all participants in the merged data set. Focussing on individual language sets, differences were less dominant and not consistent, as was presented in 3.7.1-3.7.9. Significant differences in a comparison of mono- and bilingual individuals were not present in every language set. As can be seen in Figure 4.2 depicting the difference scores of mono-and bilinguals, with positive scores representing a superior performance of monolinguals, and negative scores representing a superior performance of bilinguals, a wide variation in scores within sub-categories could be observed and homogeneity of results is not present.

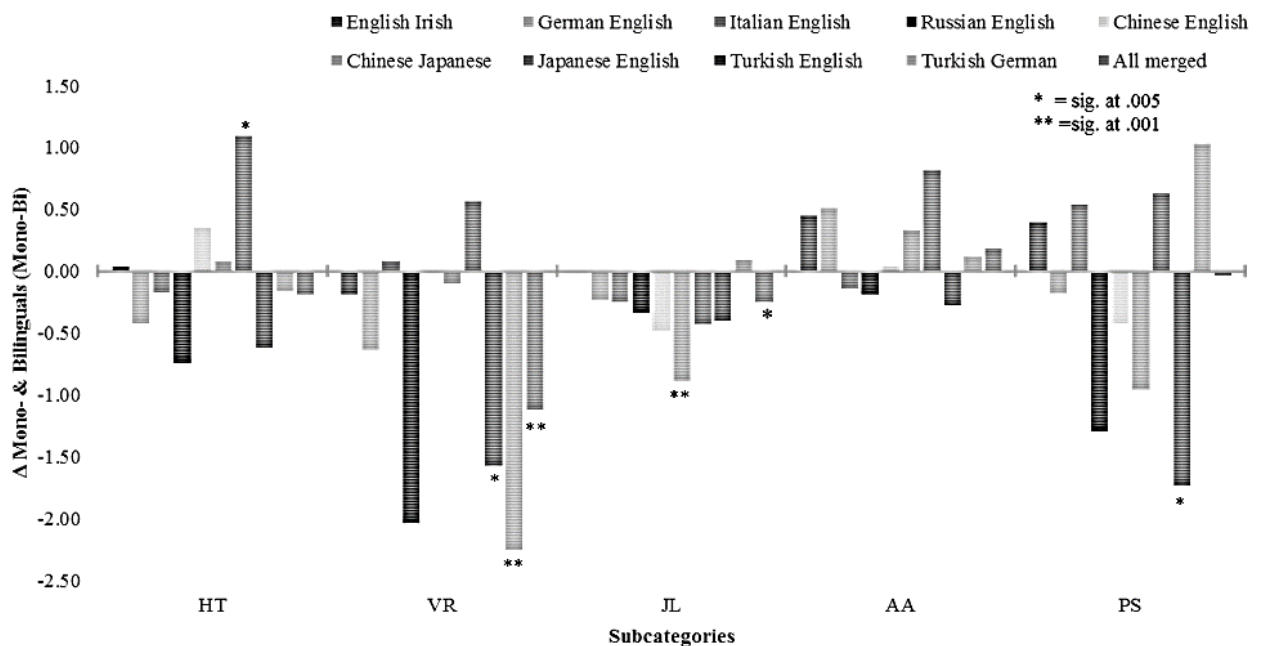


Figure 4.2. Overview of Scores of all Language Groups comparing Mono- and Bilinguals, divided into Sub-Categories (Positive Scores represent a better Outcome of Monolinguals, negative Values represent a better Outcome of Bilinguals).

Significantly different scores were only present in four out of eight groups, Chinese-Japanese, Japanese-English, Turkish-English, and Turkish-German. Differences caused by translation errors can be excluded due to the fact that monolinguals and bilinguals are both assessed in the same language. One of the language sets, Russian English, was excluded from between comparison analyses due to an insufficient

sample size of monolingual participants ($n=9$). Consistencies were not only absent within the language sets with every language set showing different trends, but also within the sub-categories that differed in the language sets containing significant results (see Table 3.11 and Table 3.12). HT was solved better by monolinguals within the Japanese-English set. Bilingual participants within two groups, Turkish-English and Turkish-German, solved VR significantly better than monolinguals. JL was solved better by bilinguals within the Chinese-Japanese language set, whereas there were no significant differences in any of the other language sets concerning this sub-category. PS was solved better by bilinguals within the Turkish-English language set. Overall, three of the four language sets confirmed a bilingual advantage (Chinese-Japanese, Turkish-English, and Turkish-German).

The inclusion of multiple language sets within this project aims to identify consistent differences, which in turn can be traced back to the influence of bilingualism on cognitive operation. With the results presented above it can be assumed that differences are not leading back to generic changes within a bilinguals' brain - originating in second language acquisition or usage - in which case they would appear consistently throughout all language sets. Differences between monolinguals and bilinguals are likely to derive from external factors other than bilingualism, such as language-specific grammar or culture. The affiliation with a culture seems to lead to a certain cognitive approach of decision-making (Briley, Morris, & Simonson, 2000). Cross-cultural comparisons report differences based on a varying perception of heuristic cues which is grounded in culture (Aaker & Maheswaran, 1997). This influence of language and culture on cognition is partly grounded in grammar-related differences, which affect an individual's categorization performance as well as judgment and choice (Schmitt & Zhang, 1998). Cultural background shapes the perception of one's environment (Briley, Morris, & Simonson, 2000). Bandura (2002) points out the diversity and dynamic nature of cultures as social systems and emphasizes the importance of the variety of cultures, even within broad categories like collectivism and individualism. Bilingual individuals are often subject to two sets of cultures, which they are surrounded by. Dependent on the language environment and setting an individual is situated in, one or the other culture dominates at each moment in time, and relevant cultural knowledge structures, like beliefs and decision principles, are activated (Briley, Morris, & Simonson, 2000; Hong, Morris, Chiu, & Benet-Martinez, 2000). Activation of specific cultural knowledge can affect an individual's emotions, judgment, and decisions (Hong et al., 2000). Nevertheless, culture being a dynamical system also entails the fact that cultural effects do not influence every individual equally and thus cannot be generalised (Vogeley & Roepstorff, 2009).

Explaining differences between monolinguals and bilinguals based on cultural differences works well in the case of bilinguals outperforming monolinguals, with the additional culture possibly carrying aspects of cognition that aid CT which are not represented in the first language. This applies to the Chinese-Japanese language set as well as Turkish-English and Turkish-German. Especially in the case of Turkish being the first language, with similar tendencies in two independent language-setups, a cultural bias could be assumed, affecting CT as implemented in this assessment. The contact with an additional second language and culture, in this case, German or English, possibly added to the bilinguals' thinking, leading to an advantage over monolinguals in solving the scenarios.

However, cultural differences do not explain differences where the opposite trend was present, as is the case with the Japanese-English language set. Monolinguals scored significantly higher than bilinguals (with both groups being assessed in the same, their first, language, Japanese). If there was an aspect within Japanese culture that nurtures those aspects of cognition that CT is built on, bilinguals would show similar results within Japanese responses, unless the effect disappears as soon as a Japanese individual is confronted with an additional language or culture. When examining the Japanese-English dataset more closely and comparing monolinguals and bilinguals, focussing on possible differences between the two groups, i.e., homogeneity of participants' age, education levels (including years of education overall and higher education levels), one distinction could be made between monolingual and bilingual groups: gender. The monolingual group consisted of 40 percent females and 60 percent males, whereas the bilingual group comprised 75 percent females and 25 percent males. Observed differences between mono- and bilinguals might not reflect language differences but may be grounded in varying scores dependent on gender. An additional analysis comparing males and females, both including all participants as well as split into mono- and bilinguals, indeed showed higher values within male participants for some subcategories, but with a non-significant overall score. Keeping in mind the gender distribution within monolinguals and bilinguals, with a higher percentage of male participants within the monolingual group, and with males seemingly scoring higher than females within this language set, this might explain the observed differences. When dividing the data set by gender, differences between groups are highest in the sub-category PS, which did not differ significantly within the overall comparison of monolinguals versus bilinguals (compared to the significantly differing sub-category HT), making this hypothesis questionable though. Similar tendencies as reported above could be observed when investigating differences based on participants' country of residency. Lower exposure to the

Japanese language might lead to a disadvantage in solving given scenarios. The exclusion of all participants whose current country of residence is not Japan did not change the outcome within HT, which still differed between monolinguals and bilinguals. Consequently, the origin of the differences cannot be explained completely, but assumptions can be made: A male advantage when solving given scenarios seems to be part of the reason for the differences, together with a lower comprehension of Japanese scenarios within bilingual individuals.

When focussing on the comparison of bilinguals' first and second language, the results showed a clear trend: Three of the five subcategories, VR, AA, and JL, differed significantly within L1 and L2 as depicted in

Figure 4.3, two of which differed highly significantly, VR and JL. In two of the three categories, participants scored higher when solving tasks in their first language, VR and AA, whereas there seemed to be an advantage for solving JL in L2.

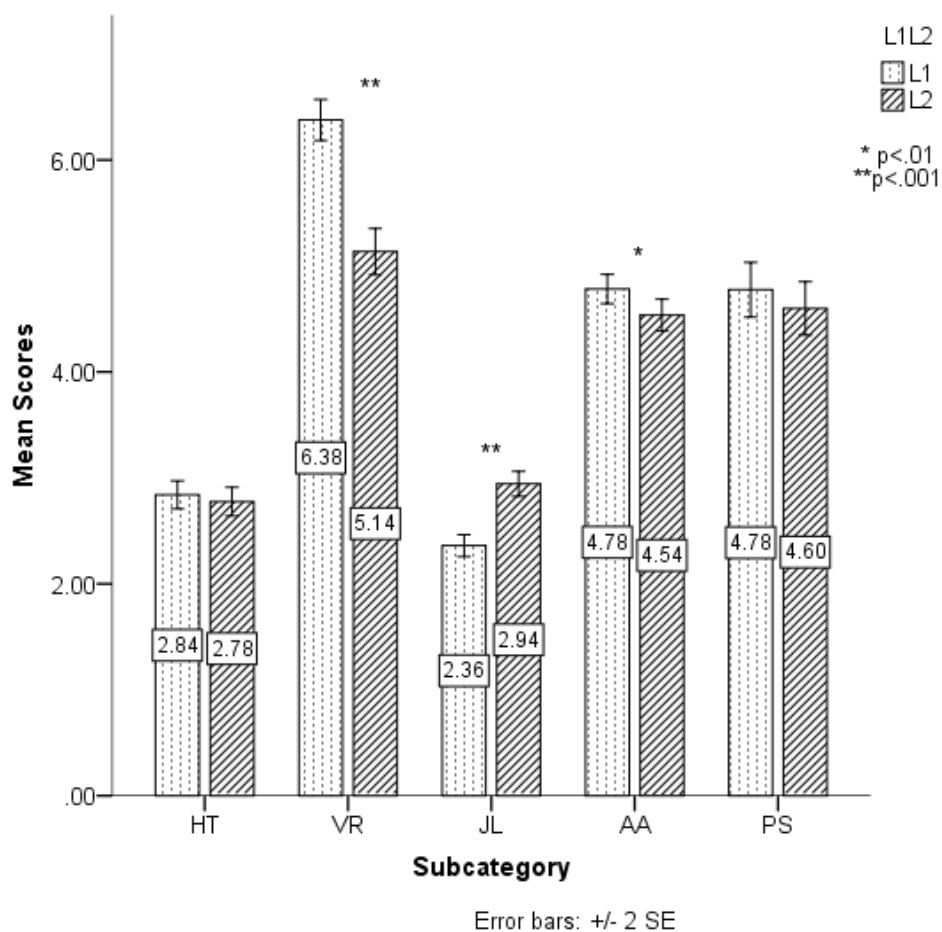


Figure 4.3. Overview of Scores (All Language Groups Merged) comparing L1 and L2, divided into Sub-Categories.

To ensure the origin of differences did not stem from language difficulties, and to control for misunderstandings due to lower second language proficiency, an additional calculation was undertaken, excluding bilingual individuals who reported only mediocre fluency in L2. Therefore, only scores of highly proficient bilinguals within the merged data set were compared in L1 and L2. Highly significant values were present within VR and JL, with bilinguals scoring better in L1 in VR but in L2 in JL. This underlined the depth of differences present in the overall comparison including all participants and put special emphasis on those two sub categories.

Similarly to above, another calculation was undertaken to assess a possible influence of age of acquisition of L2. All bilingual participants were divided dependent on the age of acquisition of their second language, with early bilinguals having acquired L2 before the age of ten. The results of both groups, early and late bilinguals, fell in line with earlier findings, showing a clear advantage of solving VR scenarios in L1 and significantly better scores within L2 for JL questions. These additional calculations helped to exclude possible covariates as origins of differences, such as language difficulties and little experience with a language and increased the expectation of the source of differences to be of a cognitive nature.

As presented above, the results of two subcategories stood out, in which participants' scores differed highly significantly depending on the language the scenarios were presented in: VR and JL (see Figure 4.4). When observed in detail divided by language set, participants of all language sets scored better in their first language when solving verbal reasoning tasks, with significant differences in six out of nine language sets. Participants responded more accurately in their first language in tasks that required very deep, detailed understanding of vocabulary. As introduced earlier, the subcategory verbal reasoning examined participants' ability to analyse the structure of language, it assessed participants' ability to withstand pervasive and misleading language. This seemed to be done more successfully in individuals' first language.

The better grasp of language concepts mentioned earlier (Jessner, 1999), seems to be restricted within L2. Higher exposure to L1 during an individuals' first contact with language, as well as L1 often being the base for learning L2, might lead to a deeper understanding of structures and relationships within L1. Wartenburger, Heekeren, Abutalebi, Cappa, Villringer, and Perani (2003) assessed the performance of bilinguals when scanning sentences for semantic and grammatical errors in L1 and L2. Behavioural differences were not observed and bilinguals scored similarly in L1 and L2, yet fMRI scans showed a significantly larger Blood Oxygen Level Dependent (BOLD) response when participants read grammatically incorrect sentences in L2. This

was found to be correlated with the age of L2 acquisition, showing increasingly less effective higher activation, the later L2 was acquired (Wartenburger et al., 2003). A difference in activation was also found by Dehaene et al (1997), who showed activation of additional brain regions when presented with material in L2, compared to processing material in L1. Activated brain regions are associated with attentive processing tasks, possibly involving a higher demand of resources when processing material in L2. Participants of this fMRI study had obtained mediocre proficiency and results for highly proficient bilinguals are not available. Indefrey (2006) summarized the findings of thirty studies focused on the comparison of L1 and L2 activation and confirmed an increased L2 processing activation in comparison to activation when processing L1. Differences vary dependent on AoA of L2, L2 proficiency, as well as exposure to L2. The earlier an individual acquired L2 and the more fluent and regular this person speaks L2, the lower brain activation seems to be (Indefrey, 2006). Explanations for higher activation might be rooted in a compensation process of lower efficiency of L2 as well as lower neural connectivity of brain structures in L2 (Indefrey, 2006). The high demand of verbal processing and understanding in verbal tasks in general, and specifically in VR presented to participants within *Mixed Measures*, might exceed resources available for processing within L2 and thus lead to lower performance in L2 compared to L1.

In the sub-category JL, all language groups achieved higher scores in their second language; six out of nine groups with significant differences (see Figure 4.4). Judging likelihood scenarios were generally maths based, specifically focussing on the probability of events, and thus required abstract thinking. This might be implemented more successfully in L2, where individuals were able to focus on given details, removing emotion and intuition from the answering process. When presented with scenarios in L2, bilinguals show a reduced aversion to loss and risk, enabling them to encode a problem in a more objective manner (Costa, Vives, & Corey, 2017; Keysar, Hayakawa, & An, 2012). Diaz-Lago and Matute (2019) assessed the influence of solving biased scenarios in a second language, asking participants to describe the contingency of independent events. Participants' replies were significantly more accurate concerning null contingency situations when answering in their L2 while controlling for correct understanding of the task. Using a second language helps to reduce causality bias, which describes the misconception of a causal relationship between two independent events (Diaz-Lago & Matute, 2019). Thus, decision-making processes might be more effective in situations which are out of the individual's control (Diaz-Lago & Matute, 2019). One explanation for this might be the detachment of a problem through increased psychological distance leading to a more utilitarian problem-solving approach (Diaz-Lago & Matute, 2019) and emotional distance leading

to increased emotional resonance (Keysar, Hayakawa, & An, 2012). Implemented scenarios in *Mixed Measures* often enhanced misinterpretations, and correct solutions might be slightly counterintuitive at first glance, not following intuition but requiring deeper thought. Second language processing seems to minimize the effect of intuition and support decision making based on deliberation and consideration (Costa, Vives, & Corey, 2017) which might be the reason for a better performance in L2 in JL. Increased arithmetic skills in L2 are not expected to be the reason for a better performance in JL due to the fact that this is not supported in the literature. When presented with arithmetic questions which are verbally phrased, bilinguals reach lower scores in L2 compared to L1, even if L2 is the language of teaching mathematical content (Bernardo, 2002; Bernardo & Calleja, 2005; Van Rinsveld et al., 2016). Following Halpern's (2010) definition of CT, JL is one of three non-verbal subcategories, together with HT and PS. It is surprising that the results showed a clear advantage in performance of JL in L2 within all language sets including the merged dataset, while results for HT and PS did not show a consistent tendency toward an advantage in either language and outcomes varied from language set to language set.

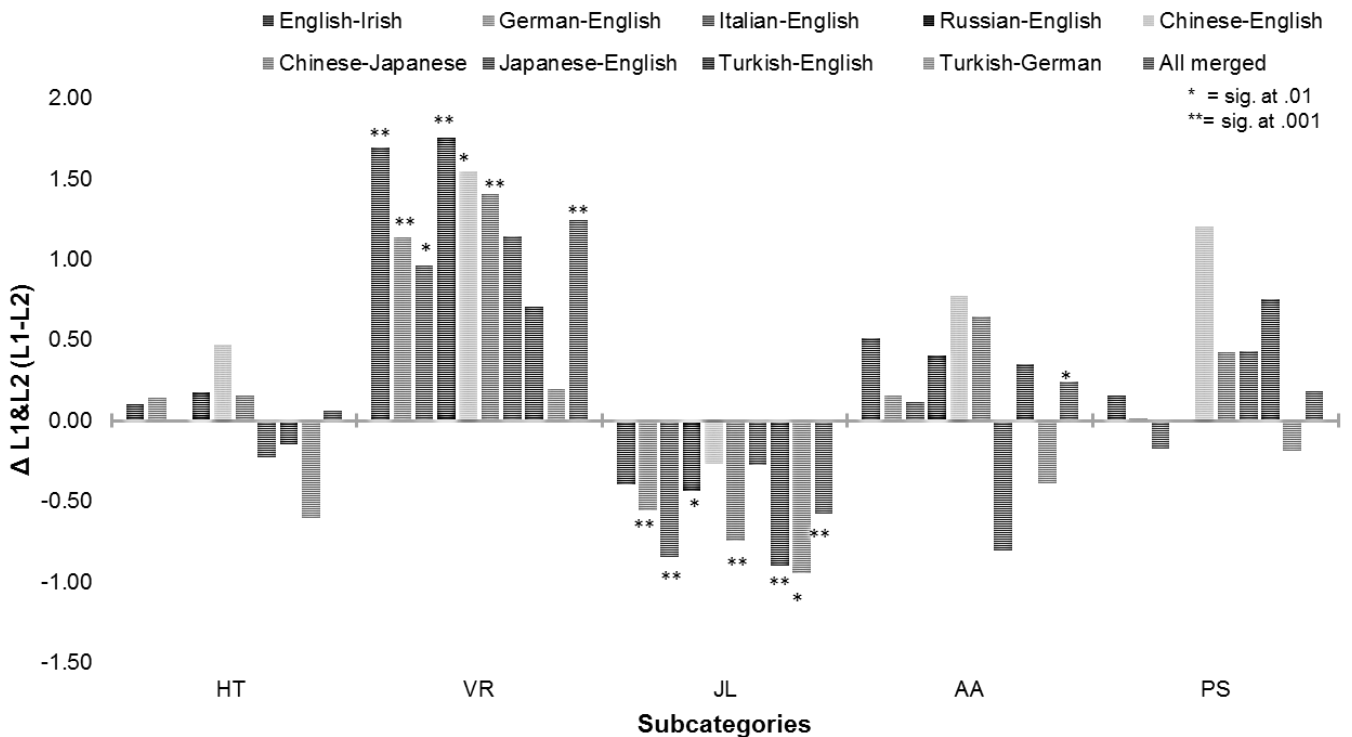


Figure 4.4. Overview of Scores of all Language Groups comparing L1 and L2, divided into Sub-Categories (Positive Scores represent a better Outcome within L1, negative Values represent a better Outcome within L2).

Final analyses were undertaken to obtain an overview of participants' performance divided by language groups. This comparison was of explorative nature and outcomes do not possess evaluative power, because group sizes and distributions of mono- and bilinguals in each group were not equal. Differences between participants' first language scores were significantly different depending on the language group, with variations in three of the five subcategories, HT, VR, and PS. Bilinguals' scores on L2 similarly varied significantly dependent on the language group, in two subcategories, VR and PS. An overview of differences between language groups on L1total (including all subcategories) can be seen in Figure 4.5, whereas scores for L2total including all bilingual participants are visualised in Figure 4.6.

Scores for participants' L1 contained less variation in comparison to those for L2. Variances might be due to comprehension difficulties. Language groups Russian-English and Japanese-English seemed to be scoring slightly higher within L1 in comparison to the other languages, whereas the two groups involving Turkish, Turkish-English and Turkish-German, scored slightly lower in L1. English-Irish bilinguals scored lower within L2 compared to other groups, similar to the Chinese-English language group.

Differences could originate in multiple sources. Slight adjustments in translations might have led to an increase or decrease of difficulty, which is not controlled for in these between language-comparisons. Moreover, these analyses did not take into account background variables, such as age, gender, and socioeconomic status of participants. As discussed above, years of education, for instance, influenced scores of CT. All language groups consisted of varying numbers of mono-and bilingual participants, and as has been shown, bilingual participants reached slightly higher scores in some subcategories. Consequently, groups with a higher percentage of bilinguals were more likely to reach higher scores, seeing as language groups were not divided by mono-and bilinguals in this comparison. However, differences might also reflect true variation between certain language sets, representing different CT skills as measured within this study.

This final set of calculations remains subject of debate and sensible conclusions cannot be drawn based on data collected, due to the above-mentioned restrictions.

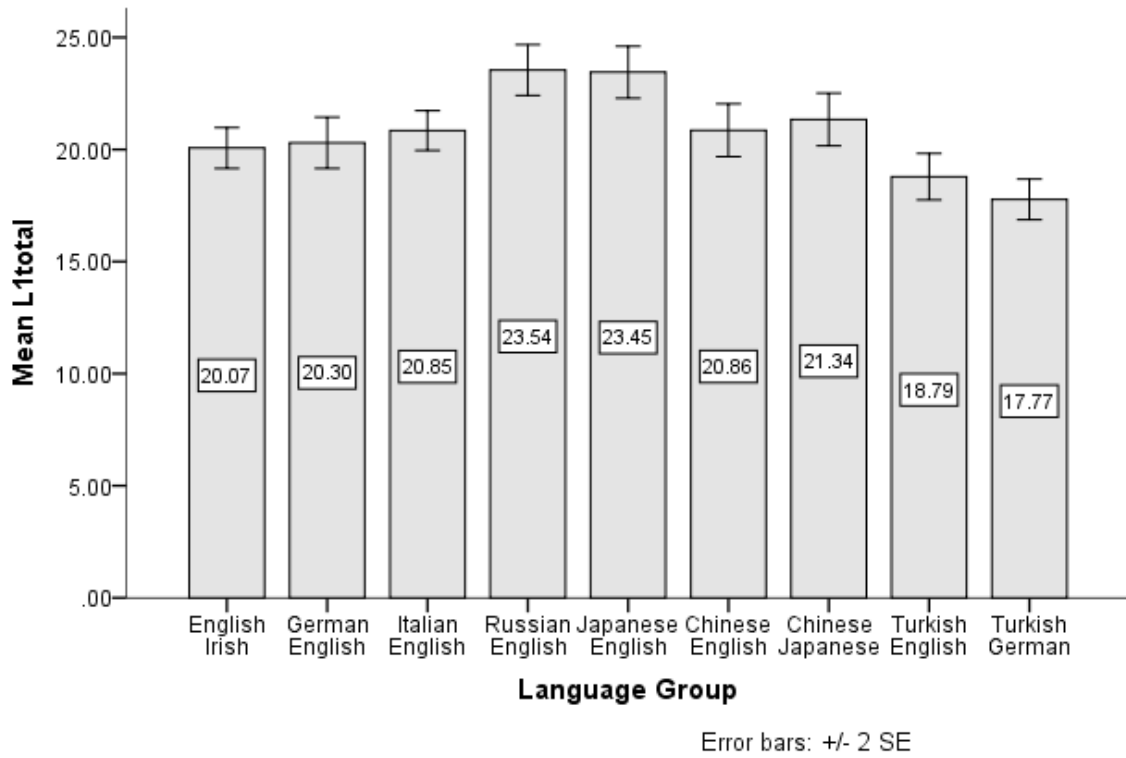


Figure 4.5. Overview of Overall Scores of CT in L1 (L1total), divided by Language Group.

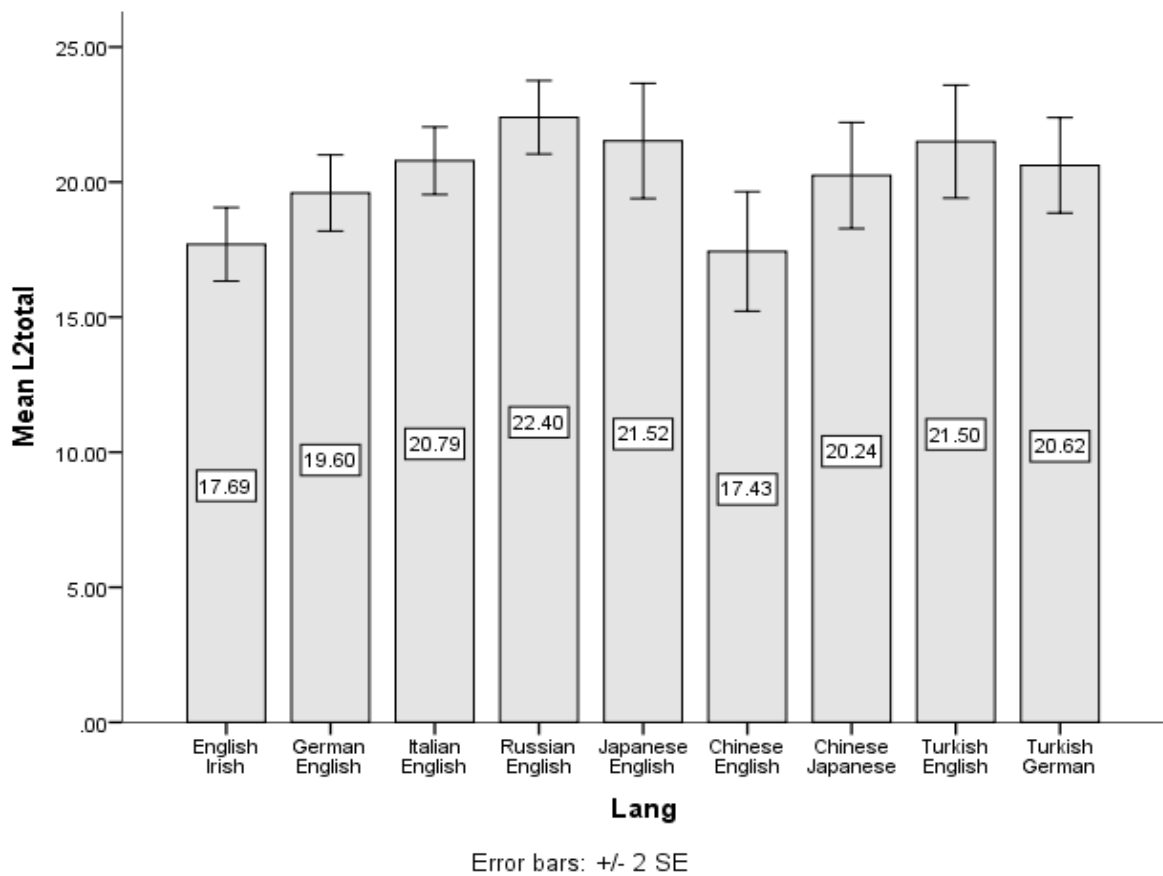


Figure 4.6. Overview of Overall Scores of CT in L2 (L2total), including all Bilinguals, divided by Language Group.

4.1. Conclusion

The purpose of the current study was to explore the possible effects of speaking two languages on higher cognition, in particular, focusing on critical thinking (CT). This was assessed in two ways, first by comparing individuals with no second language ability (monolinguals) to those using two languages in their everyday lives, representing the applied definition of bilingualism, and second by comparing the performance of bilinguals' in their first and second language. Critical thinking can be divided into five subcategories, hypothesis testing (HT), verbal reasoning (VR), judging likelihood and probability testing (JL), argumentation analysis (AA), and problem-solving (PS; Halpern, 1998; 2002). Each subcategory was assessed by a range of published tests, and results were compared. A total of nine language sets were included in the study, namely English-Irish, German-English, Italian-English, Russian-English, Japanese-English, Chinese-English, Chinese-Japanese, Turkish-English, and Turkish-German, with the first language mentioned being participants' first language respectively. Results were compared both individually for each language set as well as one overall calculation incorporating all available data. Merging all language sets was considered important to help to minimise the effect, which specific languages or cultural backgrounds might have on the responses, and to make outcomes generalizable. Results are intended to reflect two things: differences in CT evoked by cognitive changes through speaking two languages, and the less investigated research area of first and second language performance differences. Robustness of discovered differences was ensured through additional calculations only including certain group members, such as highly proficient bilinguals, to assure perfect comprehension of the assessment, controlling for possible misunderstandings.

Focussing on the comparison between monolinguals and bilinguals, significant differences within the individual language groups are rather inconsistent. However, the merged data set including all languages shows a significant difference in two of the five subcategories, VR and JL. Bilinguals outperformed monolinguals in both categories. These findings are not surprising considering the basis of CT in executive functions and higher cognition: areas in which a wide range of research reports a bilingual advantage. Questions arise around why significant differences were specifically found in those subcategories and not the others, and why these findings are only present in the overall comparison, but not in the specific language groups. Due to these uncertainties and a restricted ability to clarify them with the data collected, the emphasis put on this comparison is kept minimal. Findings of this research which are considered to be of a higher impact, and which show greater consistency are comparative calculations between bilinguals' first and second language. Participants

performed significantly better in their first language in the subcategory VR, and in their second language in the subcategory JL. These findings can be observed across all language groups as well as in the merged data set. Findings are present within highly proficient bilinguals, and even remain significant if including bilinguals who only reported a mediocre proficiency in their L2 in the calculations. An advantage of answering questions in L1 in section VR - the subcategory based on verbal understanding and reasoning, requiring very detailed knowledge of the language and its vocabulary - is less surprising. Taking into account that proficiency of L1 usually exceeds proficiency in L2, due to higher exposure and a younger age of first contact with L1 compared to L2, better scores within a verbally based subcategory can be expected. Outcomes of the subcategory JL are considered to be the most prominent findings of this project. Supported by the theory of a certain emotional detachment in L2, prior research has started to shed light on this aspect of cognition within bilinguals.

4.2. Implications

Findings between monolinguals and bilinguals in two of the five subsections suggest a slight advantage of solving critical thinking scenarios as a bilingual individual. This adds to a broad field of research including this view of the bilingual advantage. The remaining three of the subsections seem to be providing support for the opposite opinion of null findings between mono- and bilinguals. The major debate and uncertainty within the research field of bilingualism regarding the possible existence of a bilingual advantage (Paap & Greenberg, 2013) complicates communicating scientific findings to a non-scientific audience and leads to confusion and a certain incredibility of research in the field. By conducting research, following standards built upon prior findings and observations, clarity on the effects can be reached, on which future projects can be built upon (Bialystok, 2018).

Moreover, this research project highlights the importance of a clear assessment of language skills, which in turn facilitates the comparison of language groups of different proficiency (Marian, 2018). The comparison of bilinguals with different language proficiencies adds to the understanding of how bilingualism affects cognition as an adaptable system. Additionally to this, and as attested by the outcomes of this study, the comparison of first and second language performance of bilinguals can carry valuable information about processing in bilinguals (Costa & Sebastián-Gallés, 2014; Costa, Vives, & Corey, 2017; Keysar, Hayakawa, & An, 2012). An important piece of evidence, which can be taken away from this project referring to the research area of bilingualism, is the importance of the inclusion of multiple language sets when aiming to make generalisations. Differences observed in the comparison of two languages,

representing one specific language set, are not necessarily transferrable to other language sets, as represented in the between groups comparison of this study, with varying significances in different language sets. The inclusion of multiple language sets increases the scope of any project and ensures findings to originate in language or cultural differences if present across a range of language sets.

Outcomes of this study from a non-academic point of view raise the awareness of possible differences in critical thinking performance dependent on the language implemented, L1 or L2. Being bilingual is shown to be rather the norm than the exception with more than half of the world's population speaking at least two languages in their daily lives, and so this consideration is one affecting public life in many situations. Some situations might entail a certain advantage if faced in L1, relating to tasks which are verbally very highly loaded, such as scientific writing, and others which might be more advantageously absolved in L2, for instance, court hearings or any situation including decision making which involves some uncertainty, such as gambling or buying a house. Keeping this in mind and considering it when possible might lead to a more efficient decision-making process (Costa, Vives, & Corey, 2017) with higher quality outcomes, possibly including benefits for both the individual and society in general. Moreover, the outcomes of this research carry implications for the educational assessment of individuals. Results of this study show that the performance of a bilingual individual might vary depending on the language of assessment. Underlying cognition cannot always be reliably measured language-independent but might vary depending on the language setting of the assessment. Bearing this in mind might help to explore an individual's full potential and to specifically tailor education. This, in turn, can increase effectiveness, the ultimate goal of any educational undertaking.

