Norwegian University
of Life Sciences

Master's Thesis 2022 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Student's learning preferences and perception of mathematics; *How different cognitive types view contrasting approaches to statistics education*

Studenters læringspreferanser og tanker om matematikk; *Hvordan ulike kognitive typer foretrekker ulike tilnærminger til statistikk undervisning*

Mina Therese Gjefle
Bioinformatics and Applied Statistics

Preface

This master thesis is in context with my Master of Applied Statistics at the Norwegian University of Life Sciences (NMBU). My five long years at Ås are now coming to an end, though it has barely felt like a minute. When I found out I could write my thesis on the topics of personality types, I was overjoyed, as this is something I have found quite interesting for a long time. But more important than that is the people I have had the joy to acquaintance. I would like to extent a huge thank you to my main supervisor, Kathrine Frey Frøslie (associate professor, NMBU), for great feedback, your keen eye, and unwavering enthusiasm for learning and statistics. Despite my shortcomings, your belief in me never wavered and for that I am forever grateful. A huge thank you to my co-supervisor, Hilde Vinje (associate professor and advisor, NMBU), for your helpful revisions and input – Yes, I absolutely did spend way too much time dwelling on the ancient Mesopotamians. I would also like to thank my second co-supervisor Solve Sæbø (professor, NMBU). Thank you for your help in acquiring and understanding the data in sample 2. Also, thank you for your guidance with the PCA/LDA correlation scores analyses; My F's are now indeed in the upper left corner.

Thank you to Solveig Arnesen (CEO, Vitenparken) and Aiyana Shanti Hudgins (educational leader, Vitenparken) for your help with recruitment. Thanks to Nasjonalt Senter for Realfagsrekruttering (NSR) for providing data for analysis, and to Mari Augensen Mathisen (adviser, NMBU) for your help with the data processor agreement.

Finally, thank you to my friends, family, and to my boyfriend Vemund Lenschow – my wonderful love and king of the head massages. Although my statistical talk may just as well have been Greek to your ears, you never shied away from asking (hi dad).

Norwegian University
of Life Sciences

Ås, May 15th, 2022

Abstract

The Big-Four is a popular typology used to assess behavior through categorical means, and it's four dichotomous dimensions have been found to associate with the cognitive patterns of individuals. Analyses have found that certain types are disfavored by traditional education, particularly within mathematics.

This study aimed at investigating the associations between Big-Four types, work interests, and preferred learning strategies using the interest- and personality-assessment called Utdanningstesten. Several hypotheses were put forward regarding which connections were expected to be found. This study also sought to analyze the validity and reliability of Utdanningstesten as a Big-Four predictor, through test-retest and inter-rater reliability tests with both Big-Four tests and a Big-Five test.

Two samples were analyzed in this thesis: A "validation" sample consisting primarily of students from the Norwegian University of Life Sciences (NMBU), and a pre-sampled "explorative" sample containing all answer from the individuals that took Utdanningstesten in 2016-2020. In the "validation" sample, Kappa inter-rater reliability coefficients were estimated to assess Utdanningstesten's validity. Two-sample T-tests were also done to check whether the contradicting Big-Four types scored differently in the Big-Five. In the "explorative" sample, PCA and LDA was applied to explore any type-dependent tendencies, and classification models were made to assess the effectivity of each method. Statistical models were also made to assess multiple variables' ability to predict perception of mathematics and preferred learning strategies.

I found that there were differences between the cognitive types for both perception of mathematics and preferred learnings strategies. However, effects were only noticed for unidimensional types and letter pairs – particularly the function pair (Sensing/Intuitive–Thinking/Feeling type) –, and not the full 4-letter types. Intuitive and Extraverted types were often found to disfavor math. I also found that Utdanningstesten did well in test-retest analysis. However, weaker results were noted in the inter-reliability tests. Although the questionnaire seems to do a satisfactory job at predicting the Introvert/Extravert and Perceiving/Judging type, it performs poorer predicting the Sensing/Intuitive type.

My results suggests that it could be advantageous to incorporate personality type theory into education and how it is structured. Utdanningstesten may be utilized, but with caution.

Sammendrag

Big-Four er en populær typologi som kan anvendes til å kategorisk definere oppførsel, og dens fire dikotome dimensjoner har blitt funnet til å assosiere med de kognitive mønstre hos individer. Tidligere studier har funnet at noen persontyper sidelinjes av den tradisjonelle undervisningsstilen, spesielt innenfor matematikk.

Denne studien tok sikte på å undersøke assosiasjonene mellom Big-Four typer, arbeidsinteresser, og prefererte læringsstrategier, gjennom å bruke en interesse- og personlighets-undersøkelse kalt Utdanningstesten. Det ble fremsatt flere hypoteser rundt hvilke sammenhenger det var forventet å finne. Jeg søkte også etter å analysere validiteten og reliabiliteten til Utdanningstesten som en Big-Four-prediktor, gjennom test-retest og inter-rater reliabilitetstester med både Big-Four og en Big-Five test.

To utvalg ble evaluert i denne studien: Et «validering»-utvalg, som primært besto av studenter fra Norges miljø- og biovitenskapelige universitet (NMBU), samt et forhånds-innsamlet «eksplorativt» utvalg som inneholdt alle svar fra individer som tok testen i tidsrommet 2016-2020. I valideringsutvalget ble Kappa inter-rater reliabilitets koeffisienter estimert for å vurdere Utdanningstestens validitet. To-utvalgs T-tester ble også kjørt for å sjekke hvorvidt de motsatte Big-Four typene skåret forskjellig i Big-Five. I det eksplorative utvalget ble PCA og LDA utført for å undersøke eventuelle typeavhengige tendenser, og klassifikasjonsmodeller ble laget for å vurdere effektiviteten til hver metode. Statistiske modeller ble også laget for å vurdere variablers evne til å predikere matematikk-persepsjon eller foretrukne læringsstrategier.

Jeg fant at det var forskjeller mellom de kognitive typene for både matematikk-persepsjon og foretrukne læringsstrategier. Slike forskjeller ble derimot bare funnet mellom unidimensjonale og parede typer – spesielt funksjonsparet (Sansende/Intuitiv-Tenkende/Følende) –, og ikke for de fullstendige 4-bokstavtypene. Intuitive og Ekstroverte individer tenderte til å mislike matte. Jeg fant også at Utdanningstesten gjorde det bra i test-retest analyse. Det ble imidlertid notert svakere resultater på tvers av persontypetester. Selv om spørreskjemaet ser ut til å gjøre en tilfredsstillende jobb med å forutsi typene Introvert/Ekstrovert og Kontekstuell/Digital (Perceiving/Judging), presterer testen dårligere i å forutsi Sansende/Intuitiv-typen.

Mine resultater tyder på at det kan være fordelaktig å inkorporere personlighetstypeteori til den pedagogiske praksis. Utdanningstesten kan benyttes, men med forsiktighet.

Contents

List of tables

List of figures

ABBREVIATIONS AND SYMBOLS

Abbreviations used in this Thesis

Sample 1 .... Data sampled from students by the author in the period of fall 2021 – spring 2022, to validate Utdanningstesten in part one of this thesis
Sample 2 .... Historical data sampled from Utdanningstesten in the period 2016 – 2020
Utd.1 .... One of the five tests sample 1 could take; Utdanningstesten, the first time (every participant took this)
Utd.2 .... One of the five tests sample 1 could take; Utdanningstesten, the second time, i.e., 1 day or longer after they took Utd.1
Uroboros .... One of the five tests sample 1 could take; The "predecessor" that inspired Utdanningstesten
Truity .... One of the five tests sample 1 could take; The big-four test called Truity Typefinder (may also be denoted as: Big-Four)
BF .... One of the five tests sample 1 could take; The Big-Five test (alias: Big-Five)
Big Four .... Denotes personality instruments whose theories are based on the studies by Isabel Briggs Myers and Katharine Cook Briggs' MBTI; Often used to denote the Truity Typefinder.
MAE .... Mean absolute error
MSE .... Mean squared error
LDA .... Linear discriminant analysis
LD .... Linear discriminant (when talking about components from LDA)
PCA .... Principal component analysis
PC .... Principal component (when talking about components from PCA)
NSR .... Nasjonalt Senter for Realfagsrekruttering
NSD .... The Norwegian Centre for Research Data
NSR .... National Centre for Science Recruitment
NMBU .... The Norwegian University of Life Sciences
NTNU .... The Norwegian University of Science and Technology
INN .... The Inland Norway University of Applied Sciences
GLM .... Generalized Linear Models, an umbrella term
SVC .... Support Vector Classifier, a Support Vector Machine
CV .... Cross-Validation
OR .... Odds-Ratio
OvR .... One vs. Rest (alternatively: OvA)
OvA .... One vs. All (alternatively: OvR)
HiOF .... Østfold University College
BCI .... Brain-Computer Interface

Symbols used in this Thesis

N .... observations / rows of the data frame
n .... observations / rows of the data frame (for a group-subset of the data)
i .... iterator; temporary variable, vertical direction (rows) or used when there is only a single variable-iteration
j .... iterator; temporary variable, horizontal direction (variables)
h .... iterator; temporary variable, used when 3 iterators are needed
p .... the number of regular variables (before any PCA/LDA)
k .... the number of transformed variables (e.g., number of PC's)
X .... 2-dimensional matrix
X_i .... 1-dimensional matrix / vector

r .... Pearson's Correlation Coefficient
M .... sample mean matrix
m_i .... sample mean matrix, for group-subset "i"
S .... covariance matrix
S_i .... covariance matrix, for group-subset "i"
S_W .... within-group covariance matrix
S_B .... between-group covariance matrix
B .... mean-deviation form matrix
B_i .... mean-deviation form matrix, for group-subset "i"
K .... Cohen's Kappa reliability score (estimate)

1. – Introduction

What makes you your unique self? Why do we think and act the way that we do? These are but a fragment of what humans across time have pondered in relation to human behavior – from the earliest philosopher to the latest psychologist or neuroscientist. Responsible for interpreting sensory input and governing both conscious and an abundance of non-conscious counter-reactions, nothing might be as complex of a system than the human brain. It sends us on our way and directs our behavior, and though there is still much we do not know about its function, and the various links between behavioral patterns and cognitive pathways and reactions, the gap in knowledge are steadily decreasing year by year.