4.3. Directions for future research

While the current study can be considered as a valuable and extensive examination of the effects of bilingualism on CT, some questions remain unanswered. Exploring present outcomes suggests a certain natural progression of research on this topic in the eyes of the researchers, which will be presented in the following section. An issue of importance in the field of research on bilingualism is internal versus external validity, and the balance between the two constructs, especially with regards to the on-going debate within researchers in terms of the bilingual advantage. Replications of the present study with bilinguals of varying language sets as well as different L2 proficiencies can help to endorse findings. More importantly and majorly adding to the research area, would be studies based on the same definition of CT as introduced by Halpern (1998; 2002), but implementing different measures. Focussing on the external

validity of the construct and the assumptions of this research, this would help to specify the effect of bilingualism on CT further and solidify findings presented in this study.

Another area of interest, which deserves further investigation for clarification, could be the effects discovered within the Japanese-English language set, which were opposing to the outcomes of all other language sets. Results of Japanese-English participants suggest a monolingual advantage in CT. Replication of these effects would ensure that these outcomes do not originate in confounding variables and would aid the understanding of this tendency. Finally, yet importantly, and for the obvious reasons of current findings, further investigation of bilingual performance within judging likelihood and probability scenarios could lead to a clearer understanding and the confirmation of the discovered effects. Deepening the insight into bilingual cognition can lead to benefits on an individual level and possible adjustments on an educational level. Insight into the effect of the language of teaching might increase teaching quality and foster individual progress.

4.4. Concluding remarks

This thesis investigates the effects of bilingualism on higher cognition. The awareness of the importance of good critical thinking (CT) is constantly growing, with CT playing a crucial role in society today. With this in mind, and considering the fact that bilingualism is the norm rather than the exception, with over fifty per cent of the world's population speaking at least two languages, deeper insight into the effect of both concepts on each other adds valuable knowledge leading to a better understanding of modern society as a multicultural and multilinguistic construct.

References

- Aaker, J. L., & Maheswaran, D. (1997). The effect of cultural orientation on persuasion. *Journal of consumer research*, 24(3), 315-328.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207-245.
- Albert, R. T., Albert, R. E., & Radsma, J. (2002). Relationships among bilingualism, critical thinking ability, and critical thinking disposition. *Journal of professional nursing*, 18(4), 220-229.
- Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior research methods*, 50(1), 250-263.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4), 439-462.
- Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory & cognition*, 38(7), 928-940.
- Ardila, A. (1998). Bilingualism: A neglected and chaotic area. *Aphasiology*, 12(2), 131-134.
- Austin, G., Pongpaiboj, N., & Trenkic, D. (2015). Structural Competition in Second Language Production: Towards a Constraint-Satisfaction Model. *Language Learning*, 65(3), 689-722.
- Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?. *Memory & cognition*, 32(8), 1369-1378.
- Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging?. *Annals of neurology*, 75(6), 959-963.
- Bandura, A. (2002). Social cognitive theory in cultural context. *Applied psychology*, 51(2), 269-290.
- Barbeau, E. B., Chai, X. J., Chen, J. K., Soles, J., Berken, J., Baum, S., Watkins, K. E., & Klein, D. (2017). The role of the left inferior parietal lobule in second language learning: An intensive language training fMRI study. *Neuropsychologia*, 98, 169-176.
- Bataineh, R. F., & Zghoul, L. H. (2006). Jordanian TEFL graduate students' use of critical thinking skills (as measured by the Cornell Critical Thinking Test, Level Z). *International Journal of Bilingual Education and Bilingualism*, 9(1), 33-50.
- Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of memory and language*, 26(6), 658-672.
- Bellwood, P. S., & Renfrew, C. (2002). *Examining the farming/language dispersal hypothesis*. McDonald Institute for Archaeological Research.
- Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child development*, 1009-1018.
- Berger, F. R., Gupta, W. B., & Berger, R. M. (1990). *Air Force officer qualifying test (AFOQT) Form P: Test manual*. PSYCHOMETRICS INC SHERMAN OAKS CA.

- Bernardo, A. B. (2002). Language and mathematical problem solving among bilinguals. *The Journal of Psychology, 136*(3), 283-297.
- Bernardo, A. B., & Calleja, M. O. (2005). The effects of stating problems in bilingual students' first and second languages on solving mathematical word problems. *The Journal of Genetic Psychology, 166*(1), 117-129.
- Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of personality and social psychology, 45*(6), 1185.
- Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child development, 498*-510.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child development, 70*(3), 636-644.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.
- Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 60*(1), 68.
- Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism, 10*(3), 210-223.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and cognition, 12*(1), 3-11.
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological bulletin, 143*(3), 233.
- Bialystok, E. (2018). Bilingualism and executive function. *Bilingual Cognition and Language: The state of the science across its subfields, 54*, 283.
- Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological science in the public interest, 10*(3), 89-129.
- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging, 19*(2), 290.
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition, 34*(4), 859.
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in cognitive sciences, 16*(4), 240-250.
- Bialystok, E., & Hakuta, K. (1994). *In other words*. New York: Basic Books.
- Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition, 15*(2), 397-401.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525-531.
- Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment, 19*(3), 354-369.

- Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion*, 24(4), 561-595.
- Bloomfield, L. (1933). *Language history: from Language (1933 ed.)*. Holt, Rinehart and Winston.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3), 355-387.
- Bogulski, C. A., Rakoczy, M., Goodman, M., & Bialystok, E. (2015). Executive control in fluent and lapsed bilinguals. *Bilingualism: Language and Cognition*, 18(3), 561-567.
- Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3), 382-398.
- Butler, H. A. (2012). Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 26(5), 721-729.
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity*, 7(2), 112-121.
- Briley, D. A., Morris, M. W., & Simonson, I. (2000). Reasons as carriers of culture: Dynamic versus dispositional models of cultural influence on decision making. *Journal of consumer research*, 27(2), 157-178.
- Campbell, L., & Poser, W. J. (2008). *Language classification: History and method*. Cambridge University Press.
- Cao, J., & Banaji, M. R. (2016). The base rate principle and the fairness principle in social judgment. *Proceedings of the National Academy of Sciences*, 113(27), 7475-7480.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental science*, 11(2), 282-298.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology*, 54(1), 1.
- Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15(3), 139-164.
- Charles, T., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bimodal input. *Subtitles and language learning*, 173-197.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187-276.
- Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135-149.
- Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain?. *Nature Reviews Neuroscience*, 15(5), 336.
- Costa, A., Vives, M. L., & Corey, J. D. (2017). On language processing shaping decision making. *Current Directions in Psychological Science*, 26(2), 146-151.

- Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in cognitive sciences*, 10(3), 131-138.
- Crosan, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of risk and uncertainty*, 30(3), 195-209.
- Cushen, P. J., & Wiley, J. (2011). Aha! Voila! Eureka! Bilingualism and insightful problem solving. *Learning and Individual Differences*, 21(4), 458-462.
- Damian, M. F., Ye, W., Oh, M., & Yang, S. (2018). Bilinguals as "experts"? Comparing performance of mono-to bilingual individuals via a mousetracking paradigm. *Bilingualism: Language and Cognition*, 1-18.
- De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias?. *Psychological science*, 26(1), 99-107.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., Van de Moortele, P. F., Lehericy, S., & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17), 3809-3815.
- de Montes, L. G., Semin, G. R., & Valencia, J. F. (2003). Communication patterns in interdependent relationships. *Journal of Language and Social Psychology*, 22(3), 259-281.
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of personality and social psychology*, 66(5), 819.
- Dermott, B. (2002). *501 word analogy questions*. Learning Express.
- Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
- Díaz-Lago, M., & Matute, H. (2019). Thinking in a Foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, 72(1), 41-51.
- Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of bilingualism: Psycholinguistic approaches*, 179-201.
- Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42(4), 445-464.
- Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, 5(3), 175-197.
- Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta psychologica*, 43(2), 111-121.
- Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science*, 11(1), 133-155.
- Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, 15(3), 353-369.
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion*, 21(5), 1064-1076.

- Ennis, R. H. (1993). Critical thinking assessment. *Theory into practice*, 32(3), 179-186.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers*, 28(1), 1-11.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, 16(1), 143-149.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.
- Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report).
- Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal logic*, 20(1).
- Ferré, P., García, T., Fraga, I., Sánchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language?. *Cognition and Emotion*, 24(5), 760-785.
- Fischbein, S. (1980). IQ and social class. *Intelligence*, 4(1), 51-63.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive psychology*, 18(3), 253-292.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.
- Frith, C., & Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. *Cognitive brain research*, 5(1-2), 175-181.
- Gathercole, V. C. M. (2014). Bilingualism matters: One size does not fit all. *International Journal of Behavioral Development*, 38(4), 359-366.
- Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking & Reasoning*, 15(4), 355-376.
- Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology*, 11(4), 473-498.
- Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, 11(3), 279-302.
- Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and aging*, 9(3), 356.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1(2), 67-81.
- Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, 33(2), 387-396.
- Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, 13(2), 197-208.

- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & cognition*, 33(7), 1220-1234.
- Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing*, 122.
- Grosjean, F. (2010). Bilingualism, biculturalism, and deafness. *International Journal of Bilingual Education and Bilingualism*, 13(2), 133-145.
- Grosjean, F., & Byers-Heinlein, K. (2018). Speech perception and comprehension. *The Listening Bilingual: Speech Perception, Comprehension, and Bilingualism*, 25-39.
- Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature reviews neuroscience*, 11(9), 651.
- Hakuta, K. (1986). Bilingualism and Cognitive Development: Three Perspectives and Methodological Implications.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American psychologist*, 53(4), 449.
- Halpern, D. F. (2002). *Thought and knowledge: An introduction to critical thinking*. Psychology Press.
- Halpern, D. F. (2006). Is intelligence critical thinking? Why we need a new definition of intelligence. *Extending intelligence: Enhancement and new constructs*, 293310.
- Halpern, D. F. (2010). Manual HCTA: Halpern Critical Thinking Assessment. *Schuhfried*.
- Hamers, J. F., Blanc, M., & Blanc, M. H. (2000). *Bilinguality and bilingualism*. Cambridge University Press.
- Hartanto, A., & Yang, H. (2019). Does early active bilingualism enhance inhibitory control and monitoring? A propensity-matching analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(2), 360.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological science*, 15(6), 409-414.
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic bulletin & review*, 18(4), 625-658.
- Hoff, E. (2013). *Language development*. Cengage Learning.
- Hong, Y. Y., Morris, M. W., Chiu, C. Y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7), 709.
- Incera, S., & McLennan, C. T. (2016). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, 19(3), 610-620.
- Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean?. *Language learning*, 56, 279-304.

- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica*, 127(2), 277-288.
- Jackson, C. N., & Ruf, H. T. (2017). The priming of word order in second language German. *Applied Psycholinguistics*, 38(2), 315-345.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829-6833.
- Jessner, U. (1999). Metalinguistic awareness in multilinguals: Cognitive aspects of third language learning. *Language awareness*, 8(3-4), 201-209.
- Jiang, W. (2000). The relationship between culture and language. *ELT journal*, 54(4), 328-334.
- Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, 246, 150-173.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103, 582-591.
- Kantola, L., & van Gompel, R. P. (2011). Between-and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition*, 39(2), 276-290.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16(4), 705-710.
- Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, 18(1), 27-28.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, 23(6), 661-668.
- Kharkhurin, A. V. (2011). The role of selective attention in bilingual creativity. *Creativity Research Journal*, 23(3), 239-254.
- Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *Journal of personality and social psychology*, 83(4), 828.
- Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of personality and social psychology*, 63(4), 534.
- Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological review*, 107(4), 852.
- Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, memory, and cognition*, 25(6), 1534.
- Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, 109(20), 7877-7881.
- Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of cognitive psychology*, 25(5), 497-514.

- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta psychologica*, 128(3), 416-430.
- Kuhl, P. K., Stevenson, J., Corrigan, N. M., van den Bosch, J. J., Can, D. D., & Richards, T. (2016). Neuroimaging of the bilingual brain: Structural brain correlates of listening and speaking in a second language. *Brain and language*, 162, 1-9.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621-647.
- Lee, H., & Kim, K. H. (2011). Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link. *Personality and individual differences*, 50(8), 1186-1190.
- Lee, M., & Smith, G. (2002). Regression to the mean and football wagers. *Journal of Behavioral Decision Making*, 15(4), 329-342.
- Lee, C. S., Therriault, D. J., & Linderholm, T. (2012). On the cognitive benefits of cultural experience: Exploring the relationship between studying abroad and creative thinking. *Applied Cognitive Psychology*, 26(5), 768-778.
- Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, 43(6), 431.
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*, 144(4), 394.
- Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 381.
- Lin, W. L., & Lien, Y. W. (2013). The different role of working memory in open-ended versus closed-ended creative problem solving: a dual-process theory account. *Creativity Research Journal*, 25(1), 85-96.
- Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic bulletin & review*, 2(2), 174-207.
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605-621.
- Luk, G., De Sa, E. R. I. C., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control?. *Bilingualism: Language and cognition*, 14(4), 588-595.
- Luk, Z. P. S., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural-s, articles, and possessive's. *Language Learning*, 59(4), 721-754.
- Lyovin, A. (1997). *An Introduction to the Languages of the World*. Oxford University Press, USA.
- Maass, A., Karasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic analysis of person representation. *Journal of personality and social psychology*, 90(5), 734.

- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, 109(2), 163.
- Maier, N. R., & Burke, R. J. (1967). Response availability as a factor in the problem-solving performance of males and females. *Journal of Personality and Social Psychology*, 5(3), 304.
- Maier, N. R., & Solem, A. R. (1952). The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions. *Human Relations*, 5(3), 277-288.
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *science*, 341(6149), 976-980.
- Marian, V. (2018). Bilingual research methods. In *An Introduction to Bilingualism* (pp. 28-52). Routledge.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940-967.
- Marian, V., Lam, T. Q., Hayakawa, S., & Dhar, S. (2018). Top-down cognitive and linguistic influences on the suppression of spontaneous otoacoustic emissions. *Frontiers in neuroscience*, 12.
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97-115.
- Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6(1), 1-13.
- Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: language and cognition*, 11(1), 81-93.
- Matsuoka, J., & Lepage, Y. (2014). Measuring Similarity from Word Pair Matrices with Syntagmatic and Paradigmatic Associations. In *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex)* (pp. 77-86).
- Matthews, P. H. (2003). *The Concise Oxford Dictionary of Linguistics: Over 3000 Entries*. Oxford University Press.
- Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(2), 288.
- Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & cognition*, 15(3), 238-246.
- Mishra, R. K. (2015). Let's not forget about language proficiency and cultural variations while linking bilingualism to executive control. *Bilingualism: Language and Cognition*, 18(1), 39-40.
- Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology*, 46(4), 759-778.
- Nagin, D. S., & Pogarsky, G. (2003). An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity. *Criminology*, 41(1), 167-194.
- Newstead, S. E., Pollard, P., Evans, J. S. B., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, 45(3), 257-284.

- Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In Carroll, J. S. & Payne, J. W. (Eds.), *Cognition and social behaviour*, 113-134, Hillsdale, NJ: Erlbaum.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological review*, 90(4), 339.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Upper Saddle River, NJ: Prentice Hall.
- Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental science*, 15(4), 516-527.
- O'Hare, L., & McGuinness, C. (2009). Measuring critical thinking, intelligence, and academic performance in psychology undergraduates. *The Irish Journal of Psychology*, 30(3-4), 123-131.
- Oyserman, D., & Lee, S. W. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological bulletin*, 134(2), 311.
- Paap, K. R. (2016). The neuroanatomy of bilingualism: will winds of change lift the fog?. *Language, Cognition and Neuroscience*, 31(3), 331-334.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, 66(2), 232-258.
- Palmer, S. D., van Hooff, J. C., & Havelka, J. (2010). Language representation and processing in fluent bilinguals: Electrophysiological evidence for asymmetric mapping in bilingual memory. *Neuropsychologia*, 48(5), 1426-1437.
- Paul, R., & Elder, L. (2006). *Critical thinking competency standards*. Dillon Beach, CA: Foundation for Critical Thinking.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: general and applied*, 76(27), 1.
- Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 313.
- Posner, M. I. (1973). *Cognition: An introduction*. Glenview, IL: Scott, Foresman, and Co.
- Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive psychology*, 33(1), 43-63.
- Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 295-301.
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology*, 79(1), 49-71.
- Raven, J. C. (1936). Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. *Unpublished master's thesis, University of London*.

- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, 41(1), 1-48.
- Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review*, 11(4), 317-336.
- Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, 16(4-6), 635-645.
- Saer, D. J. (1923). The effect of bilingualism on intelligence. *British Journal of Psychology. General Section*, 14(1), 25-38.
- Sánchez, G. I. (1934). Bilingualism and mental measures. A word of caution. *Journal of Applied Psychology*, 18(6), 765.
- Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *PloS one*, 12(4), e0176151.
- Schmitt, B. H., & Zhang, S. (1998). Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research*, 25(2), 108-122.
- Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of experimental psychology: General*, 122(2), 166.
- Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, 56(2), 153-171.
- Semin, G. R. (2000). Agenda 2000—Communication: Language as an implementational device for cognition. *European Journal of Social Psychology*, 30(5), 595-612.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, 30(4), 510-532.
- Sörman, D. E., Hansson, P., & Ljungberg, J. K. (2019). Different Features of Bilingualism in Relation to Executive Functioning. *Frontiers in Psychology*, 10(269).
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57(1), 31-95.
- SPSS (2015). IBM SPSS statistics for windows, Version 23.0.
- Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. *Innovations in educational psychology: Perspectives on learning, teaching and human development*, 195-237.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of experimental psychology: general*, 127(2), 161.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological review*, 84(4), 353.
- Sternberg, R. J. (1986). *Critical Thinking: Its Nature, Measurement, and Improvement*. National Institute of Education.
- Sternberg, R. J. (2004). Culture and intelligence. *American psychologist*, 59(5), 325.

- Sternberg, R. J., & Davidson, J. E. (1983). Insight in the gifted. *Educational Psychologist*, 18(1), 51-57.
- Surrain, S., & Luk, G. (2017). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, 1-15.
- Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in higher education*, 36(1), 23-39.
- Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104(30), 12530-12535.
- Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International journal of bilingual education and bilingualism*, 21(2), 197-206.
- Toplak, M. E., Liu, E., MacPherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making*, 20(2), 103-124.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, 39(7), 1275.
- Tran, C. D., Arredondo, M. M., & Yoshida, H. (2018). Early executive function: The influence of culture and bilingualism. *Bilingualism: Language and Cognition*, 1-19.
- Trenkic, D., Mirkovic, J., & Altmann, G. T. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, 17(2), 237-257.
- Trenkic, D., & Warmington, M. (2018). Language and literacy skills of home and international university students: How different are they, and does it matter?. *Bilingualism: Language and Cognition*, 1-17.
- Turney, P. D. (2008, August). A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics*, 1, 905-912. Association for Computational Linguistics.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4), 293.
- Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, 18(1), 3-24.
- Van Rinsveld, A., Schiltz, C., Brunner, M., Landerl, K., & Ugen, S. (2016). Solving arithmetic problems in first and second language: Does the language context matter?. *Learning and instruction*, 42, 72-82.
- Vives, M. L., Aparici, M., & Costa, A. (2018). The limits of the foreign language effect on decision-making: The case of the outcome bias and the representativeness heuristic. *PLoS one*, 13(9), e0203528.
- Vogel, K., & Roepstorff, A. (2009). Contextualising culture and social cognition. *Trends in cognitive sciences*, 13(12), 511-516.

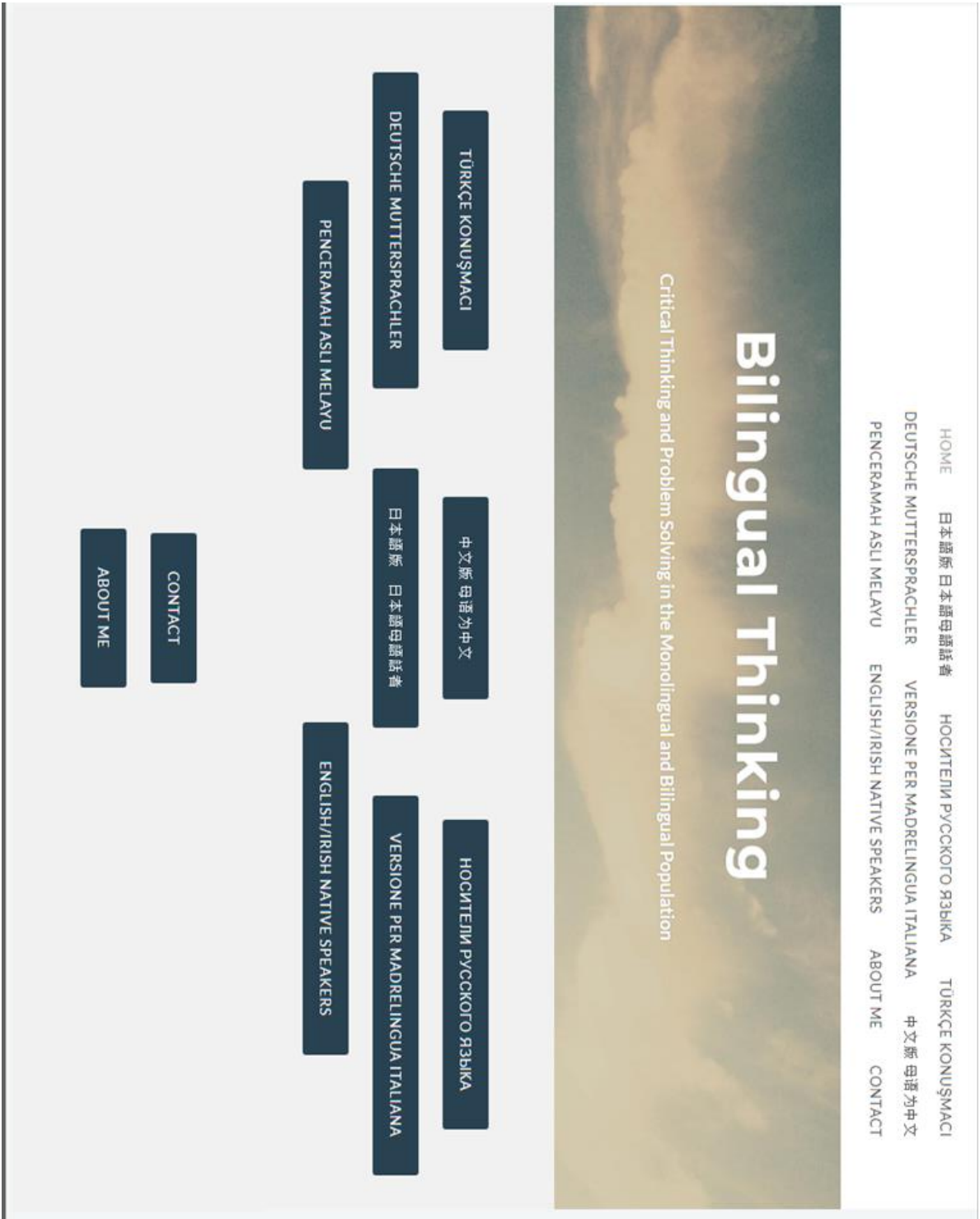
- Walsh, C. M., & Hardy, R. C. (1999). Dispositional differences in critical thinking related to gender and academic major. *Journal of Nursing Education, 38*(4), 149-155.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron, 37*(1), 159-170.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology, 20*(3), 273-281.
- Watson, J. B. (1924). The Unverbalized in Human Behavior. *Psychological Review, 31*(4), 273.
- Webb, M. E., Little, D. R., Cropper, S. J., & Roze, K. (2017). The contributions of convergent thinking, divergent thinking, and schizotypy to solving insight and non-insight problems. *Thinking & Reasoning, 23*(3), 235-258.
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological bulletin, 120*(1), 3.
- Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition, 14*(3), 412-422.

Appendices

Appendix 1: Website Used for Data Collection

<https://bilingualthinking.weebly.com/>

Home Page



English/Irish Native Speakers

HOME 日本語版 日本語母語話者 НОСИТЕЛИ РУССКОГО ЯЗЫКА TÜRKÇE KONUŞMACI
DEUTSCHE MUTTERSPRACHLER VERSIONE PER MADRELINGUA ITALIANA 中文版 母语为中文
PENCERAMAH ASLI MELAYU ENGLISH/IRISH NATIVE SPEAKERS ABOUT ME CONTACT

English/Irish Native Speakers

General Information

Invitation

You are invited to take part in this research study. Before you decide, it is important that you understand why the research is being done and what it will involve. This *Participant Information* tells you about the purpose, risks and benefits of this research study. If there is anything that you are not clear about, we will be happy to explain it to you.

Purpose of the Study

This study is concerned with the critical thinking skills and how this is being influenced by living and working in one or two languages. Participants will be tested on their critical thinking skills from a variety of scenarios. The questionnaires will take approximately 30-50 minutes to complete.

[LINK TO STUDY](#)

Taking Part – What it involves

Do I have to take part?

It is up to you to decide whether or not to take part. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect your rights in any way.

What will happen to me if I take part?

If you chose to take part in this study you will be presented with a series of scenarios with questions we wish you to answer. We ask that you answer these questions honestly and to the best of your ability.

How long will my part in the study last?

The questionnaire will take around 30-50 minutes to complete.

What are the possible benefits in taking part?

Answering questions about your general critical thinking skills may make you more aware of your abilities and may in turn improve them.

What are the possible disadvantages and risks of taking part?

There are no foreseeable risks attached to taking part in this study. If however during the course of this study you would like to speak to someone about the issues it raises we will be happy to recommend someone to you.

What happens if I change my mind during the study?

You are entitled to change your mind about participating in this at any time without disadvantage or penalty.

Who do I contact for more information or if I have further concerns?

Ph.D. Student Researcher:

Sophia Arndt email: s.arndt2@nuigalway.ie (or click on "contact form")

Supervisor:

Dr. Mark Elliott email: markelliott@nuigalway.ie

[CONTACT FORM](#)

Appendix 2: SES Background Questionnaire

- (1.1) Age
- (1.2) Gender: Male – Female – Other
- (1.3) What type of school did you mainly attend between the ages of 13 and 17?
State run/state funded school – independent/fee paying school – home schooled – prefer not to say – don't know
- (1.4) How many people are currently living in your household, including yourself?
Only me – family home/parent's house – my partner and me – shared flat/house/student accommodation
- (1.5) Parents occupation: Did any of your parent(s)/guardian(s) complete a university degree course or equivalent (for example Bachelor)?
Yes – no – prefer not to say
- (1.6) What describes your financial situation best?
Own income - parental/family support – student loan/grant – savings – benefits/scholarship
- (1.7) Including wages, salaries, self-employment, and any other source of income we just talked about, was your total combined income during the last 12 months?
Less than 5000€ - less than 10000 – less than 20000 – 20000 or more – no income – prefer not to say

Do you speak Irish?

Do you speak any other language except English and Irish?

Please rate your fluency in Irish on a scale from one to ten with one being none and ten being fluent.

Appendix 3: Language Questionnaire (LEAP-Q)

<https://bilingualism.northwestern.edu/leapq/>; last checked 09/04/2019

- (2.1) Please list all the languages you know IN ORDER OR DOMINANCE (your strongest language first):
- (2.2) Please list all the languages you know IN ORDER OF ACQUISITION (your native language first)
- (2.3) Please list what percentage of the time you are CURRENTLY and ON AVERAGE exposed to each language. (your percentages should add up to 100%)
Language:
Percent:
- (2.4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. (your percentages should add up to 100%)
Language:
Percent:
- (2.5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time. (your percentages should add up to 100%)
Language:
Percent:
- (2.6) Please name the cultures with which you identify. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc;
- (2.7) How many years of formal education do you have?
- (2.8) Please check your highest education level (or the approximate equivalent to a degree)
Currently doing a Bachelor Degree – Bachelor – Master – PhD – other form of higher education
- (2.9) What's your country of origin?
- (2.10) What's your parent's country of origin?
- (2.11) What's your current country of residency?
- (2.12) Have you ever had a: vision problem – hearing impairment – language disability – learning disability (check all applicable).
If yes, please explain (including any corrections).

LANGUAGE:

This is my native/**first** – second – third language

- (3.1) Age when you:
Began acquiring:
Became fluent in:
Began reading in:
- (3.2) Please list the number of years and months you spent in each language environment:
A country where your **first** language is spoken: years – months
A family where your **first** language is spoken: years – months
A school and/or working environment where your **first** language is spoken: years – months
A circle of friends where your **first** language is spoken
- (3.3) On a scale from one to ten, please select your level of PROFICIENCY in speaking, understanding, and reading your **first** language: (1=none, 2=very low, 3=low, 4=slightly less than adequate, 5=adequate, 6=slightly more than adequate, 7=good, 8=very good, 9=excellent, 10=perfect)
Speaking:
Understanding spoken language:
Reading:
- (3.4) On a scale from one to ten, please select how much the following factors contributed to you learning (1=not a contributor, 1= minimal contributor, 5= moderate contributor, 10= most important contributor)
Interacting with friends:
Interacting with family:
Reading:
Language Course/School:
Self instruction:
Media (TV, Radio, Internet, ...):
- (3.5) Please rate to what extend you are currently exposed to your **first** language in the following contexts: (1=never, 2= almost never, 5= half of the time, 10= always):
Interacting with friends:
Interacting with family:
Studying/working:
Media (TV, Radio, Internet, ...)
Reading:
- (3.6) In your perception, how much of a foreign accent do you have in your **first** language? (1= none, 2= almost none, 3= light, 4= some, 5= moderate, 6=considerable, 7= heavy, 8= very heavy, 9= extremely heavy, 10= pervasive).
- (3.7) Please rate how frequently others identify you as a non-native speaker based on your accent in your **first** language: (1=never, 2=almost never, 5=half the time, 10= always).

Did any of the questions seem familiar to you? (If yes, please explain)
Would you like to leave a comment on anything you noticed during the survey?

Appendix 4: Assessment *Mixed Measures*

English

Hypothesis Testing

1.1.1 Version A

A high school student had to choose between two colleges A and B. The student had several friends, who were similar to himself in values and abilities, at each college. All of his friends at college A liked it on both educational and social grounds; all of them at college B had problems on both grounds. The student visited both colleges for a day, and his impressions were the opposite of their reports.

Which college should the high school student choose?

- College A
- College B

1.1.2 Version B

A mother and a father had to choose between two kindergartens A and B for their first child. The couple had several friends, who were similar to them in values and beliefs, who had their children at each kindergarten. All of their friends with a child in kindergarten A liked it on both educational and social grounds; all of them with children at kindergarten B had problems on both grounds. The couple visited both kindergartens for a few hours, and their impressions were the opposite of their reports.

Which kindergarten should the mother and the father choose?

- Kindergarten A
- Kindergarten B

1.2.1 Version A

A city called Kingston has had an unpopular police chief for a year and a half. He is politically active and a colleague of the mayor, and he had little previous experience in police administration when he was appointed. The mayor has recently defended the chief in public, announcing that in the time since he took office, crime rates decreased by 12%. Which of the following pieces of evidence would most weaken the mayor's claim that his chief is competent?

- The crime rates of the two cities closest to Kingston in location and size have decreased by 18% in the same period.
- An independent survey of the citizens of Kingston shows that 40% more crime is reported by respondents in the survey than is reported in police records.
- Common sense indicates that there is little a police chief can do to lower crime rates. These are for the most part due to social and economic conditions beyond the control of officials.
- The police chief has been discovered to have personal contacts with people who are known to be involved in organized crime.

1.2.2 Version B

A high school has had an unpopular teacher for a year and a half. He is a friend of the headmaster, and he had little previous experience in teaching when he was appointed. The headmaster has recently defended the teacher in front of the school, announcing that in the time since he started teaching, his student's misbehaviour decreased by 12%. Which of the following pieces of evidence would most deflate the headmaster's claim that this teacher is competent?

- Student's misbehaviour rates of the two classes not taught by this teacher have decreased by 18% in the same period.
- A survey of the students shows that 40% more misbehaviour is reported by respondents in the survey than is reported by the headmaster.
- Common sense indicates that there is little a teacher can do to lower misbehaviour. These are for the most part due to the student's mood and behaviour.
- The teacher has been discovered to have penalized students which have misbehaved.

1.3.1 Version A

Consider a decision about whether a car is of the brand X or of the brand Y. Assume that two types of information are potentially available about each alternative-the percentage of car X and

car Y that have a consumption of more than 10 l/100km (liters per 100 km) and the percentage of car X and car Y that have had no major mechanical problems in the first two years of ownership.

Assume that an unknown car (either of the brand X or Y) does over 10 l/100 km and that it has had no major mechanical problems in the first two years of ownership. Also assume that 65% of car brand X consume more than 10 l/100 km.

If you could find out one of the three following types of information, which one would help you to decide which one the unknown car is, car X or car Y?

- Percentage of cars of brand Y that get over 10 l/100 km
- Percentage of cars of brand X that have had no major mechanical problems in the first two years
- Percentage of cars of brand Y that have had no major mechanical problems in the first two years

1.3.2 Version B

Consider a decision about whether an unknown laptop is brand X or brand Y. Assume that two types of information are potentially available about each alternative-the percentage of brand X and brand Y that have a battery life of more than 10 hours and the percentage of brand X and brand Ys that have had no major technical problems in the first two years of ownership.

Assume that an unknown laptop (either X or Y) had a battery life over 10 hours and that it has had no major technical problems in the first two years of ownership. Also assume that 65% of brand X laptops have a battery life over 10 hours.

If you could find out one of the three following types of information, which one would help you to decide which one the unknown laptop is, brand X or brand Y?

- Percentage of brand Y that get over 10 hours battery life.
- Percentage of brand X that have had no major technical problems in the first two years
- Percentage of brand Y that have had no major technical problems in the first two years

1.4.1 Version A

Mr. Maxwell attended a party to which only university professors and business executives were invited (whereas guests would be either or, not both of the positions combined). The only thing you know about Mr. Maxwell is that he is a member of the Bear's Club. You are asked to judge the probability that Mr. Maxwell is a university professor by asking 1, 2, 3, or 4 of the questions given below. You are asked to check their relevancy. A relevant question is one the answer to which will help you in your judgement. Please consider each of the questions separately and indicate whether it is relevant or irrelevant for your task.

- What percentage of the people at the party are university professors?
- What percentage of the Bear's Club members are at the party?
- What percentage of the university professors at the party are members of the Bear's Club?
- What percentage of the business executives at the party are members of the Bear's Club?

1.4.2 Version B

Mrs. Crowe attended a guided tour which is only attended by American tourists and Canadian tourists (whereas guests would have either or nationality, not both of them combined). The only thing you know about Mrs. Crowe is that she is a guest of hotel ABC. You are asked to judge the probability that Mrs. Crowe is an American tourist by asking 1, 2, 3, or 4 of the questions given below. You are asked to check their relevancy. A relevant question is one the answer to which will help you in your judgement. Please consider each of the questions separately and indicate whether it is relevant or irrelevant for your task.

- What percentage of the people at the tour are American tourists?
- What percentage of guests of hotel ABC are attending the tour?
- What percentage of the Canadian tourists from the tour are guests of hotel ABC?
- What percentage of American tourists from the tour are guests of hotel ABC?

1.5.1 Version A

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Which of two alternatives is more probable:

- Linda is a bank teller
- Linda is a bank teller and is active in the feminist movement

1.5.2 Version B

Jack is 45, he is married and has four children. He is generally careful and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Which of two alternatives is more probable:

- Jack is an engineer
- Jack is an engineer and is active in the local club for water sports

1.6.1 Version A

You wish to buy a new car. Today you must choose between two alternatives: to purchase either a car from company A or B. You use only one criterion for that choice, the car's life expectancy. You have information from Consumer Reports that in a sample of several hundred cars the car A has the better record. Yesterday a neighbour told you that his new car of brand A broke down. Which car do you buy?

- Car A
- Car B

1.6.2 Version B

You wish to buy a new computer. Today you must choose between two alternatives: to purchase either a computer A or computer B. You use only one criterion for that choice, the computer's life expectancy. You have information from Consumer Reports that in a sample of several hundred computers the computer A has the better record. Yesterday a neighbour told you that his new computer of brand A broke down. Which car do you buy?

- computer A
- computer B

Verbal Reasoning

Analogies

A1

_____ : horse :: board : train

- stable
- shoe
- ride
- mount

B1

poetry : rhyme :: philosophy : _____

- imagery
- music
- bi-law
- theory

A2

mend : sewing :: edit : _____

- darn
- repair
- manuscript

- makeshift

B2

alphabetical : _____ :: sequential : files

- sort
- part
- list
- order

A3

mouse : _____ :: flash : camera

- rat
- computer
- cord
- dessert

B3

cushion : sofa :: shelf : _____

- ledge
- bookcase
- storage
- frame

A4

moisten : _____ :: cool : freeze

- water
- soak
- oven
- grow

B4

ribbon : _____ :: icing : cake

- present
- cut
- bow
- typewriter

A5

kitten : _____ :: soldier : army

- cat
- litter
- puppy
- meow

B5

scientist : experiment :: _____ : play

- beaker
- rehearsal
- actor
- lab

A6

dictionary : definition :: _____ : map

- direction
- south
- atlas
- longitude

B6

thumbtack : _____ :: hook : coat

- nail
- poster
- wall
- hammer

A7

unusual : novelty :: _____ : standard

- odd
- novel
- familiar
- poem

B7

_____ : college :: mechanic : garage

- book
- learning
- professor
- engine

A8

drum : instrument :: drill : _____

- hammer
- oven
- tool
- crescendo

B8

peak : mountain :: _____ : house

- maximize
- roof
- porch
- bungalow

A9

_____ : rainfall :: condensation : humidity

- erosion
- cloud
- ground
- forecast

B9

_____ : frame :: mosaic : tile

- film
- engraving

- bathroom
- decoration

Judging Likelihood and Probability

3.1.1 Version A

A game of squash can be played to either 9 or 15 points. If A is a better player than B, which scoring scheme would give player A a better chance of winning?

- 9 point game
- 15 point game

3.1.2 Version B

Table tennis can be won with best score out of 5 games or best score out of 9 games. If A is a better player than B, which scoring scheme would give player A a better chance of winning?

- Best out of 5 games
- Best out of 9 games

3.2.1 Version A

After the first 2 weeks of the major league baseball season, newspapers begin to print the top 10 batting averages. Typically, after 2 weeks, the leading batter often has an average of about .450. However, no batter in major league history has ever averaged .450 at the end of the season. Why do you think this is? Tick one:

- When a batter is known to be hitting for a high average, pitchers bear down more when they pitch to him.
- Pitchers tend to get better over the course of a season, as they get more in shape. As pitchers improve, they are more likely to strike out batters, so batters' averages go down.
- A player's high average at the beginning of the season may be just luck. The longer season provides a more realistic test of a batter's skill.
- A batter who has such a hot streak at the beginning of the season is under a lot of stress to maintain his performance record. Such stress negatively affects his playing.
- When a batter is known to be hitting for a high average, he stops getting good pitches to hit. Instead, pitchers "play the corners" of the plate because they don't mind walking him.

3.2.2 Version B

After the first 2 weeks of the academic term, a teacher of a school tries to predict the top 10 students of the year. Typically, after 2 weeks, the leading student often has an average of about 92%. However, no student in the school's history has ever averaged 92% at the end of the term. Why do you think this is? Circle one:

- When a student is known to be aiming for a high average, teachers increase their expectations when grading this student's papers.
- Papers tend to get more difficult over the course of a term, as the content accumulates. As papers get more difficult, students are less likely to score high marks, so student's averages go down.
- A student's high average at the beginning of the term may be just luck. The longer academic year provides a more realistic test of a student's knowledge and ability.
- A student who has such a high average at the beginning of the term is under a lot of stress to maintain his average. Such stress adversely affects his results.
- When a student is known to be aiming for a high average, he stops getting asked easy questions by the teachers. Instead, teachers "increase the intensity" of the questions for this student because they want to test the student's ability.

3.3.1 Version A

When playing slot machines, people win something about 1 in every 10 times. Julie, however, has just won on her first three plays. What are her chances of winning the next time she plays? _____ out of 10

3.3.2 Version B

When participating in a local lottery, people have the chance to win a prize about 1 in every 20 lottery tickets. Jack, however, has just won on his first three tickets. What are his chances of winning the next time he plays?

_____ out of 20

3.4.1 Version A

Imagine that we are tossing a fair coin (a coin that has a 50/50 chance of coming up heads or tails) and it has just come up heads 5 times in a row. For the 6th toss do you think that:

- It is more likely that tails will come up than heads.
- It is more likely that heads will come up than tails.
- Heads and tails are equally probable on the sixth toss.

3.4.2 Version B

Imagine that we are throwing a fair die (a die that has a 1 in 6 chance of coming up with each number) and number 6 has just come up 5 times in a row. For the 6th toss do you think that:

- It is more likely than the first five times that a different number will come up than 6.
- It is more likely than the first five times that 6 will come up than a different number.
- 6 and any other number are equally probable on the sixth toss.

3.5.1 Version A

Assume that you are presented with two trays of black and white marbles: a large tray that contains 100 marbles and a small tray that contains 10 marbles. You must draw out one marble from either tray without looking. If you draw a black marble, you win \$2. Consider a condition in which the small tray contains 1 black marble and 9 white marbles, and the large tray contains 8 black marbles and 92 white marbles. From which tray would you prefer to select a marble in a real situation?

- Large tray
- Small tray

3.5.2 Version B

Assume that you are presented with two bags of \$10 and \$20 notes: a large bag that contains 100 different notes and a small bag that contains 10 different notes. The notes are mixed randomly in the bags. You must draw out one note from either bag without looking and you will get to keep the note. Consider a condition in which the small bag contains one \$20 note and nine \$10 notes, and the large bag contains eight \$20 notes and ninety-two \$10 notes. From which bag would you prefer to select a note in a real situation?

- Large bag
- Small bag

Argumentation Analysis

Please tick: valid - invalid

A1

- No cigarettes are inexpensive.
- Some addictive things are inexpensive.
- Therefore, some addictive things are not cigarettes.

B1

- Some buildings are wooden buildings
- no wooden buildings are skyscrapers
- Therefore some buildings are not skyscrapers.

A2

- some items of office equipment are not computers
- some electrical appliances are items of office equipment
- Therefore Some electrical appliances are not computers.

B2

- Some drinks are non-alcoholic
- non-alcoholic drinks are not liquors
- Therefore Some drinks are not liquors.

A3

- Chairs are not tables.
- some pieces of furniture are chairs
- Therefore some pieces of furniture are not tables

B3

- Some fish are inedible
- trout are edible
- Therefore some fish are not trout.

B4

- No addictive things are inexpensive.
- Some cigarettes are inexpensive.
- Therefore, some addictive things are not cigarettes.

A4

- some wooden buildings are skyscrapers
- no buildings are wooden buildings
- Therefore Some buildings are not skyscrapers.

B5

- some electrical appliances are not items of office equipment
- some items of office equipment are computers
- Therefore Some electrical appliances are not computers.

A5

- non-alcoholic drinks are liquors
- no drinks are non-alcoholic
- Therefore Some drinks are not liquors.

B6

- No pieces of furniture are chairs
- chairs are tables.
- Therefore some pieces of furniture are not tables

A6

- No trout are edible
- fish are edible
- Therefore some fish are not trout.

A7

- Some tools are not manual
- screwdrivers are manual tools
- Therefore some screwdrivers are not tools.

B7

- Some musical instruments are wind instruments
- wind instruments are not violins
- Therefore some violins are not musical instruments.

A8 // B8

- Some green things are not grasses.

- some plants are green.
- Therefore some grasses are not plants.

B9

- Screwdrivers are not manual tools
- No tools are manual
- Therefore some screwdrivers are not tools.

A9

- Wind instruments are violins
- No musical instruments are wind instruments
- Therefore some violins are not musical instruments.

Problem Solving

5.1.1 Version A

A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

- 5 cents
- 10 cents
- 55 cents
- 100 cents

5.1.2 Version B

A loaf of bread and an apple cost \$2.20 in total. The bread costs \$2 more than the apple. How much does the apple cost?

- 10 cents
- 20 cents
- 110 cents
- 200 cents

5.2.1 Version A

If it takes 5 machines 5 minutes to make 5 plastic parts, how long would it take 100 machines to make 100 plastic parts?

- 1 minute
- 5 minutes
- 25 minutes
- 100 minutes

5.2.2 Version B

If it takes 2 bakers 2 minutes to make 2 loafs of bread, how long would it take 50 bakers to make 50 loafs of bread?

- 1 minute
- 2 minutes
- 25 minutes
- 50 minutes

5.3.1 Version A

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 24 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

- 2 days
- 4 days
- 12 days
- 23 days

5.3.2 Version B

A container is being filled with rocks. The amount of rocks doubles in number every day. The container will be full in 10 days. In how many days will it be half full?

- 2 days
- 4 days
- 5 days
- 9 days

5.4.1 Version A

There are black and brown loose socks in a drawer mixed in a ratio of 4 to 5. How many separate socks would you have to take out without looking to be sure of getting a pair of the same colour?

- 1 sock
- 2 socks
- 3 socks
- 4 socks
- 5 socks

5.4.2 Version B

There are green and blue marbles in a bag mixed in a ratio of 6 to 8. How many marbles would you have to take out without looking to be sure of getting a two of the same colour?

- 2 marbles
- 3 marbles
- 4 marbles
- 5 marbles
- 6 marbles

5.5.1 Version A

Two trains, 50 km apart start towards each other at 25 km/h (km per hour) each. As the trains start a bird flies from the front of one train towards the second. On reaching the second train the bird turns round and flies back to the first train and so on until the trains meet. If the bird flies at 60 km/h how many km will the bird have flown before the trains meet?

- 15 km
- 30 km
- 45 km
- 60 km

5.5.2 Version B

Two pedestrians, 10 km apart start walking towards each other at 5 km/h. As the pedestrians start walking a dog runs from one person towards the second. On reaching the second person the dog turns round and runs back to the first person and so on until the pedestrians meet. If the dog runs at 20 km/h how many kilometres will the dog have run before the pedestrians meet?

- 2 km
- 5 km
- 10 km
- 20 km

5.6.1 Version A

A man buys a horse for £60, sells it for £70, buys it back for £80 and sells it finally for £90. How much has he made?

- £10
- £20
- £30
- £40

5.6.2 Version B

A lady buys shares of a company at the stock market for \$600, sells them for \$700, buys them back for \$800 and finally sells them for \$900. How much has she made?

- \$100
- \$200
- \$300

- \$400

5.7.1 Version A

The police were convinced that either A, B, C, or D had committed a crime. Each of the suspects, in turn, made a statement, but only one of the four statements was true.

- A said, "I didn't do it."
- B said, "A is lying."
- C said, "B is lying."
- D said, "B did it."

Who is telling the truth? Who committed the crime?

Truth: A – B – C – D

Crime: A – B – C – D

5.7.2 Version B

The judge knows that either person A, B, C, or D came up with an invention. Each of them, in turn, made a claim, but only one of the four statements was true.

- A said, "I invented it."
- B said, "A is lying."
- C said, "B is lying."
- D said, "B didn't invent it."

Who is telling the truth? Who came up with the invention?

Truth: A – B – C – D

Invention: A – B – C – D

5.8.1 Version A

You are shown a set of four cards placed on a table, each of which has a number on one side and a colour on the other side. The visible faces of the cards show 3, 8, red and brown. Which card(s) do you have to turn over in order to test the truth of the proposition that if a card shows an even number on one face, then its opposite face is red? Try to work in the most efficient way by turning around the least amount of cards necessary.

Mark as many as apply: 3 – 8 – red – brown

5.8.2 Version B

You are shown a set of four cards placed on a table, each of which has a number on one side and a colour on the other side. The visible faces of the cards show 5, 6, green and yellow. Which card(s) must you turn over in order to test the truth of the proposition that if a card shows an odd number on one face, then its opposite face is green? Try to work in the most efficient way by turning around the least amount of cards necessary.

Mark as many as apply: 5 – 6 – green – yellow

Irish

Hypothesis Testing

1.1.1 Leagan A

Bhí ar mhac léinn ardscoile rogha a dhéanamh idir dhá choláiste, A agus B. Bhí go leor cairde ag an mac léinn mar é féin ó thaobh luachanna agus cumais, ag gach coláiste. Thaitin coláiste A lena chairde uile a bhí ann ar chúinsí oideachais agus sóisialta araon; bhí fadhbanna ag a chairde uile ag coláiste B ar na cúinsí sin. Chaith an mac léinn lá amháin ag gach aon choláiste, agus bhí a mhalairt de thuairim aigesean ná mar a bhí ag a chairde.

Cén choláiste ar chóir don mhac léinn ardscoile a roghnú?

- Coláiste A
- Coláiste B

1.1.2 Leagan B

Bhí ar mháthair agus athair rogha a dhéanamh idir dhá náiscoil, A agus B, dá gcéad pháiste. Bhí roinnt cairde ag an lánúin, a bhí mar iad ó thaobh luachanna agus creidiúintí, a raibh a gcuid páistí i ngach náiscoil. Thaitin náiscoil A lena gcairde uile a raibh páiste acu ann ar chúinsí oideachais agus sóisialta araon; bhí fadhbanna ag a gcairde uile a raibh páistí acu i náiscoil B ar na cúinsí sin. Chaith an lánúin cúpla uair an chloig ag gach aon náiscoil, agus bhí a mhalairt de thuairim acu ná mar a bhí ag a gcairde.

Cén náiscoil ar chóir don mháthair agus don athair a roghnú?

- Náiscoil A
- Náiscoil B

1.2.1 Leagan A

Bhí ardoifigeach póilíneachta nach raibh tóir air i gcathair ar a dtugtar Kingston le bliain go leith. Tá sé gníomhach i gcúrsaí polaitíochta agus is comhghleacaí den mhéara é, agus is beag taithí a bhí aige i gcúrsaí riaracháin póilíneachta sular ceapadh é. Chosain an méara an t-ardoifigeach go poiblí le déanaí, agus d'fhógair sé go bhfuil rátaí coireachta íslithe 12% ó ghlac sé an post. Cé acu de na giotaí fianaise seo a leanas is mó a lagódh éileamh an mhéara go bhfuil an t-ardoifigeach inniúil?

- Tá rátaí an dá chathair is gaire do Kingston ó thaobh suímh agus méide íslithe 18% sa tréimhse chéanna.
- Léirítear le suirbhé neamhspleách ar shaoránaigh Kingston go bhfuil 40% níos mó coireachta tuairiscithe ag freagróirí sa suirbhé ná mar atá i dtaifid na bpóilíní.
- Léiríonn ciall gur beag is féidir le hardoifigeach póilíní a dhéanamh le rátaí coireachta a ísliú. Tarlaíonn sé sin den chuid is mó mar thoradh ar chúinsí sóisialta agus eacnamaíochta nach bhfuil aon smacht ag oifigigh orthu.
- Fuarthas amach go bhfuil ceangal príobháideach ag ardoifigeach na bpóilíní le daoine a bhfuil a fhios go bhfuil baint acu le coireacht eagraithe.

1.2.2 Leagan B

Tá múinteoir nach bhfuil tóir air in ardscoil le bliain go leith. Is cara leis an ardmháistir é, agus is beag taithí múinteoireachta a bhí aige sular ceapadh é. Chosain an t-ardmháistir an múinteoir os comhair na scoile le déanaí, agus d'fhógair gur ísligh mí-iompraíocht a chuid mac léinn 12% ón am ar thosaigh sé ag múineadh. Cé acu de na giotaí fianaise seo a leanas is mó a thraoithfeadh éileamh an ardmháistir go bhfuil an múinteoir seo inniúil?

- Tá rátaí mí-iompraíochta mac léinn an dá rang nach bhfuil á múineadh ag an múinteoir seo íslithe 18% sa tréimhse chéanna.
- Léirítear le suirbhé ar na mic léinn go bhfuil 40% níos mó mí-iompraíochta tuairiscithe ag freagróirí sa suirbhé ná mar atá ag an ardmháistir.
- Léiríonn ciall gur beag is féidir le múinteoir a dhéanamh le leibhéil mhí-iompraíochta a ísliú. Tarlaíonn sé sin den chuid is mó mar thoradh ar ghiúmar agus iompraíocht an mhic léinn.
- Fuarthas amach gur ghearr an múinteoir pionós ar mhic léinn a bhí i mbun mí-iompar.

1.3.1 Leagan A

Smaoinigh ar chinneadh maidir cibé an branda X nó branda Y é carr. Glac leis go bhféadfadh dhá chineál píosa eolais a bheith ar fáil faoi gach aon cheann-an céatadán de charr X agus de charr Y a bhaineann níos mó ná 10l/100km (lítear in aghaidh 100km) amach agus an céatadán

de charr X agus de charr Y nach raibh aon fhadhbanna móra meicniúla acu sna chéad dá bhliain d'úinéireacht.

Glac leis go ndéanann carr anaithnid (branda X nó Y) níos mó ná 10l/100km agus nach raibh aon fhadhbanna móra meicniúla leis sna chéad dá bhliain d'úinéireacht. Glac leis freisin go n-úsáideann 65% de charranna branda X níos mó ná 10l/100km.

Dá bhféadfá ceann de na trí chineálacha seo a leanas de ghiotaí eolais a fháil amach, cé acu a chuideodh leat cinneadh an carr X nó carr Y é an carr anaithnid?

- Céatadán na gcarranna branda Y a bhaineann níos mó ná 10l/100km amach
- Céatadán na gcarranna branda X nach raibh aon fhadhbanna móra meicniúla acu sna chéad dá bhliain
- Céatadán na gcarranna branda Y nach raibh aon fhadhbanna móra meicniúla acu sna chéad dá bhliain

1.3.2 Leagan B

Smaoinigh ar chinneadh maidir le cibé an branda X nó branda Y é ríomhaire glúine anaithnid. Glac leis go bhféadfadh dhá chineál píosa eolais a bheith ar fáil faoi gach aon cheann-an céatadán de bhranda X agus de bhranda Y a bhfuil saolré cadhnra níos mó ná 10 n-uair an chloig acu agus an céatadán de bhranda X agus de bhranda Y nach raibh aon fhadhbanna móra teicniúla acu sna chéad dá bhliain d'úinéireacht.

Glac leis go raibh saolré cadhnra os cionn 10 n-uair an chloig ag ríomhaire glúine anaithnid (branda X nó Y) agus nach raibh aon fhadhbanna móra teicniúla leis sna chéad dá bhliain d'úinéireacht. Glac leis freisin go bhfuil saolré cadhnra os cionn 10 n-uair an chloig ag 65% de ríomhairí glúine branda X.

Dá bhféadfá ceann de na trí chineálacha seo a leanas de ghiotaí eolais a fháil amach, cé acu a chuideodh leat cinneadh an branda X nó carr Y é an ríomhaire glúine anaithnid?

- Céatadán branda Y a bhaineann saolré cadhnra os cionn 10 n-uair an chloig amach.
- Céatadán branda X nach raibh aon fhadhbanna móra teicniúla acu sna chéad dá bhliain
- Céatadán branda Y nach raibh aon fhadhbanna móra teicniúla acu sna chéad dá bhliain

1.4.1 Leagan A

D'fhreastail an tUasal Maxwell ar chóisir nár tugadh cuireadh ach d'ollaimh ollscoile agus d'fheidhmeannaigh gnó (d'fhéadfadh aíonna a bheith ina n-ollamh nó ina bhfeidhmeannach, ach ní fheadfaidís an dá phost a bheith acu). Is é an t-aon rud atá ar eolas agat faoin Uasal Maxwell ná gur ball é don *Bear's Club*. Iarrtar ort an dóchúlacht gur ollamh ollscoile é an tUasal Maxwell a mheas trí cheist 1, 2, 3, nó 4 a thugtar thíos a chur. Iarrtar ort a n-ábharthacht a sheiceáil. Is ionann ceist ábhartha agus ceist a gcuideoidh a freagra leat le do bhreithiúnas. Breithnigh gach ceist ar leithligh le do thoil agus léirigh cibé an bhfuil sé ábhartha nó neamhábhartha maidir leis an tasc atá romhat.

- Cén chéatadán de na daoine ag an gcóisir ar ollaimh ollscoile iad?
- Cén chéatadán de bhaill an *Bear's Club* atá ag an gcóisir?
- Cén chéatadán d'ollaimh ollscoile ag an gcóisir ar baill den *Bear's Club* iad?
- Cén chéatadán de na feidhmeannaigh gnó ag an gcóisir ar baill den *Bear's Club* iad?

1.4.2 Leagan B

D'fhreastail Bean Mhic Chonchrú ar thuras treoraithe nach mbíonn ach turasóirí Meiriceánacha agus turasóirí ó Ceanada ag freastal air (bheadh aíonna ina Meiriceánaigh nó ina gCeanadaigh, ní bheadh an dá náisiúnacht acu). Is é an t-aon rud atá ar eolas agat faoi Bhean Mhic Chonchrú ná gur aoí in óstán ABC. Iarrtar ort an dóchúlacht gur turasóir Meiriceánach í Bean Mhic Chonchrú a mheas trí cheist 1, 2, 3, nó 4 a thugtar thíos a chur. Iarrtar ort a n-ábharthacht a sheiceáil. Is ionann ceist ábhartha agus ceist a gcuideoidh a freagra leat le do bhreithiúnas. Breithnigh gach ceist ar leithligh le do thoil agus léirigh cibé an bhfuil sé ábhartha nó neamhábhartha maidir leis an tasc atá romhat.

- Cén chéatadán de na daoine ar an turas ar turasóirí Meiriceánacha iad?
- Cén chéatadán d'aíonna óstán ABC atá ag freastal ar an turas?
- Cén chéatadán de na turasóirí ó Ceanada ar an turas ar aíonna iad d'óstán ABC?
- Cén chéatadán de na turasóirí Meiriceánacha ar an turas ar aíonna iad d'óstán ABC?

1.5.1 Leagan A

Tá Linda 31 bliain d'aois, tá sí singil, neamhbhalbh agus an-chliste. Bhain sí céim amach i bhfealsúnacht. Agus i ina mac léinn, bhí an-suim aici i gcúrsaí idirdhealaithe agus ceartais shóisialta, agus ghlac sí páirt i léirsithe frithnúicléacha.

Cé acu díobh seo a leanas is dóchúla:

- Is áiritheoir bainc í Linda
- Is áiritheoir bainc í Linda agus tá sí gníomhach i ngluaiseacht na bhfeimineach

1.5.2 Leagan B

Tá Jack 45, tá sé pósta agus ceathrar páistí aige. Tríd is tríd, tá sé cúramach agus uailmhianach. Ní léiríonn sé aon suim i gcúrsaí polaitíochta ná sóisialta agus caitheann sé an chuid is mó dá am saor ar an iliomad caitheamh aimsire atá aige, lena n-áirítear siúinéireacht bhaile, seoltóireacht agus puzail mhatamaitice.

Cé acu díobh seo a leanas is dóchúla:

- Is innealtóir é Jack
- Is innealtóir é Jack agus tá sé gníomhach sa chlub áitiúil le haghaidh spóirt uisce

1.6.1 Leagan A

Is mian leat carr nua a cheannach. Caithfidh tú piocadh idir dhá rogha inniu: carr a cheannach ó chuideachta A nó B. Ní úsáideann tú ach critéar amháin don rogha sin, ionchas saoil an chairr. Tá eolas agat ó Thuairiscí Tomhaltóirí i sampla de roinnt céadta carranna go bhfuil taifead níos fearr ag carr A. Dúirt comharsa leat inné gur bhris a charr nua ar Branda A é síos. Cén carr a cheannaíonn tú?

- Carr A
- Carr B

1.6.2 Leagan B

Is mian leat ríomhaire nua a cheannach. Caithfidh tú piocadh idir dhá rogha inniu: ríomhaire a cheannach ó chuideachta A nó B. Ní úsáideann tú ach critéar amháin don rogha sin, ionchas saoil an ríomhaire. Tá eolas agat ó Thuairiscí Tomhaltóirí i sampla de roinnt céadta ríomhairí go bhfuil taifead níos fearr ag ríomhaire A. Dúirt comharsa leat inné gur bhris a ríomhaire nua ar Branda A é síos. Cén ríomhaire a cheannaíonn tú?

- Ríomhaire A
- Ríomhaire B

Verbal Reasoning

Analáí

A1

_____ : capall :: dul ar bord : traein :

- stábla
- crú capaill
- marcaíocht
- téigh in airde

B1

filíocht : rann :: fealsúnacht : _____

- íomhánna
- ceol
- fodhlí
- teoiric

A2

deisigh : fúáil :: cuir in eagar : _____

- dearnáil

- deisigh
- lámhscríbhinn
- aithris

B2

aibítreach : _____ :: seicheamhach : comhaid

- sórtáil
- scar
- liostaigh
- cuir in ord

A3

luch : _____ :: flais : ceamara

- francach
- ríomhaire
- corda
- milseog

B3

cúisín : tolg :: seilf : _____

- sliastán
- cás leabhar
- stóráil
- fráma

A4

tais: _____ :: fuaraigh : reoigh

- uisce
- fliuch
- oigheann
- fás

B4

ribín : _____ :: reoán : cáca

- bronntanas
- gearr
- cuachóg
- clóscríobhán

A5

piscín : _____ :: saighdiúir : arm

- cat
- ál
- coileáinín
- meamhlach

B5

eolai : turgnamh : _____ : dráma

- eascra
- cleachtadh
- aisteoir
- saotharlann

A6

foclóir : sainmhíniú :: _____ : mapa

- treo
- ó dheas
- atlas
- domhanfhad

B6

tacóid ordóg : _____ :: crúca : cóta

- tairne
- póstaer
- balla
- casúr

A7

neamhghnách : úrnuacht :: _____ : caighdeánach

- aisteach
- úrnua
- aithnidiúil
- dán

B7

_____ : coláiste :: meicneoir : garáiste

- leabhar
- foghlaim
- ollamh
- inneall

A8

druma : ionstraim:: druil : _____

- casúr
- oigheann
- uirlis
- crescendo

B8

binn : sliabh :: _____ : teach

- uasmhéadaigh
- díon
- póirse
- bungaló

A9

_____ : báisteach :: comhdhlúthú : bogthaise

- creimeadh
- scamall
- talamh
- réamhaisnéis

B9

_____ : fráma :: mósáic : tíl

- scannán
- greanadh
- seomra folctha
- maisiúchán

Judging Likelihood and Probability

3.1.1 Leagan A

Is féidir cluiche scuaise a imirt go dtí 9 nó 15 pointe. Má tá imreoir A ina imreoir níos fearr ná imreoir B, cén scéim scórála a thabharfadh seans níos fearr d'imreoir A an cluiche a bhuanachan?

- cluiche 9 bpointe
- cluiche 15 pointe

3.1.2 Leagan B

Is féidir leadóg boird a bhuanachan leis an scór is fearr as 5 chluiche nó an scór is fearr as 9 gcluiche. Má tá imreoir A ina imreoir níos fearr ná imreoir B, cén scéim scórála a thabharfadh seans níos fearr d'imreoir A buachan?

- An scór is fearr as 5 chluiche
- An scór is fearr as 9 gcluiche

3.2.1 Leagan A

Tosaíonn nuachtáin na 10 meánscór slachta is fearr a chur i gcló tar éis an chéad dá sheachtain sa mhórséasúr daorchluiche. De ghnáth, tar éis 2 sheachtain, bíonn meánscór de thuairim is .450 go minic ag an slacaí a bhíonn chun tosaigh. Mar sin féin, ní raibh meánscór de .450 ag aon slacaí riamh is stair an mhórshéasúir ag deireadh an tséasúir. Cén chúis atá leis sin i do thuairim? Cuir tic le ceann díobh seo a leanas.

- Nuair a tá a fhios go bhfuil slacaí ag bualadh ar mhaithe le meánscór ard a bhaint amach, leagann caiteoirí anuas níos mó agus iad ag caitheamh ina threo.
- Éiríonn caiteoirí níos fearr i rith an tséasúir, de réir mar a fhaigheann siad níos aclaí. De réir mar a fheabhsaíonn siad, bíonn sé níos dealraithe go gcuirfidh siad slacaithe amach agus dá bhrí sin téann a meánscóranna síos.
- D'fhéadfadh díreach gur ádh a bhaineann le hardscór imreoirí ag tús an tséasúir. Is tástáil níos réalaíche é an séasúr níos faide ar scil an tslacaí.
- Bíonn slacaí a n-éiríonn chomh maith leis ag tús an tséasúir faoi go leor brú a thairfead feidhmíochta a chothabháil. Bíonn drochthionchar ag strus mar sin ar a chuid imeartha.
- Nuair a bhíonn fios go bhfuil slacaí ag bualadh le haghaidh ardscór, ní chaitear liathróidí maithe chuige le bualadh. Imríonn na caiteoirí "coirnéil" an phláta mar nach miste leo é a chur ag siúl.

3.2.2 Leagan B

Déanann múinteoir scoile iarracht na 10 mac léinn is fearr don bhliain a dhéanamh amach tar éis an chéad dá sheachtain den téarma acadúil. De ghnáth, tar éis 2 sheachtain, bíonn 92% ar an meán go minic ag an mac léinn a bhíonn chun tosaigh. Mar sin féin, ní raibh 92% ar an meán riamh ag mac léinn i stair na scoile ag deireadh an téarma. Cén chúis atá leis sin i do thuairim? Cuir ciorcal ar cheann díobh seo a leanas:

- Nuair a bhíonn a fhios go bhfuil mac léinn ag díriú ar mheán ard a bhaint amach, méadaíonn múinteoirí a n-ionchais agus iad ag marcáil páipéir an mhic léinn sin.
- Éiríonn páipéir níos casta i rith a téarma, de réir mar a mhéadaíonn an líon ábhair. De réir mar a éiríonn páipéir níos casta, ní bhíonn sé chomh dealraitheach go mbainfidh mic léinn marcanna arda amach agus dá bhrí sin téann meánscóranna na mac léinn síos.
- D'fhéadfadh díreach gur ádh a bhaineann le hardscór mac léinn ag tús an téarma. Is tástáil níos réalaíche í an bhliain acadúil níos faide ar eolas agus cumas mhic léinn.
- Bíonn mac léinn a n-éiríonn chomh maith leis ag tús an téarma faoi go leor brú a mheánscór a chothabháil. Bíonn drochthionchar ag strus mar sin ar a chuid torthaí.
- Nuair a bhíonn fios go bhfuil mac léinn ag díriú ar mheánscór ard, ní chuireann múinteoirí ceisteanna éasca air. Cuireann múinteoirí "le deacracht" na gceisteanna don mhac léinn sin mar go mbíonn siad ag iarraidh cumas an mhic léinn a thástáil.

3.3.1 Leagan A

Agus iad ag imirt ar mheaisíní sliotáin, éiríonn le daoine rud éigin a bhuachan tuairim is 1 as gach 10 n-uaire. Tá Julie áfach, tar éis buachan sna trí huair tosaigh ar imir sí. Cad é an seans go mbuafaidh sí an chéad uair eile a imreoidh sí?

_____ as 10

3.3.2 Leagan B

Agus iad rannpháirteach i gcrannchur áitiúil, bíonn seans ag daoine duais a bhuachan le 1 as gach 20 ticéad crannchuir. Tá Jack áfach, tar éis buachan lena chéad trí thicéad. Cad é an seans go mbuafaidh sé an chéad uair eile a imreoidh sé?