One might say that to be an educator is to be a scientific professional within the human psyche. Your job is to make order out of chaos, and present information in such a manner as to successfully convey its sentiment to their intended audience. To do so, it is important to know your audience – not least their theoretical and/or practical background, but the way they are wired as well. Not all methods will suit everyone, and though there is strength in variety, a certain level of personalized education can be key to achieve long-term success in your students and mediate loss of focus or retention.

The first documented form of institutionalized education dates to ancient Mesopotamia, a civilization which existed in parallel with the Egyptian. The formal education was aimed at scribes and priests, its discipline was harsh, and it was characterized by practical training, using memorization, oral repetition, and copying of texts (Britannica, 2022). For the longest of time, these core elements of discipline and repetition were defining of education in the west. Though huge changes have been made in the last century to how educators teach, universities and colleges have to a large extent held on to traditional "one-way" lecture-based teaching.

Due to massification, higher education today is generally widespread and available for most. This has in turn resulted in an overwhelming number of students, from various socioeconomic backgrounds and cognitive dispositions, entering university (IGI Global, 2022). There is no longer a set blueprint for the "typical student". This newfound cognitive heterogeneity has only helped illuminate that the human mind is complex and synthesizes information in different manners; thus, there is a need for variety in how educators approach their profession (Bruni & Catalano & Daraio & Gregori & Moed, 2020)(Thomsen, 2012). Behavioral cerebrum patterns are thought to play a large part in how information is interpreted by the individual. This is part of the reason why, as an educator, you need to know your audience to give a good presentation. Not only their theoretical or sociological background, but their cognitive backgrounds as well. Research into an individual's personality – in relation to cognitive reactions – holds great power and might be a steppingstone towards changing the education system for the better and to help increase student's success rates.

To support the theories that will be presented at the end of this chapter, I must first delve into the seas of personality theory and psychology, and previous research. The subsequent sub-chapters will present topics such as research into personality, different typologies and instruments, and educational styles, before arriving at the core of this thesis: Its background, hypotheses, and overall aim.

1.1 – A brief history of psychology and personality research

Personality is multifaceted and defined as the relative stable individual differences in how people think, act, and feel across different social situations (Kennair, 2018). Though personality psychology as a discipline is fragmented, most share the belief that personality begins with the biological, and that "the innate tendencies are channeled by the influence of many factors, such as family or culture." (Corr & Matthews, 2009, p. 6). Most researchers acknowledge personality as being shaped by both our genetic makeup and environmental factors; these components interact perpetually to create neurological links and behavioral pathways, which in turn is expressed by the individual in how they react to, and interact with, the world and its components, both consciously and unconsciously (Brovold, 2014) (Gjefle, STIN300).

Before delving further into the discussion of personality and introducing elements such as types and traits, it is vital to present the theme of temperaments. The study of temperament has an ancient history with roots back to the Greco-Roman physicians (Corr & Matthews, 2009). A temperament can be described as an overarching categorization of behavior, that is largely dependent on the deterministic biology of the individual rather than situational context. As such, behavioral analysis emphasizing the theory of temperaments tends to generalize across situations. Thus, it is relatively constant across time, although lacking in predictive power for situation-specific behavior. Personality however is considered more mendable and nuanced and it is shaped around the core temperament. Thus, while personality may change during an individual's life span, their overarching temperament will in most cases remain fixed.

Numerous attempts have been made to categorize and quantify behavior. As described by Cloninger in (Corr & Matthews, 2009, ch1), there are 6 major perspectives within personality and behavioral theory. One such perspective is called "biological" and is represented by concepts such as temperament and evolutional adaptation. Two other perspectives, "Cognitive" and "Learning", respectively prioritize themes such as reciprocal determinism (the mutual influence between the environment, behavior, and the individual (APA 1, 2020)) and conditioning. Thus, they both emphasize that behavior is contextual. A view that has become quite popularized in modern times, however, is the "Trait"-perspective, whose focal point is that personality and behavior might be measured on a continuum. Herein lies concepts such as trait, factors, as well as neuroticism / emotional instability.

A multitude of models – both scientific and pseudoscientific – have been made, some of which are summarized in figure 1.1. Among the earliest such models was astrology, invented by the Babylonians nearly four thousand years ago (Beck, 2007). However, astrology is considered pseudoscience due to consistently failing attempts at experimental and theoretical validation (Silverman, 1971). Some hundred years BC, the Greek physician Hippocrates (460-370 BC) created the four-types model called "the four humors". This model described behavior as being determined by the bodily fluids of the individual (MedicineNet, 1, 2019). Half a millennium later, the roman physician and philosopher Galen built upon his theories, creating the "four temperaments" model. This model further described the four humors / temperaments (Choleric, Phlegmatic, Melancholic, and Sanguine) as being determined by the abundance of warmth and dryness, or the lack thereof (NIH, 2012). Seen in context to what we know today regarding behavior, about the importance of the production and presence of chemical compounds like hormones and neurotransmitters in the body (McEwen, 2020)(Atzil & Hendler & Feldman, 2011), this was a rather modern take on behavioral psychology. This four-factor biological model would later become an indirect inspiration to a multitude of physicians and psychologists throughout history (Gjefle, STIN300).

In 1910, Swiss psychiatrist and psychoanalyst Carl Jung published “The association method”. In this publication, Jung presented his typology which consisted of three dichotomous dimensions: Introvert/Extravert, Sensing/Intuition, and Thinking/Feeling (Jung, 1910). Jung elaborated on his theories in 192: By combining Introvert/Extravert with one of the four other dimensional types, 8 personality types emerged. These would be known as the “Jungian 8”, and are as follows: I/E Sensing, I/E Intuition, I/E Thinking, and I/E Feeling.

While Jung’s typology failed to gain success in the mainstream, the theories by his one-time collaborator Sigmund Freud – the Austrian neurologist and founder of psychoanalysis – were embraced. According to Freud, the mind consisted of the conscious and the unconscious, while personality was divided into three: The Id, representing the instinctual tendencies of the brain; the Super-Ego, i.e., the morality of the mind; and the Ego, acting as a sort of mediator of the two (MacLeoud, 2021). Multiple explanations might be possible as to why Jung’s theories did not gain a foothold. Two key factors might have been its lack of a sound mathematical foundation, and the aura of elitism surrounding his work. Jung’s typology had a focus on encompassing types rather than gradient traits, the latter of which was becoming quite popular, as exemplified by Freud. Jung also mainly established his theories through private seminars and his own closed-doors psychological club (Geyer, 1995).

However, Jung’s 1921 “Psychological types” would wind up the muse of Katharine Cook Briggs and Isabel Briggs Myers, the mother-daughter duo credited with constructing the Big-Four model and the cognitive instrument known as the MBTI. The first iteration of their model was initialized in 1962, when they published their instrument’s manual, titled “The Myers-Briggs type indicator”. In their typology they introduced a fourth dichotomous dimension to Jung’s original three, titled Perceiving/Judging. By combining the letters, 16 personality types emerged. Since its introduction, the instrument has been maintained and updated, and the framework of this instrument is widely used today, e.g., in popular media and recruiting. Though the Big-Four model were not met without critique. For further in-depth descriptions about the Big-Four model, see chapter 1.2 onwards.

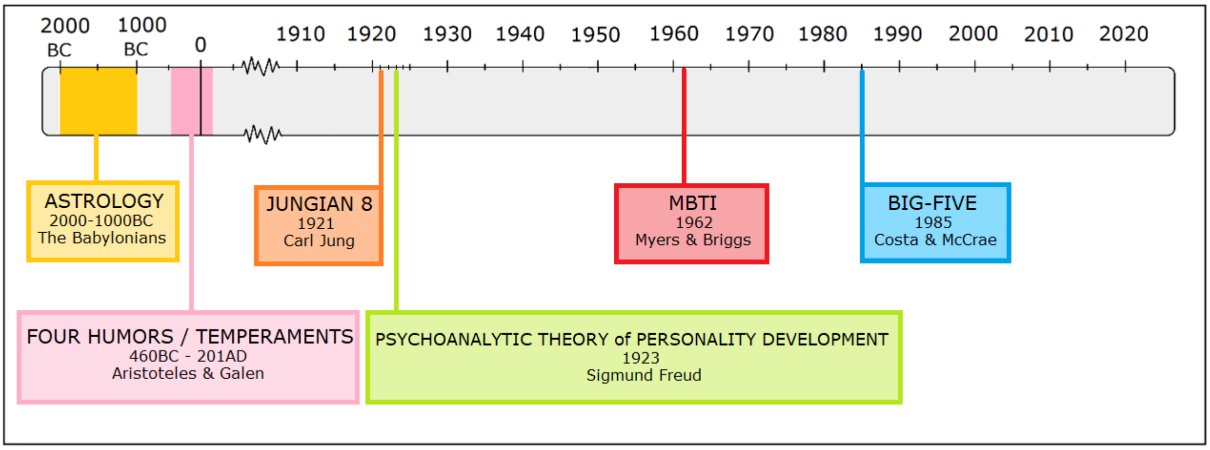


Figure 1.1: *Notable behavioral and personality models or theories throughout history. The models or theories were chosen either for their relevancy to this thesis, or because of their popularity in the modern age, e.g., Astrology.*

The newest personality model presented in figure 1.1 is the Big-Five model. Officially introduced in 1985 by Costa and McCrae, the Big-Five is a trait-based cognitive instrument. Although the number of items varies across various iterations of the questionnaire, the Big-Five encompasses 5 dimensions with 30 traits (/facets) total. Its inception can be considered a collaborative effort that resulted from rounds of factor analyses and computational clustering processes by several independent researchers and practitioners (Pervin & John, 1999). For further details, see chapter 1.2.2.

The Big-Four algorithm and the Big-Five model are two of the most used models today. The next chapter will discuss the two.

1.2 – The Big-Four, The Big-Five, and their differences

In this segment, we will take a closer look into the history behind the two most well-known personality type theories today: the Big-Four model, by virtue of Myers and Briggs, as well as the Big-Five model. Developed in the 1900s, these models gained notable traction during the 1980s and onwards. Both models measure personality across four or more dimensions, with striking resemblances. However, in terms of theoretical background and how they are scaled and intended to be used, the differences between the two are quite notable.

Before I continue: Though some of the personality tests analyzed in this thesis are heavily based on the theories by Myers and Briggs, I will refer to these as “Big-Four” /

"Jungian" tests in place of "MBTI". This is because the latter refers to the trademarked Myers-Briggs Type Indicator personality test distributed by the Myers-Briggs foundation – the true Big-Four test. The MBTI have however not been utilized in this thesis due to economic constraints, as will be further discussed in chapter 3.

1.2.1 – Big-Four

As mentioned in the previous chapter (1.1), the emergence of the Big-Four model was due to the joined – albeit not directly collaborative – effort of psychiatrist and psychoanalyst Carl Jung, and the mother-daughter duo Briggs and Briggs-Myers. In 1907-1912, Jung was a close collaborator to Sigmund Freud, and held important positions in the psychoanalytic movement (Fordham, 2021). Their collaboration however ended, in part due to different viewpoints relating to Freud's "insistence on the sexual bases of neurosis" (Fordham, 2021). For the empiricist Freud, biological determinism was defining of his work. Jung however believed that man was also a "self-creating subject" (Jung, 1923, foreword). In the following years, Jung successfully differentiated two classes of people – the extraverted and the introverted –, as well as four functions relating to the mind: sensation, intuition, thinking and feeling. His results culminated in the publication of "Psychologische Typen" in 1921, which was published in English in 1923.

Katharine Briggs began researching topics within behavioral psychology in 1919. Her inspiration was her daughter's husband, of which she found expressed particularly interesting behavior (Myers Briggs Company, 2022). In her unpublished early work, Briggs described a type-classification regarding one's attitude towards the outer world, that would become known as the Judging / Perceiving dimension in the Big-Four typology. In 1923 she discovered the newly translated "Psychological types" by Carl Jung, which Briggs – then joined by her daughter, denoted Myers – took inspiration from, and began synthesizing Jung's work.

The methods used by Myers and Briggs when creating the MBTI instrument was rather subjective in nature. They defined the types and their characteristics through observation and made paired item-descriptors stereotyping their respective letter-types to force preferential responses from their participants (Myers et al., 1998). During its early developmental stages, evaluation of the MBTI was mainly done through individual feedback rather than by mathematical calculations; Myers and

Briggs saw the self as best equipped to attest to any inaccuracies in typings, and they based any changes in type description or item definition on this feedback (Myers et.al., 1998). As told by Myers et.al.: "Individual respondents are viewed as experts who are best qualified to judge the accuracy of the type descriptions that result from their self-report." (p.5). As such , the construction of the original Big-Four instrument was rooted in subjectivity, observation, and dialogue, rather than pure objective theory and mathematics.

A Big-Four model consists of 8 paired letters, encompassing four dichotomous dimensions. These are: Introvert/Extravert (I/E), Sensing/Intuitive (S/N), Thinking/Feeling (T/F), and Perceiving/Judging (P/J). The first dimension, I/E, describes how the individual directs their energy; Extraverts are categorized as social and gain energy from interacting with others. They are turned towards the outer world of things and people (Sæbø & Almøy & Brovold, 2015). Introverts more often prefer solitude and feel more easily drained from excessive amounts of socialization. Contrary to popular belief, the I/E dimension is not intrinsically linked to shyness or gregariousness (Boyle, 1995). The S/N dimension relates to how you process information; Sensers are often grounded and take thinks at face value, attending to their sensory stimuli. Intuitives however might be more detached and insightful and prefer to interpret and add meaning (Boyle, 1995)(Myers & Briggs foundation, 2022). T/F is concerned with the decision-making process; Thinkers first look at the consistency and they mainly base their decisions in logic. Feelers contrastingly are more oriented towards the details and the subjective – their decision-making process are "based on their values rather than pure logic" (Sæbø & Almøy & Brovold, 2015, p.277). Lastly, the P/J dichotomy references how the individual prefer to structure information about the world; Judgers are focused with details and swift decisions, while Perceivers are contextual, like to "stay open to new information and options", and prefer flexibility "regarding how and when to reach a preset goal" (Myers & Briggs foundation, 2022, p.277).

The 16 types on the Big-Four scale are unevenly distributed. The American sample in (Myers et.al.1998) reported that while only 1.5% were categorized as INFJ, more than 13.8% of the population were categorized as ISFJ. Furthermore, while the division were approximately 50:50 in both the Extraversion (E/I) and the Conscientiousness

(J/P) dimension, almost 75% of the population were Sensors (S) compared to Intuitives (N). This is summarized in table 1.1 below.

At the core of the Big-Four model are the concepts of attitudes (also called orientations) and functions. The two **attitude-dimensions** are: Introversion / Extraversion, and Perceiving / Judging. The two **function-dimensions** are: Sensing/Intuition and Thinking/Feeling. “Although everyone has access to and uses all four mental functions, each type prefers to use these functions in a specific order.” (Myers & Briggs Foundation 1, 2022). Table 1.1 show the 16 types grouped by their respective function and attitude pairs.

Table 1.1: *Distribution of Big-Four personality types in the national (American 1972-2002) representative sample from (Myers et.al., 1998). Percentwise estimates are shown for single-letter types and the full 4-letter types (N=3009). For 4-letter types, estimates are also shown for each gender: Females (n=1531) are shown below the total, to the right of the slash, and males (n=1478) to the left. The cumulative percentwise score is shown for each function / attitude in the yellow cells.*

Single letters				Function						
				ST	SF	NT	NF			
				30%	43.4%	16.5%	10.4%			
I	50.7%	E	49.3%	ISTP 5.4 8.5 / 2.3	ISFP 8.8 7.6 / 9.9	INTP 3.3 4.8 / 1.7	INFP 4.4 4.1 / 4.6	21.9%	IP	Attitude
S	73.3%	N	26.7%	ISTJ 11.6 16.4 / 6.9	ISFJ 13.8 8.1 / 19.4	INTJ 2.1 3.3 / 0.9	INFJ 1.5 1.2 / 1.6	29%	IJ	
T	40.2%	F	59.8%	ESTP 4.3 5.6 / 3.0	ESFP 8.5 6.9 / 10.1	ENTP 3.2 4.0 / 2.4	ENFP 8.1 6.4 / 9.7	24.1%	EP	
P	45.9%	J	54.1%	ESTJ 8.7 11.2 / 6.3	ESFJ 12.3 7.5 / 16.9	ENTJ 1.8 2.7 / 0.9	ENFJ 2.5 1.6 / 3.3	25.3%	EJ	

As stated in chapter 1.1, the Four Temperaments model inspired other researchers within the field of behavioral analytics. The Big-Four types can be linked to the four types in this typology, by grouping them by their primary (/predominant) and secondary type (16personalities C, 2022)(Psychologia, 2022). This grouping is illustrated in Table 1.2.

Table 1.2: *The 16 Jungian personality types grouped by their predominant and secondary temperament. The “Common”-row and -column summarizes the common personality type letters elements across their respective columns/rows; The letters are put in brackets if only 3 out of the 4 type-elements included the letter, and further separated by comma if not all 3*

row/column elements including letter "a" also included letter "b". (16personalities C, 2022), (Psychologia, 2022).

		Secondary				Common
		Choleric	Phlegmatic	Melancholic	Sanguine	
Predominant	Choleric	ENTJ	ENFJ	ESTJ	ESTP	E, (J)
	Phlegmatic	INTP	INFP	ISFJ	ISFP	I, (P)
	Melancholic	INTJ	INFJ	ISTJ	ISTP	I, (T, J)
	Sanguine	ENTP	ENFP	ESFJ	ESFP	E, (F, P)
Common		NT	NF	SJ	SP	

1.2.2 – Critique and praise; The predictive power of Big-Four

The MBTI instrument, and thereby the Big-Four algorithm, have met a significant deal of adversity since its inception. There are questions about its predictive power. (Pittenger, 2005) argued its several limitations. For one, the theoretical foundation of the MBTI suggests that there are distinct groups of people with relative intra-group homogeneity of variance and high inter-group heterogeneity, which Pittenger argues is not as clear-cut as stated. The Jungian model also view personality more quite deterministically, which does not perfectly align with the overarching consensus today that the expression of behavior is both due to nature and nurture.

(Boyle, 1995) stated that the MBTI appeared to measure 30-35% of the trait variance in Big-Five, which could be extrapolated to mean that the MBTI is only a third as good at predicting behavior than the Big-Five. However, "[Big-Four] clearly places a proportionately greater emphasis on cognitive styles than do most other personality instruments" (Boyle, 1995, p. 2). This would make educational settings an exceptionally good area of usage for the Big-Four. Evidence also exists for its power to predict general characteristics, as well as educational- and job-related preferences, of its subjects (Murray, 1990). However, the characteristics "that most accurately identified each group of students seemed to fall into broad categories such as orientation to life, interest patterns, and behavior in social contexts" (Murray, 1990, p.1195)

1.2.3 – Big-Five

In the early-to-mid-1950s the trait-based approach to personality assessment was steadily becoming more popular; Multiple practitioners and researchers within the field were working towards creating their own trait-based models (Pervin, 1994)(Pervin & John, 1999). However, there were no established consensus about which scales, and traits were important. And to make matters worse, “scales with the same name often measure[d] concepts that [were] not the same, and scales with different names often measure[d] concepts that [were] quite similar” (Pervin & John, 1999, p. 102).

Researchers and practitioners banded together to establish a common “taxonomy”. In 1936, Gorwon W. Allport and Henry S. Odberg published a study listing 17954 unique terms that were identified to describe behavior. Raymond B. Cattell then further limited the number to 35 traits in the 1940s, which he used to create his 16PF model. This was a huge feat owed in large part to the increase of computational power allowing for more advanced semantic and empirical clustering. Cattell’s drastic simplification stimulated other researchers, of these Paul Costa and Robert McCrae, who in 1985 published their NEO PI trait-model. With 30 traits across five factors, this became known as the Big-Five model (Pervin & John, 1999).

The Big-Five model is divided into five dimensions: Extraversion, Openness, Agreeableness, Conscientiousness, and Neuroticism. Each dimension is further segmented into 6 unique sub-dimensions – or facets –, to yield more nuanced results. In general, Big-Five questionnaires are designed using questions on a Likert scale, using a system which sorts from “least/worst” to “most/best”. The points received across all facets within each respective dimension are then tallied. Notable facets for each of the five dimensions includes: warmth & assertiveness (extraversion), fantasy & feelings (openness), trust & altruism (agreeableness), self-discipline & order (conscientiousness), and anxiety & anger (neuroticism). All 30 facets are summarized in table C2 in the appendix.