_____ as 20

3.4.1 Leagan A

Samhlaigh go bhfuilimid ag caitheamh bonn cothrom (bonn a bhfuil seans 50/50 aige go dtiocfaidh sé anuas ceann nó cruit) in airde agus tagann sé anuas ar an taobh céanna, ceann, 5 uair i ndiaidh a chéile. Don 6ú huair an gceapann tú:

- Gur dóichí gur cruit a thiocfaidh aníos seachas ceann.
- Gur dóichí gur ceann a thiocfaidh aníos seachas cruit.
- Go bhfuil sé chomh dóichí céanna go dtiocfaidh ceann agus cruit aníos ar an séú babhta.

3.4.2 Leagan B

Samhlaigh go bhfuilimid ag catheamh díse cothrom (díse a bhfuil seans 1 as 6 ann go dtiocfaidh gach uimhir suas) agus tá uimhir 6 tar éis teacht suas 5 uair i ndiaidh a chéile. Don 6ú huair an gceapann tú:

- Gur dóichí ná na chéad cúig huair go dtiocfaidh uimhir eile seachas 6 suas.
- Gur dóichí atá sé ná na chéad cúig huair go dtiocfaidh uimhir 6 seachas uimhir eile suas.
- Go bhfuil sé chomh dóichí céanna go dtiocfaidh 6 agus aon uimhir eile suas ar an séú babhta.

3.5.1 Leagan A

Glac leis go dtugtar dhá thráidire de mhirlíní dubha agus bána duit: tráidire mór le 100 mirlín agus tráidire beag le 10 mirlín. Caithfidh tú mirlín amháin a thógáil as tráidire amháin díobh gan breathnú. Má thógann tú mirlín dubh, buann tú \$2. Smaoinigh ar chás ina bhfuil 1 mirlín dhubh agus 9 mirlín bhána ar an tráidire beag, agus go bhfuil 8 mirlín dhubha agus 92 mirlín bhána ar an tráidire mór. Cén tráidire a mb'fhearr leat mirlín a thógáil as i bhfíorchás?

- Tráidire mór
- Tráidire beag

3.5.2 Leagan B

Glac leis go dtugtar dhá mhála de nótaí \$10 agus \$20 duit: mála mór ina bhfuil 100 nóta éagsúil agus mála beag ina bhfuil 10 nóta éagsúil. Tá na nótaí measctha go randamach sna málaí. Caithfidh tú nóta amháin a thógáil as mála díobh gan breathnú agus féadfaidh tú an nóta a choinneáil. Smaoinigh ar chás ina bhfuil nóta amháin \$20 agus naoi nóta \$10 sa mhála beag, agus go bhfuil ocht nóta \$20 agus nócha a dó \$10 sa mhála mór. Cén mála a mb'fhearr leat nóta a thógáil as i bhfíorchás?

- Mála mór
- Mála beag

Argumentation Analysis

Cuir tic le do thoil: bailí - neamhbhailí

A1

- Níl aon toitíní neamhchostasach

- Tá roinnt rudaí andúile neamhchostasach.
- Dá bhrí sin, ní toitíní iad roinnt rudaí andúile.

B1

- Is foirgnimh adhmaid roinnt foirgneamh
- Ní foirgnimh adhmaid iad aon ilstóraigh spéire
- Dá bhrí sin ní ilstóraigh spéire iad roinnt foirgneamh.

A2

- Ní ríomhairí iad roinnt píosaí de threalamh oifige
- Is píosaí de threalamh oifige iad roinnt feistí leictreacha
- Dá bhrí sin ní ríomhairí iad roinnt feistí leictreacha.

B2

- Tá roinnt deochanna neamh-mheisciúil
- Ní licéir iad deochanna neamh-mheisciúla
- Dá bhrí sin ní licéir iad roinnt deochanna.

A3

- Ní boird iad cathaoireacha
- Is cathaoireacha iad roinnt píosaí troscáin
- Dá bhrí sin ní boird iad roinnt píosaí troscáin.

B3

- Tá roinnt iasc do-ite
- Tá breac inite
- Dá bhrí sin ní bric iad roinnt iasc.

B4

- Níl aon rudaí andúile neamhchostasach.
- Tá roinnt toitíní neamhchostasach.
- Dá bhrí sin, ní toitíní iad roinnt rudaí andúile.

A4

- Is foirgnimh adhmaid iad roinnt ilstórach spéire
- Ní foirgnimh adhmaid aon fhoirgnimh
- Dá bhrí sin ní ilstóraigh spéire iad roinnt foirgneamh.

B5

- Ní píosaí de threalamh oifige iad roinnt feistí leictreacha
- Is roinnt píosaí de threalamh oifige iad ríomhairí
- Dá bhrí sin ní ríomhairí iad roinnt feistí leictreacha.

A5

- Is licéir iad deochanna neamh-mheisciúla
- Níl aon deochanna neamh-mheisciúil
- Dá bhrí sin ní licéir iad roinnt deochanna.

B6

- Ní cathaoireacha iad aon pháosaí troscáin
- Is boird iad cathaoireacha.
- Dá bhrí sin ní boird iad roinnt píosaí troscáin.

A6

- Níl aon bhreac inite
- Tá iasc inite
- Dá bhrí sin ní bric iad roinnt iasc.

A7

- Ní uirlisí láimhe iad roinnt uirlisí
- Is uirlisí láimhe iad scriúirí
- Dá bhrí sin ní uirlisí iad roinnt scriúirí.

B7

- Is gaothuirlisí iad roinnt uirlisí ceoil
- Ní gaothuirlisí iad veidhlíní
- Dá bhrí sin ní uirlisí ceoil iad roinnt veidhlíní.

A8 // B8

- Ní féara iad roinnt rudaí glasa.
- Tá roinnt plandaí glas.
- Dá bhrí sin ní plandaí iad roinnt féar.

B9

- Ní uirlisí láimhe iad scriúirí
- Ní uirlisí láimhe iad aon uirlisí
- Dá bhrí sin ní uirlisí iad roinnt scriúirí.

A9

- Is gaothuirlisí iad veidhlíní
- Ní gaothuirlisí iad aon uirlisí ceoil
- Dá bhrí sin ní uirlisí ceoil iad roinnt veidhlíní.

Problem Solving

5.1.1 Leagan A

Cosnaíonn slacán agus liathróid \$1.10 ina iomlán. Cosnaíonn an slacán \$1.00 níos mó ná an liathróid. Cá mhéad a chosnaíonn an liathróid?

5 cent – 10 cent – 55 cent – 100 cent

5.1.2 Leagan B

Cosnaíonn builín aráin agus úll \$2.20 ina iomlán. Cosnaíonn an t-arán \$2.00 níos mó ná an t-úll. Cá mhéad a chosnaíonn an t-úll?

10 cent – 20 cent – 110 cent – 200 cent

5.2.1 Leagan A

Má thógann sé 5 nóiméad ar 5 mheaisín 5 chomhpháirt phlaisteach a dhéanamh, cá fhad a thógfadh sé ar 100 meaisín 100 comhpháirt phlaisteach a dhéanamh?

1 – 5 – 25 – 100 nóiméad

5.2.2 Leagan B

Má thógann sé 2 nóiméad ar 2 bháicéir 2 bhuilín aráin a dhéanamh, cá fhad a thógfadh sé ar 50 báicéir 50 builín aráin a dhéanamh?

1 – 2 – 25 – 50 Nóiméad

5.3.1 Leagan A

Tá paiste duillí na duilleoige báite ar loch. Dúbláíonn an paiste i méid gach lá. Má thógann sé 24 lá ar an bpaiste an loch iomlán a chlúdach, cá fhad a thógfadh sé ar an bpaiste leath don loch a chlúdach?

2 – 4 – 12 – 23 lá

5.3.2 Leagan B

Tá coimeádán á líonadh le clocha. Dúbláíonn líon na gcloch gach lá. Beidh an coimeádán lá i 10 lá. Cá mhéad lá a thógfaidh sé air a bheith leath-lán?

2 – 4 – 5 – 9 lá

5.4.1 Leagan A

Tá stocáí dubha agus donna scaoilte i dtarraiceán agus iad measctha i gcóimheas 4 le 5. Cá mhéad stoca ar leithligh a chaithfeá a thógáil amach gan breathnú le bheith cinnte go bhfaighfeá péire ar an dath céanna?

1 - 2 - 3 - 4 - 5 stoca

5.4.2 Leagan B

Tá mirlíní glasa agus gorma i mála agus iad measctha i gcóimheas 6 le 8. Cá mhéad mirlín a chaithfeá a thógáil amach gan breathnú le bheith cinnte go bhfaighfeá péire ar an dath céanna?

2 - 3 - 4 - 5 - 6 mirlín

5.5.1 Leagan A

Tosaíonn dhá thraein, 50km ó chéile ag dul i dtreo a chéile ag 25 km/u (km san uair) an ceann. De réir mar a thosaíonn na traenacha, eitlíonn éan ó cheann traein amháin i dtreo an dara ceann. Nuair a shroicheann sé an dara traein, iompaíonn an t-éan thart agus éitlíonn sé ar ais chuig an chéad traein agus mar sin ar aghaidh nó go gcasann na traenacha ar a chéile. Má eitlíonn an t-éan ag 60 km/u, cá mhéad km a eitleoidh an t-éan sula gcasfaidh na traenacha ar a chéile?

15 km - 30 km - 45 km - 60 km

5.5.2 Leagan B

Tosaíonn beirt choisithe atá 10 km ó chéile ag siúl i dtreo a chéile ag 5 km/u. Agus iad ag tosú ag siúl ritheann madra ó dhuine amháin i dtreo an duine eile. Nuair a shroicheann an madra an dara duine, iompaíonn sé agus ritheann sé go dtí an chéad duine agus mar sin ar aghaidh nó go gcasann na choisithe ar a chéile. Má ritheann an madra ag 20 km/u, cá mhéad ciliméadar a rithfidh an madra sula gcasfaidh na choisithe ar a chéile?

2 - 5 - 10 - 20 km

5.6.1 Leagan A

Ceannaíonn fear capall ar £60, díolann sé é ar £70, ceannaíonn sé arís é ar £80 agus díolann sé ar deireadh é ar £90. Cá mhéad a dhéanann sé?

£10 - £20 - £30 - £40

5.6.2 Leagan B

Ceannaíonn bean scaireanna cuideachta sa stocmhargadh ar \$600, díolann sí iad ar \$700, ceannaíonn sí arís iad ar \$800 agus díolann sí ar deireadh iad ar \$900. Cá mhéad a dhéanann sí?

\$100 - \$200 - \$300 - \$400

5.7.1 Leagan A

Bhí na póilíní cinnte go ndearna A, B, C, nó D coir. Rinne gach duine de na daoine faoi amhras ráiteas, ar a sheal, ach ní raibh ach ceann de na ceithre ráiteas fíor.

Dúirt A, "Ní dhearna mise é."

Dúirt B, "Tá A ag insint bhréaga."

Dúirt C, "Tá B ag insint bhréaga."

Dúirt D, "Rinne B é."

Cé atá ag insint na fírinne? Cé a rinne an choir?

Fírinne: A - B - C - D

Coir: A - B - C - D

5.7.2 Leagan B

Tá a fhios ag an mbreitheamh gur tháinig duine A, B, C, nó D suas le haireagán. Rinne gach duine díobh éileamh, ar a sheal, ach ní raibh ach ceann de na ceithre ráiteas fíor.

Dúirt A, "Is mise a cheap é."

Dúirt B, "Tá A ag insint bhréaga."

Dúirt C, "Tá B ag insint bhréaga."

Dúirt D, "Níor cheap B é."

Cé atá ag insint na fírinne? Cé a tháinig suas leis an aireagán?

Fírinne: A - B - C - D

Aireagán: A - B - C - D

5.8.1 Leagan A

Taispeántar ceithre cárta duit ar bhord, tá uimhir ar thaobh amháin de gach ceann agus dath ar an taobh eile. Ar na haghaidheanna atá le feiceáil de na cartaí, taispeántar 3, 8, dearg agus donn. Cén cárta(i) a chaithfidh tú a iompú chun fírinne na ceiste a thástáil má thaispeántar ré-uimhir ar aghaidh amháin ar chárta gur dath dearg a bheidh ar aghaidh eile an chárta sin? Déan iarracht oibriú ar an tslí is éifeachtúla trí a laghad cártaí is féidir a iompú.

Marcáil an méid is infheidhme: 3 – 8 – dearg – donn

5.8.2 Leagan B

Taispeántar ceithre cárta duit ar bhord, tá uimhir ar thaobh amháin de gach ceann agus dath ar an taobh eile. Ar na haghaidheanna atá le feiceáil de na cartaí, taispeántar 5, 6, glas agus buí. Cén cárta(i) a chaithfidh tú a iompú chun fírinne na ceiste a thástáil má thaispeántar corruimhir ar aghaidh amháin ar chárta gur dath glas a bheidh ar aghaidh eile an chárta sin? Déan iarracht oibriú ar an tslí is éifeachtúla trí a laghad cártaí is féidir a iompú.

Marcáil an méid is infheidhme: 5 - 6 - glas - buí

German

Hypothesis Testing

1.1.1 Version A

Ein Schüler muss sich zwischen zwei Universitäten A und B entscheiden. Der Schüler hat in beiden Einrichtungen einige Freunde, die Werte mit ihm teilen und ähnliche Fähigkeiten besitzen. Alle seine Freunde in A sind zufrieden, sowohl auf Bildungsebene, als auch bezüglich sozialer Aspekte. Alle Freunde in B haben Zweifel diesbezüglich. Der Schüler sieht sich beide Einrichtungen für einen Tag an und sein Eindruck ist geteilt zu den Berichten.

Welche Universität sollte der Schüler wählen?

- Universität A
- Universität B

1.1.2 Version B

Eine Mutter und ein Vater müssen sich zwischen zwei Kindergärten A und B für ihr erstes Kind entscheiden. Sie haben einige Freunde, die ihnen bezüglich Werten und Einstellungen ähnlich sind, mit Kindern in beiden Kindergärten. Alle ihre Freunde mit einem Kind in Kindergarten A sind sowohl mit den Bildungs- als auch mit den sozialen Aspekten zufrieden. Alle Freunde mit einem Kind in Kindergarten B haben Zweifel diesbezüglich. Das Paar besichtigt beide Kindergärten für ein paar Stunden und haben einen gegenteiligen Eindruck zu den Berichten. Welchen Kindergarten sollten die Mutter und der Vater wählen?

- Kindergarten A
- Kindergarten B

1.2.1 Version A

Die Stadt Kingston hat seit eineinhalb Jahren einen unbeliebten Polizeipräsidenten. Er ist politisch aktiv und ein Kollege des Bürgermeisters und er hatte wenig vorgehende Erfahrungen in der Polizeiverwaltung als er ernannt wurde. Der Bürgermeister verteidigte kürzlich den Polizeipräsidenten öffentlich und gab bekannt, dass, seit dieser sein Amt angetreten hat, die Kriminalitätsraten um 12% gesunken sind. Welche der folgenden Aussagen würde die Behauptung, dass der Polizeipräsident kompetent ist, am stärksten schwächen?

- Die Kriminalitätsraten der umliegenden Städte, die Kingston bezüglich Lage und Größe am meisten ähneln, haben in derselben Zeit ihre Kriminalitätsraten um 18% gesenkt.
- Eine unabhängige Umfrage der Bewohner von Kingston zeigt, dass 40% mehr Verbrechen gemeldet wurden, als in den Umfragen der Polizei berichtet wurde.
- Gesunder Menschenverstand zeigt, dass ein Polizeipräsident wenig tun kann, um die Kriminalitätsraten zu verringern. Diese sind hauptsächlich das Resultat sozialer und ökonomischer Bedingungen außerhalb der Kontrolle von Beamten.
- Es wurde festgestellt, dass der Polizeipräsident Geschäftskontakte zu Personen pflegt, die im organisierten Verbrechen involviert sind.

1.2.2 Version B

Eine Schule hat seit eineinhalb Jahren einen unbeliebten Lehrer. Er ist ein Freund des Schulleiters und hatte wenig Erfahrung als Lehrer als er den Job antrat. Der Schulleiter verteidigte den Lehrer vor der Schule und verkündete, dass seit er den Job als Lehrer antrat, die Rate auffälligen Verhaltens seiner Schüler um 12% gesunken sei. Welche der folgenden Aussagen entkräftet die Behauptung des Schulleiters, der Lehrer sei kompetent, am stärksten?

- Das Fehlverhalten der zwei Klassen, die nicht von dem Lehrer unterrichtet werden, sind im selben Zeitraum um 18% gesunken.
- Eine Umfrage unter den Schülern zeigt, dass 40% mehr Fehlverhalten berichtet wird, als der Umfrage des Schulleiters angegeben wird.
- Gesunder Menschenverstand zeigt, dass sein Lehrer wenig tun kann, um Fehlverhalten zu reduzieren. Diese hängen hauptsächlich von der Laune und dem Verhalten der Schüler ab.
- Es wurde festgestellt, dass Schüler, die sich schlecht verhielten, von dem Lehrer bestraft wurden.

1.3.1 Version A

Prüfe die Entscheidung, ob ein unbekanntes Auto ein Auto der Marke X oder der Marke Y ist. Nehme an, dass folgende Informationen potentiell zu beiden Autos verfügbar sind – die

Prozentangaben von Autos der Marke X und Autos der Marke Y welche mehr als 10 Liter pro 100 km verbrauchen, und die Prozentangaben von Autos der Marke X und Autos der Marke Y welche keine schwerwiegenden mechanischen Probleme in den ersten zwei Jahren der Benutzung hatten.

Gehe davon aus, dass ein unbekanntes Auto (entweder Marke X oder Y) über 10 L/100 km benötigt und dass es keine großen mechanischen Probleme in den ersten zwei Jahren der Benutzung hat. Gehe ebenfalls davon aus, dass 65% der Autos von Marke X über 10 L/100 km verbrauchen.

Wenn du eine der drei folgenden Informationen erfahren könntest, welche würde dir helfen zu entscheiden, welches das unbekanntes Auto ist, Auto X oder Auto Y?

- Prozentsatz der Autos von Marke Y die über 10 L/100 km verbrauchen.
- Prozentsatz der Autos der Marke X, die keine schwerwiegenden mechanischen Probleme in den ersten zwei Jahren hatten
- Prozentsatz der Autos der Marke Y, die keine schwerwiegenden mechanischen Probleme in den ersten zwei Jahren hatten

1.3.2 Version B

Prüfe die Entscheidung, ob ein unbekannter Laptop der Marke X oder Marke Y zugehörig ist. Nehme an, dass potentiell zwei Informationen zu jeder Alternative zur Verfügung stehen – der Prozentsatz von Laptops der Marke X und Marke Y, die eine Akkulaufzeit von mehr als 10 Stunden hat und der Prozentsatz von Laptops der Marke X und Marke Y, die keine gravierenden technischen Probleme in den ersten zwei Jahren der Benutzung hatten.

Gehe davon aus, dass ein unbekannter Laptop (entweder der Marke X oder Y) eine Akkulaufzeit von über 10 Stunden hat und dass er keine gravierenden technischen Probleme in den ersten zwei Jahren der Benutzung aufwies. Nehme zudem an, dass 65% der Marke X Laptops eine Akkulaufzeit von über 10 Stunden haben.

Wenn du eine der folgenden drei Informationen herausfinden könntest, welche würde dir helfen herauszufinden ob der Laptop der Marke X oder Marke Y zugehörig ist?

- Prozent der Marke Y, die eine Akkulaufzeit von über 10 Stunden haben.
- Prozent der Marke X, die keine gravierenden technischen Probleme in den ersten zwei Jahren aufweisen.
- Prozent der Marke Y, die keine gravierenden technischen Probleme in den ersten zwei Jahren aufweisen.

1.4.1 Version A

Herr Maxwell nimmt an einer Feier teil, zu der nur Universitätsprofessoren Unternehmensleiter eingeladen sind (wobei Gäste eins von beiden sein können, nicht aber beide Positionen kombiniert). Das Einzige, was du über Herrn Maxwell weißt, ist, dass er ein Mitglied des Bear's Club ist. Deine Aufgabe ist die Wahrscheinlichkeit abzuschätzen, ob Herr Maxwell ein Universitätsprofessor ist, indem du 1, 2, 3 oder 4 der unten angegebenen Fragen stellst. Prüfe ihre Relevanz. Eine relevante Frage kennzeichnet, dass sie dir dabei helfen wird deine Entscheidung zu treffen. Bitte bewerte jede Frage individuell und gebe an, ob sie relevant oder irrelevant für die Aufgabe ist.

- Wie viel Prozent der Personen auf der Feier sind Universitätsprofessoren?
- Wie viel Prozent der Mitglieder des Bear's Clubs sind auf der Feier?
- Wie viel Prozent an Universitätsprofessoren auf der Feier sind Mitglieder des Bear's Clubs?
- Wie viel Prozent an Unternehmensleiter auf der Feier sind Mitglieder des Bear's Club?

1.4.2 Version B

Frau Crowe nimmt an einer Tour teil, bei der nur amerikanische und kanadische Touristen anwesend sind (wobei keiner der Teilnehmer eine doppelte Staatsbürgerschaft besitzt). Das Einzige, was du über Frau Crowe weißt, ist, dass sie ein Gast im Hotel ABC ist. Du wirst gebeten die Wahrscheinlichkeit, dass Frau Crowe eine amerikanische Touristin ist abzuschätzen, indem du 1, 2, 3 oder 4 der unten angegebenen Fragen stellst. Eine relevante Frage kennzeichnet, dass sie dir dabei helfen wird deine Entscheidung zu treffen. Bitte bewerte jede Frage individuell und gebe an, ob sie relevant oder irrelevant für die Aufgabe ist.

- Wie viel Prozent der Personen in der Tour sind amerikanische Touristen?
- Wie viel Prozent der Gäste im Hotel ABC nehmen an der Tour teil?

- Wie viel Prozent an kanadischen Touristen bei der Tour sind Gäste des Hotels ABC?
- Wie viel Prozent an amerikanischen Touristen bei der Tour sind Gäste des Hotels ABC?

1.5.1 Version A

Linda ist 31 Jahre alt, alleinstehend, offen und sehr klug. Sie studierte Philosophie. Als Studentin beschäftigte sie sich intensiv mit den Themen Diskriminierung und soziale Gerechtigkeit und nahm an Anti-Atomkraft-Demonstrationen teil.

Welche der beiden Alternativen ist wahrscheinlicher:

- Linda ist Kassiererin
- Linda ist Kassiererin und aktiv in der Frauenbewegung

1.5.2 Version B

Jack ist 45, verheiratet und hat vier Kinder. Er ist generell konservativ, vorsichtig und ambitioniert. Er zeigt kein Interesse an politischen und sozialen Probleme und verbringt den Großteil seiner Freizeit mit seinen vielen Hobbies, wie z.B. dem Heimwerken, Segeln und mathematischen Rätseln.

Welche der zwei Alternativen ist wahrscheinlicher:

- Jack ist ein Ingenieur.
- Jack ist ein Ingenieur und aktiv im lokalen Club für Wassersportarten.

1.6.1 Version A

Du möchtest ein neues Auto kaufen. Du musst dich zwischen zwei Alternativen entscheiden: Auto A oder Auto B. Du benutzt nur ein Kriterium, um dich zu entscheiden, und zwar die erwartete Lebenszeit des Autos. Du hast Informationen von Kundenberichten, dass in einer Stichprobe von hundert Autos das Auto A die beste Bewertung erhielt. Genau gestern erzählte dir dein Nachbar, dass sein neues Auto der Marke A eine Panne hatte. Welches Auto kaufst du?

- Auto A
- Auto B

1.6.2 Version B

Du möchtest einen neuen Computer kaufen. Heute musst du dich zwischen zwei Alternativen entscheiden: Einen Computer A oder einen Computer B zu erwerben. Du benutzt nur ein Kriterium für deine Entscheidung, die Lebenserwartung des Computers. Du hast Informationen von Kundenberichten, dass in einer Stichprobe von mehreren hundert Computern der Computer A die beste Bewertung erhielt. Gestern erzählte dir ein Nachbar, dass sein neuer Computer der Marke A kaputt ging. Welchen Computer kaufst du?

- Computer A
- Computer B

Verbal Reasoning

Analogies

A1

_____ : Pferd :: Einsteigen : Zug

- Stall
- Schuh
- reiten
- aufsitzen

B1

Dichtung : Reim :: Philosophie : _____

- Symbolik

- Musik
- Gesetz
- Theorie

A2

flicken : Näharbeit :: editieren : _____

- stopfen
- reparieren
- Manuskript
- Provisorium

B2

alphabetisch : _____ :: sequentiell : Akten

- Sorte
- Teil
- Liste
- Ordnung

A3

Maus : _____ :: Blitz : Kamera

- Ratte
- Computer
- Kabel
- Dessert

B3

Polsterung : Sofa :: Brett : _____

- Sims
- Bücherregal
- Lager
- Rahmen

A4

befeuchten : _____ :: kühlen : frieren

- Wasser
- einweichen
- Ofen
- wachsen

B4

Schleife : _____ :: Glasur : Kuchen

- Geschenk
- Schnitt
- Bogen
- Schreibmaschine

A5

Welpen : _____ :: Soldat : Armee

- Hund
- Wurf
- Kätzchen
- Wau

B5

Wissenschaftler : Experiment :: _____ : Stück

- Messbecher
- Vorsprechen
- Schauspieler
- Labor

A6

Lexikon : Definition :: _____ : Karte

- Richtung
- Süden
- Atlas
- Längengrad

B6

Reißzwecke : _____ :: Haken : Mantel

- Nagel
- Poster
- Wand
- Hammer

A7

ungewöhnlich : Neuheit :: _____ : Standard

- eigenartig
- neu
- familiär
- Gedicht

B7

_____ : Universität :: Mechaniker : Werkstatt

- Buch
- Lernen
- Professor
- Motor

A8

Trommel : Instrument :: Bohrer : _____

- Hammer
- Ofen
- Werkzeug
- Crescendo

B8

Spitze : Berg :: _____ : Haus

- Maximum
- Dach
- Veranda
- Bungalow

A9

_____ : Regen :: Kondensation : Feuchtigkeit

- Erosion
- Wolke
- Boden
- Prognose

B9

_____ : Bild :: Mosaik : Kachel

- Film
- Gravierung
- Badezimmer
- Dekoration

Judging Likelihood and Probability

3.1.1 Version A

Ein Squash-Spiel kann entweder mit 9 oder 15 Punkten gespielt werden. Wenn A ein besserer Spieler als B ist, welches Punktesystem gibt A eine größere Chance zu gewinnen?

- Das 9-Punkte-Spiel
- Das 15-Punkte-Spiel

3.1.2 Version B

Tischtennis kann entweder mit 5 oder mit 9 Sätzen gespielt werden. Wenn A ein besserer Spieler als B ist, welches Punktesystem gibt A eine bessere Chance zu gewinnen?

- 5 Sätze
- 9 Sätze

3.2.1 Version A

Nach den ersten 2 Wochen der Baseball Saison in der oberen Spielklasse beginnen die Zeitungen, die Top 10 Ranglisten der durchschnittlichen Leistungen der Spieler zu drucken. Üblicherweise hat der führende Schlagmann nach zwei Wochen einen Durchschnitt von ungefähr 0,450. Es hat jedoch noch nie ein Spieler der oberen Spielklasse am Ende der Saison einen Durchschnitt von 0,450 gehabt. Warum denkst du ist das so? Kreuze eine Antwort an:

- Wenn bekannt ist, dass der Schläger oft trifft, spielen Werfer ihm den Ball mit mehr Wucht zu
- Werfer verbessern sich über die Season hinweg, da sie fitter werden. Dadurch ist es wahrscheinlicher, dass sie die Schläger schlagen, sodass deren Durchschnitt sinkt.
- Der hohe Durchschnitt zu Beginn einer Season könnte bloß Glück sein. Die gesamte Season spiegelt die Fähigkeiten eines Schlägers realistischer wider.
- Ein Schläger, der eine Glückssträhne zu Beginn der Season hat, steht unter großem Druck, diese Leistung aufrecht zu halten. Dieser Druck beeinflusst seine Leistung negativ.
- Wenn ein Schläger dafür bekannt ist, einen guten Durchschnitt anzustreben, bekommt er keine einfachen Pässe mehr zugespielt. Statt dessen zielen Werfer in die Ecken um den Schläger zu fordern.

3.2.2 Version B

Nach den ersten zwei Wochen des Schuljahres, versucht ein Lehrer die 10 besten Schüler des Jahrgangs vorherzusagen. Normalerweise hat der beste Schüler nach 2 Wochen oft einen Durchschnitt von ungefähr 92%. Allerdings hat kein Schüler in der Geschichte der Schule je einen Durchschnitt von 92% am Ende eines Schuljahres gehabt. Warum denkst du ist das so? Kreuze eine Antwort an:

- Wenn ein Schüler einen hohen Durchschnitt anpeilt, steigern Lehrer ihre Erwartungen, wenn sie die Arbeiten des Schülers bewerten.
- Arbeiten werden im Laufe des Jahres schwieriger, da sich Inhalt anhäuft. Da die Arbeiten schwieriger werden, wird es für Schüler unwahrscheinlicher, gute Noten zu bekommen, deswegen sinkt die Notendurchschnitte.

- Der hohe Durchschnitt eines Schülers zu Beginn eines Jahres ist vielleicht nur Glück. Das längere Schuljahr bietet eine realistischere Prüfung des Wissens und der Fähigkeiten eines Schülers.
- Ein Schüler, der so hohe Durchschnittsnoten zu Beginn eines Semesters hat, steht unter viel Druck diesen Durchschnitt aufrecht zu erhalten. Dieser Druck beeinflusst seine Ergebnisse negativ.
- Wenn bekannt ist, dass ein Schüler einen hohen Durchschnitt anpeilt, dann bekommt er keine einfachen Fragen mehr von den Lehrern gestellt. Stattdessen erhöhen Lehrer die Schwierigkeit der Fragen für diesen Schüler, um dessen Wissen zu testen.

3.3.1 Version A

Bei Spielautomaten gewinnen Leute ungefähr in 1 von 10 Fällen. Julia dagegen hat gerade bei jedem ihrer ersten drei Spiele gewonnen. Was sind ihre Chancen beim nächsten Spiel zu gewinnen?

_____ von 10 Fällen

3.3.2 Version B

Bei einer lokalen Lotterie gewinnen etwa 1 aus 20 Lotterielosen einen Preis. Jakob dagegen hat gerade mit seinen ersten 3 Tickets gewonnen. Was sind seine Chancen bei seinem nächsten Spiel zu gewinnen?

_____ zu 20

3.4.1 Version A

Stelle dir vor, dass wir eine faire Münze werfen (eine Münze, bei der die Wahrscheinlichkeit Kopf oder Zahl zu bekommen bei 50/50 liegt) und die letzten fünf Mal hintereinander Kopf geworfen wurde. Was denkst du über den 6. Wurf:

- Es ist wahrscheinlicher, dass Zahl statt von Kopf geworfen wird.
- Es ist wahrscheinlicher, dass Kopf statt Zahl geworfen wird.
- Kopf und Zahl sind beim sechsten Wurf gleich wahrscheinlich.

3.4.2 Version B

Stelle dir vor, dass wir einen fairen Würfel werfen (jede Seite hat eine Chance von 1 zu 6 geworfen zu werden) und die Nummer 6 ist gerade fünfmal hintereinander gewürfelt worden. Was denkst du über den 6. Wurf:

- Es ist wahrscheinlicher, dass eine andere Nummer als 6 gewürfelt wird.
- Es ist wahrscheinlicher, dass eine 6 statt einer anderen Nummer gewürfelt wird.
- 6 und jede andere Nummer sind beim 6. Wurf gleich wahrscheinlich.

3.5.1 Version A

Stelle dir vor, dass dir zwei Tablett mit schwarzen und weißen Murmeln gezeigt werden: Ein großes Tablett mit 100 Murmeln und ein kleines Tablett mit 10 Murmeln. Die Murmeln sind auf beiden Tablett verteilt. Du musst eine Murmel von einem der Tablett ziehen, ohne hinzusehen. Wenn du eine schwarze Murmel ziehst, gewinnst du 2€. Auf dem kleinen Tablett ist eine schwarze Murmel und 9 weiße Murmeln und auf dem großen Tablett sind 8 schwarze und 92 weiße Murmeln. Von welchem Tablett würdest du in eine Murmel ziehen?

- Vom großen Tablett
- Vom kleinen Tablett

3.5.2 Version B

Stelle dir vor, dass dir zwei Tüten mit 10€ und 20€ Scheinen gezeigt werden: Eine große Tüte mit 100 verschiedenen Scheinen und eine Tüte, die 10 verschiedene Scheine beinhaltet. Du musst von einer der beiden Tüten einen Schein ohne hinzuschauen ziehen und darfst den Schein behalten. In der kleinen Tüte befindet sich ein 20€ Schein und neun 10€ Scheine und der großen Tüte befinden sich acht 20€ Scheine und zweiundneunzig 10€ Scheine. Von welcher Tüte würdest du lieber einen Schein ziehen?

- Von der großen Tüte

- Von der kleinen Tüte

Argumentation Analysis

Bitte kreuze an: richtig – falsch

A1

- Keine Zigarten sind billig.
- Manche suchterzeugenden Produkte sind billig.
- Demzufolge sind manche suchterzeugenden Produkte keine Zigaretten.

B1

- Manche Gebäude sind aus Holz.
- Keine hölzernen Gebäude sind Wolkenkratzer.
- Demzufolge sind manche Gebäude keine Wolkenkratzer.

A2

- Manche Bürogegenstände sind keine Computer.
- Manche elektronischen Geräte sind Bürogegenstände.
- Demzufolge sind manche elektronischen Geräte keine Computer.