The Big-Five model is considered by many a nuanced model, due to its solid mathematical foundation of objective factor analysis and clustering, focus on traits versus types, and its continuous scales. The Big-Five model does quite well in predicting various aspects in their subjects, such as psychological well-being (Anglim et.al., 2020), trends in current behavior (Fleeson & Gallagher, 2009), as well as future behavior, e.g., academic performance (Kappe & Flier, 2010)(Otero & Cuadrado & Martínez, 2020) and training success (Otero & Cuadrado & Martínez, 2020)(Dean & Conte & Blakenhorn, 2006). Although the numerical nature of the Big-Five model is one of its main strengths as a behavioral predictor, it can also be considered a negative. It is less interpretable as opposed to the Big-Four; Though the Big-Four model do loose details through the usage of either-or generalizations, its formulations are quite easy to understand.

1.2.4 – Big-Four and Big-Five; Differences and Similarities

Big-Four differ from other personality models by its theoretical focus on dichotomies. "These dichotomies are believed to reflect innate psychological or mental dispositions." (Myers et.al., 1998, p.4). The MBTI seeks to analyze its participants' stances on opposite personality categories, "both of which are regarded as neural in relation to emotional health, intellectual functioning, and psychological functioning." (Myers et.al, 1998, p.5). By pitting opposites against each other, the respondents are forced to favor one. Trait-based models instead measure personality on a continuum using traits, and individuals are defined as having either a deficit or surplus of a certain trait. E.g., an individual might have a deficit of extraversion. Contrary to this, Big-Four does not use these notions. For example, Extraverts are not seen as having excessive amounts of Extraversion, nor is the opposite believed for Introverts. Rather, they are two distinct sides to the same thematization (Myers et.al., 1998).

Another important distinction between the Big-Four and trait-based models lies in the way the questionnaires' accuracy are measured or defined. While the various iterations of the Big-Five models are grounded in pure mathematical calculations and its results are only used to describe the behavioral inclinations of the participant, the Big-Four models focus more on how the individual is expected to respond in certain situations.

The Big-Four and Big-Five models are closely related; In fact, 4 of the overarching dimensions of the Big Five model were found to significantly correlate with the four dimensions of the Big-Four model. Both (Furnham, 1996) and (McCrae & Costa, 1989) observed the strongest intercorrelation to be between the I/E- and the Extraversion-dimension (abs. avg.=-0.71). For the remaining letter-pairs, the highest correlation-scores observed were found between S/N and Openness (abs. avg.= 0.6), T/F and Agreeableness (abs. avg.= 0.46), and P/J and Conscientiousness (abs. avg.= 0.51). All correlations the two respective studies found significant are found in table C3 in the appendix. In lieu of the similarities measured between the Big-Four and the Big-Five model, the Big-Four dimensions will henceforth be linked to – and sometimes referred to as – Extraversion, Openness, Agreeableness, and Conscientiousness. High extraversion will refer to Extravert (E), high openness refers to Intuitive (N), high agreeableness to Feeling (F), high Conscientiousness with Judging (J). This is purely done for the sake of simplicity, as the pairwise linked Big-Five and Big-Four dimensions should in no way be blindly considered one and the same.

1.3 – Quantifying cognitive processes

I have now touched upon personality types, and how both Big-Four models and Big-Five models can be viewed in relation to how a person might think and act. Big-Five is a nuanced model on a continuous scale, that could be used to predict future behavior, while Big-Four models tend to generalize, and are best suited to describe the past and the now. In essence, personality type tests are tools used to quantify cognitive processes.

A previous master's study at NMBU used the personality type test Utdanningstesten, in conjunction with various other instruments, to measure the finer cognitive processes that happens during learning. The main aim of the study was to assess how the different Myers-Briggs types processed and integrated information when presented differently (Aspheim, 2020). Utdanningstesten was found to be flawed by (Aspheim, 2020), which is what prompted my work. The next chapters will thus discuss her thesis, as well as experimental design and the importance of validity and reliability.

1.3.1 – Detecting thought processes using biometric measures

In her 2020 thesis, Aspheim sought to use eye-tracking technology and galvanic measurements to examine students whilst watching educational statistical videos. A goal was to identify and quantify cognitive reactions. An interest- and personality questionnaire called Utdanningstesten was used to label the partakers with a personality type prior to the analyzes, so that the results could be linked to type. Aspheim however identified a problem with the Utdanningstesten questionnaire; Only one question – consisting of two opposite claims – was used to distinguish between the types in each dimension. This has, in turn, raised questions about the validity and reliability of Utdanningstesten as a Big-Four personality type predictor. No previous documentation regarding this exists, thus rendering any subsequent results uncertain in their implications. If we do not know the precision, recall, and/or accuracy of the instruments we use during analysis – i.e., if we do not know whether it is statistically feasible or validated –, the experiment may be rendered meaningless in its uncertainty.

This validation process has been the defining grounding stone of which my thesis objectives emerged. The following chapter will introduce Utdanningstesten, and the process of which validity and reliability can be assessed.

1.3.2 – Utdanningstesten and the importance of Validity

Before (Aspheim, 2020) raised questions about Utdanningstesten's ability to predict Big-Four types, previous studies at the University of Life Sciences (NMBU) used the questionnaire (Vinje et.al., 2021)(Sæbø & Brovold & Almøy, 2015). Utdanningstesten was created by Ph.D Psychol Helge Brovold and MD. IT Olaf Valeur for the National Center for Science Recruitment (NSR). Besides providing a four-letter personality code inspired by the MBTI terminology, this questionnaire also asks questions related to preferred learning styles and career choices. For further details about the Utdanningstesten questionnaire, see chapter 2.1.3. However, Utdanningstesten was never properly compared to other Big-Four questionnaires. Utilizing instruments or models that we know are reliable and accurate is a prerequisite when designing a statistically feasible experiment.

Validation is the act of comparing an item – this can be an instrument, model, etc. – with another, to determine whether the item is suitable for its intended purpose (Mayer & Butler, 1993). This is an important technique within the statistical field, and it presupposes that one of the items being compared is validated. Say for example that you want to compare something that you believe is a form of cheese with another product, to determine whether the first product might be categorized as such. However, for this to be a validation of the first item rather than just a comparison, the other item needs to have been acknowledged as a cheese – i.e., be validated. However, as with statistical analyzes in general, a validation might only be used to increase or decrease the items' credibility, as you can never prove something valid, only invalid (Mayer & Butler, 1993, p.22).

A wide range of approaches are available to use in validation, and which techniques to use depends on what you are seeking to validate. In this thesis the focus will be on questionnaires and rating the agreement between outputs (i.e., between tests) for the same individuals, also called the inter-specific agreement. The commonly used Cohen's kappa was chosen to assess the reliability.

1.4 – Education

In this thesis, data were mainly collected from university students. How they answered the personality- and interest questionnaire Utdanningstesten must thus be seen in context to the educational system they are part of. This chapter will therefore give a brief overview of the educational system in Norway, expanding on theory previously provided.

Due to the various advancements within technology, and the increasing accessibility to information and education, the field of education – and not least its student mass – is rapidly changing. 50 to 100 years ago, the amount of schooling mandated was severely limited compared to today, and higher education was mainly a privilege reserved for the rich or exceptionally capable. This naturally led to a rather small and arguably quite homogenous student population in organized higher education. With the increased financial living conditions that has since followed, and the increased demand for completed higher education within the work market, more people are entering higher education.

1.4.1 – The Norwegian education system and NMBU

Norway was ranked in an article by (Edsys, 2019) as one of the top educational systems in the world; some important factors stated were education being free, and a higher teacher-to-student-ratio. But what kind of educational system does Norway follow? Surely, such a well-received educational system must be quite heterogenous in what if offers regarding educational aids and methods, to successfully reach its cognitively wide audience? The next couple of paragraphs will outline Norway's educational system, first at the lower level (primary- and secondary school) and then at the upper (university- and college-) level.

The Norwegian directorate for Education and Training (UDIR) is responsible for kindergartens, primary school, and upper secondary education in Norway. They construct frameworks, exams, as well as curricula for these institutions. Several curriculum reforms have been initialized through the years, the two newest being

LK06 (2006-2020) and LK20 (2020-). Most Norwegian undergraduate students at higher level education today were largely educated following the LK06 model – assuming the median age at matriculation to be between 18 and 22 years old.

Compared to its predecessor, LK06 introduced clearer learning goals with higher focus on basic skills, longer school days (the core subjects, e.g., math, received more weekly hours), and new textbooks. Schools and teachers received more freedom of action regarding learning methods and organization. LK20 was introduced in the fall of 2020, and it presented several interesting updates and differences compared to the LK06. For example, adapted education was emphasized more, and the learning objectives were reformulated to provide room for such adaptations. E.g., the learning objectives for math was changed to emphasize discovery, exploration, and the detection and analyzing of patterns. LK20 also has a more holistic view of the subject, where math is not considered to be an isolated subject, but more so a natural part of everyday life.

Higher education, i.e., universities and colleges, have a higher degree of self-governance. Thus, the educational systems and academic structure used may vary between institutions. From my own experiences as a student within both education, biotechnology, and applied statistics & data science, certain educational methods are favored more depending on course type. In many biology- and chemistry courses practical work in the lab is an important and natural part of the teaching. Within data science, projects or submissions are usually done in limited groups, and the correct answers are mostly streamlined – at least in the early parts of the studies (DAT200, 2022)(INF200, 2022). On the other hand, courses related to engineering often encourages creativity, teamwork, and interdisciplinary cooperation (IMRT100, 2022). Contradictory, within the field of mathematics group work is usually limited and there is little room for experimentation.

The Norwegian University of Life Sciences (NMBU) offers studies across a wide spectrum. There are seven faculties, and most of the studies are centered around nature- and environmental science. Many of the courses offered at NMBU follow the traditional blackboard form of lecturing. This is cost-effective, albeit homogenous and possibly cognitively excluding. The types that are usually the most comfortable with this "traditional" form of teaching are the Introverted Judgers (Myers et.al., 1998). Most students will be able to pass in such a "IJ"-specialized class, but the lack of educational options and optimization might offset the future aspirations and achievements of some of its students, mainly creative, altruistic, and extraverted students (Vinje et.al., 2021).

However, professors' and teachers' approach to education are changing. This is both a response to massification, as well as to prepare the students for the ever-changing job market. More and more courses are incorporating more student-active activities and digitalized teaching (e.g., in the form of online lectures and the flipped classroom style of teaching), increased requirements for programming skills in studies (Sevik, 2016)(STIN100, 2022, introduced to biotechnology at NMBU in 2018), or the use of software to help illustrate various concepts (Haanæs, 2021)(KJB310, 2022). Alternate ways of structuring courses are also on the rise, a process that was accelerated (and perhaps somewhat initialized) by the pandemic.