B2

- Manche Getränke sind nicht alkoholisch.
- Nicht alkoholische Getränke sind nicht Spirituosen.
- Demzufolge sind manche Getränke keine Spirituosen.

A3

- Stühle sind keine Tische.
- Manche Möbel sind Stühle.
- Demzufolge sind manche Möbel keine Tische.

B3

- Manche Fische sind nicht essbar.
- Forelle ist essbar.
- Demzufolge sind manche Fische nicht Forelle.

B4

- Keine suchterzeugenden Produkte sind billig.
- Manche Zigarten sind billig.
- Demzufolge sind suchterzeugende Produkte keine Zigaretten.

A4

- Manche Gebäude aus Holz sind Wolkenkratzer.
- Keine Gebäude sind Gebäude aus Holz.
- Demzufolge sind manche Gebäude keine Wolkenkratzer.

B5

- Manche elektronischen Geräte sind keine Bürogegenstände.
- Manche Bürogegenstände sind Computer.
- Demzufolge sind manche elektronischen Geräte keine Computer.

A5

- Nicht alkoholische Getränke sind Spirituosen.
- Keine Getränke sind nicht alkoholisch.
- Demzufolge sind manche Getränke keine Spirituosen.

B6

- Keine Möbelstücke sind Stühle.

- Stühle sind Tische.
- Demzufolge sind manche Möbelstücke nicht Tische.

A6

- Keine Forellen sind essbar.
- Fisch ist essbar.
- Demzufolge sind manche Fische keine Forellen.

A7

- Manche Werkzeuge sind nicht manuell.
- Schraubenzieher sind manuelle Werkzeuge.
- Demzufolge sind manche Schraubenzieher keine Werkzeuge.

B7

- Manche Musikinstrumente sind Blasinstrumente.
- Blasinstrumente sind nicht Geigen.
- Demzufolge sind manche Geigen keine Musikinstrumente.

A8 // B8

- Manche grünen Dinge sind nicht Gräser.
- Manche Pflanzen sind grün.
- Demzufolge sind manche Gräser keine Pflanzen.

B9

- Schraubenzieher sind keine manuellen Werkzeuge.
- Keine Werkzeuge sind manuell.
- Demzufolge sind manche Schraubenzieher keine Werkzeuge.

A9

- Blasinstrumente sind Geigen.
- Keine Musikinstrumente sind Blasinstrumente.
- Demzufolge sind manche Geigen keine Musikinstrumente.

Problem Solving

5.1.1 Version A

Ein Schläger und ein Ball kosten zusammen 1,10\$. Der Schläger kostet 1,00\$ mehr als der Ball. Wie viel kostet der Ball?

5 Cent – 10 Cent – 55 Cent – 100 Cent

5.1.2 Version B

Ein Laib Brot und ein Apfel kosten zusammen 2,20\$. Das Brot kostet 2 \$ mehr als der Apfel. Wie viel kostet der Apfel?

10 Cent – 20 Cent – 110 Cent – 200 Cent

5.2.1 Version A

Wenn 5 Maschinen in 5 Minuten 5 Plastikstücke herstellen, wie lange würden 100 Maschinen brauchen 100 Plastikstücke herzustellen?

1 – 5 – 25 – 100 Minuten

5.2.2 Version B

Wenn 2 Bäcker in 2 Minuten 2 Laibe Brot herstellen, wie lange würden 50 Bäcker brauchen um 50 Laibe Brot herzustellen?

1 – 2 – 25 – 50 Minuten

5.3.1 Version A

Auf einem See wachsen Seerosen. Jeden Tag verdoppelt sich die Größe der Pflanzen. Wenn es 24 Tage dauert, bis der gesamte See bedeckt ist, wie lange würde es dauern, die Hälfte des Sees zu bedecken?

2 – 4 – 12 – 23 Tage

5.3.2 Version B

Ein Container ist mit Steinen gefüllt. Die Menge der Steine verdoppelt sich jeden Tag. Der Container wird in 10 Tagen voll sein. In wie vielen Tagen wird er halbvoll sein?

1 – 4 – 5 – 9 Tage

5.4.1 Version A

In einer Schublade sind schwarze und braune lose Socken mit einem Verhältnis von 4 zu 5 gemischt. Wie viele Socken würdest du ohne hinzusehen herausnehmen müssen, um sicher ein Paar derselben Farbe zu bekommen?

1 – 2 – 3 – 4 – 5 Socken

5.4.2 Version B

In einer Tüte sind grüne und blaue Murmeln in einem Verhältnis von 6 zu 8 gemischt. Wie viele Murmeln müsstest du ohne hinzusehen herausnehmen müssen, um zwei in derselben Farbe zu erhalten?

2 – 3 – 4 – 5 – 6 Murmeln

5.5.1 Version A

Zwei Züge, welche 50 km voneinander entfernt sind, beginnen mit einer Geschwindigkeit von 25 km pro Stunde (kmh) aufeinander zuzufahren. Als die Züge losfahren, fliegt ein Vogel von der Spitze des einen Zuges auf den anderen Zug zu. Sobald der Vogel beim anderen Zug ankommt, fliegt er wieder zurück und so weiter, bis sich die Züge treffen. Wenn der Vogel mit einer Geschwindigkeit von 60 kmh fliegt, wie viele km wird der Vogel geflogen haben, bis die beiden Züge sich treffen?

15 km – 30 km – 45 km – 60 km

5.5.2 Version B

Zwei Fußgänger, welche 10 Kilometer voneinander entfernt stehen, beginnen mit einer Geschwindigkeit von 5 km/h aufeinander zuzugehen. Als die Fußgänger losgehen, rennt ein Hund von einer Person zur anderen. Sobald der Hund die andere Person erreicht, dreht er um und rennt zurück zur ersten Person und so weiter bis sich die Fußgänger treffen. Wenn der Hund mit einer Geschwindigkeit von 20 km/h läuft, wie viele Kilometer läuft der Hund, bis sich die Fußgänger treffen?

2 – 5 – 10 – 20 km

5.6.1 Version A

Ein Mann kauft ein Pferd für £60, verkauft es für £70, kauft es zurück für £80 und verkauft es letztendlich für £90. Wie viel Gewinn hat er gemacht?

£10 – £20 – £30 – £40

5.6.2 Version B

Eine Dame kauft Aktien einer Firma auf dem Börsenmarkt für \$600, verkauft sie für \$700, kauft sie zurück für \$800 und verkauft sie letztendlich für \$900. Wie viel Gewinn hat sie gemacht?

\$100 – \$200 – \$300 – \$400

5.7.1 Version A

Die Polizei ist überzeugt davon, dass entweder A, B, C oder D ein Verbrechen begangen hat. Jeder der Verdächtigen hat eine Aussage gemacht, aber nur eine der Aussagen ist wahr.

- A sagte: „Ich war es nicht.“
- B sagte: „A lügt.“
- C sagte: „B lügt.“
- B sagte: „B war es.“

Wer sagt die Wahrheit? Wer hat das Verbrechen begangen?

Wahrheit: A – B – C – D

Verbrechen: A – B – C – D

5.7.2 Version B

Der Richter weiß, dass eine der Personen A, B, C oder D eine Erfindung gemacht hat. Jeder von ihnen hat ein Statement abgegeben, aber nur eines der vier Aussagen ist wahr:

- A sagte: „Ich habe es erfunden.“
- B sagte: „A lügt.“
- C sagte: „B lügt.“
- D sagte: „B hat es nicht erfunden.“

Wer sagt die Wahrheit? Wer hat die Erfindung gemacht?

Wahrheit: A – B – C – D

Erfindung: A – B – C – D

5.8.1 Version A

Dir wird ein Set von vier Karten präsentiert, jede Karte hat eine Zahl auf der einen Seite und eine Farbe auf der anderen Seite. Die Zahlen sind 3 und 8, und die Farben rot und braun.

Welche Karte(n) musst du umdrehen, um die Aussage zu testen, dass eine Karte, welche auf der einen Seite eine gerade Zahl hat, auf der Rückseite rot ist? Versuche die effizienteste Art herauszufinden, indem du so wenige Karten wie möglich umdrehst.

Markiere so viele wie nötig: 3 – 8 – rot – braun

5.8.2 Version B

Dir wird ein Set von vier Karten präsentiert, jede Karte hat eine Zahl auf der einen Seite und eine Farbe auf der anderen Seite. Die Zahlen sind 5 und 6 und die Farben grün und gelb.

Welche Karte(n) musst du umdrehen, um die Aussage zu testen, dass eine Karte, welche auf der einen Seite eine ungerade Zahl hat, auf der Rückseite grün ist? Versuche die effizienteste Art herauszufinden, indem du so wenige Karten wie möglich umdrehst.

Markiere so viele wie nötig: 5 – 6 – grün – gelb

Italian

Hypothesis Testing

1.1.1 Versione A

Uno studente di scuola superiore deve scegliere tra due università A o B. Lo studente ha diversi amici in entrambe le università. Gli amici che frequentano l'Università A sono soddisfatti sia dal punto di vista accademico che sociale; Gli amici che frequentano l'università B hanno riscontrato problemi sia dal punto di vista accademico che sociale. Lo studente ha visitato entrambe le università per qualche ora, e le sue impressioni sono opposte a quelle riportate dagli amici.

Quale università dovrebbe scegliere lo studente?

Università A - Università B

1.1.2 Versione B

Una coppia di genitori deve scegliere tra due asili per decidere dove iscrivere il loro primogenito. La coppia ha diversi amici che hanno figli che frequentano entrambi gli asili. Gli amici che hanno un figlio che frequenta la scuola materna A sono contenti sia dal punto di vista educativo che sociale; gli amici che hanno figli che frequentano l'asilo B hanno riscontrato problemi su entrambi i fronti. La coppia ha visitato entrambi gli asili per qualche ora, e le loro impressioni sono opposte a quelle riportate dagli amici.

Quale asilo dovrebbero scegliere i genitori?

Asilo A - Asilo B

1.2.1 Versione A

Nella città di Kingston è in carica da un anno e mezzo un capo della polizia, non ben visto agli occhi dei cittadini. Egli è politicamente attivo ed è un collega del sindaco, ma quando è stato nominato aveva poca esperienza sul campo. Il sindaco ha recentemente difeso il capo della polizia in pubblico, affermando che da quando è entrato in carica, i tassi di criminalità sono diminuiti del 12%. Quale dei seguenti elementi di prova indebolirebbe l'affermazione del sindaco riguardo alla competenza del capo della polizia?

- Anche i tassi di criminalità delle città più vicine a Kingston sono diminuiti del 18% nello stesso periodo.
- un sondaggio fatto dai cittadini mostra che gli intervistati riportano il 40% in più di crimine di quanto riportato negli archivi della polizia.
- Il buon senso indica che c'è poco che un capo della polizia possa fare per diminuire i tassi di criminalità. Questi sono per la maggior parte dovuti alle condizioni sociali ed economiche che quindi esulano dalla funzione del capo della polizia.
- Si è scoperto che il capo della polizia ha rapporti stretti con persone che sono noti per essere coinvolti nel crimine organizzato.

1.2.2 Versione B

In una scuola superiore è in carica da un anno e mezzo un insegnante poco popolare. L'insegnante è un amico del preside, e aveva poca esperienza nell'insegnamento prima di essere nominato. Il preside ha recentemente difeso l'insegnante di fronte alla scuola, annunciando che da quando è entrato in carica, la cattiva condotta dei suoi allievi è diminuita del 12%. Quale delle seguenti opzioni farebbe vacillare l'affermazione del preside riguardo alla competenza dell'insegnante ?

- Anche i tassi di cattiva condotta di altre due classi non insegnate da questo docente sono diminuite del 18% nello stesso periodo.
- un sondaggio degli studenti dimostra che è presente il 40% in più di cattiva condotta rispetto a quello riportato dal preside.
- Il buon senso ci dice che c'è poco che un insegnante possa fare per diminuire la cattiva condotta. Questa è in gran parte correlato al comportamento dello studente.
- Si è scoperto che l'insegnante ha penalizzato gli studenti che avevano una cattiva condotta.

1.3.1 Versione A

Devi capire se una macchina è di marca X o di marca Y. Supponi che tu possa potenzialmente avere due tipi di informazioni su ciascuna alternativa, la percentuale di auto X e Y che hanno un

consumo di oltre 10 l / 100 km (litri per 100 km) e la percentuale di auto X e Y auto che hanno avuto grossi problemi meccanici nei primi due anni di proprietà.

Supponi che una macchina (sia di marca X o Y) faccia più di 10 l / 100 km e che abbia avuto grossi problemi meccanici nei primi due anni di proprietà. Inoltre considera che il 65% del marchio di auto X consuma più di 10 l / 100 km.

Se potessi scoprire uno dei seguenti tre tipi di informazioni, quale ti aiuterebbe a decidere se l'auto è di marca X o Y?

- la percentuale di autovetture di marca Y che ottengono più di 10 l / 100 km
- la percentuale di autovetture di marca X che hanno avuto grossi problemi meccanici nei primi due anni
- la percentuale di autovetture di marca Y che hanno avuto grossi problemi meccanici nei primi due anni

1.3.2 Versione B

Devi capire se un computer portatile è di marca X o Y. Supponi che tu possa potenzialmente avere due tipi di informazioni su ciascuna alternativa, la percentuale di computer di marca X e di marca Y che hanno una durata della batteria di oltre 10 ore e la percentuale di marca X e di marca Y che non hanno avuto grossi problemi tecnici nei primi due anni di proprietà.

Supponi che un computer (X o Y) abbia avuto una durata della batteria di oltre 10 ore e che non abbia avuto grossi problemi tecnici nei primi due anni di proprietà. Tieni conto anche che il 65% dei computer portatili di marca X hanno avuto una durata della batteria di oltre 10 ore.

Se potessi scoprire uno dei seguenti tre tipi di informazioni, quale ti aiuterebbe a decidere se il computer è di marca X o Y?

- La percentuale di computer marca Y che ha oltre 10 ore di durata di batteria.
- La Percentuale di computer marca X, che non ha avuto grossi problemi tecnici nei primi due anni
- La Percentuale di computer marca Y, che non ha avuto grossi problemi tecnici nei primi due anni

1.4.1 Versione A

Mr. Maxwell partecipa ad una festa a cui sono stati invitati solo professori universitari e dirigenti d'azienda (gli invitati sono professori universitari o dirigenti d'azienda ma non possono fare entrambe le professioni insieme). L'unica cosa che sai di Mr. Maxwell è che è un membro del Club dell'Orso. Ti viene chiesto di stimare la probabilità che il signor Maxwell sia un professore universitario, chiedendo 1, 2, 3, o 4 delle domande di seguito riportate. Ti viene chiesto di verificare la loro pertinenza. Una delle domande è la risposta che ti aiuterà nel rispondere. Considera ciascuna delle domande separatamente e indica se sia rilevante o irrilevante a ciò che ti viene richiesto.

- Qual è la percentuale di professori universitari presenti alla festa?
- Qual è la percentuale di membri del Club dell'orso presente alla festa?
- Qual è la percentuale di professori universitari che sono membri del Club dell'Orso?
- Qual è la percentuale di dirigenti d'azienda che sono membri del Club dell'Orso presenti alla festa?

1.4.2 Versione B

La signora Crowe partecipa ad un tour guidato che è frequentato solo da turisti americani e turisti canadesi (i turisti possono essere di una sola nazionalità, non tutte e due insieme).

L'unica cosa che si sa riguardo alla signora Crowe è che lei è un ospite dell'hotel ABC. Ti viene chiesto di stimare la probabilità che la signora Crowe è una turista americana chiedendo 1, 2, 3, o 4 delle domande di seguito riportate. Ti viene chiesto di verificare la loro pertinenza. Una delle domande è la risposta che ti aiuterà nel rispondere. Si prega di considerare ciascuna delle domande separatamente e indicare se sia rilevante o irrilevante per rispondere a ciò che ti è stato richiesto.

- Qual è la percentuale di turisti americani che sono presenti al tour?
- Qual è la percentuale di ospiti dell'hotel ABC che partecipano al tour?
- Qual è la percentuale di turisti canadesi presenti al tour che sono ospiti dell'hotel ABC?
- Qual è la percentuale di turisti americani presenti al tour che sono ospiti dell'hotel ABC?

1.5.1 Versione A

Linda ha 31 anni, è single, senza peli sulla lingua, è una ragazza molto sveglia e si è laureata in filosofia. Da studentessa era particolarmente appassionata di questioni riguardanti discriminazione e giustizia sociale, e ha anche partecipato a manifestazioni antinucleari. Quale delle due alternative è più probabile:

- Linda è un'impiegata di banca
- Linda è un'impiegata di banca ed è attiva nel movimento femminista

1.5.2 Versione B

Jack ha 45 anni, è sposato e ha quattro figli. E' generalmente attento e ambizioso, non mostra alcun interesse per le questioni politiche e sociali e trascorre la maggior parte del suo tempo libero mettendo in pratica i suoi hobby, tra cui lavorare il legno, andare in barca e risolvere enigmi matematici.

Quale delle due alternative è più probabile:

- Jack è un ingegnere
- Jack è un ingegnere ed è attivo nel club locale per gli sport acquatici

1.6.1 Versione A

Vuoi acquistare una nuova auto. Oggi devi scegliere tra due alternative: l'acquisto di un' utilitaria A o B. È possibile utilizzare un solo criterio per la scelta, ovvero l'aspettativa di vita della vettura. Avete informazioni da Consumer Reports, che riporta il record migliore per l'auto A su un campione di diverse centinaia di autovetture. Ieri un vicino di casa ti ha detto che la sua nuova auto di marca A si è rotta. Quale macchina acquisti?

Auto A - Auto B

1.6.2 Versione B

Vuoi acquistare un nuovo computer. Oggi devi scegliere tra due alternative: per l'acquisto di un computer A o di un computer B. È possibile utilizzare un solo criterio per la scelta, ovvero l'aspettativa di vita del computer. Avete informazioni da Consumer Reports, che riporta il record migliore per il computer A su un campione di diverse centinaia di computer. Ieri un vicino di casa ti ha detto che il suo nuovo computer di marca A si è rotto. Quale computer acquisti?
computer A - computer B

Verbal Reasoning

Analogie

A1) _____: cavallo: salire a bordo: treno

- stalla
- Ferro di cavallo
- Cavalcare
- Salire in sella

B1) Poesia: Rima :: filosofia: _____

- linguaggio figurato
- Musica
- Legge
- Teoria

A2) rammendare: cucito :: correggere: _____

- rammendo
- aggiustare
- Manoscritto
- rattoppo

B2) alfabetico: _____ :: sequenziale: file

- ordinare
- Parte
- elenco
- ordine

A3) mouse: _____ :: Flash: macchina fotografica

- ratto
- Computer
- cavo
- dolce

B3) cuscino: divano :: Mensola: _____

- ripiano
- libreria
- magazzino
- cornice

A4) inumidire: _____ :: rinfrescare: congelare

- acqua
- immergere
- Forno
- crescere

B4) nastro: _____ :: ciliegina: torta

- regalo
- taglio
- fiocco
- macchina da scrivere

A5) gattino: _____ :: soldato: esercito

- gatto
- cucciolata
- Cucciolo
- Miao

B5) scienziato: esperimento :: _____: opera

- bicchiere
- prova
- Attore
- laboratorio

A6) Dizionario: Definizione :: _____: Mappa

- direzione
- b. Sud
- c. Atlante
- d. longitudine

B6) puntina: _____ :: Gancio: cappotto

- chiodo
- poster
- Parete
- martello

A7) insolito: novità :: _____: normale

- bizzarro
- nuovo
- comune
- poesia

B7) _____: università :: meccanico: officina

- libro
- istruzione
- Professore
- motore

A8) tamburo: strumento :: trapano: _____

- Martello
- forno
- attrezzo
- crescendo

B8) cima: montagna :: _____: casa

- massimizzare
- tetto
- Veranda
- bungalow

A9) _____: pioggia:: condensazione: Umidità

- erosione
- nuvola
- terreno
- previsione

B9) _____: Telaio :: mosaico: piastrelle

- pellicola
- incisione
- Bagno
- decorazione

Judging Likelihood and Probability

3.1.1 Versione A

Squash può essere giocato a 9 o 15 punti. Se A è un giocatore migliore di B, quale schema di punteggio darebbe al giocatore A una migliore possibilità di vincere?

- Partita da 9 punti • Partita da 15 punti

3.1.2 Versione B

Una partita di Pin Pong può essere vinta con il miglior punteggio su 5 partite o miglior punteggio su 9 partite. Se A è un giocatore migliore di B, quale schema di punteggio darebbe al giocatore A una migliore possibilità di vincere?

- punteggio migliore su 5 partite • Punteggio migliore su 9 partite

3.2.1 Versione A

Dopo le prime 2 settimane dall'inizio del più importante campionato di baseball, i giornali cominciano a stampare le prime 10 medie di battuta. In genere, dopo 2 settimane, il battitore leader ha spesso una media di circa 450. Tuttavia, nessun battitore nella storia della Major League ha mai avuto media di 450 alla fine della stagione. Perché pensi che questo accada?

Spunta un'alternativa:

- Quando un battitore è conosciuto per raggiungere una media alta, i lanciatori tendono a pressarlo di più quando lanciano su di lui
- I lanciatori tendono a migliorare nel corso della stagione grazie ad una migliore forma fisica. Migliorando, i lanciatori tirano lanci più difficili sui battitori, così le medie dei battitori scendono.
- Un'alta media di un giocatore all'inizio della stagione può essere dovuta solo a fortuna. A fine stagione si hanno dati più realistici sulla qualità di gioco dei battitori.
- Un battitore che ha una striscia positiva all'inizio della stagione è sotto molta pressione per mantenere la sua prestazione. Tale stress influisce negativamente sul suo modo di giocare.
- Quando un battitore è noto per essere forte, smette di ricevere buoni tiri dai lanciatori. Al contrario, i lanciatori cercano di metterlo di più in difficoltà.

3.2.2 Versione B

Dopo le prime 2 settimane dall'inizio della scuola, un insegnante di una scuola cerca di prevedere i primi 10 miglior studenti dell'anno. In genere, dopo 2 settimane, spesso lo studente

che ha una media più alta ha una media di circa il 92%. Tuttavia, nessuno studente nella storia della scuola ha mai avuto una media di 92% alla fine dell'anno. Per quale motivo pensi che questo accada? Spunta un' alternativa:

- Gli insegnanti aumentano le loro aspettative quando valutano i compiti di uno studente che è risaputo avere una media alta.
- I compiti tendono a diventare più difficili nel corso dell'anno, come il materiale da studiare aumenta. Come i compiti diventano più difficili, gli studenti hanno meno probabilità di raggiungere un punteggio elevato, in questo modo le medie degli studenti scendono.
- Una media alta di uno studente all'inizio della scuola può essere dovuta solo a fortuna. L'anno accademico prevede una valutazione più realistica di conoscenza e capacità di uno studente.
- Uno studente che ha una media così alta all'inizio dell'anno è sotto pressione per cercare di mantenere la sua media. Tale stress influisce negativamente sui suoi risultati.
- Quando uno studente ha una media alta, smette di ricevere semplici domande dagli insegnanti. Al contrario, gli insegnanti aumentano la difficoltà delle domande per questo studente perché vogliono mettere alla prova le sue capacità.

3.3.1 Versione A

Giocando con le slot machine, le persone vincono circa 1 volta su 10. Julie, però, ha appena vinto i suoi primi tre giochi. Qual è la sua probabilità di vincere anche la prossima partita?

___ Su 10

3.3.2 Versione B

Partecipando ad una lotteria locale, le persone hanno la possibilità di vincere circa 1 volta ogni 20 biglietti della lotteria. Jack, invece, ha appena vinto i suoi primi tre biglietti. Quali sono le sue probabilità di vincere la prossima volta che gioca?

___ Su 20

3.4.1 Versione A

Immagina di aver lanciato una moneta (una moneta che ha un 50/50 di possibilità di uscire testa o croce) ed è uscita la testa 5 volte di fila. Per il 6° lancio pensi che:

- È più probabile che esca croce che testa.
- È più probabile che esca testa che croce
- testa e croce sono ugualmente probabili al sesto lancio.

3.4.2 Versione B

Immagina di aver lanciato (un dado che ha un 1 possibilità su 6 di uscire con ogni numero) e il numero 6 è appena uscito 5 volte di fila. Per il 6° lancio pensi che:

- È più probabile che esca un numero diverso da 6.
- È più probabile che esca 6 piuttosto che un numero diverso
- 6 e qualsiasi altro numero sono ugualmente probabili al sesto lancio.

3.5.1 Versione A

Supponi che ti vengano presentati due contenitori di biglie bianche e nere: un contenitore grande che contiene 100 biglie e un contenitore piccolo che contiene 10 biglie. Devi prendere una biglia da uno dei contenitori senza guardare. Se prendi una biglia nera vinci 2 euro. Considera che il contenitore piccolo contiene 1 biglia nera e 9 biglie bianche, e il contenitore grande contiene 8 biglie nere e 92 biglie bianche. Da quale contenitore prenderesti la biglia in una situazione reale?

- Contenitore Piccolo • Contenitore Grande

3.5.2 Versione B

Supponi che ti vengano presentati due sacchi di banconote da 10 euro e da 20 euro: un grande sacco che contiene 100 diverse banconote e un sacco piccolo che contiene 10 banconote diverse. Le banconote sono mescolate casualmente nei sacchi. Devi tirare fuori una banconota da entrambi i sacchetti senza guardare e poi potrai tenere la banconota. Considera che il sacco piccolo contiene una banconota da 20 euro e nove banconote da 10 euro, e il sacco grande

contiene otto banconote da 20 euro e novantadue banconote da 10 euro. Da quale sacco prenderesti la banconota in una situazione reale?

- Sacco Piccolo • Sacco Grande

Argumentation Analysis

Scegli: valido - non valido

A1

- Nessun tipo di sigarette è economico
- Alcune dipendenze sono economiche
- Pertanto, alcune dipendenze non sono sigarette

B1

- Alcuni edifici sono edifici in legno
- Nessun edificio in legno è un grattacielo
- Pertanto alcuni edifici non sono grattacieli

A2

- alcuni articoli di apparecchiature per ufficio non sono computer
- alcuni apparecchi elettrici sono elementi di apparecchiature per ufficio
- Pertanto alcuni apparecchi elettrici non sono computer.

B2

- Alcune bevande sono analcoliche
- Le bevande analcoliche non sono liquori
- Pertanto alcune bevande non sono liquori.

A3

- Le sedie non sono tavoli.
- Alcuni mobili sono sedie.
- Perciò alcuni mobili non sono tavoli.

B3

- Alcuni pesci sono commestibili
- Le trote sono commestibili
- Perciò alcuni pesci non sono trote.

B4

- Le dipendenze sono economiche.
- Alcuni tipi di sigarette sono economiche.
- Pertanto, alcune dipendenze non sono le sigarette.

A4

- Alcuni edifici in legno sono grattacieli.
- Nessun edificio è un edificio in legno.
- Quindi alcuni edifici non sono grattacieli.

B5

- alcuni apparecchi elettrici non sono elementi di apparecchiature per ufficio
- alcuni articoli di apparecchiature per ufficio sono i computer
- Pertanto Alcuni apparecchi elettrici non sono computer.

A5

- Le Bevande analcoliche sono liquori.
- Non ci sono bevande analcoliche.
- Pertanto Alcune bevande non sono liquori.

B6

- Nessun mobile è una sedia.
- Le sedie sono tavoli.
- Perciò alcuni mobili non sono tavoli

A6

- Le trote non sono commestibili.
- I pesci sono commestibili.
- Perciò alcuni pesci non sono trote.

A7

- Alcuni attrezzi non sono manuali.
- I cacciaviti sono attrezzi manuali.
- Pertanto alcuni cacciaviti non sono attrezzi.

B7

- Alcuni strumenti musicali sono strumenti a fiato
- gli strumenti a fiato non sono violini
- Perciò alcuni violini non sono strumenti musicali.

A8 // B8

- Le cose verdi non sono erba.
- Alcune piante sono verdi.
- Pertanto l'erba non è una pianta.

B9

- I cacciaviti non sono attrezzi manuali
- non ci sono attrezzi manuali
- Pertanto alcuni cacciaviti non sono attrezzi.

A9

- Strumenti a fiato sono violini.
- Nessuno strumento musicale è uno strumento a fiato.
- Perciò i violini non sono strumenti musicali.

Problem Solving

5.1.1 Versione A

Una racchetta e una palla costano 1,10 \$ in totale. La racchetta costa 1,00 \$ più della palla.
Quanto costa la palla?

5 centesimi - 10 centesimi - 55 centesimi - 100 centesimi

5.1.2 Versione B

Un pezzo di pane e una mela costano 2,20 \$ in totale. Il pane costa \$ 2 più della mela. Quanto costa la mela?

10 centesimi - 20 centesimi - 110 centesimi - 200 centesimi

5.2.1 Versione A

Se ci vogliono 5 macchine e 5 minuti per fare 5 pezzi di plastica, quanto tempo impiegano 100 macchine per fare 100 pezzi di plastica?

1 - 5 - 25 - 100 minuti

5.2.2 Versione B

Se ci vogliono 2 panettieri e 2 minuti per fare 2 pagnotte di pane, quanto tempo impiegano 50 fornai per fare 50 pagnotte di pane?

1 - 2 - 25 - 50 minuti

5.3.1 Versione A

In un lago, c'è un'aiuola di ninfee. Ogni giorno, l'aiuola raddoppia. Se ci vogliono 24 giorni all'aiuola per coprire tutto il lago, quanto tempo ci vorrebbe all'aiuola per coprire la metà del lago?

2 - 4 - 12 - 23 giorni

5.3.2 Versione B

Un contenitore viene riempito di sassi. La quantità di sassi raddoppia in numero ogni giorno. Il contenitore sarà pieno in 10 giorni. In quanti giorni sarà mezzo pieno?

2 - 4 - 5 - 9 giorni

5.4.1 Versione A

Ci sono calzini neri e marroni sciolti in un cassetto mescolati in un rapporto di 4 a 5. Quanti calzini sciolti dovresti prendere senza guardare per essere sicuro di ottenere una coppia dello stesso colore?

1 - 2 - 3 - 4 - 5 calze

5.4.2 Versione B

Ci sono biglie verdi e blu mescolate in un sacchetto in un rapporto di 6 a 8. Quante biglie dovresti prendere senza guardare per essere sicuro di ottenerne due dello stesso colore?

2 - 3 - 4 - 5 - 6 biglie

5.5.1 Versione A

Due treni, a 50 km di distanza partono l'uno nella direzione dell'altro ad una velocità di 25 km / h (km all'ora) ciascuna. Come partono i treni, un uccello vola dalla parte anteriore di un treno verso il secondo. Al raggiungimento del secondo treno l'uccello si gira e vola di nuovo al primo treno e così via fino a quando i treni si incontrano. Se l'uccello vola a 60 chilometri all'ora, quanti km deve volare l'uccello prima che i treni si incontrino?

15 km - 30 km - 45 km - 60 km

5.5.2 Versione B

Due pedoni a 10 km di distanza camminano uno verso l'altro a 5 km / h. Mentre i pedoni iniziano a camminare, un cane corre da una persona verso la seconda. Raggiunta la seconda persona il cane si gira e corre di nuovo alla prima persona e così via fino a quando i pedoni si incontrano. Se il cane corre a 20 chilometri all'ora, quanti chilometri dovrà fare il cane prima che i pedoni si incontrino?

2-5 - 10 - 20 km

5.6.1 Versione A

Un uomo acquista un cavallo per £ 60, lo vende per £ 70, lo acquista di nuovo per £ 80 e lo vende, infine, per £ 90. Quanto ha guadagnato l'uomo?

£ 10 - £ 20 - £ 30 - £ 40

5.6.2 Versione B

Una signora acquista azioni di una società al mercato azionario per \$ 600, li vende per \$ 700, li riacquista per \$ 800 e, infine, li vende per \$ 900. Quanto ha guadagnato lei?

\$ 100 - \$ 200 - \$ 300 - \$ 400

5.7.1 Versione A

La polizia è convinta che uno tra A, B, C e D abbia commesso un crimine. Ciascuno dei sospettati, a sua volta, fa una dichiarazione, ma solo una delle quattro dichiarazioni è vera.

- A ha detto: "Non sono stato io."
- B ha detto: "A sta mentendo".

- C ha detto: "B sta mentendo".
- D ha detto: "è stato B."

Chi dice la verità? Chi ha commesso il crimine?

Verità: A - B - C - D

Crimine: A - B - C - D

5.7.2 Versione B

Il giudice sa che uno tra A, B, C o D sta mentendo. Ciascuno di essi fa un'affermazione, ma solo uno delle quattro dichiarazioni è vera.

- A ha detto: "ho mentito."
- B ha detto: "A sta mentendo".
- C ha detto, "B sta mentendo".
- D ha detto: "B non ha mentito."

Chi dice la verità? Chi ha mentito?

Verità: A - B - C - D

Menzogna: A - B - C - D

5.8.1 Versione A

Ti viene mostrato un gruppo di quattro carte posizionate su un tavolo, ciascuna delle quali ha un numero su un lato e un colore sull'altro. Le facce visibili delle carte mostrano 3, 8, rosso e marrone. Quale carta/e devi girare al fine di testare la verità dell'affermazione "se una carta presenta un numero pari su una faccia, la sua faccia opposta è quindi rossa"? Prova a trovare la soluzione in modo più efficace utilizzando il minor numero di carte necessarie.

Segna come molti appaiono: 3 - 8 - rosso - marrone

5.8.2 Versione B

Ti viene mostrato un gruppo di quattro carte posizionate su un tavolo, ciascuno delle quali ha un numero su un lato e un colore sull'altro lato. Le facce visibili delle carte mostrano 5, 6, verde e giallo. Quale carta/e devi girare al fine di verificare la verità della proposizione "se una carta presenta un numero dispari su una faccia, poi la sua faccia opposta è verde"? Prova a trovare la soluzione in modo più efficace utilizzando il minor numero di carte necessarie.

Segna come molti appaiono: 5 - 6 - verde - giallo

Russian

Hypothesis Testing

1.1.1 Версия А

Выпускник средней школы должен выбрать между двумя колледжами А и В. У ученика есть несколько друзей в каждом колледже, у которых похожие интересы. Друзья из колледжа А выбрали колледж за его образовательную и социальную системы. В то время, как его друзья из колледжа В негативно высказывались о его образовательной и социальной системе. Выпускник посетил оба колледжа в течение дня, и его впечатления были противоположны отчетам его друзей.

Какой колледж должен выбрать старшеклассник?

- Колледж А
- Колледж В

1.1.2 Версия В

Мать и отец должны выбрать между двумя детскими садами А и В для своего первенца. У пары есть несколько друзей, у которых похожие ценности и убеждения и чьи дети посещали один из этих садиков.

Друзья, чьи дети посещали детский сад А, хвалили садик за его образовательную и социальную системы; однако, друзья, дети которых посещали детский сад В, остались недовольны детсадом по тем же критериям. Пара посетила оба детских сада в течение нескольких часов и их впечатления были противоположны впечатлениям их друзей.

Какой детский сад должны выбрать родители?

- Детский сад А
- Детский сад В

1.2.1 Версия А

В городе под названием К в течение полутора лет работает непопулярный начальник полиции. Он политически активен и является коллегой мэра. Когда он был назначен начальником полиции, у него был небольшой опыт работы. Мэр недавно защитил начальника публично, заявив, что за время, прошедшее с момента его вступления в должность, уровень преступности снизился на 12%. Какие из следующих доказательств поставили бы под сомнение заявление мэра о том, что его начальник компетентен?

- Уровень преступности в двух городах, очень схожих с городом К по расположению и размеру, за тот же период уменьшился на 18%.
- (Б) Независимый опрос граждан К показывает, что опрошенное население в опросе сообщает о 40% больше преступлений, чем сообщается в полицейских отчетах.
- Здравый смысл указывает на то, что полицейский мало что может сделать, для того чтобы снизить уровень преступности. Это по большей части связано с социальными и экономическими условиями, не зависящими от должностных лиц.
- (Г) У начальника полиции обнаружены личные контакты с людьми, которые, как известно, участвуют в организованной преступности.

1.2.2 Версия В

В средней школе непопулярный учитель преподавал полтора года. Он является другом директора школы. На момент назначения на должность, у него был небольшой опыт преподавания. Директор школы недавно защитил учителя перед школой, объявив, что за время, прошедшее с того момента, как он начал преподавать, плохое поведение его учеников уменьшилось на 12%. Какое из следующих доказательств наименее правдоподобное и могут опровергнуть утверждение директора о том, что учитель компетентен?

- (А) Уровень плохого поведения учеников двух классов, где этот учитель не преподает, за тот же период уменьшился на 18%.
- (Б) Опрос студентов показывает, что в опросе о плохом поведении они сообщают на 40% больше случаев, чем сообщает директор.
- (С) Здравый смысл указывает на то, что учителя мало что могут сделать, чтобы уменьшить процент неправильного поведения. Это по большей части связано с настроением и поведением учеников.
- (D) Было обнаружено, что учитель штрафовал студентов, которые плохо себя вели.

1.3.1 Версия А

Допустим, Вам нужно определить, относится ли неизвестный автомобиль к торговой марке X или к торговой марке Y. Предположим, что доступны два типа информации по каждой марке: процент автомобилей X и автомобилей Y, которые расходуют более 10 л / 100 км (Литров на 100 км) и процент автомобилей X и автомобилей Y, которые не имели серьезных механических проблем в течение первых двух лет эксплуатации.

Предположим, что неизвестный автомобиль (марки X или Y) расходует более 10 л / 100 км и что он не имел серьезных механических проблем в течение первых двух лет эксплуатации. Также предположим, что 65% автомобилей марки X потребляют более 10 л / 100 км.

Если бы Вам были бы представлены три следующих типа информации, какой из них помог бы Вам определить, какой марки неизвестный автомобиль, X или Y?

- Процент автомобилей марки Y, которые потребляют более 10 л / 100 км
- Процент автомобилей марки X, у которых не было серьезных механических проблем в первые два года эксплуатации.
- Процент автомобилей марки Y, у которых не было серьезных механических проблем в первые два года эксплуатации

1.3.2 Версия В

Допустим, Вам нужно определить, является ли неизвестный ноутбук представителем марки X или марки Y. Предположим, что для каждой альтернативы потенциально доступны два типа информации: процент марки X и марки Y, срок службы батареи которых превышает 10 часов, и процентное соотношение марки X и марки Y, которые не имели серьезных технических проблем в течение первых двух лет эксплуатации.

Предположим, что неизвестный ноутбук (X или Y) имел срок службы батареи более 10 часов и что он не имел серьезных технических проблем в течение первых двух лет эксплуатации. Также предположим, что 65% ноутбуков марки X имеют срок службы батареи более 10 часов.

Выберите из трех типов предоставленной информации ту информацию, которая вам поможет решить, какой марки неизвестный ноутбук: X или Y?

- Процент марки Y, который обеспечивает более 10 часов автономной работы.
- Процент марки X, у которого не было серьезных технических проблем в первые два года эксплуатации.
- Процент марки Y, у которого не было серьезных технических проблем в первые два года эксплуатации.

1.4.1 Версия А

Господин М присутствовал на вечеринке, на которую были приглашены преподаватели университетов и бизнес-руководители (гости не могли иметь две этих должности одновременно). Единственное, что вы знаете о мистере М, это то, что он является членом Клуба М. Вам необходимо оценить вероятность того, что Господин М является профессором университета, задав вопросы, приведенные ниже. Сперва вам нужно проверить, какие из вопросов уместны и помогут вам в вашем решении. Затем, рассмотрите каждый из вопросов отдельно и укажите, какой из них **He** относится к вашему заданию.

- Какой процент людей на вечеринке являются профессорами университета?
- Какой процент членов Клуба М находится на вечеринке?
- Какой процент преподавателей университета на вечеринке являются членами Клуба М?
- Какой процент бизнес-руководителей на вечеринке являются членами Клуба М?

1.4.2 Версия В

Госпажа К посетила экскурсию, на которой присутствовали только американские и канадские туристы (Туристы принадлежали либо к одной, либо к другой из национальностей). Единственное, что вы знаете о госпоже К, это то, что она постоялец отеля ABC. Вам необходимо оценить вероятность того, что госпожа К является американским туристом, задав вопросы, приведенные ниже. Вам нужно проверить, какие

из вопросов уместны и помогут вам в вашем решении. Пожалуйста, рассмотрите каждый из вопросов отдельно и укажите, какой из них Не подходит для решения вашей задачи.

- Какой процент людей в туре - американские туристы?
- Какой процент постояльцев отеля ABC посещает тур?
- Какой процент канадских туристов из тура являются постояльцами отеля ABC?
- Какой процент американских туристов из тура являются постояльцами отеля ABC?

1.5.1 Версия А

Лили 31 год, одинока, открыта и очень умна. Она изучала философию. Будучи студентом, она была глубоко обеспокоена проблемами дискриминации и социальной справедливости, а также участвовала в антиядерных демонстрациях.

Какая из двух альтернатив более вероятна:

- Лиля - банковский служащий
- Б) Лиля - банковский служащий и активист феменистического движения

1.5.2 Версия В

Жене 45, он женат и имеет четверо детей. Он осторожен и амбициозен. Он не проявляет интереса к политическим и социальным вопросам и большую часть своего свободного времени проводит за своими увлечениями: домашние плотничные работы, парусный спорт и математические головоломки.

Какая из двух альтернатив более вероятна:

- А) Женя - инженер
- Б) Женя - инженер и принимает активное участие в местном клубе по водным видам спорта

1.6.1 Версия А

Вы хотите купить новый автомобиль. Сегодня вы должны выбрать между двумя альтернативами: купить автомобиль компании А или В. Для этого вы используете только один критерий - ожидаемую продолжительность службы автомобиля. У вас есть информация из опроса потребителей, о том что в выборке из нескольких сотен автомобилей автомобиль А имеет лучший рейтинг. Вчера сосед сказал вам, что его новый автомобиль марки А сломался. Какой автомобиль вы купите?

- Автомобиль А
- Автомобиль В

1.6.2 Версия В

Вы хотите купить новый компьютер. Сегодня вы должны выбрать между двумя альтернативами: купить компьютер А либо компьютер В. Вы используете только один критерий для этого выбора - продолжительность службы компьютера. У вас есть информация из опроса потребителей, о том что в выборке из нескольких сотен компьютеров компьютер А имеет лучший рейтинг. Вчера сосед сказал вам, что его новый компьютер марки А сломался. Какой компьютер вы купите?

- Компьютер А
- Компьютер В

Verbal Reasoning

Аналогии

Выберите один из четырех ответов, который лучше всего подходит для завершения логического смысла. Например: слово, следующее до знака “:” должно указывать на отношение к рядом стоящему слову, такое же, как уже сформированная по смыслу пара после знака “ ::”

A1

_____ : лошадь :: посадка: поезд

- а. конюшня
- б. подкова
- с. езда
- d. взбираться

В1

поэзия: рифма :: философия: _____

- а. образы
- б. музыка
- с. правовой акт
- д. теория

А2

поправлять: шитье :: редактировать : _____

- а. штопать
- б. ремонтировать
- с. рукопись
- д. импровизированный

В 2

В алфавитном порядке: _____ :: последовательно: файлы

- а. сортировать
- б. часть
- с. перечень
- д. заказ

А3

мышь: _____ :: вспышка: камера

- а. крыса
- б. компьютер
- с. шнур
- д. десерт

В3

подушка: диван :: полка: _____

- а. выступ
- б. книжный шкаф
- с. склад
- д. рамка

А 4

увлажнять: _____ :: охлаждать: замораживать

- а. вода
- б. пропитывать
- с. духовка
- д. расти

В4

лента: _____ :: сахарная глазурь: торт

- а. подарок
- б. порез
- с. бантик
- д. печатная машинка

А5

Дерево : _____ :: солдат: армия

- а. лист
- б. лес
- с. мох
- д. дорога

B5

Ученый: эксперимент :: _____: пьеса

- а. мерный стакан
- б. репетиция
- с. актер
- д. лаборатория

A6

словарь: определение :: _____: карта

- а. направление
- б. юг
- с. атлас
- д. долгота

B6

чертежная кнопка: _____ :: крючок: пальто

- а. гвоздь
- б. плакат
- с. стена
- д. молоток

A7

необычный: новинка :: _____: стандарт

- а. странный
- б. роман
- с. привычный
- д. стих

B7

_____ : колледж :: механик: гараж

- а. книга
- б. обучение
- с. профессор
- д. двигатель

A8

барабан: инструмент :: дрель: _____

- а. молоток
- б. духовой шкаф
- с. инструмент
- д. крещендо

B8

вершина: гора :: _____: дом

- максимизировать
- крыша
- крыльцо
- бунгало

A9

_____ : осадки :: конденсация: влажность

- а. эрозия
- б. облако
- с. земля
- д. прогноз

B9

_____ : кадр :: мозаика: плитка

- а. фильм
- б. гравюра
- с. ванная
- д. украшение

Judging Likelihood and Probability

3.1.1 Версия А

Игра в сквош может быть сыграна в 9 или 15 очков. Если А лучший игрок, чем В, какая схема подсчета даст игроку А лучший шанс выиграть?

- 9-очковая игра
- 15-очковая игра

3.1.2 Версия В

Настольный теннис можно выиграть с лучшим счетом из 5 игр или с лучшим счетом из 9 игр. Если А лучший игрок, чем В, какая схема подсчета даст игроку А лучший шанс выиграть?

- лучший шанс из 5 игр
- лучший шанс из 9 игр

3.2.1 Версия А

После первых 2 недель основного бейсбольного сезона лиги, газеты начинают печатать 10 лучших средних значений. Как правило, через 2 недели лучший бьющий обычно имеет количество очков в среднем около 0,450. Тем не менее, ни один бьющий в истории высшей лиги получал в среднем 0,450 в конце сезона. Как Вы считаете, почему?

Отметьте один ответ :

- а. Когда известно, что бьющий хочет достичь высокого счета, подающие больше выкладываются, когда ему подают.
- б. Подающие, как правило, становятся лучше в течение сезона, так как они приходят в форму. По мере того, как навыки подающих улучшаются, они, скорее всего, чаще делают страйк-аут, что ограничивает бьющих, поэтому средние коэффициенты падают.
- с. Высокий средний показатель игрока в начале сезона может быть просто удачей. Более продолжительный сезон обеспечивает более реалистичное испытание навыков.
- д. Бьющий, у которого такие высокие показатели в начале сезона, находится под давлением, чтобы сохранить свой показатель в игре. Такое напряжение отрицательно влияет на его игру.
- е. Когда бьющий хочет достичь высокого счета, он перестает получать хорошие удары. Вместо этого, подающие «разыгрывают углы» поля, потому что они не против погонять бьющего.

3.2.2 Версия В

После первых 2 недель академического семестра преподаватель пытается предопределить 10 лучших учеников года. Как правило, через 2 недели лучшие ученики получают в среднем около 92 процента. Тем не менее, ни один ученик в истории школы не устанавливал в среднем 92 процента в конце семестра. Как Вы думаете, почему?

- а. Когда ученик стремится к высокому среднему показателю, учителя превышают свои ожидания при оценке работ этого ученика.
- б. Работы, как правило, усложняются в течение семестра, поскольку накапливается содержание материала. По мере усложнения, ученики с меньшей вероятностью набирают высокие оценки, поэтому средние показатели учеников снижаются.
- с. Высокие показатели учеников в начале семестра могут быть просто удачей. Более длительный учебный год обеспечивает более реалистичный тест знаний и умений учащихся.
- д. Ученики, которые имеют высокие средние показатели в начале семестра, находятся под большим давлением и стрессом, чтобы поддерживать их. Такое напряжение отрицательно сказывается на их результатах.

- е. Как известно, когда ученики стремятся к высокому среднему показателю, учителя перестают задавать им простые вопросы. Вместо этого, учителя «увеличивают интенсивность» вопросов для учеников, потому что они хотят проверить их способность.

3.3.1 Версия А

При игре в игровые автоматы люди выигрывают примерно 1 раз в 10 раз. Юлия, однако, только что выиграла в своих первых трех играх. Каковы ее шансы на победу в следующий раз?

_____ из 10

3.3.2 Версия В

При участии в местной лотерее у людей есть шанс выиграть примерно в 1 из 20 лотерейных билетах. Женя, однако, только что выиграл в своих первых трех билетах. Каковы его шансы на победу в следующий раз?

_____ из 20

3.4.1 Версия А

Представьте, что Вы бросаете монету (которая имеет шанс 50/50 выпасть орлом или решкой, и только что выпала решка 5 раз подряд. Для 6-го броска вы думаете, что:

- Скорее выпадет орел, чем решка.
- Б) Скорее выпадет решка, чем орел.
- Шансы решки и орла одинаково вероятны на шестом броске.

3.4.2 Версия В

Представьте, что Вы играете в кости (и может выпастьот 1 до 6 при каждом броске), а число 6 появилось 5 раз подряд. Для 6-го броска Вы думаете, что:

- Большая вероятность того, что после пяти бросков в 6-й раз выпадет другое число
- Б) Большая вероятность того, что после пяти бросков в 6-й раз выпадет то же число
- Число 6 или любое другое число имеют равные шансы выпасть на шестом броске.

3.5.1 Версия А

Предположим, что Вам представлены два подноса черных и белых стеклянных шариков : большой поднос, на котором 100 шариков и небольшой поднос, на котором 10 шариков. Вы должны взять один шарик из любого подноса, не глядя. Если вы вытащите черный шарик, Вы выиграете 2 доллара. Рассмотрим условие, в котором на маленьком подносе 1 черный шарик и 9 белых шариков, а на большом подносе 8 черных шариков и 92 белых шариков. С какого подноса Вы предпочли бы выбрать шарик в реальной ситуации?

- Большой поднос
- Маленький поднос

3.5.2 Версия В

Предположим, что Вам представлены две сумки с долларовыми банкнотами, номиналом в 10 и 20 долларов: большую сумку, содержащую 100 разных купюр и маленькую сумку, содержащую 10 разных купюр. Купюры смешаны случайным образом. Вы должны вытащить одну купюру из любой сумки, не глядя, и оставить себе. Рассмотрим условие, в котором маленькая сумка содержит одну купюру номиналом в 20 долларов и девять купюр в 10 долларов, а большая сумка содержит восемь купюр номиналом в 20 долларов и девяносто две в 10 долларов. Из какой сумки Вы предпочли бы выбрать купюру в реальной ситуации?

- Большая сумка
- Маленькая сумка

Argumentation Analysis

Пожалуйста, внимательно прочтите первые два заявления. Если они верны, действительный ли вывод?

Отметьте: действительный – недействительный

A1

- Нет сигарет, которые не стоят дорого.
- Некоторые плохие привычки не стоят дорого.
- Поэтому некоторые плохие привычки не являются сигаретами.

B1

- Некоторые здания являются деревянными
- Нет деревянных зданий, которые были бы небоскребами.
- Поэтому, некоторые здания не являются небоскребами.

A2

- некоторые предметы офисной техники не являются компьютерами
- некоторые электроприборы - это предметы офисной техники
- Поэтому некоторые электроприборы не являются компьютерами.

B2

- Некоторые напитки безалкогольные
- безалкогольные напитки не являются ликерами
- Поэтому некоторые напитки не являются ликерами.

A3

- Стулья не являются столами.
- Некоторые предметы мебели являются стульями
- Поэтому некоторые предметы мебели не являются столами

B3

- Некоторая рыба несъедобна
- Форель съедобна
- Поэтому некоторая рыба не является форелью.

B4

- Нет плохих привычек, которые не стоят дорого.
- Некоторые сигареты не стоят дорого.
- Поэтому некоторые плохие привычки не являются сигаретами.

A4

- Некоторые деревянные здания являются небоскребами
- Нет зданий, которые являются деревянными
- Поэтому некоторые здания не являются небоскребами.

B5

- Некоторые электроприборы не являются предметами офисной техники
- Некоторые предметы офисной техники - компьютеры
- Поэтому некоторые электроприборы не являются компьютерами.

A5

- Безалкогольные напитки являются ликерами
- Напитки не являются безалкогольными
- Поэтому, некоторые напитки не являются ликерами

B6

- Нет предметов мебели, которые были бы стульями
- Стулья являются столами.
- Поэтому некоторые предметы мебели не являются столами

A6

- Форель не съедобна
- Рыба съедобна
- Поэтому некоторая рыба не форель.

A7

- Некоторые инструменты не являются ручными
- Отвертки - это ручные инструменты
- Поэтому некоторые отвертки не являются инструментами.

B7

- Некоторые музыкальные инструменты являются духовыми инструментами
- Духовые инструменты не являются скрипками
- Поэтому некоторые скрипки не являются музыкальными инструментами.

A8 // B8

- Некоторые зеленые вещи не являются травами.
- Некоторые растения зеленые.
- Поэтому некоторые травы не являются растениями.

B9

- Отвертки не являются ручными инструментами
- Инструменты не являются ручными
- Поэтому некоторые отвертки не являются инструментами.

A9

- Духовые инструменты являются скрипками
- Нет музыкальных инструментов, которые являются духовыми инструментами
- Поэтому некоторые скрипки не являются музыкальными инструментами.

Problem Solving

5.1.1 Версия А

Бита и мяч вместе стоят 1,10 доллара. Бита стоит на \$1,00 больше, чем мяч. Сколько стоит мяч?

5 центов - 10 центов - 55 центов - 100 центов

5.1.2 Версия В

Хлеб и яблоко вместе стоят 2,20 доллара. Хлеб стоит на 2 доллара больше, чем яблоко. Сколько стоит яблоко?

10 центов - 20 центов - 110 центов - 200 центов

5.2.1 Версия А

Если для изготовления 5 пластиковых деталей 5 машинам требуется 5 минут, сколько времени потребуется 100 машинам для изготовления 100 пластиковых деталей?

1 - 5 - 25 - 100 минут

5.2.2 Версия В

Если для приготовления 2 буханок хлеба 2 пекарям требуется 2 минуты, сколько времени потребуется 50 пекарям, чтобы приготовить 50 буханок?

1 - 2 - 25 - 50 минут

5.3.1 Версия А

В озере есть плот из лилий. Каждый день плот удваивается в размере. Если для покрытия всего озера потребуется всего 24 дня, сколько времени потребуется, чтобы плот покрыл половину озера?

2 - 4 - 12 - 23 дня

5.3.2 Версия В

Контейнер заполняется камнями. Количество камней удваивается с каждым днем. Контейнер заполнится через 10 дней. Сколько дней потребуется, чтобы заполнить контейнер наполовину ?

2 - 4 - 5 - 9 дней

5.4.1 Версия А

В ящике черные и коричневые непарные носки смешаны в соотношении 4:5. Сколько непарных носков вам нужно вынуть наугад, не глядя, чтобы получить пару одинакового цвета?

1 - 2 - 3 - 4 - 5 носков

5.4.2 Версия В

В сумке зеленые и голубые стеклянные шарики смешаны в соотношении 6:8. Сколько шариков Вам нужно вынуть наугад, не глядя, что бы получить два одинаковых цвета?

2 - 3 - 4 - 5 - 6 шариков

5.5.1 Версия А

Два поезда, расположенные на расстоянии 50 км друг от друга, едут по направлению друг к другу со скоростью 25 км/ч (километров в час) каждый. Когда поезда тронулись, птица полетела от одного поезда ко второму поезду. По достижении второго поезда, птица развернулась и полетела к первому поезду и так далее, пока поезда не встретились. Если птица летит со скоростью 60 км/ч, сколько километров пролетит птица до того, как поезда встретятся?

15 км - 30 км - 45 км - 60 км

5.5.2 Версия В

Два пешехода, расположенные на расстоянии 10 км друг от друга, направляются друг к другу со скоростью 5 км/ч. Когда пешеходы начали движение, собака побежала от одного пешехода ко второму. По достижении второго пешехода, собака повернула и побежала обратно по направлению к первому и так далее, пока пешеходы не встретились. Если собака бежит со скоростью 20 км/ч, сколько километров должна пробежать собака прежде чем пешеходы встретятся?

2 - 5 - 10 - 20 км

5.6.1 Версия А

Покупатель покупает лошадь за 60 фунтов, продает ее за 70 фунтов стерлингов. Затем, покупает ее за 80 фунтов и продает за 90 фунтов. Сколько он заработал?

£ 10 - £ 20 - £ 30 - £ 40

5.6.2 Версия В

Дама покупает акции компании на фондовом рынке за 600 долларов и продает их за 700 долларов. Затем, покупает их за 800 долларов и продает их за 900 долларов. Сколько она заработала?

\$ 100 - \$ 200 - \$ 300 - \$ 400

5.7.1 Версия А

Полиция была убеждена, что один из людей А, В, С или D совершил преступление. Каждый из подозреваемых, в свою очередь, сделал заявление. Но только одно из четырех утверждений было правдой:

- А сказал: «Я не делал этого».
- Б сказал: «А лжёт».
- С сказал: «В лжёт».
- Д сказал: «В сделал это».

Кто говорит правду? Кто совершил преступление?

Правду: А - В - С - D

Преступление: А - В - С – D

5.7.2 Версия В

Судья знает, что один из людей А, В, С или D придумал изобретение. Каждый из них, в свою очередь, сделал заявление. Но только одно из четырех утверждений было правдой.

- А заявил: «Я это придумал».
- Б заявил: «А лжёт».
- С заявил: «В лжёт».
- D заявил: «В этого не изобретал».

Кто говорит правду? Кто придумал изобретение?

Правда: А - В - С - D

Изобретение: А - В - С – D

5.8.1 Версия А

Вам показан набор из четырех карт, расположенных на столе, каждая из которых имеет цифру с одной стороны и цвет с другой стороны. Видимые стороны карт показывают 3, 8, красный и коричневый. Какую карту(ы) Вам нужно перевернуть, чтобы проверить теорию о том, что, если на карте изображено четное число с одной стороны, то ее обратная сторона красного цвета? Постарайтесь использовать наиболее эффективный способ, перевернув наименьшее количество карт, которое необходимо. Отметьте необходимые карты: 3 - 8 - красный – коричневый

5.8.2 Версия В

Вам показан набор из четырех карт, расположенных на столе, каждая из которых имеет цифру с одной стороны и цвет с другой стороны. Видимые стороны карт показывают 5, 6, зеленый и желтый. Какую карту(ы) вы должны перевернуть, чтобы проверить гипотезу о том, что, если карта показывает нечетное число с одной стороны, то ее обратная сторона зеленого цвета? Постарайтесь использовать наиболее эффективный способ, перевернув наименьшее количество карт которое необходимо. Отметьте необходимые карты: 5 - 6 - зеленый – желтый

Japanese

Hypothesis Testing

1.1.1 Aバージョン

ある高校生が進学先に A 大学、B 大学のどちらかを選ぶことになった。彼には、A 大学、B 大学の両方に、自分に似た価値観と能力を持っている友人が数名ずついた。A 大学にいる友人はみな教育面と人間関係面の両方において A 大学を気に入っており、B 大学の友人はみな教育面と人間関係面の両方において B 大学には問題があると感じていた。彼が両大学を一日ずつ見学した結果、彼の得た印象は友人たちから聞いていたものと正反対であった。

彼はどちらの大学を選んだほうがよいだろうか。

- A 大学
- B 大学

1.1.2 Bバージョン

ある父親と母親が第一子の入園先に A 幼稚園、B 幼稚園のどちらかを選ぶことになった。その夫婦には、A 幼稚園、B 幼稚園の両方に、自分たちに似た価値観と考え方をもち、またそこに子供を通わせている友人が数名ずついた。子供を A 幼稚園に通わせている友人はみな教育面と人間関係面の両方において A 幼稚園を気に入っており、子供を B 幼稚園に通わせている友人はみな教育面と人間関係面の両方において B 幼稚園には問題があると感じていた。その夫婦が両幼稚園を数時間ずつ見学した結果、彼らの得た印象は友人たちから聞いていたものと正反対であった。

その夫婦はどちらの幼稚園を選んだほうがよいだろうか。

- A 幼稚園
- B 幼稚園

1.2.1 Aバージョン

キングストン市の警察署長は 1 年半署長の職に就いているが、評判がよくない。彼は政治に熱心であり、また市長の古くからの仲間でもある。彼は署長に任命された時点では警察を運営した経験がほとんどなかった。市長は最近、市民に対して「彼が署長に就任してから犯罪率が 12%減少した」と公表し彼を弁護している。次に挙げる証拠のうち、この警察署長が有能であるという市長の主張を最も弱めてしまうものはどれか。

- キングストンに隣接しており、規模もキングストンに近い二つの市の犯罪率は同じ時期に 18%減少した。
- キングストン市民に対して別の調査を行ったところ、犯罪発生件数は警察の報告よりも 40%高いという結果になった。
- 一般常識的には、警察署長が犯罪率を下げるためにできることはほとんどない。犯罪率はほとんどの場合が社会的および経済的な条件によって決まるものであるため、警察当局があればこれできるものではないからである。
- その警察署長は組織犯罪に関係したと知られている人物と業務上の接触があったことが判明した。

1.2.2 Bバージョン

ある高校のある教師は 1 年半その職に就いているが、評判がよくない。彼は校長の友人であり、その職に就いた時点では教師としての経験がほとんどなかった。校長は最近、学校関係者に対して「彼がこの学校の教師になってから、担当しているクラスの生徒の非行が 12%減少した」と公表し彼を弁護している。次に挙げる証拠のうち、この教師が有能であるという校長の主張を最も弱めてしまうものはどれか。

- この教師の担当ではない他の二つのクラスの生徒の非行率は同じ時期に 18%減少した。
- この学校の生徒に対して調査を行ったところ、非行率は校長の報告よりも 40%高いという結果になった。
- 一般常識的には、教師が非行を減らすためにできることはほとんどない。非行はほとんどの場合、生徒の気分や態度に起因するものだからである。
- その教師が、非行をした生徒に罰を与えたことがあることが分かった。

1.3.1 A バージョン

ある車が X 社の製品であるか Y 社の製品であるかを推定したいとする。いま、X 社の車、Y 社の車について次の 2 種類の情報を得ることができる。X 社、Y 社の車のうち何%が 100 km あたり 10 リットル以上のガソリンを消費するかという情報と、これまでに X 社、Y 社の車のうち何%が購入後 2 年間無故障であったかという情報である。

ある車(X 社または Y 社の製品)は 100 km あたり 10 リットル以上のガソリンを消費し、購入後 2 年間無故障であった。また X 社の車のうち 65%が 100 km あたり 10 リットル以上のガソリンを消費することが判っている。

- 次に挙げる 3 つの情報のうち、どれか 1 つを得られるとする。この車が X 社、Y 社のどちらの製品であるかを判断するのに役立つ情報はどれか。
- Y 社の車の何%が 100 km あたり 10 リットル以上のガソリンを消費するか。
- これまで、X 社の車のうち何%が最初の 2 年間無故障であったか
- これまで、Y 社の車のうち何%が最初の 2 年間無故障であったか

1.3.2 B バージョン

あるノートパソコンが X 社の製品であるか Y 社の製品であるかを推定したいとする。いま、X 社のノートパソコン、Y 社のノートパソコンのそれぞれについて次の 2 種類の情報を得ることができる。X 社、Y 社のノートパソコンのうち何%が 10 時間以上バッテリーがもつかという情報と、これまでに X 社、Y 社のノートパソコンのうち何%が購入後 2 年間無故障であったかという情報である。

あるノートパソコン(X 社または Y 社の製品)は 10 時間以上バッテリーがもち、購入後 2 年間無故障であった。また、X 社のノートパソコンのうち 65%が 10 時間よりも長くバッテリーがもつ。

- 次に挙げる 3 つの情報のうち、どれか 1 つを得られるとする。このノートパソコンが X 社、Y 社のどちらの製品であるかを判断するのに役立つ情報はどれか。
- Y 社のノートパソコンの何%が 10 時間以上バッテリーがもつか。
- これまで、X 社のノートパソコンのうち何%が最初の 2 年間無故障であったか
- これまで、Y 社のノートパソコンのうち何%が最初の 2 年間無故障であったか

1.4.1 A バージョン

マクスウェル氏は大学教授とビジネスマンだけが招待されるパーティーに出席した(招待された客は大学教授かビジネスマンのどちらかであり、両方の肩書を持っている者はいなかった)。マクスウェル氏については、彼がベアーズクラブのメンバーであるということだけが判っている。いま、あなたは以下の 1.2.3.4 の問いのいずれかに対する答えを元にマクスウェル氏が大学教授である確率を推定することを求められている。また各質問がどの程度判定に関係があるか(それに対する答えがマクスウェル氏が大学教授である可能性を判定する際に役立つか)を調べなければならない。各質問があなたの判定に関係しているかしていないかを答えよ。

- パーティーに出席している人の何%が大学教授であるか?
- ベアーズクラブのメンバーの何%がパーティーに出席しているか。
- パーティーに参加している大学教授の何%がベアーズクラブのメンバーであるか。
- パーティーに参加しているビジネスマンの何%がベアーズクラブのメンバーであるか。

1.4.2 B バージョン

クロウ夫人はアメリカ人観光客とカナダ人観光客だけが参加するガイドツアーに参加した(参加者はアメリカ国籍かカナダ国籍のどちらかを持っており、両方の国籍を持つ者はいなかった)。クロウ夫人については、彼女が ABC ホテルの宿泊客であるということだけが判っている。いま、あなたは以下の 1.2.3.4 の問いのいずれかに対する答えを元にクロウ夫人がアメリカ人である確率を推定することを求められている。また各質問がどの程度判定に関係があるか(それに対する答えがクロウ夫人がアメリカ人観光客である可能性を判定する際に役立つか)を調べなければならない。各質問があなたの判定に関係しているかしていないかを答えよ。

- ガイドツアーに参加している人の何%がアメリカ人であるか？
- ABC ホテルの宿泊客の何%がガイドツアーに参加しているか。
- ガイドツアーに参加しているカナダ人観光客の何%が ABC ホテルの宿泊客であるか。
- ガイドツアーに参加しているアメリカ人観光客の何%が ABC ホテルの宿泊客であるか。

1.5.1 A バージョン

リンダは 31 歳独身ではっきりとものを言う賢い女性である。彼女は大学で哲学を専攻していた。また学生時代には差別問題と社会正義問題に深く取り組み、反核デモにも参加した。

次の 2 つのうち、リンダはどのような女性である確率が高いか。

- リンダは銀行員である
- リンダは銀行員であり、かつフェミニスト運動に熱心である。

1.5.2 B バージョン

ジャックは 45 歳の既婚者で子供が 4 人いる。彼は穏健で注意深い、活動的でもある。彼は政治や社会問題に興味がなく、自由な時間のほとんどを日曜大工、ヨット(帆船)、数学パズルなど数々の趣味に費やしている。

次の 2 つのうち、ジャックはどのような男性である確率が高いか。

- ジャックは技師である。
- ジャックは技師であり、かつ地域のウォータースポーツクラブで活動している。

1.6.1 バージョン

あなたは新しい車を購入しようとしている。今日あなたは A 社の車か B 社の車のどちらを購入するかを決めなければならない。あなたの選択基準はその車の耐用年数がどのくらいであるかという点だけである。何百台もの車を調べた消費者調査の結果によると、A 社の車のほうが耐用年数が長かった。しかしまさに昨日、あなたの近所の人 A 社で買った新車が壊れたと言っていた。あなたは A 社、B 社どちらの車を購入するか。