There is uncapped potential in introducing themes of personality (/cognitive predispositions) in educational settings. In 2015, a study was performed by Sæbø, Almøy, and Brovold at NMBU which found significant differences in academic performance between certain personality types. Students were randomly selected from nine subjects, and though all these courses followed the traditional approach to lecturing, there was a mixture of mathematical and non-mathematical centered courses. In conclusion, the extraverted (E) and the more contextually focused (P) individuals showed significant tendencies of poorer performance than their counterparts. The Intuitives and Feelers were also, to some extent, found to perform poorer than their respective counterparts. Therefore, in the following chapter, I will discuss themes of personality type preferences in regard to learning strategies.

1.4.2 – Personality based preferences

Personality type discriminations can be used to help us understand how we or other people think and act in different situations. Personality is intrinsically linked to both the conscious and unconscious ways in which we think. Considering this, it is natural to assume that personality type also can be seen in relation to how the individual prefers information to be presented to them in an educational setting. By linking types and cognitive processes together, it is assumed that different leaning strategies will benefit different personality types (Sæbø & Almøy & Brovold, 2015)(Vinje et.al, 2021)(Felder & Felder & Dietz, 2013)(Myers et.al., 1998). This chapter will exemplify some of the previously noted inclinations between types and preferences.

The two studies by (Sæbø & Almøy & Brovold, 2015) and (Vinje et.al., 2021) addressed how different approaches to education may affect how well the different types learn. The latter found researched the effects of restructuring the introductory course STAT100 from standard blackboard lecturing to a flipped-classroom style. Of the previously disfavored Extraverted, Feeling and Perceiving types (Sæbø & Almøy & Brovold, 2015), Extraverts received net positive exam scores, while differences were still found for the two other dimensions. (Vinje et.al., 2021) further provided a table outlining learning styles representative for each individual letter type. (Myers et.al., 1998) also did the same. A reworked version of these two tables can be found in table C1 in the appendix.

Extraverts were described as best fit to work in an environment where cooperative behavior and discussion were encouraged. They preferred goal-oriented and experimental methods but needed to be activated by the teacher. This opposed Introverts' need for quiet introspection and their ability for self-activation. The Extrovert was also the type who benefitted the most from the flipped classroom restructuring of STAT100 (Vinje et.al., 2021). Feelers were also somewhat predisposed to like collaborative learning, and they were also experimental and dependent learners. Conversely, Thinkers were less interpersonally inclined, being data-driven and wanting a logical – albeit more abstract – flow to the subject. Thinkers learn best from goal-oriented work, and they might like tasks relating to fact-retention and repetition. Intuitives were conceptual thinkers concerned with general concepts, associations, and meanings, preferring self-directed and goal-oriented activities. Their counterpart however, Sensers, preferred to start with details and facts, moving slowly towards the abstract. S-individuals prefer experimental but concrete teaching, fact-retention – and like Extraverts and Feelers they also thrive in collaborative environments. Next, Perceivers were global ("bottom-down") thinkers, thriving in environments where they were allowed to be open to new possibilities and experiences. Together with Intuitives, F was the only type defined as innovative. Judgers on the other hand, liked to start small and work upwards (details first) in a logical and sequential manner. They might like solo-centered work with clear goals.

1.5 – Objective of thesis

Although some courses at NMBU are taking steps towards diversifying, as exemplified by the introductory statistics course STAT100, recent studies suggest that the educational structure in general does not allow for all types of students to reach their full potential. As an educator, being aware about the variation in personality types can be valuable and help make appropriate measures to how we might structure the classes, to ensure the best learning environment for a heterogenous student population. Likewise, by making students aware of their own cognitive dispositions, they may become better equipped to make conscious choices regarding their own learning.

In this thesis, I will first seek to examine the validity and reliability of the personality and interest questionnaire called "Utdanningstesten", through inter-comparative tests. Secondly, historical data sampled by the questionnaire will be analyzed to investigate the possible links between the personality types, interests, learning styles, and perception of mathematics. Previous research has found links between specific preferred learning styles and personality types (Vinje et.al., 2021)(Myers et.al., 1998)(Fairhurst & Fairhurst, 1995)(Sæbø & Almøy & Brovold, 2015). My working hypotheses are as follows:

A. There is an association between personality types and perceived difficulties in mathematics, for at least one of the four dichotomies and one (or more) letter-pairs. Intuitives, Feelers, and perhaps Perceivers may have a more negative view towards mathematics than their respective counterparts. The biggest differences may be observable for the T/F dichotomy, followed by the S/N dichotomy.
B. Different function-pairs tend to prefer certain fields of work or learning styles. E.g., group-focused work is believed to be preferred by Extraverted and Feeling individuals – perhaps also Sensers –, as these direct their focus to the observable outer world of objects and people. The T/F dimension will show the most notable inclination for altruistic or public-service related work (positively weighted for F's). Also, it is though that N's will be predisposed to creative-theoretical methods of education and fields of work. Such preferences might be detected using PCA correlation score plots or other statistical models (see point c).
C. There is believed to be an association between Big-Four letters and Big-Five scores. Any correlations will in large part support previous findings and be

between these dimensions: I/E and Extraversion, S/N and Openness to new experiences, T/F and Agreeableness, and P/J and Conscientiousness.

As this is a thesis within applied statistics, I will also:

a) Compare Utdanningstesten (Utd1) to itself (Utd2), Uroboros, Big-Five, and another Big-Four test. Inter-rater reliability is expected to be the highest between Utd1 and Utd2, and between Utd1 and Uroboros. However, comparisons using single-letter dimensions will yield much higher scores than that of full 4-letter types.
b) Compare the unsupervised PCA-method with the supervised LDA. Assuming Utdanningstesten does an acceptable job at typing the individuals, and that there are indeed differences between types' interests, LDA is assumed to do a better job at segmenting the types.
c) Create various statistical models with learning styles or math perceptions as the response. These will be compared using model statistics such as mean squared error (MSE) and mean absolute error (MAE). Variable importance's will also be analyzed. Where applicable, I expect these variables to be important predictors: Gender, age, function pair type (for math perception), and one or more of the single-dimensional types (e.g., S/N).

It is important to state that the aim of this thesis is not to find ways in which education *must* be structured for the various personality types. Rather, I will try to find out whether different approaches to education might be preferred differently by the different personality types, through analyzing the data from Utdanningstesten using various statistical methods. If these theoretical approaches were to be incorporated into the curricula, and if students are made aware of their cognitive predispositions, the students themselves would find themselves with (more) options, and the ability to make their own decisions as to how they want to learn.

2. – Methods and materials

There was two parts to this experiment. First, I wanted to assert the validity and reliability of the online personality type and learning style questionnaire Utdanningstesten. This was done by sampling data mainly from university students. Secondly, historical data from the questionnaire was analyzed, with the aim to glean further information about the connection between personality types and learning strategies, among other things. These samples were treated separately, as sample1 and sample2.

2.1 – Defining the data and outlining the project

2.1.1 – Selecting the questionnaires

When validating a questionnaire, several aspects might be interesting to look at. For example, how reliable is the test when compared to itself, or how reliable is the test when compared to another, similar test? As such four experiments, using various questionnaires, were suggested to properly validate Utdanningstesten (See table 2.1). First the test (Utd1) would be compared to itself to assert its reproductive power; This was done by participants taking the test multiple times. Next, it was compared to a test called Uroboros, also created by Brovold and Valeur. Utdanningstesten may be thought of as a smaller version of the Uroboros questionnaire. By that definition, it was important to assert whether individuals are expected to get similar test scores using both the “original” and the “derivative”. Furthermore, cross validation was performed, by comparing the questionnaire to another Big-Four and one Big-Five questionnaire respectively.

With two out of four questionnaires pre-selected, the last step was to determine which Big-Four and Big-Five test to use. The most important qualifications to select a test as a candidate were as follows: 1) Online availability; participants must be able fill out the questionnaire online and get their results immediately. 2) Well documented and feasible; documentation about how the questionnaire was developed, e.g., which theories it builds upon, as well as information about the instruments’ reliability must be readily available. 3) Affordability.

Initially, three questionnaires were chosen as Big-Four candidates: 16personalities.com, the personality type test offered at jobbsafari.no, and the official

Myers-Briggs Type Indicator test at myersbriggs.org. The first is free and is quite popular in (American-centered) social media. It denotes personality types using classic Big-Four typing, although with a fifth dimension included: Assertive/Turbulent, yielding 32 possible types. The other test, Jobbsafari, is originally a Danish questionnaire. It is also free and yields a four-letter code using the Big-Four notation. The last questionnaire, which was the most well-documented instrument of the lot, was the MBTI. However, it is MBTI is paywalled.

The approach used by 16personalities differs from Big-Four in many ways. Among them, the scales are based on reworked Big-Five dimensions rather than Jungian concepts, hence the focus on traits and the absence of cognitive functions. (16personalities B, 2021). Thus, 16personalities was discarded as an option. On another note, the personality test by Jobbsafari were based on the Myers-Briggs Type Indicator. However, there was no public documentation available to offer insight into its validity as a Big-Four test. Email correspondence with the creator of the test, Fredi Falk Vogelius, revealed that the questionnaire gives a typing that matches the official MBTI in 85% of the cases. This result was gleaned from a very small sample (n=20) of people close to Vogelius (i.e., a non-random sample). Due to this, Jobbsafari was also discarded.

The MBTI was the pioneering Big-Four model. As such, this test was the natural first choice for a Big-Four tests, if not for the unavoidable steep paywall. To use the instrument, one of two options were offered by Alexandra Schlimmer, a spokesperson for the Myers-Briggs foundation, through email-correspondence: Either one project supervisor must be MBTI certified, and we could pay 17£ per test, or the cost would be 55£ per test. Factoring in the price of becoming MBTI certified, we were looking at a cost of approximately 32000 – 45000 NOK for a subset of 50 participants. This was not feasible, and the MBTI was discarded as well, leaving no viable Big-Four options.