- A 社の車
- B 社の車

1.6.2 B バージョン

あなたは新しいコンピューターを購入しようとしている。今日あなたは A 社のコンピューターか B 社のコンピューターのどちらを購入するかを決めなければならない。あなたの選択基準はそのコンピューターの耐用年数がどのくらいであるかという点だけである。何百台ものコンピューターを調べた消費者調査の結果によると、A 社のコンピューターのほうが耐用年数が長かった。しかしまさに昨日、あなたの近所の人 A 社で買った新しいコンピューターが壊れたと言っていた。あなたは A 社、B 社どちらのコンピューターを購入するか。

- A 社のコンピューター
- B 社のコンピューター

Verbal Reasoning

アナロジー

A1

けちな : 貧乏性の :: 同質の : _____

- 贅沢な
- 不親切な
- よく似た
- 人なつっこい

B1

詩: 韻 :: 哲学 : _____

- イメージ
- 音楽
- 法律
- 理論

A2

繕う : 布製品 :: 編集する : _____

- つぎあて
- 修理
- 原稿
- 一時しのぎ

B2

パン粉: パン :: _____ : 分子

- 破片
- 原子
- 跡
- イオン

A3

マウス : _____ :: フラッシュ : カメラ

- ハムスター
- コンピューター
- コード
- デザート

B3

_____ : 抜粋 :: 練習 : 訓練

- 例外
- 一節
- 日課
- 原因

A4

濡らす : _____ :: 冷やす : 凍らせる

- 水
- 浸す
- オープン
- 育てる

B4

リボン : _____ :: クリーム : ケーキ

- プレゼント
- 切る
- 蝶ネクタイ
- タイプライター

A5

_____ : 手首 :: ベルト : 腰

- 腕
- 手
- 曲げる
- ブレスレット

B5

科学者：実験 :: _____：劇

- ビーカー
- リハーサル
- 俳優
- 実験室

A6

辞書：定義 :: _____：地図

- 方向
- 南
- 地図帳
- 経度

B6

押しピン：_____ :: ハンガー：コート

- 釘
- ポスター
- 壁
- ハンマー

A7

珍しい：新規 :: _____：標準

- 変な
- 斬新な
- よくある
- 詩

B7

_____：大学 :: 整備士：自動車整備工場

- 本
- 勉強
- 教授
- エンジン

A8

ドラム：楽器 :: ドリル：_____

- ハンマー
- オープン
- 工具
- クレッシェンド

B8

頂上：山 :: _____：家

- 最大化する
- 屋根
- 玄関
- 小屋

A9

_____：降雨 :: 結露：湿気

- 浸食
- 雲
- 地面
- 天気予報

B9

_____ : コマ :: モザイク : タイル

- 映画
- 彫刻
- 浴室
- 装飾

Judging Likelihood and Probability

3.1.1 A バージョン

スカッシュは 9 点制か 15 点制のどちらかを選ぶことができる。A が B よりも強いプレーヤーである場合、どちらで行った方が A の勝つ確率が高くなるか。

- 9 点制
- 15 点制

3.1.2 B バージョン

卓球は 5 セット制と 9 セット制のどちらかを選ぶことができる。A が B よりも強いプレーヤーである場合、どちらで行った方が A の勝つ確率が高くなるか。

- 5 セット制
- 9 セット制

3.2.1 A バージョン

大リーグのシーズンが始まって 2 週間経つと、新聞紙に打率上位 10 名が発表され始める。シーズン開始から 2 週間の時点では、首位打者の打率が約 0.450 であることは珍しくない。しかし、大リーグ史上においてシーズン終わりに打率が 0.450 に達していた打者はいない。なぜこのようになるのか。その理由として考えられるものを以下から一つ選べ。高い打率を記録している打者には、ピッチャーが一層力投するようになるから。

- ピッチャーはシーズンが進むにつれて調子上がる傾向にあるため打者の三振が出やすくなり、打率が下がるから。
- シーズンの初めにある打者の打率が高いのは単に彼の運が良かっただけである可能性があるが、シーズン全体を通して見ると、打率はその打者の実力に沿った数字で現れるから。
- シーズン初めに打率の高かった打者にはその記録を維持しなければならないというストレスがかかり、プレーに悪影響が出るから。
- よい打率を記録している打者には打ちやすい球が来なくなり、ピッチャーは四球を気にせずストライクゾーンぎりぎりの球を投げるようになるから。

3.2.2 B バージョン

ある学校では、学年が始まって 2 週間経つと教師の一人がその年の成績上位 10 人の予想を試みる。学年の開始から 2 週間の時点では成績トップの生徒が 100 点満点中平均 92 点を取ることは珍しくない。しかしその学校が始まって以来、学年を通して平均 92 点に達した生徒はいない。なぜこのようになるのか。その理由として考えられるものを以下から一つ選べ。

ある生徒が高い点を取ろうとしていることが分かると、教師たちはこの生徒の回答の採点基準を厳しくするから。

- 学年の後半になるにしたがって、勉強しなければならないことが増えて試験が難しくなる傾向があり、高得点が取れなくなって平均点が下がるから。
- 学年の初めにある生徒の平均点が高いのは単にその生徒の運が良かっただけである可能性があるが、学年全体を通して見ると、点数がその生徒の知識量や能力に沿った数字で現れるから。
- 学年初めに平均点の高かった生徒にはその点数を維持しなければならないというストレスがかかり、テストの結果に悪影響が出るから。
- 高得点を取ろうとしている生徒には簡単な問題が出されなくなり、教師はその生徒の実力を測るためにより難易度の高い問題を出すようになるから。

3.3.1 A バージョン

あるスロットマシンでは、およそ 10 回に 1 回の確率で当たりが出る。しかしジュリーの場合、最初の 3 回がすべて当たりだった。4 回目に彼女が当たりを出す確率はどのくらいか。

10 回中 ___ 回(1 から 10 までのいずれかで回答)

3.3.2 B バージョン

ある宝くじは、およそ 20 枚に 1 枚の確率で当たる。しかしジャックの場合は最初の 3 枚がすべて当たりだった。4 枚目に彼が当たりを引く確率はどのくらいか。

20 枚中 ___ 枚

3.4.1 A バージョン

歪みのない(表と裏が半々の確率で出る)コインを使ってコイントスをする。いま、5 回連続で表が出ている。6 回目のコイントスの結果を予想せよ。

- 表よりも裏の出る確率が高い。
- 裏よりも表の出る確率が高い。
- 6 回目の試行においても裏と表は同じ確率で出る。

3.4.2 B バージョン

歪みのない(すべての目が同じ確率で出る)さいころを投げる s 。いま、5 回連続で 6 の目が出ている。6 回目にさいころを投げた結果を予想せよ。

- 6 以外の目が出る確率が高い。
- 6 の目が出る確率が高い。
- 6 回目の試行においてもすべての目は同じ確率で出る。

3.5.1 A バージョン

2 つの盆それぞれに白いビー玉と黒いビー玉が載っている。大きい方の盆には 100 個のビー玉が、小さい方の盆には 10 個のビー玉が載っており、どちらの盆においてもビー玉は積み重なっていない。ここで、どちらかの盆から目を瞑ってビー玉を 1 個取り、黒のビー玉を取ると 2 ドル貰えるとする。小さい方の盆には黒いビー玉が 1 個と白いビー玉が 9 個、大きい方の盆には黒いビー玉が 8 個と白いビー玉が 92 個載っているとする。この状況で実際にビー玉を取るならば、あなたはどちらの盆からビー玉を取るか。

- 大きい盆
- 小さい盆

3.5.2 B バージョン

2 つの鞆それぞれに 10 ドル紙幣と 20 ドル紙幣が入っている。大きい方の鞆には 100 枚の紙幣が、小さい方の鞆には 10 枚の紙幣が入っている。紙幣は鞆の中で混ざっている。ここで、どちらかの鞆から目を瞑って紙幣を 1 枚取り、取った紙幣を貰えるとする。小さい方の鞆には 20 ドル紙幣が 1 枚と 10 ドル紙幣が 9 枚、大きい方の鞆には 20 ドル紙幣が 8 枚と 10 ドル紙幣が 92 枚入っているとする。この状況で実際に紙幣を取るならば、あなたはどちらの鞆から紙幣を取るか。

Argumentation Analysis

有効な
無効

A1

- 安い煙草はない。
- 中毒性のあるものの中には安いものがある。
- したがって、中毒性のあるものの中には煙草ではないものがある。

B1

- 建物の中には木造のものがある。

- 木造の高層ビルはない。
- したがって、建物の中には高層ビルではないものがある。

A2

- オフィス用品の中にはコンピューターでないものがある。
- 電化製品の一部はオフィス用品である。
- したがって、電化製品の一部はコンピューターではない。

B2

- 飲み物の中にはノンアルコールのものがある。
- ノンアルコールの飲み物は酒ではない。
- したがって、飲み物の中には酒でないものがある。

A3

- 椅子はテーブルではない。
- 家具には椅子が含まれている。
- したがって、家具にはテーブルではないものがある。

B3

- 魚の中には食べられないものがある。
- マスは食べられる魚である。
- したがって、マスではない魚もある。

B4

- 中毒性のあるものには安いものはない。
- 煙草の中には安いものがある。
- したがって、中毒性のあるものの中には煙草でないものがある。

A4

- 木造建造物の中には高層ビルがある。
- どの建物も木造建築ではない。
- したがって、建物の中には高層ビルではないものがある。

B5

- 電化製品の中にはオフィス機器でないものがある。
- オフィス機器の一部はコンピューターである。
- したがって、電化製品の中にはコンピューターでないものがある。

A5

- ノンアルコールの飲み物は酒である。
- どの飲み物もノンアルコールではない。
- したがって、飲み物の中には酒でないものがある。

B6

- 家具が椅子であることはない。
- 椅子はテーブルである。
- したがって、家具の中にはテーブルではないものがある。

A6

- マスが食用には適さない。
- 魚は食用である。
- したがって、魚の中には食用に適さないものがある。

A7

- 工具には手動でないものがある。
- ドライバーは手動の工具である。

- したがって、ドライバーの中には工具でないものがある。

B7

- 楽器には管楽器であるものがある。
- 管楽器はバイオリンではない。
- したがって、バイオリンには楽器でないものがある。

A8 // B8

- 緑色のものには草ではないものがある。
- 植物には緑色のものがある。
- したがって、草には植物でないものがある。

B9

- ドライバーは手動の工具ではない。
- 工具が手動であることはない。
- したがって、ドライバーには工具でないものもある。

A9

- 管楽器はバイオリンである。
- どの楽器も管楽器ではない。
- したがって、バイオリンの中には楽器でないものがある。

Problem Solving

5.1.1 A バージョン

バット 1 本とボール 1 個で合計 110 セントである。バットはボールよりも 100 セント高い。ボールの値段はいくらか。

5 セント - 10 セント - 55 セント - 100 セント

5.1.2 B バージョン

パン 1 つとリンゴ 1 個で合計 220 セントである。パンはリンゴより 200 セント高い。リンゴの値段はいくらか。

10 セント - 20 セント - 110 セント - 200 セント

5.2.1 A バージョン

機械を 5 台同時に動かしてプラスチック製の部品を 5 つ作るのに 5 分かかるとする。このとき、この機会を 100 台同時に動かしてプラスチック製の部品を 100 個作るのにはどれくらいの時間がかかるか。

1 分 - 5 分 - 25 分 - 100 分

5.2.2 B バージョン

2 人のパン職人が同時に働いてパンを 2 つ作るのに 2 分かかるとする。このとき、パン職人が 50 人同時に働いてパンを 50 個作るのにはどれくらいの時間がかかるか。

1 - 2 - 25 - 50 分

5.3.1 A バージョン

大きな池の一部分がスイレンに覆われている。その覆われている部分の面積は 1 日たつごとに 2 倍の大きくなる。池全体が覆われるのには 24 日かかるとすると、池の面積の半分がスイレンで覆われるためには何日かかるか。

4 日 - 6 日 - 12 日 - 23 日

5.3.2 B バージョン

コンテナに石を入れていくとする。石の量は毎日 2 倍になる。コンテナは 10 日でいっぱいになる。石の量がコンテナの半分になるには何日かかるか。

2 日 - 4 日 - 5 日 - 9 日

5.4.1 A バージョン

引き出しに黒い靴下と茶色の靴下が 4:5 の割合で片足ずつ入っている。靴下をひとつずつ取り出すとき、目をつぶって同じ色の靴下を確実に 2 つ取り出すにはいくつ取り出す必要があるか。

1 つ - 2 つ - 3 つ - 4 つ - 5 つ

5.4.2 B バージョン

靴に緑のビー玉と青いビー玉が 6:8 の割合で入っている。ビー玉をひとつずつ取り出すとき、目をつぶって同じ色のビー玉を確実に 2 個取り出すには何個取り出す必要があるか。

2 - 3 - 4 - 5 - 6 個

5.5.1 A バージョン

2 つの列車が互いに 50 km 離れたところからそれぞれ時速 25 km で近づいている。列車が発車すると同時に 1 羽の鳥が一方の列車の先頭からもう一方の列車の方に飛んでいく。もう一方の列車に到達すると同時にその鳥は元の列車の方に飛んでゆき、2 つの列車が出会うまで同じことを繰り返す。この鳥が時速 60 km で飛ぶとする。2 台の列車が出会うまでに合計何 km 飛んだことになるか。

15 km - 30 km - 45 km - 60 km

5.5.2 B バージョン

2 人が互いに 10 km 離れたところからそれぞれ時速 5 km で近づいている。彼らが歩き始めると同時に 1 匹の犬が片方の人のところからもう片方の人の元へ走っていく。もう片方の人のところに到達すると同時にその犬は元の人のもとに走っていき、2 人が出会うまで同じことを繰り返す。この犬が時速 20 km で走るとする。2 人が出会うまでにこの犬は合計何 km 走ったことになるか。

2 - 5 - 10 - 20 km

5.6.1 A バージョン

ある男性が馬を 60 ポンドで買い、それを 70 ポンドで売り、さらにそれを 80 ポンドで買い戻し、最後に 90 ポンドで売った。彼は何ポンド儲けたか。

10 ポンド - 20 ポンド - 30 ポンド - 40 ポンド

5.6.2 B バージョン

ある女性がある会社の株を 600 ドルで買い、それを 700 ドルで販売し、さらにそれを 800 ドルで買い戻し、最後に 900 ドルで販売した。彼女は何ポンド儲けたか。

100 ドル - 200 ドル - 300 ドル - 400 ドル

5.7.1 A バージョン

警察は容疑者 A.B.C.D のうち誰かが犯罪を犯したという確証をつかんでいる。4 人の容疑者はそれぞれ次のように供述したが、本当のことを言っているのはこのうち一人だけである。

- A の供述「私はやっていません。」
- B の供述「A はうそをついています。」
- C の供述「B はうそをついています。」
- D の供述「犯人は B です。」

真実を言っているのは誰か。また、犯人は誰か。

真実を言っている人: A - B - C - D

犯人 : A - B - C - D

5.7.2 B バージョン

裁判官は A.B.C.D のうち誰かが証拠を捏造していることを知っている。4 人はそれぞれ次のように証言したが、本当のことを言っているのはこのうち一人だけである。

- A の証言「私が捏造しました。」
- B の証言「A はうそをついています。」

- Cの証言「Bはうそをついています。」
- Dの証言「Bは捏造していません。」

真実を言っているのは誰か。また、証拠を捏造したのは誰か。

真実を言っている人: A - B - C - D

捏造した人 : A - B - C - D

5.8.1 Aバージョン

テーブルに4枚のカードが並べられている。どのカードも片面には数字が書かれており、もう片方の面には色が塗られている。いま表になっているのは「3」「8」「赤色」「茶色」の面である。あるカードに書かれている数字が偶数ならば、もう片方の面は赤色であるという命題が常に真であることを確かめるためにひっくり返さなければならないカードはどれか(2枚以上である場合もある)。ひっくり返すカードの数が最小になるように答えよ。

該当するすべてにしるしをつけよ。

3-8-赤-茶色

5.8.2 Bバージョン

テーブルの上に4枚のカードが並べられている。どのカードも片面には数字が書かれており、もう片方の面には色が塗られている。いま表になっているのは「5」「6」「緑色」「黄色」の面である。あるカードに書かれている数字が奇数ならば、もう片方の面は緑色であるという命題が常に真であることを確かめるためにひっくり返さなければならないカードはどれか(2枚以上である場合もある)。ひっくり返すカードの数が最小になるように答えよ。

該当するすべてにしるしをつけよ。

5-6-緑-黄色

Chinese

Hypothesis Testing

1.1.1 Version A

一名高中生要从大学 A 和大学 B 中进行选择。在这两所大学里有一些和他价值观和能力相近的朋友。所有在 A 大学的朋友都喜欢这个学校的教育方式和社会背景；所有在 B 学校的朋友认为在这两方面都存在一些问题。这个学生有一天参观了这两所大学，然而他对这两所大学的印象正和他的朋友相反。

这个高中生应该选择哪所大学？

- 大学 A
- 大学 B

1.1.2 Version B

一对父母要为他们的第一个孩子在 A 幼儿园和 B 幼儿园中选择一所幼儿园。这对夫妇有几个在价值观和信仰上和他们相似的朋友，他们的孩子分别在这两个幼儿园。所有孩子在 A 幼儿园的朋友都喜欢这个幼儿园的教育方式和社会背景。所有孩子在 B 幼儿园的朋友认为在这两方面都存在一些问题。这对父母用几个小时参观了这两所幼儿园，然而他们对幼儿园的印象和他们朋友所说的正相反。

这对父母应该选择哪所幼儿园？

- 幼儿园 A
- 幼儿园 B

1.2.1 Version A

金斯敦市里有一位任职一年半不受欢迎的警察局长。他热衷于政治活动并且是市长的同事，而且当他被任命的时候几乎没有警务行政经验。市长最近在公共场合维护这位局长，宣布自从他任职以来，犯罪率下降了 12%。下面的哪个证据最能削弱市长的关于警察局长有能力的发言。

在同一时期，在地理位置上和面积上和金斯敦最接近的两个城市的犯罪率下降了 18%

关于金斯敦的一份独立调查显示，这份调查中的犯罪率比警察局报告的数据多了 40%

常识告诉我们在减少犯罪率问题上，警察局长能起的作用很小。社会和经济因素作用要超过了官方作用

警察局长被发现和一些众所周知的卷入集团犯罪的人有个人联系。

1.2.2 Version B

一所高中有一位任职一年半不受欢迎的老师。他是校长的朋友，并且在他被任命时在教学方面几乎没有什么经验。校长最近在全校师生面前维护这个老师，公开宣称在他任职期间，他的学生的犯过错率下降了 12%。下面哪个证据最能削弱校长的关于这个老师很能干的发言？

在同一时期，没有被这个老师教的两个班级学生的犯过错率下降了 18%

一份学生调查显示，和校长报告相比，这份调查中的学生犯过错行为多了 40%。

常识表明在减少学生犯过错方面一个老师能做的很少。这些更多是取决于学生的心理状况和行为。

人们发现这个老师体罚过犯错的学生。

1.3.1 Version A

现有自动车是 X 牌， Y 牌。现提供以下信息， X 牌和 Y 牌汽车每 100km 的耗油量超过 10L 的百分比， x 牌和 Y 牌汽车在使用头两年没有主要机械问题的百分比。

现有一辆汽车（X 牌或者 Y 牌）的耗油量超过了 10 l/100 km，并且在使用的头两年里没有主要的机械问题。65% X 牌汽车的耗油量超过了 10 l/100 km。

以下三个条件中，哪个能够帮助你决定这辆汽车的牌子？

- Y 牌子汽车的耗油量超过 10 l/100 km 的百分比
- X 牌子汽车在使用头两年里面没有主要技术问题的百分比
- Y 牌子汽车在使用头两年里面没有主要技术问题的百分比

1.3.2 Version B

现有笔记本电脑 X 牌, Y 牌。现提供一下信息, X 牌和 Y 牌笔记本电脑的电池使用时间超过十个小时的百分比, X 牌和 Y 牌笔记本电脑在使用的头两年里不会出现主要技术方面问题的百分比。

现有一台笔记本电脑 (X 牌或者 Y 牌) 的电池使用时间超过十个小时并且在购买后的头两年里没有出现主要的技术方面的问题。65% x 牌笔记本电脑电池的使用时间超过了 10 小时。

以下三个条件中, 哪个能帮助你决定这台电脑的牌子?

- Y 牌电脑中电池使用时间超过 10 小时的电脑的百分比
- X 牌电脑中在购买后的两年内没有出现主要技术问题的电脑的百分比
- Y 牌电脑中在购买后的两年内没有出现主要技术问题的电脑的百分比

1.4.1 Version A

麦克斯韦先生参加了一个聚会, 这个聚会只邀请了大学教授和商业上的高级行政人员 (然而来宾只能有其中一种身份, 不能同时具有两种身份)。您知道唯一一件关于麦克斯韦先生的事情是他是一个啤酒俱乐部的会员。请您判断麦克斯韦先生的身份是一位大学教授的可能性。一个相关问题可以对您的判断提供帮助。请分别判断下列提问是否和您的任务相关。

不相关:

- 参加聚会的是大学教授的百分比是多少?
- 参加聚会的啤酒俱乐部的会员百分比是多少?
- 参加聚会的人中大学教书是啤酒俱乐部的会员的百分比是多少?
- 参加聚会的人中商业上的高管是啤酒俱乐部的会员的百分比是多少?

1.4.2 Version B

克罗小姐参加了一次旅行, 这次旅程中只有美国和加拿大的游客 (然而游客只能有其中一个国籍, 不能是美国和加拿大双重国籍)。您知道的关于克罗小姐的唯一一件事情是她住在 ABC 旅店。请您判断克罗小姐是一名美国游客的可能性。一个相关问题可以对您的判断提供帮助。请分别判断下列提问是否和您的任务相关。

不相关:

- 在这个观光团中美国人的比例是多少?
- 旅馆 ABC 的客人中有多少百分比的人参加了这个旅行团?
- 加拿大游客中有多少百分比的人是旅馆 ABC 的客人?
- 美国游客中有多少百分比的人是旅馆 ABC 的客人?

1.5.1 Version A

琳达 31 岁, 单身, 心直口快很开朗。她主修哲学。作为一个学生, 她非常关心歧视问题和社会公正问题, 并且还参加过反对使用核武器的游行。

下面两个选项中哪个更有可能性

- 琳达是一个银行出纳员
- 琳达是一个银行出纳员并且在男女平等主义活动方面很活跃

1.5.2 Version B

杰克 45 岁, 他已经结婚并且有四个孩子。他谨慎行事且有雄心。他对政治和社会问题不感兴趣, 并且他的大部分自由时间都花在他的一些爱好上了, 包括在家里做木工, 帆船运动和数字猜谜游戏。

下面两个选项中哪个更有可能?

- 杰克是一个工程师
- 杰克是一个工程师并且经常去一家当地的水上运动俱乐部

1.6.1 Version A

您想从 A 公司或 B 公司买一辆新车。您只考虑汽车的预期寿命。您从消费者调查报告中了解到几百辆车的调查结果显示 A 牌子汽车的使用寿命更长。昨天你的一个邻居告诉你他新买的 A 牌子的汽车坏掉了。你会买哪个牌子的汽车?

- 汽车 A
- 汽车 B

1.6.2 Version B

你想从电脑 A 或电脑 B 中购入一台电脑。您只考虑电脑的预期寿命。你从一份消费者调查中了解到几百台电脑的调查结果显示 A 电脑的预期寿命更长。就在昨天你的一个邻居告诉你他的 A 牌子的新电脑坏了。你会买哪个电脑？

- 电脑 A
- 电脑 B

Verbal Reasoning

文字推理
类推

A1

_____ : 马 :: 登上 (船, 车或者飞机) : 火车

- 马棚
- 蹄铁
- 骑
- 骑上

B1

诗: 韵律 :: 哲学: _____

- 意象, 比喻
- 音乐
- 法
- 理论

A2

缝补 : 衣物 :: 编辑 : _____

- 绣花
- 修理
- 手稿, 原稿
- 临时替代品

B2

按字母排序的: _____ :: 按数字排序的: 文件夹

- 种类
- 部分
- 清单
- 次序

A3

鼠标 : _____ :: 闪光灯: 相机

- 大老鼠
- 计算机
- 绳
- 甜点

B3

垫子 : 沙发 :: 架子 : _____

- 壁架
- 书柜
- 储藏柜
- 框架

A4

湿润 : _____ :: 凉爽 : 冻结

- 水

- 湿透
- 烤箱
- 生长

B4

缎带: _____ :: 奶油: 蛋糕

- 礼物
- 剪掉
- 弓
- 打字机

A5

小猫: _____ :: 士兵: 军队

- 猫
- 一窝猫崽
- 小狗
- 猫叫

B5

科学家: 实验 :: _____ : 戏剧

- 面包师
- 排练
- 演员
- 实验室

A6

字典: 定义 :: _____ : 地图

- 方向
- 南方
- 地图集
- 经度

B6

图钉: _____ :: 挂衣钩: 外套

- 钉子
- 海报
- 墙
- 锤子

A7

与众不同的: 新奇的 :: _____ : 普遍的

- 古怪的
- 新颖的
- 通晓的
- 诗歌

B7

_____ : 大学 :: 园艺师: 花园

- 书
- 学习
- 教授
- 种植

A8

鼓: 乐器 :: 钻头: _____

- 锤子
- 烤箱
- 工具
- 乐谱

B8

山峰: 山脉 :: _____ : 房子

- 最大化
- 屋顶
- 走廊
- 平房

A9

_____ : 阵雨 :: 冷凝: 湿气

- 腐蚀
- 云
- 地面
- 预测

B9

_____ : 框架 :: 镶嵌工艺 : 瓷砖

- 电影
- 雕刻
- 浴室
- 装饰品

Judging Likelihood and Probability

判断的可能性和概率

3.1.1 Version A

壁球比赛中有 9 分制和 15 分制。如果 A 比 B 玩得好，哪种得分制度会使 A 赢得比赛的几率变大？

- 9 分制
- 15 分制

3.1.2 Version B

乒乓球比赛中有 5 局制和 9 局制。如果 A 的球技比 B 好，哪种比赛制度会使 A 赢得比赛的几率变大？

- 五局取胜
- 九局取胜

3.2.1 Version A

在棒球职业联赛赛季的前两周之后，一些报纸开始预测前十名打手的平均成绩。两周之后，领先打手的平均分大约为 450。然而当赛季结束后，联赛历史上没有出现过平均分 450 的打手。您认为原因是什么？选出一个答案：

当大家都期待打手打出高平均分时，投手给他投球时会更加努力。

投手在整个赛季会随着状态变好而投球技术提高。随着投手的提高，他们会更有可能将打手淘汰出局，因此打手的平均成绩下降。

一名打手在赛季初的高平均值可能只是运气。长期的赛季才能反映出打手的真实水平。

在赛季初被带上光环的选手在维持成绩时可能会倍感压力。这样的压力给他的球技带来负面影响。

当大家都期待打手打出高平均分时，投手会提高投球难度，往刁钻的方向投球，以使打手上垒。

3.2.2 Version B

在一个学期的前两周过后，一所学校的一个老师试着去预测年级前 10 名的学生。通常，2 周过后，拔尖的学生经常会得到 92% 的平均分。然而当学期结束后，学校历史上没有学生能达到 92% 的平均分数。您认为原因是什么？选出一个答案：

当大家期待一名学生得到高平均分时，老师在给这名学生的试卷打分的时会有更高的期待。期末时，随着内容的累积试卷难度加大，学生得到高分的可能性降低，因此学生的平均分下降。

一名学生的在学期初的高平均分可能只是比较幸运。更长的一学年才能真实反映学生的知识和能力水平。

学期初得到高平均分的学生在维持成绩时可能会倍感压力。这样的压力会给他的成绩带来负面影响。

当大家期待一名学生得到高平均分时，老师不会就简单的问题提问。相反，老师为了测试这个学生的能力而提高问题的难度。

3.3.1 Version A

人们在玩老虎机时每 10 次会赢一次，然而，朱莉在她玩的前三次里赢了。那么她在下次玩的时候赢的概率多少？

10 分之_____

3.3.2 Version B

当参加抽奖时，大约每 10 个抽奖券中会有一个人中奖。然而，杰克在他的前三张抽奖券中就中了奖。那么杰克下一次的抽奖时中奖概率是多少？

10 分之_____

3.4.1 Version A

投掷硬币时（一个正面朝上或反面朝上的机会是 50/50 的硬币），连续出现了五次正面。您认为第六次投掷时会是：

- 很有可能是反面而不是正面
- 很有可能是正面而不是反面
- 第六次正面和反面的概率是一样的

3.4.2 Version B

投掷筛子时（这个筛子每个数字出现的概率是六分之一），数字 6 接连出现了 5 次。您认为第 6 次投掷时会出现数字几：

- 更有可能出现不是 6 的数字
- 更有可能出现 6
- 第六次扔的时候 6 和其他数字出现的可能性是一样的

3.5.1 Version A

现有两个托盘，托盘里有黑白两色石子。大托盘有 100 颗石子，其中 8 颗黑色石子和 92 颗白色石子。小的托盘里有 10 颗石子，其中 1 颗黑色石子和 9 颗白色石子。在看不到托盘里面的情况下，从托盘里拿出 1 颗石子。如果你拿出的是黑色石子，你会赢得 2 美元。选择哪个托盘，赢的概率更大？

- 大托盘
- 小托盘

3.5.2 Version B

现有两个包，包里装着 10 美元和 20 美元的笔记本。小包里放着 10 个笔记本，其中 1 个 20 美元的笔记本和 9 个 10 美元的笔记本。大包里面放着 100 个笔记本，其中 8 个 20 美元的笔记本和 92 个 10 美元的笔记本。在看不到包里的情况下，拿出一个笔记本。您更愿意选择哪个包？

- 大包
- 小包

Argumentation Analysis

论证分析

请判断论证有效或无效

A1

- 没有香烟是便宜的
- 有些容易上瘾的东西是便宜的
- 因此，有些容易上瘾的东西不是香烟

B1

- 有些建筑是木制的建筑
- 没有木制建筑是摩天高楼
- 因此有些建筑不是摩天高楼

A2

- 有些办公室设备不是电脑
- 有些电器是办公室设备
- 因此有些电器不是电脑

B2

- 有些饮品是不含酒精的
- 不含酒精的饮品不是白酒
- 因此有些饮品不是白酒

A3

- 椅子不是桌子
- 有些家具是椅子
- 因此有些家具不是桌子

B3

- 有些鱼是不能吃的
- 鳟鱼是能吃的
- 因此有些鱼不是鳟鱼

B4

- 没有上瘾的东西是便宜的
- 有些香烟是便宜的
- 因此，有些容易上瘾的东西不是香烟

A4

- 有些木制楼是摩天高楼
- 没有楼是木制楼
- 因此有些楼房不是摩天高楼

B5

- 有些电器不是办公设备
- 有些办公设备是电脑
- 因此一些电器不是电脑

A5

- 不含酒精的饮料是酒
- 没有饮料是没有酒精的
- 因此有些饮料不是酒

B6

- 没有家具是椅子
- 椅子是桌子
- 因此有些家具不是桌子。

A6

- 没有鳟鱼是可以吃的
- 鱼是可以吃的
- 因此有些鱼不是鳟鱼

A7

- 有些工具不是手动的
- 螺丝刀是手动工具
- 因此有些螺丝刀不是工具。

B7

- 有些乐器是管乐器
- 管乐器不是小提琴
- 因此有些小提琴不是乐器。

A8 // B8

- 有些绿色的东西不是草。
- 有些植物是绿色的。
- 因此有些草不是植物。

B9

- 螺丝刀不是手动工具
- 没有工具是手动的
- 因此有些螺丝刀不是工具。

A9

- 管乐器是小提琴
- 没有音乐乐器是管乐器
- 因此有些小提琴不是音乐乐器

Problem Solving

解决问题

5.1.1 Version A

一个球杆和一个球一共花费 1.10 美元。杆比球多花费 1.00 美元。那么球花费多少钱?
5 美分 - 10 美分 - 55 美分 - 100 美分

5.1.2 Version B

一块面包和一个苹果一共花费 2.20 美元。面包比苹果多 2 美元。那么苹果花费多少钱?
10 美分 - 20 美分 - 110 美分 - 200 美分

5.2.1 Version A

如果做 5 个塑料制品零件需要 5 个机器花费 5 分钟，那么 100 个机器做 100 个塑料制品零件需要多久?
1 - 5 - 25 - 100 分钟

5.2.2 Version B

如果做 2 块面包需要 2 个面包师花费 2 分钟，那么 50 个面包师做 50 个面包需要多久?
1 - 2 - 25 - 50 分钟

5.3.1 Version A

湖里，有一片百合花丛。每天，这块花丛都成倍扩大。如果这块花丛把整个湖都覆盖上需要 24 天的话，当这个湖被覆盖了一半的时候需要多久?
- 4 - 12 - 23 天

5.3.2 Version B

往一个容器里面装石头。石头的数量每天都在成倍的增加。这个容器会在 10 天的时候被装满。当容器被装满一半时需要多少天？

1 - 4 - 5 - 9 天

5.4.1 Version A

在一个抽屉里以 4 比 5 的比例混有黑色和棕色的袜子。在看不到抽屉里面的情况下你必须拿出多少只的袜子才能拿出颜色一样的袜子？

1 - 2 - 3 - 4 - 5 只袜子

5.4.2 Version B

在一个袋子里以 6 比 8 的比例混有绿色和蓝色的珠子。在看不到袋子里面的情况下你必须拿出多少个珠子才能拿出颜色一样的珠子？

2 - 3 - 4 - 5 - 6 个珠子

5.5.1 Version A

两辆火车，在相距 50km 处以 25 km/h 的速度向对方行驶。当两辆火车出发时，一只鸟从一辆火车处飞向另一辆火车。当鸟和第二辆火车相遇后掉头飞向第一辆火车，如此反复直至两辆火车相遇。如果这只鸟的速度是 60 km/h，那么直至两辆火车相遇，鸟飞了多少公里？

15 km - 30 km - 45 km - 60 km

5.5.2 Version B

两名步行者在相距 10km 处以 5 km/h 的速度向对方行走。当两人开始行走时，一条狗从一名步行者处跑向另一名步行者。当这条狗和另一名步行者相遇时掉头跑向第一名步行者，如此反复直至两人相遇。如果这条狗的速度为 20 km/h，那么直到两人相遇狗跑了多少公里？

2 - 5 - 10 - 20 km

5.6.1 Version A

一个男人花了 60 法郎买了一匹马，后以 70 法郎卖出。然后又花 80 法郎买回来，最后卖了 90 法郎。他赚了多少钱？

£10 - £20 - £30 - £40

5.6.2 Version B

一位女士在股票市场花了 600 美元买了一个公司的股份，后以 700 美元的价格售出。然后又花 800 美元购入，并且最后卖了 900 美元，她赚了多少钱？

\$100 - \$200 - \$300 - \$400

5.7.1 Version A

警察确信在 A, B, C, D 中有人犯了罪。每位嫌疑人轮流做陈述，且四项陈述中只有一个是正确的。

- A 说：“我没有犯罪。”
- B 说：“A 在说谎。”
- C 说：“B 在说谎。”
- D 说：“B 犯的罪。”

谁在讲真话？谁犯了罪

真话: A - B - C - D

犯罪: A - B - C - D

5.7.2 Version B

法官知道 A, B, C, D 中的一个人做出了一个发明。每人轮流做陈述，且四项陈述中只有一个是正确的。

- A 说：“是我发明的。”
- B 说：“A 在说谎。”
- C 说：“B 在说谎。”
- D 说：“B 没有发明它。”

谁说的是真话？发明是谁做出的？

真话: A - B - C - D

发明: A – B – C – D

5.8.1 Version A

在桌子上有四张牌，每一张牌的一面为数字，另一面为颜色。四张牌以 3, 8, 红色, 棕色的顺序向上摆放。如需证明一张卡片在一面显示的是偶数，那么另一面是红色为真命题，您需要翻开哪张或哪几张牌？以最有效的方法，翻开最少的牌。

你认为有几个就标出几个: 3 – 8 – 红色 – 棕色

5.8.2 Version B

在桌子上有四张牌，每一张牌的一面为数字，另一面为颜色。四张牌以 5, 6, 绿色, 黄色的顺序向上摆放。。哪一张（或者哪几张）牌是你必须要翻过来如需证明如果一张卡片在一面显示的是奇数，那么另一面是绿色为真命题，您需要翻开哪张或哪几张牌？以最有效的方法，翻开最少的牌。

– 6 – 绿色 – 黄色

Turkish

Hypothesis Testing

1.1.1 A Versiyonu

Bir lise öğrencisi A ve B üniversitelerinden birisini seçmek zorundadır. Öğrencinin, her iki üniversitede de değerleri ve yetenekleri kendisinininkine benzeyen birçok arkadaşı vardır. A üniversitesindeki arkadaşlarının tümü üniversitelerini hem eğitim hem de sosyal açıdan beğenmektedir; B üniversitesindeki arkadaşlarının hepsi ise her iki açıdan da B üniversitesinde sorunlar olduğunu düşünmektedir. Öğrenci, her iki üniversiteyi birer gün ziyaret etmiş ve iki üniversite için de arkadaşlarının söylediklerinin tersine izlenimler edinmiştir.

(Sizce) Lise öğrencisi hangi üniversiteyi seçmelidir?

- A Üniversitesi
- B Üniversitesi

1.1.2 B Versiyonu

Bir anne ve baba, ilk çocuklarını göndermek için A ve B anaokullarından birisini seçmek zorundadır. Bu çiftin, değerler ve inanışlar açısından kendilerine benzeyen birçok arkadaşının çocukları A ve B anaokullarına gitmektedir. Çocuklarını A anaokuluna gönderen arkadaşları bu okulu hem eğitim hem de sosyal açıdan beğenmektedir. Çocuklarını B anaokuluna gönderen arkadaşları ise her iki açıdan B anaokulunda sorunlar olduğunu düşünmektedir. Çift her iki okulu birkaç saatliğine ziyaret etmiş ve iki anaokulu için de arkadaşlarının söylediklerinin tersine izlenimler edinmiştir.

(Sizce) Anne-baba hangi anaokulunu seçmelidir?

- A Anaokulu
- B Anaokulu

1.2.1 A Versiyonu

Kingston adlı bir şehirde bir buçuk yıldır görev yapan ve pek sevilmeyen bir emniyet müdürü vardır. Aktif bir şekilde politika içinde yer alan ve belediye başkanının meslektaşı olan bu emniyet müdürünün polis yöneticisi olarak atandığında çok az yöneticilik tecrübesi vardır. Belediye başkanı, yakın bir zaman önce emniyet müdürünü kamuoyu önünde savunmuş ve o göreve başladığından bu yana şehirdeki suç oranlarının %12 oranında düştüğünü söylemiştir. Aşağıdaki kanıtlardan hangisi belediye başkanının emniyet müdürünün yetkin olduğuna dair iddiasını en çok zayıflatır?

- Aynı dönemde, Kingston şehrine yakın ve aynı büyüklükteki iki komşu şehirdeki suç oranlarında %18'lik bir azalma olmuştur.
- Kingston şehrinde yaşayanlarla yapılan bağımsız bir anket, polis kayıtlarında rapor edilenden %40 oranında daha fazla suç işlendiğini göstermiştir.
- Mantıklı düşünülecek olursa, bir polis müdürünün tek başına suç oranlarını azaltması pek olası değildir. Suç oranlarındaki bu düşüş büyük oranda yetkililerin kontrolü dışında olan sosyal ve ekonomik durumlardan kaynaklanır.
- Polis müdürünün, organize suçlara karışmış olduğu bilinen kişilerle kişisel temasının olduğunun ortaya çıkarılmış olması.

1.2.2 B Versiyonu

Bir lisede bir buçuk yıldır sevilmeyen bir öğretmen çalışmaktadır. Bu öğretmen müdürün arkadaşıdır ve işe başladığında çok az öğretmenlik tecrübesi vardır. Müdür, yakın bir zaman önce öğrencilere hitaben yaptığı konuşmada, bu öğretmeni tüm okulun önünde savunmuş ve öğretmenin işe başladığından bu yana öğrencilerin yaramazlığının %12 oranında azaldığını duyurmuştur.

Aşağıdaki kanıtlardan hangisi, okul müdürünün bu öğretmenin yetkin olduğuna dair iddiasını en iyi şekilde boşa çıkartır?

- Bu öğretmenin girmedikleri iki sınıfın öğrencilerinin yaramazlık oranları aynı dönemde % 18 oranında azalmıştır.
- Öğrencilerle yapılan bir anket çalışmasına göre müdürün kayıtlarında belirtilenden %40 oranında daha fazla yaramazlık rapor edilmiştir.
- Mantıklı bir şekilde düşünülürse bir öğretmenin yaramaz davranışları azaltması için yapabileceği şeyler kısıtlıdır. Öğrencilerin duygu durumu ve davranışı çok daha büyük bir etkiye sahiptir.
- Öğretmenin yaramazlık yapan öğrencileri cezalandırdığı ortaya çıkmıştır.

1.3.1 A Versiyonu

Bir arabanın X veya Y markalı olup olmadığı hakkında bir karar vermeniz gerektiğini düşünün. Her iki marka için de şu iki bilgiyi elde etme olanağınız olduğunu varsayın: her 100 km'de 10 litreden fazla yakıt tüketim oranı ve ilk iki yılda büyük bir mekanik problem çıkarmama olasılığı.

Diyelim ki iki markadan BİLİNMEYEN bir tanesi (X veya Y) 100 km'de 10 litreden fazla yakıt harcamakta ve ilk iki yılda hiçbir mekanik arıza vermemektedir. Ayrıca, X markasının ürettiği arabaların %65'inin 100 km'de 10 litreden fazla yakıt harcadığını varsayın.

Aşağıdaki üç bilgiden hangisini öğrenirseniz, bir önceki parafta bahsedilen BİLİNMEYEN markanın X veya Y olduğuna daha iyi karar verirsiniz?

- Y markasının ürettiği araçların yüzde kaçının 100 km'de 10 litreden fazla yakıt harcadığı bilgisi.
- İlk iki yıl içinde X markasının ürettiği araçların yüzde kaçının önemli mekanik arıza vermediği bilgisi.
- İlk iki yıl içinde Y markasının ürettiği araçların yüzde kaçının önemli mekanik arıza vermediği bilgisi.

1.3.2 B Versiyonu

Bir dizüstü bilgisayarın X veya Y marka olup olmadığı hakkında bir karar vermeniz gerektiğini düşünün. Her iki marka için de iki konuda bilgiye ulaşma olasılığınız var: pil ömrü 10 saatten fazla olan X ve Y marka bilgisayar yüzdesi ve bu iki markanın ilk iki yılda büyük bir teknik problem çıkarmama olasılığı.

Diyelim ki iki markadan BİLİNMEYEN bir tanesinin (X veya Y) pil süresi 10 saatten fazla ve ilk iki yılda herhangi bir teknik problem çıkarmıyor. Ayrıca, X markasının ürettiği bilgisayarların %65'inin 10 saatten fazla pil süresine sahip olduğunu varsayın.

Aşağıdaki üç bilgiden hangisini öğrenirseniz, bir önceki parafta bahsedilen BİLİNMEYEN markanın X veya Y olduğuna daha iyi karar verirsiniz?

- Y markasının 10 saatten fazla batarya süresine sahip bilgisayar oranı.
- İlk iki yılda X markasının büyük bir teknik problem çıkarmama yüzdesi.
- İlk iki yılda Y markasının büyük bir teknik problem çıkarmama yüzdesi.

1.4.1 A Versiyonu

Bay Maxwell sadece profesörlerin ve iş adamlarının davet edildiği bir partiye katılmıştır (katılımcılar profesör veya iş adamı olabilir ama her ikisi birden olamaz). Bay Maxwell hakkında bildiğiniz tek şey onun Bear adlı bir kulübe üye olmasıdır. Sizden, aşağıdaki soruları sorarak Bay Maxwell'in bir profesör olma olasılığını değerlendirmeniz isteniyor. Yani sizden bu soruların ne derecede alakalı olduğunu değerlendirmeniz isteniyor. Alakalı soru, cevabını aldığınız zaman karar vermede faydalı olan sorudur. Lütfen aşağıdaki her bir soruyu tek tek düşünün ve hangisinin size verilen görevle alakasız bir soru olduğunu belirtin.

- Partideki kişilerden yüzde kaç üniversite profesörüdür?
- Partidekilerin % kaç Bear Kulübü üyesidir?
- Partideki üniversite profesörlerinin yüzde kaç Bear Kulübü üyesidir?
- Partideki iş adamlarının yüzde kaç Bear Kulübü üyesidir?

1.4.2 B Versiyonu

Bayan Crowe sadece Amerikalı ve Kanadalı turistlerin katıldığı rehberli bir geziye katılmıştır (turistler ya Amerikalı ya Kanadalıdır, her iki ülke vatandaşı birden değildir). Bayan Crowe hakkında tek bildiğiniz şey ABC otelinde konakladığıdır. Sizden, aşağıdaki soruları sorarak Bayan Crowe'un bir Amerikalı turist olma ihtimalini değerlendirmeniz istenmektedir. Diğer bir deyişle, sizden bu soruların ne derecede alakalı olduğunu değerlendirmeniz istenmektedir. Alakalı soru, cevabını aldığınız zaman karar vermede faydalı olan sorudur. Lütfen aşağıdaki her bir soruyu tek tek düşünün ve hangisinin size verilen görevde alakasız olduğunu belirtin.

- Turdaki kişilerin yüzde kaç Amerikalı turisttir?
- ABC otelinde konaklayanların yüzde kaç tura katılmaktadır?
- Turdaki Kanadalı turistlerin yüzde kaç ABC otelinde konaklamaktadır?
- Turdaki Amerikalı turistlerin yüzde kaç ABC otelinde konaklamaktadır?

1.5.1 A Versiyonu

Linda 31 yaşında, bekar, açıksözlü ve çok zeki birisidir. Felsefe bölümünden mezun olmuştur. Öğrenciliğinde ayrımcılık ve sosyal adalet konularıyla son derece ilgilenmiş ve anti-nükleer protestolara katılmıştı.

Aşağıdakilerden hangisi daha olasıdır?

- Linda bankada memurdur.
- Linda bankada memurdur ve feminist hareket içinde aktif olan bir bireydir.

1.5.2 B Versiyonu

Jack 45 yaşında, evli ve dört çocuk babasıdır. Jack genellikle dikkatli ve hırslıdır. Politik ve sosyal konulara hiç ilgisi yoktur ve boş zamanlarının çoğunu marangozluk, yelkencilik ve matematik problemleri çözme gibi birçok hobisine ayırmaktadır.

Aşağıdakilerden hangisi daha olasıdır?

- Jack bir mühendistir.
- Jack bir mühendistir ve su sporları ile ilgili yerel bir kulübün aktif üyesidir.

1.6.1 A Versiyonu

Yeni bir araba almak istiyorsunuz. Bugün iki alternatiften birisini seçmeniz gerekiyor: ya A firmasından ya da B firmasından bir araba satın almak. Bu kararınız için tek bir ölçütünüz var; o da arabanın beklenen kullanım ömrü. Yüzlerce arabalık bir örnekleme içeren Tüketici Raporu'na göre A'nın daha iyi olduğu bilgisini edindiniz. Dün bir komşunuz A markalı yeni arabasının bozulduğunu söyledi. Hangi arabayı alırsınız?

- A markalı araba
- B markalı araba

1.6.2 B Versiyonu

Yeni bir bilgisayar almak istiyorsunuz. Bugün iki alternatiften birisini seçmeniz gerekiyor: ya A firmasından ya da B firmasından bir bilgisayar satın almak. Bu kararınız için tek bir ölçütünüz var; o da bilgisayarın beklenen kullanım ömrü. Yüzlerce bilgisayarın incelendiği bir Tüketici Raporu'na göre A'nın daha iyi olduğu bilgisini edindiniz. Dün bir komşunuz A markalı yeni bilgisayarının bozulduğunu söyledi. Hangi bilgisayarı alırsınız?

- A markalı bilgisayar
- B markalı bilgisayar

Verbal Reasoning

Sözel Muhakeme

Benzerlikler

Aşağıdaki iki analogi ile en iyi uyum gösteren yanıtı seçiniz.

(":" ilişkili anlamına gelmektedir; ":::" benzer anlamına gelmektedir)

A1

_____ : at :: yolcu almak : tren

- sabit
- ayakkabı
- binmek
- üzerine çıkmak

B1

şiiir : kafiye :: felsefe : _____

- imge
- müzik
- kanun
- kuram

A2

onarmak : dikiş :: düzeltmek : _____

- yama
- tamir
- makale
- eğreti

B2

alfabetik : _____ :: sıralı (dizisel) : dosya

- sıralamak
- parça
- liste
- düzen

A3

fare : _____ :: flaş : fotoğraf makinesi

- sıçan
- bilgisayar
- kordon
- tatlı

B3

minder : koltuk :: raf : _____

- pervaz
- kitaplık
- depo
- çerçeve

A4

nemlendirmek : _____ :: serinletmek : dondurmak

- sulamak
- sırlıslamak etmek
- fırınlamak
- büyümek

B4

kurdele : _____ :: krema : pasta

- hediye
- kesmek
- fiyonk
- daktilo

A5

kedi yavrusu : _____ :: asker : ordu

- kedi
- bir batında doğanlar
- yavru köpek
- miyav

B5

bilim insanı : deney :: _____ : rol yapmak

- deney şişesi
- prova
- aktör
- laboratuvar

A6

sözlük : tanım :: _____ : harita

- yön
- güney
- atlas
- boylam

B6

raptiye : _____ :: çengel : palto

- çivi
- poster
- duvar
- çekiç

A7

sıradışı : yenilik :: _____ : standart

- tuhaf
- orijinal
- tanıdık
- şiir

B7

_____ : üniversite :: tamirci : tamirhane

- kitap
- öğrenme
- profesör
- makine

A8

davul : enstrüman :: matkap : _____

- çekiç
- fırın
- alet
- kreşendo (müziğin gittikçe yükselmesi)

B8

- zirve : dağ :: _____ : ev
- en yükseğe çıkarmak
- çatı
- veranda
- bungalov

A9

_____ : sağanak :: yoğunlaşma : nem

- erozyon
- bulut
- yer
- hava tahmini

B9

_____ : fotoğraf karesi :: mozaik : karo

- film
- oyma
- banyo
- dekorasyon

Judging Likelihood and Probability

Olabilirliği ve Olasılığı Belirleme

3.1.1 A Versiyonu

Bir duvar tenisi (squash) oyunu 9 veya 15 puan üzerinden oynanabilir. A, B'den daha iyi bir oyuncu ise, hangi oyunda A'nın kazanma şansı daha yüksek olur?

- 9 puanlık maç
- 15 puanlık maç

3.1.2 B Versiyonu

Masa tenisi 5 veya 9 setlik oyunlarla oynanabilir. A, B'den daha iyi bir oyuncu ise, hangi oyunda A'nın kazanma şansı daha yüksek olur?

- 5 setlik bir maç
- 9 setlik bir maç

3.2.1 A Versiyonu

Profesyonel beyzbol liginin ilk iki haftasından sonra gazeteler en iyi vuruş yapan 10 oyuncunun ortalama puanlarını yayınlamaya başladı. Genellikle, iki hafta sonunda en iyi vuruş yapan oyuncunun puan ortalaması .450 civarındadır. Ancak, şimdiye kadar hiçbir profesyonel beyzbol sezonunda bir oyuncu .450 ortalama ile sezonu bitirememiştir. Bunun sebebi sizce ne olabilir? Birini işaretleyin:

- Bir vuruşçu oyuncunun yüksek ortalama ile oynadığı biliniyorsa, topu ona fırlatan rakip oyuncular topa vurmasını diye daha fazla gayret gösterirler.
- Sezon ilerledikçe topu vuruşçuya fırlatan oyuncular daha iyi form tuttukları için daha iyi oynamaya başlarlar. Topu fırlatan rakip oyuncular form tuttukça vuruşçu oyuncunun ortalaması düşer.
- Bir vuruşçunun sezon başındaki yüksek ortalaması şans eseri olabilir. Uzun bir sezon vuruşçunun yetenekleri hakkında daha gerçekçi bir ölçüttür.
- Sezon başında böylesine yüksek bir ortalama yakalayan bir vuruşçu, ortalamasını korumak için çok yoğun stres yaşar. Böyle bir stres de oyun performansını olumsuz etkiler.
- Bir vuruşçunun yüksek bir ortalama yakalamak istediği biliniyorsa, artık ona kolay toplar fırlatılmaz. Onun yerine, onu daha da zorlayacak bir oyun tarzı tercih ederler.

3.2.2 B Versiyonu

Akademik dönem başladıktan 2 hafta sonra, bir öğretmen yılın en iyi 10 öğrencisi kim olacak diye tahmin yürütmeye çalışmaktadır. Genellikle, ilk iki haftanın ardından en iyi öğrencinin ortalama performansı %92'dir. Ancak okul tarihinde dönem sonunda hiçbir öğrenci %92 ortalama alamamıştır. Bunun sebebi sizce ne olabilir? Birini yuvarlak içine alın:

- Bir öğrencinin yüksek ortalama yapmak istediği biliniyorsa, öğretmenlerin bu öğrencinin sınavlarını okurken beklentileri yükselmektedir.
- Dönem ilerledikçe, konular arttığı için sınavlar daha zor bir hale gelmektedir. Sınavlar zorlaştıkça da öğrencilerin yüksek not alma olasılığı azalmakta ve ortalamaları düşmektedir.
- Bir öğrencinin dönem başındaki yüksek ortalaması şans eseri olabilir. Uzun bir akademik yıl bir öğrencinin bilgi ve yeteneklerini ölçmek için daha gerçekçidir.
- Dönem başında böylesine yüksek bir ortalama yakalayan öğrenci, ortalamasını korumak için çok stres yaşar. Böyle bir stres de sonraki sınavlarını olumsuz etkiler.
- Bir öğrencinin yüksek ortalama yapmak istediği biliniyorsa, artık ona kolay sorular sorulmamaya başlanır. Hatta, öğretmenler bu öğrenciye sordukları soruların zorluğunu artırırılar, çünkü öğrencinin becerilerini test etmek isterler.

3.3.1 A Versiyonu

Kollu kumar makinesinde kumar oynayan insanlar, her 10 seferden birinden birşeyler kazanırlar. Ama Julie ilk üç seferinde de kazanmıştır. Bir sonraki seferde kazanma olasılığı nedir? 10 üzerinden_____

3.3.1 B Versiyonu

Yerel bir piyango çekilişinde katılanların bir ödül kazanma şansları her 20 biletten birinedir. Ama Jack, aldığı ilk üç bilete ödül kazanmıştır. Bir sonraki sefer Jack'in ödül kazanma olasılığı nedir? 20 üzerinden_____

3.4.1 A Versiyonu

Hilesiz bir madeni parayla yazı-tura attığımızı varsayın (yazı ve tura gelme olasılıkları %50'dir) ve 5 kez üst üste yazı gelmiştir. Madeni para altıncı kez havaya atıldığında:

- Tura gelme olasılığı yazı gelme olasılığından daha yüksektir.
- Yazı gelme olasılığı tura gelme olasılığından daha yüksektir.
- Altıncı atışta yazı ve tura gelme olasılığı eşittir.

3.4.2 B Versiyonu

Hilesiz bir zar attığımızı varsayın (her sayının gelme olasılığı altıda birdir) ve ard arda beş kez 6 gelmiştir. Bir sonraki zar atışımızda:

- İlk beş zar atışına kıyasla, altıncı atışta 6'dan farklı bir sayı gelme olasılığı daha fazladır.
- İlk beş zar atışına kıyasla, altıncı atışta 6'nın gelme olasılığı, başka bir sayı gelme olasılığından daha fazladır.
- Altıncı atışta 6 veya başka bir sayının gelme olasılığı aynıdır.

3.5.1 A Versiyonu

İki fıçı içerisinde siyah ve beyaz bilyelerin size sunulduğunu düşünün. Büyük fıçının içinde 100 bilye, küçük fıçının içinde ise 10 bilye vardır. Fıçılardan birinden, içine bakmadan bir bilye seçmeniz gerekiyor. Eğer seçtiğiniz bilye siyah çıkarsa 2 dolar kazanacaksınız. Varsayın ki, küçük fıçıda 1 siyah ve 9 beyaz bilye var ve büyük fıçıda 8 siyah ve 92 beyaz bilye var. (Gerçek hayatta bir seçim yapmanız gerekirse) hangi fıçıyı tercih edersiniz?

- Büyük fıçı
- Küçük fıçı

3.5.2 B Versiyonu

İki torba içerisinde 10 ve 20 dolarlık banknotların size sunulduğunu düşünün. Büyük torbanın içinde 100 farklı banknot, küçük torbanın içinde ise 10 farklı banknot var. Banknotlar torbaların içinde rastgele bir şekilde konulmuştur. Torbalardan birinden, içine bakmadan bir banknot seçmeniz gerekiyor ve seçtiğiniz banknot sizin olacak. Varsayın ki, küçük torbada bir adet 20 dolarlık banknot ve dokuz adet 10 dolarlık banknot var ve büyük torbada ise 8 adet 20 dolarlık ve 92 adet 10 dolarlık banknot var. (Gerçek hayatta bir seçim yapmanız gerekirse) hangi torbayı tercih edersiniz?

- Büyük torba
- Küçük torba

Argumentation Analysis

Lütfen aşağıda verilen cümlelerden ilk ikisini dikkatle okuyun. Sonra bu iki cümlede verilen bilgilerin doğru olduğunu varsayın ve üçüncü cümlede ifade edilen çıkarımın geçerli olup olmadığını belirtin.

Lütfen seçiniz: geçerli – geçerli değil

A1

- Hiçbir sigara ucuz değildir.
- Bazı bağımlılık yaratan şeyler ucuzdur.
- Dolayısıyla, bazı bağımlılık yaratan şeyler sigara değildir.

B1

- Bazı binalar ahşaptan yapılmıştır.
- Hiçbir ahşap bina gökdelen değildir.
- Dolayısıyla, bazı binalar gökdelen değildir.

A2

- Bazı ofis malzemeleri bilgisayar değildir.
- Bazı elektrikli aletler ofis malzemesidir.
- Dolayısıyla, bazı elektrikli aletler bilgisayar değildir.

B2

- Bazı içecekler alkolsüzdür.
- Alkolsüz içecekler likör değildir.
- Dolayısıyla, bazı içecekler likör değildir.

A3

- Sandalyeler masa değildir.
- Bazı mobilyalar sandalyedir.
- Dolayısıyla, bazı mobilyalar masa değildir.

B3

- Bazı balıklar yenmez.

- Alabalık yenilebilir.
 - Dolayısıyla, bazı balıklar alabalık değildir.
- B4
- Bağımlılık yaratan hiçbir madde ucuz değildir.
 - Bazı sigaralar ucuzdur.
 - Dolayısıyla, bazı bağımlılık yaratan maddeler sigara değildir.
- A4
- Bazı ahşap binalar gökdelerdir.
 - Hiçbir bina ahşap değildir.
 - Dolayısıyla, bazı binalar gökdelen değildir.
- B5
- Bazı elektrikli aletler ofis malzemesi değildir.
 - Bazı ofis malzemeleri bilgisayardır.
 - Dolayısıyla, bazı elektrikli aletler bilgisayar değildir.
- A5
- Alkolsüz içecekler likördür.
 - Hiçbir içecek alkolsüz değildir.
 - Dolayısıyla, bazı içecekler likör değildir.
- B6
- Hiçbir mobilya sandalye değildir.
 - Sandalyeler masadır.
 - Dolayısıyla, bazı mobilyalar masa değildir.
- A6
- Hiçbir alabalık yenilmez.
 - Balık yenilebilir.
 - Dolayısıyla, bazı balıklar alabalık değildir.
- A7
- Bazı aletler el ile kullanılır.
 - Tornavidalar el ile kullanılan aletlerdir.
 - Dolayısıyla, bazı tornavidalar alet değildir.
- B7
- Bazı müzik enstrümanları üflemelidir.
 - Üflemeli enstrümanlar keman değildir.
 - Dolayısıyla, bazı kemanlar müzik enstrümanı değildir.
- A8 // B8
- Bazı yeşil şeyler ot değildir.
 - Bazı bitkiler yeşildir.
 - Dolayısıyla bazı otlar bitki değildir.
- B9
- Tornavidalar elle kullanılan aletler değildir.
 - Hiçbir alet elle kullanılmaz.
 - Dolayısıyla, bazı tornavidalar alet değildir.
- A9
- Üflemeli enstrümanlar kemanlardır.
 - Hiçbir müzik enstrümanı üflemeli değildir.
 - Dolayısıyla, bazı kemanlar müzik enstrümanı değildir.

Problem Solving

5.1.1 A Versiyonu

Bir beyzbol sopası ve topu toplam 1.10 dolardır. Sopa, topdan 1 dolar daha pahalıdır. Topun fiyatı nedir?

5 sent – 10 sent – 55 sent – 100 sent

5.1.2 B Versiyonu

Bir somun ekmek ve bir elma toplam 2.20 dolardır. Ekmek elmadan 2 dolar pahalıdır. Elmanın fiyatı nedir?

10 sent – 20 sent – 110 sent – 200 sent

5.2.1 A Versiyonu

Eğer bir makine 5 plastik parçayı 5 dakikada yaparsa, 100 makinenin 100 plastik parçayı yapması ne kadar sürer?

1 - 5 - 25 - 100 dakika

5.2.2 B Versiyonu

Eğer 2 fırıncı 2 ekmeği 2 dakikada yapıyorsa, 50 fırıncının 50 ekmeği yapması ne kadar sürer?

1 - 2 - 25 - 50 dakika

5.3.1 A Versiyonu

Bir gölde bir öbek su zambağı vardır. Bu öbek her gün iki katı büyüklüğüne ulaşmaktadır. Su zambağı öbeğinin tüm gölü kaplaması 24 gün sürüyorsa, gölün yarısını kaplaması kaç gün sürer?

2 – 4 – 12 – 23 gün

5.3.2 B Versiyonu

Bir kap taşlarla doldurulmaktadır. Her gün kabın içindeki taş sayısı iki katına çıkmaktadır. Kap 10 günde tamamen dolacaktır. Kap kaç günde yarısına kadar dolmuş olur?

2 – 4 – 5 – 9 gün

5.4.1 A Versiyonu

Bir çekmecedeki siyah ve kahverengi teklilerden biri almaktadır ve çoraplardan birisinden 4 tek diğerinden 5 tek vardır. Çekmecenin içine bakmadan teker teker kaç tane tek çorap çekerseniz aynı renkte bir çift çoraba sahip olduğunuzdan emin olursunuz?

1 – 2 – 3 – 4 – 5 tek çorap

5.4.2 B Versiyonu

Bir çantada yeşil ve mavi renkte 6'ya 8 oranında bilye vardır. Aynı renkte iki bilyeye sahip olduğunuzdan emin olabilmek için, torbanın içine bakmadan, kaç tane bilye çekmeniz gerekir?

2 - 3 - 4 - 5 - 6 bilye

5.5.1 A Versiyonu

Birbirlerinden 50 km uzaklıkta yer alan iki tren, birbirlerine doğru saatte 25 km hızla yola çıkıyorlar. Trenler birbirlerine doğru yola çıkınca bir kuş bir trenin önünden diğer trene doğru uçmaya başlıyor. Kuş diğer trene ulaştığında ilk trene doğru tekrar uçmaya başlıyor, ta ki iki tren buluşuncaya kadar. Kuş saatte 60 km hızla uçuyorsa, iki tren karşılaşmadan önce kuş kaç km uçmuş olacaktır?

15 km - 30km - 45 km - 60km

5.5.2 B Versiyonu

Birbirlerinden 10 km uzaklıkta yer alan iki yaya, birbirlerine doğru saatte 5 km hızla yürümeye başlıyorlar. Yayalar birbirlerine doğru yürümeye başlar başlamaz bir köpek bir yayadan diğerine doğru koşmaya başlıyor. Köpek diğer yayaya ulaştığında ilk yayaya doğru tekrar koşmaya başlıyor, ta ki iki yaya buluşuncaya kadar. Köpek saatte 20 km hızla koşuyorsa, iki yaya buluşmadan önce köpek kaç km koşmuş olacaktır?

2 - 5 - 10 - 20 km

5.6.1 A Versiyonu

Bir adam 60 sterline bir at alır, 70 sterlin karşılığında satar, 80 sterline geri alır ve en sonunda 90 sterline satar. Bu adam ne kadar kazanmıştır?

10 Sterlin – 20 Sterlin – 30 Sterlin – 40 Sterlin

5.6.2 B Versiyonu

Bir kadın, borsada bir şirketin hisselerini 600 dolar karşılığında satın alıyor, 700 dolar karşılığında satıyor, 800 dolar karşılığında geri alıyor ve sonunda 900 dolar karşılığında satıyor. Ne kadar kazanmıştır?

100 dolar - 200 dolar - 300 dolar - 400 dolar

5.7.1 A Versiyonu

Polis A, B, C veya D'den birisinin suç işlediğine inanmaktadır. Dört şüpheli, sırasıyla aşağıdaki ifadeleri vermiştir, ama bu ifadelerden sadece bir tanesi doğrudur.

- A "Ben yapmadım" dedi.
- B "A yalan söylüyor" dedi.
- C "B yalan söylüyor" dedi.
- D "Suçu B işledi" dedi.

Hangisi doğru söylüyor? Suçu kim işledi?

Doğruyu söyleyen: A – B – C – D

Suç işleyen: A – B – C – D

5.7.2 B Versiyonu

Bir hakim A, B, C veya D'den birinin bir şey icat ettiğini bilmektedir. Dört kişi, sırasıyla aşağıdaki iddialarda bulunmuştur, ama içlerinden sadece bir tanesi doğrudur.

- A "Ben icat ettim" dedi.
- B "A yalan söylüyor" dedi.
- C "B yalan söylüyor" dedi.
- D "B icat etmedi" dedi.

Hangisi doğru söylüyor? Kim icat etti?

Doğruyu Söyleyen: A – B – C – D

İcat eden: A – B – C – D

5.8.1 A Versiyonu

Her birinin bir yüzünde bir rakam, diğer yüzünde bir renk bulunan dört kart bir masanın üstünde size gösteriliyor. Kartların her birinin görünen yüzlerinde 3 ve 8 rakamları ile kırmızı ve kahverengi renkleri vardır. "Bir kartın üzerindeki rakam çift sayı ise öbür yüzü kırmızıdır" iddiasının doğruluğunu test etmek için hangi kart(lar)ı çevirmeniz gerekir? En etkili olacak ve en az sayıda kartın yüzünü çevirecek şekilde düşünmeye çalışın.

İstedığınız kadar çok sayıda kartı işaretleyebilirsiniz: 3 – 8 – kırmızı – kahverengi

5.8.2 B Versiyonu

Her birinin bir yüzünde bir rakam, diğer yüzünde bir renk bulunan dört kart bir masanın üstünde size gösteriliyor. Kartların her birinin görünen yüzlerinde 5 ve 6 rakamları ile yeşil ve sarı renkleri vardır. "Bir kartın üzerindeki rakam çift sayı ise öbür yüzü yeşildir" iddiasının doğruluğunu test etmek için hangi kart(lar)ı çevirmeniz gerekir? En etkili olacak ve en az sayıda kartın yüzünü çevirecek şekilde düşünmeye çalışın.

İstedığınız kadar çok sayıda kartı işaretleyebilirsiniz: 5 – 6 – yeşil – sarı

Appendix 5: Example of Raven's Standard Progressive Matrices