After further research, the Typefinder personality test at truity.com was selected as a new candidate. This test is based on the personality theory by Isabel Myers and Katharine Briggs, i.e., on Jungian concepts. (Owens, 2021) reported that, from a subset of N > 200.000 people, each of the four dimensions received Cronbach's alpha values between 0.886 and 0.937. One note of importance is that this test is based on the newer MBTI Type II, which was created by Isabel Myers in the late 1990s as an extension of the original type-indicator. MBTI Type II integrated the principles of traits from the Big-Five, adding factor-analysis selected traits to each of the four dimensions (Myers Briggs Company, 2022). Like Big-Five, the Truity test uses questions on a Likert scale and produces both the 4-letter personality code as well as information about 23 traits – the latter of which is protected behind a paywall. Due to its documentation, the Truity Typefinder was chosen as the other Big-Four test.

Lastly, the questionnaire offered at bigfive-test.com was chosen as the Big-Five test. The maintainers behind Bigfive-test are highly transparent, and have links both to the documentation, as well as their code found at GitHub. The bigfive-test was lifted from (Johnson, 2014)'s 120-item 30 facet-scales IPIP NEO-PI-R, which itself had been developed from 5 different samples and "the subsequent testing of its psychometric properties in Goldberg's (2008) Eugene-Springfield community sample" (Johnson, 2014). Though the 120-item version received alpha reliability coefficients lower than its larger counterparts, the 240-item (Costa & McCrae, 1992) and the 300-item (Goldberg, 1999), most individual facet kappa's were a maximum of 0.15 off when compared to IPIP 300. Thus, the search for questionnaires was concluded – see table 2.1 for an overview.

Table 2.1: *Tests used to compare with Utdanningstesten*

No.	Test	URL
1	Utdanningstesten	www.utdanningstesten.no
2	Big-Four	www.truity.com/test/type-finder-personality-test-new
3	Big-Five	www.bigfive-test.com
4	Uroboros	www.uroboros.as *

*Requires login credentials

As shown above, Uroboros requires login credentials to take the test. I contacted Olaf Valeur, who provided me with login credentials and introduced the system. In short, Uroboros.as has a pipeline implemented that can receive the email-address of a prospective test-taker, make a user, and send the username and password to the email provided. After logging in and finishing the test, their results are automatically sent by mail. I therefore added an "input email" function on my Simplesite website that would funnel the mails to Uroboros.

2.1.2 – Defining the samples & Application to NSD

Before any scientific work could begin, a reporting form had to be sent to and accepted by the Norwegian Centre for Research Data (NSD). This process was initiated early summer 2021, and the application was finalized some months later, in September. The two different sample groups were thoroughly defined in the reporting form. The first group (sample1) were imagined consisting of students across various fields of study and universities in Norway: from NMBU, the Inland Norway University of Applied Sciences (INN), and Østfold University College (HiOF). However, the most intense recruiting would be done to students enrolled at NMBU. Recruitment across universities was done to increase the sample size, and it was possible as this thesis would not be an intrinsic cognitive analysis of the NMBU student population, but rather a validation process that could simultaneously be used to explore student's personality types.

There were multiple reasons why students were targeted specifically. For one, my supervisors and I believe that at their age, they have come far enough in their cognitive development to be able to reflect upon the various questions maturely. Also, Utdanningstesten was created for individuals that are either in the process of, or thinking about, studying, and thus they would be more inclined to participate in our study than people with an already established career. And lastly, any changes to the educational system that I might be able to suggest in this thesis, if implemented, would be presented to the university level.

Data collected in sample1 would be age and gender, Big-Four type, in addition to numerical scores in the following Big-Five dimensions: Extraversion, Openness to new experiences, Agreeableness, and Conscientiousness. Participation would be on a voluntary basis, and the students would only have to take a minimum of 2 tests (Utd1, + Utd2/Uroboros/BigFive/Truity). The fifth Big-Five dimension of neuroticism was excluded, due to its lack of connection to any of the Big-Four dimensions, as well as to avoid the deeper territory of medical data and mental health, as neuroticism is closely linked to health, both physical and psychological (Shipley et.al., 2007)(Lahey, 2009)(Mitchell, 2016). An online form was then made at www.nettskjema.no to anonymously collect the data. An outline of the online form can be found in chapter 6.1.4.

After thorough discussion with the caseworker from NSD, it was concluded that the historical data (sample 2) did not need to be officially included in the NSD application. Although sample2 contain information about its participants – like gender, county, age, and personality type – the dataset was too large, and the information not too detailed. The data had not been gathered by us or through any systematic fetching, but rather by individuals who discovered the quiz and took it on their own free volition. All of this made it practically impossible to identify and locate any individuals in the sample.

2.1.3 – Utdanningstesten questions

Looking at the questions asked, Utdanningstesten is divided into 6 segments: Demographics, personality type, work / career choices, education and learning techniques, preference of school-related subjects, and lastly perceptions on mathematics. The latter was not initially included in Utdanningstesten, and thus the earliest rows in sample2 do not include data for these variables.

A. DEMOGRAPHICS

Demographic data gathered is gender (Male, Female, Other), age (1-12, 13-15, 16-18, 19-30, 30+), and county.

B. PERSONALITY TYPE DETERMINANTS

Before 2021, personality type was determined by four "VS"-questions, one for each dimension. The paired questions or statements represented one of the two personality-types for each respective dichotomy. From 2021 the number of questions per dimension were increased from 1 to 3. The typing is now determined by majority vote; The type that received the fewest votes (out of 3) is eliminated. The participants in sample 1 got to take the modern version of the test, while sample 2 (who only includes data from 2016 to 2020) took the test pre-expansion. To see all questions, see table D1 in the appendix.

C. WORK-RELATED INTERESTS ("part 2")

5 segments of questions, each consisting of 4 claims, are presented to the test-taker. All pertain to career aspirations and work interests, e.g., would you want a desk job or a practical job outside? The test-taker must assign a score to two of the four claims; Which claim was the most or best fitting (score = 2), and which was the least fitting (score = 0). The remaining two are then given a score of 1 (neutrality). Thus, the total set of claims in each section are not linearly independent.

D. EDUCATION AND TEACHING ("part 3")

This part of the questionnaire is comparable with the previous as it includes 3 segments of 4 claims each that are rated the same way. The statements relate to how the partaker likes education to be structured, e.g., how should lessons begin, with theory or with examples?

E. SCHOOL-RELATED SUBJECTS ("part 4")

Six statements are iteratively given to the participants, which asks for their opinion on various topics. Each statement is on the "I like …" type of format. Although not necessarily stated directly, these describes the following topics: Chemistry (e.g., product analysis, toxicology), Biology (e.g., Botany, animal sciences), Geosciences, Mathematics, Physics (e.g., Space, the laws of nature), and Structural sciences (e.g., culinary sciences, IT, physiology). Opinions are given on a 1-6 scale, where 1 means "strongly disagree" and 6 means "strongly agree". Multiple topics / statements might be rated the same.

F. PERCEPTION OF MATHEMATICS ("part 5")

This part is comprised of two segments, each with 5 claims in a Likert-style format. The partaker must choose one. The first segment is about whether you dread or look forward to math class, and the second segment is about what feelings you have about solving math problems.

For an overview of all the questions in the questionnaire, see table D2 in the appendix.

2.1.4 – The road to Sample1

Although the nature of this analysis allowed student to be recruited from a variety of courses, the ones targeted were mostly introductory courses. There were multiple reasons behind this choice. For one, the classes are usually larger as they are mandatory for several studies, which could increase the chance of student participation. Also, by focusing on courses that most students take in the beginning stages of their studies, I might capture a more heterogenous student population. This is owed to the fact that some people are bound to drop out during their first or second year at university. Students dropping out could be explained by a variety of reasons, but it often occurs disproportionately across the different personality types. Examples of types that have been linked with a higher predisposition towards dropping out are

Extraverts and Feelers (Myers, 1998, p.278), and male Perceivers (Hull, 2007)(Uslainer, 1990). Another reason is that introductory courses are usually mandatory for all students in a specific field of study. This means that the student masses are captured before differentiating into their respective elective disciplines. As a side note, one course at master-level was also added into the mix. The goal of this choice was to hopefully recruit individuals that might be more motivated and master-focused, the latter of which might make them amenable to participate in this study.

In conjunction with the NSD application, an informational letter was made, intended to be read by any prospective participant. To spread awareness about the project, a Simplesite-website was created, of which screenshots can be found in appendix's chapter 6.1.3. The informational pdf-letter was added to the website, alongside links to the various tests and the Nettskjema web form. Afterwards a short announcement message was written, introducing the project, and linking to the website (See chapter 6.1.2). This was sent to the various teachers, with requests to publish it on e.g., Canvas for their respective students to see. See Table A1 in appendix A for a table of all the courses at NMBU that was contacted, together with descriptions of the student audiences intended to be captured. A total of 8 course-responsible teachers at NMBU answered affirmative.

As my thesis was not institutionally fixated, recruitment was also done at INN (Recruitment was also attempted at HiOF, but after multiple failed attempts at contacting the school via email, this was dismissed). I contacted the administrative leaders and deans from the following INN-faculties with a request for assistance in conveying our research:

- Audiovisual Media and Creative Technologies (AMEK)
- Applied Ecology, Agricultural Sciences and Biotechnology
- Education

The vice dean of research at the latter, Susan Lee Nacey, as well as the head of studies at AMEK answered affirmative, and published the announcement through their online channels. To my knowledge, some 4th and 5th year pedagogy-students at INN (dep. Hamar) also partook in the study.

In november 2021 I was contacted by Solveig Arnesen, the CEO of Vitenparken campus Ås. Arnesen proposed a collaboration between my thesis and one of

Vitenparken's ongoing projects. After meeting, it was concluded that direct cooperation would not be directly feasible for my thesis. However, they offered to let us recruit their student staff (n = approx. 20) at their February staff meeting and have them take all tests. Now, at this stage, data had already been attempted collected for months, though sample1 still only consisted of less than 50 individuals. Of these, less than 10 had taken Utd2, and less than 15 had taken Uroboros. In the end, 9 of Vitenparken's student staff partook in the study.

2.1.5 – Get rights to Sample 2 dataset

The historical Utdanningstesten data (sample 2) was in the ownership of The National Centre for Science Recruitment (NSR) at NTNU. To receive and be permitted to work on the data, a data processor agreement was established. This agreement shall ensure that personal data is processed in accordance with the regulations and set a clear framework for how a data processor can process information. It also regulates how the various responsibilities are to be divided (Rostad-NMBU, 2022). All companies that use a subcontractor are required to have a data processor agreement (Datatilsynet, 2022). However, after informing NSR about NSD's choice to omit sample 2 from the official application, the data processor agreement was simplified to account for the fact that no specific actions needed to be taken regarding the identity protection of the participants.

2.1.6 – General data preprocessing

SAMPLE 1; STUDENT DATA

The excel-file that was generated from Nettskjema was loaded into Jupyter Notebook, and any non-consenting rows ("Samtykke"-column = "Nei") were removed. The four "Tilleggstest_" columns were combined into a new column counting the number of tests each person took in addition to Utd1. Variables with the 4-letter personality codes were made for each Big-Four type. Columns with both the percentwise and the counted number of mismatches (between Utd1 and Utd2/Truity/Uroboros) were added as well.

SAMPLE 2; HISTORICAL DATA

A csv file containing the sample2 dataset was loaded with latin1 encoding, and any duplicate rows were dropped. Wrongly coded variables were fixed (e.g., in the column "Aldersgruppe", the age "1-12" was wrongly translated into a date). Furthermore, a new column was added that combined all single letter types into one 4-letter code, before each of these single-letter columns were numerically binarized. Some additional variable tweaking was done to the sample 2 data. The personality type letters were synthesized in different manners to create six additional type-variables:

- Function pair (S/N and T/F) & Attitude pair (I/E and P/J)
- AxBx (I/E and T/F) & xAxB (S/N and P/J)
- xxAB (T/F and P/J) & ABxx (I/E and S/N)

A binary variable "semester" was added, which scored zero if the test was taken in the first half of the year and 1 if not. Furthermore, all rows dated "2010" was removed. This was done after initial visual inspection of the data showed that ~16000 rows (observations) were allegedly collected at midnight January 1st, 2010. In comparison, all other data were dated from March 2016 and later. One of my supervisors, Sæbø, informed me that these rows were ineligible, and simply a product of the creators testing the questionnaire prior to taking it online.

When modelling on data, the resulting models might be sub-optimal, also known as either overfitted or underfitted. Overfitting happens when the model is too complex, with high variance and low bias, as it starts to model the data's noise. While such a model predicts well – perhaps even perfectly – on the data used to train the model, it does not generalize well on new data (Raschka & Mirjalili, 2017). Conversely, underfitting happens when the model is so simple it is unable to effectively detect the underlying patterns of the data – neither in the dataset used for training, or on new unseen data. Underfitted models suffer from high bias and low variance. In essence, statistical modelling is about variance and bias control; The best model is the one in which variance and bias are balanced through appropriate trade-offs (Raschka & Mirjalili, 2017).

The model fitting process can be controlled using train and test datasets. This can be done either through manual data splitting or through loop-based iterative cross-validation; The latter is commonly used in combination with grid search techniques when combing for viable hyperparameter combinations in supervised learning. The sample 2 dataset was first randomly split using a set random state to ensure reproducibility. As the sample2 data was very large, “only” 60% of the rows were used when training (i.e., as train dataset). No steps were taken to ensure all types would be equally represented in train and test, due to the dataset’s size. However, manual inspection found that all train-test percentages were nearly identical.

The categorical variables were then one-hot encoded to ensure all data was on a numerical form to allow for scaling and modelling. These variables included Gender, Age group, Semester, Year, and the various personality type columns. This way, all categorical variables were treated as unordered / nominal. Despite the innate order of age, this categorical variable was also included in the “dummy coding” process, as the data was not paired.

Any columns with >= 90% missing data was henceforth removed, as they would not contain enough information to be usable (Raschka & Mirjalili, 2017). This affected only one column, of which was empty. Among the remaining variables, only the two segment 5 questions had a missingness of above 1% (arbitrarily chosen threshold); As Segment 5 was not added until December 2017, the first 41706 rows = 23.8% of the (total) data was missing. Missing data was then imputed, however Del5_1 and Del5_2 was excluded this treatment. Following the assumption that there is a link between type and math “anxiety” (of which these variables measure), I did not want to “corrupt” the variables, particularly not when they had such a high number of missing rows. In later analyses, to be discussed shortly, the rows with missing segment 5 data will either have been automatically removed before analysis by the algorithm (e.g., in the Poisson models), or segment5-data will be delegated to separate data sets (e.g., in PCA and LDA correlation score plots).

Imputation was done by using column mean values (instead of median) as no columns had problems with outliers (Raschka & Mirjalili, 2017). I.e., no columns contained Z-scores (see formula 2.1.6.1) less than -3 or greater than 3.

$$Zscore = \frac{x - \hat{\mu}_i}{\hat{\sigma}_i} = \frac{observed - mean_{columni}}{SE_{columni}} \qquad (2.1.6.1)$$

Normally when imputing this way, the mean values selected is decided on a column-to-column basis. However, as the personalities were expected to answer differently in the questionnaire, imputing the variables with no regard to typing would not be optimal. Thus, the preprocessor function I had written were changed to accommodate a grouper-variable (the 4-letter personality type) when mean imputing.

Code E2.1 and E2.2 in the appendix show the train-test splitting formula I wrote.

2.2 – Validating Utdanningstesten (Sample 1)

2.2.1 – Descriptive statistics

A function was constructed which produced a table describing the distribution of age and gender, the number of tests taken, as well as each personality type letter counted. Due to the small sample size, mean and standard deviation was used for the continuous variables, and percent scores were presented without any decimals. As a supplement to the abovementioned table, a multi-grid pie chart of letter-distributions for each Big-Four test was joined with histogram figures for the Big-Five scores. For Utd1, a bar plot was added to show the 4-letter personality types. Another table was made, grouping gender, age and types by how they answered the question "did you agree to your typing?".

2.2.2 – Comparing the questionnaires

After initial preprocessing and descriptive statistics, the various questionnaires were pairwise compared to Utd1. Cohen's Kappa was chosen to assess the inter-rater reliability. This can be thought of as the test-retest reliability of a single test-form (Cohen, 2017). The Kappa is a better alternative to e.g., correlation or accuracy as it controls for noise due to chance (McHugh, 2012). To quote (Degnan, 2017), *"Part of the idea of adjusting for chance agreement is that if some categories are naturally more likely than others, then there might be a lot of agreement due to chance"*. This statistic can be used to analyze how 2 raters, or the results from 2 questionnaires, compare.

Estimating the Kappa coefficient parameter is done by first constructing a pivot table, Z, with all possible answers along both the column and the row axis. One questionnaire ("rater") is set along the row axis, the other along the column axis.

Each cell of the table is filled with the number of individuals that they rated the same (on the diagonal) and differently (non-diagonal). Below is an example of how this Z-table could look like:

		R2		
		C1	C2	C3
R1	C1	Z_{11}	Z_{12}	Z_{13}
	C2	Z_{21}	Z_{22}	Z_{23}
	C3	Z_{31}	Z_{32}	Z_{33}

The probability of agreement, P_0, is then found by dividing the number of agreements for all "c" number of rating classes, by the total number of rated observations (Pykes, 2020). See formula 2.2.2.1.

$$P_0 = \frac{\sum_{i=1}^{c} Z_{ii}}{\sum_{i=1}^{c}\sum_{j=1}^{c} Z_{ij}} = \frac{1}{N}\sum_{i=1}^{c} Z_{ii} \qquad (2.2.2.1)$$

Next, the probability of agreement due to randomness, P_e, is calculated. This is equal to the sum of the "c" number of individual class-label probabilities, across both raters (/questionnaires) – regardless of whether they agreed or not. This is found by:

$$P_e = \sum_{i=1}^{c} P_i \qquad (2.2.2.2)$$

Whereas, for our 2 raters:

$$P_i = \frac{1}{N^2} \times \sum_{h=1}^{c} Z_{i,h} \times \sum_{h=1}^{c} Z_{h,i} \qquad (2.2.2.3)$$

The Cohen's Kappa can then be estimated using formula 2.2.2.4.

$$K = \frac{P_0 - P_e}{1 - P_e} \qquad (2.2.2.4)$$

The Kappa can take any negative value, but scores are generally only interesting when between 0 and 1. K=1 indicates perfect agreement between the raters, and K=0 indicates that any observed agreement is due to chance. Most studies would consider a kappa score of above 0.79 as strong. However, the definition can vary significantly between situations, as it depends on the maximum margin of error allowed (McHugh, 2012). For example, within the medical field the maximum margin of error accepted is usually low, as any conclusions made could directly affect the life and health of individuals. However more leniencies might be accepted in other fields, such as education – i.e., lower Cohen's Kappa values could be deemed significantly high. See table 2.2 for suggested interpretations of Cohens Kappa.

Table 2.2: *Suggested values for interpretation of Cohen's Kappa, reworked from (McHugh, 2012). The true coefficient of determination (COD, i.e., the percent of data that is reliable) is calculated by Pearson's R but can be estimated by squaring the Kappa value.*

Kappa	Cohen's suggested Level of Agreement	Level of Agreement	COD
0.00 – 0.20	None – Slight	None	0 – 4%
0.21 – 0.39	Fair	Minimal	4 – 15%
0.40 – 0.59	Moderate	Weak	16 – 35%
0.60 – 0.79	Substantial	Moderate	36 – 63%
0.80 – 0.90	Almost perfect	Strong	64 – 81%
0.91 – 1.00	Almost perfect	Almost perfect	82 – 100%

Various functions were thus made to automate the comparison process – estimating the Kappa values, constructing confidence intervals, and performing two-sided hypothesis tests on the Kappa-estimates. Retrieved from (Degnan, 2017), the confidence interval (CI) was calculated using formula (2.2.2.5).

$$CI = K \pm SE(K) \times Z_{\frac{\alpha}{2}} \qquad (2.2.2.5)$$

Whereas the Z-score is retrieved from the Z-score table, and the standard error for the Kappa is calculated by the following formula:

$$SE(K) = \frac{1}{(1 - P_e)\sqrt{N}} \times \sqrt{P_e + {P_e}^2 - \sum_{i=1}^{c} P_{i.}P_{.i}(P_{i.} + P_{.i})} \qquad (2.2.2.6)$$

As an additional note, other techniques for validating questionnaires are also available depending on intended usage. Examples of other methods are Bland-Altman plots, which can reveal if serious bias exists in questionnaires; Pearson R, which is fitting when the model is assumed to be parametric; And Spearman's Rho, for when the model violates the requirements for parametric analysis. However, estimating correlation coefficients is not always recommended, and it has been criticized for more than 20 years; a common argument is that this correlation coefficients generally does not identify systematic bias (Schmidt & Steindorf, 2006). Another parameter that could be estimated is the Intra-class correlation coefficient (ICC), which assesses how strongly units in the same group resemble each other.

See Code E1.1 and E1.2 in the appendix for how kappa-values were found and CI's calculated. See also Code E1.3 for the unidimensional kappa values, and Code E1.4 for examples.

2.2.3 – Relationship between Big-Four and Big-Five

To assess whether opposite Big-Four types (as decided by Utd1) scored differently in Big-Five, two-sample T-tests were performed. The sample consisted of all participants that took both Utd1 and the Big-Five test. By plotting the distributions as histograms and performing Shapiro-tests (data not shown), all comparisons were assumed to be normally distributed. Due to the small sample size, visual inspection was weighted most. Equal variances were also assumed, after checking the above-mentioned plots and the variance-related p-values obtained from using the "levene" function found in the "stats" Python package.

A two-sided hypothesis test was performed, with the alternate hypothesis that there was a significant difference in mean between the independent (dimension-paired) groups. See notation below.

$$H_0 : \mu_{letter1} = \mu_{letter2}$$

$$H_1 : \mu_{letter1} \neq \mu_{letter2}$$

The formula for a two-sided T-test (Løvås, 2018) is:

$$T = \frac{(\widehat{\mu_1} - \widehat{\mu_2}) - (\mu_{letter1} - \mu_{letter2})}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \qquad (2.2.3.1)$$

If equal variances, s_1^2 and s_2^2 are replaced with s_p^2, found by:

$${s_p}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad (2.2.3.2)$$

When a two-sided T-test is performed, if the absolute value of the T-statistic is above the T-table value with (1-alpha/2), and (n1 + n2 - 2) degrees of freedom – again, assuming equal variances –, then H1 can be accepted.

To perform the T-test, the "ttest_ind" formula from the Python "stats" package was used, with the argument "equal_var" set to True and "alternative" set to "two-sided".

Same-dimension pairs were compared in this manner across all four of the selected Big-Five dimensions, and a table was printed.

2.3 – Historical data analysis (Sample 2)

2.3.1 – Descriptive statistics

Various functions were made, producing summative tables from the non-imputed and unscaled data. ‘Summary_byage’ looped through the different age groups, genders, and type-letters, to subset the data and find the percentwise number of people within each age group and gender who scored E, N, F and J respectively. Another table were manually made, illustrating the distribution within age groups and by personality type letter separately, though still divided by gender. Furthermore, a map of Norway was copied from (Norgeskart, 2022), and edited to show how many test-takers resided in each county. Lastly, two figures were constructed; One histogram for the 4-letter types, and a grid of pie charts for the single-letter types.

2.3.2– Principal component analysis

The train and test datasets were then subset to only include the question-variables relating to work, education preferences, and courses. A PCA analysis was subsequently run on the train data after mean-scaling. Principal component analysis (PCA) is a widely used unsupervised statistical technique. It can mediate problems of multicollinearity, a problem which arises when two or more of our explanatory (X) variables correlate and thus are not independent (Mendenhall & Sincich, 2014). Multicollinearity can be detected by estimating the variance inflation factor (VIF), a measurement of “how much the behavior (variance) of an independent variable is influenced or inflated by its interaction / correlation with the other independent variables” (Potters, 2021). Though there is no consensus about the maximum accepted VIF, many studies agree that VIF > 5 is indicative of multicollinearity problems, and that VIF > 10 is the ultimate limit that can be accepted (Vittinghoff et.al., 2011)(James et.al., 2017)(Menard, 2001). Other usages of PCA are dimensionality reduction (of X) and trend detection.

In short, PCA transforms the original variables and creates a new subset of data. These new variables are linear combinations of the original. They are designed to be iteratively orthogonal and thus are uncorrelated (((PC1 ⊥ PC2) ⊥ PC3) ⊥ PC4 ...), whereas each component has different weights assigned to each of the original variables. The principal components are designed to maximize the variability of the data points along their respective axes, to explain as much of the trends in the original data as possible using as few axes as possible; Preserving as much of the original variability in the data is important to avoid information loss.

Obtaining principal components are done through solving an eigenvalues/eigenvectors problem, and there are two main approaches to the topic: PCA based on the covariance matrix, and PCA based on the correlation matrix. The former is used when the variable scales are similar, and the latter is used when they are not. In both methods, the data is centered. The following explanations is reworked from (Lay & Lay & McDonald, 2016, p.443-446).

Say we have a (N * p) dataset. If we denote each observational row as an X-vector of length p, our data can be written as a matrix of observations with dimension p * N (2.3.3.1). The sample mean of the observation vectors (i.e., the column means) is given by M (2.3.3.2).

$$X = [X_1 \dots X_N] \in (p \times N) \qquad (2.3.3.1)$$

$$M = \frac{1}{N}(X_1 + \cdots + X_N) \qquad (2.3.3.2)$$

The (sample) covariance matrix S is then calculated by dividing the product of the mean-deviation form (B, a p x N matrix with 0 sample mean) and the transposed mean-deviation form (B^T), with (N-1), as follows:

$$\mathrm{S} = \frac{1}{\mathrm{N}-1}\mathrm{BB}^{\mathrm{T}} \in (\mathrm{p} \times \mathrm{p}) \qquad (2.3.3.3)$$

Whereas:

$$B = \left[\widehat{X_1} \dots \widehat{X_N}\right] \in (p \times N) \qquad (2.3.3.3a)$$

$$\widehat{X_i} = X_i - M \qquad (2.3.3.3b)$$

Or in other words: To get the covariance matrix of our sample, the dataset must first be centered (by subtracting the column mean values from each observation), and then matrix multiplication is done with its transposed variant, before division by (N-1). The resulting covariance matrix, S, will then consist of the products of the (sample) standard deviations between the respective variables. For example, element $s_{12} = \widehat{\sigma_1 \sigma_2}$. The non-diagonal elements (i ≠ j) are called the sample covariances, while the diagonal elements (i = j; s_{11}, s_{22}, etc.) represent the sample variances, of each variable. The total sample variance of the data is the sum of the diagonal variances, called the trace of the matrix (2.3.3.4).

$$\widehat{\sigma^2} = tr(S) = \sum_{i=1}^{i=p} \widehat{\sigma_i}\,\widehat{\sigma_i} \qquad (2.3.3.4)$$

As a side note, such (sample) standard deviations can be calculated in Python using the "variance" function from the statistics-module (the popular numpy.var function simply calculates the population variances – a result of division by N instead of N-1).

The goal of PCA is to find an orthogonal matrix, P, that transform the original variables into a new variable-matrix Y where the new y_i-variables are "uncorrelated and arranged in order of decreasing variance." (Lay & Lay & McDonald, 2016, p.445). An orthogonal matrix satisfies the expression: $AA^T = I$, where I is the identity matrix. The transformational expression is as follows:

$$X_i = PY_i \leftrightarrow Y_i = P^{-1}X_i = P^T X_i \quad for\ i = 1, \dots, N \qquad (2.3.3.5)$$

To find the eigenvalues associated with the covariance matrix of our dataset, this equation must be solved: $det(S - \lambda I) = 0$. The generalized method is shown below.

$$det(S - \lambda I) = det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = det\left(\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}\right)$$

$$= \left| \begin{array}{cc|c} a-\lambda & b & a-\lambda \\ c & d-\lambda & c \end{array} \right. = (a-\lambda)(d-\lambda) + (b)(c) - (a-\lambda)(d-\lambda) - (b)(c) = 0$$

After finding the p different eigenvalues solving for the polynomial, the eigenvectors (v) can be found through solving the following: $(A - \lambda_i I)\vec{v_i} = \vec{0}$. Eigenvector number "i" is found by replacing the lambda variance with eigenvalue "i". After the eigenvectors are found, they are sorted in decreasing order by their eigenvalue's size.

The first eigenvector contains then the weights of the first principal component. The variance explained by each eigenvector (/principal component) is equal to the individual eigenvalue divided by the sum of all (2.3.3.6).

$$VarExpl_{(PC_i)} = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i} \qquad (2.3.3.6)$$

The Y-transformation of X (called the score) for observation "i", that has values within the "j" number of original x-variables, using the h^{th} principal component, can be found by (2.3.3.7).

$$y_{i,PC_h} = \overrightarrow{v_h} \times \widehat{X_i} = \sum_{j=1}^{p} \overrightarrow{v_{j,h}} \times \widehat{X_{i,j}} \qquad (2.3.3.7)$$

Luckily, the PCA function from the Python Sklearn.decomposition-package can do all the heavy lifting for its users. A PCA, solving for the maximum number of k=p-1 principal components, was fit on the standard scaled train dataset. Standard scaling (i.e., centering and subsequently dividing by the column standard deviations to get variables that are ~ N(0, 1)) is important to do before applying PCA, especially when the scales across variables are different, as the PCA directions are highly sensitive to data scaling; Not doing so would cause the PCA algorithm to unfairly assign unequal importance's to the variables (Raschka & Mirjalili, 2017).

After the analysis, which yielded a scores dataframe of the transformed data and a loadings dataframe with the eigenvectors, correlation scores and -weights were calculated. Whereas scores are calculated from linear combinations between the observational values and eigenvector-values ($score_{obs.i,PCk} = \sum_{j=1}^{p} u_{j,k} * x_{i,j}$), correlation scores are found by taking the correlation between the respective eigenvector and the non-transformed observational vector ($corr.score_{obs.i,PCk} = corr(\overrightarrow{u_k}, \overrightarrow{x_i})$), both with p length. The resulting dataframe will have N rows and the same number of columns as PC's chosen to evaluate. Correlation scores assesses the strength between the observations and the respective principal components. Below is mock Python-code to show how this can be calculated. Pearson's correlation coefficient was used.

```
for PC_columnvector in loadings_dataframe:
    for obs_rowvector in original_dataframe:

        r = Corr(PC_col, obs_row)
```

Correlation weights are quite similar. They are found by taking the correlation between the PC-scores vectors and the vector containing the observational values for all observations for each variable. Both vectors have length N, and the resulting dataframe will have q rows and the same number of columns as PC's chosen to evaluate. If the original dataframe is used (i.e., the same dataframe that was transformed during the PCA), the number of rows will be q = p. However, any Z-dataframe with the same N observations can be used in place of the original dataframe. To calculate, use the mock code above and change "loadings_dataframe" to "scores_dataframe", and "obs_rowvector" to "variable_columnvector".

2.3.3 – Linear Discriminant Analysis

Often used in conjunction with classification, LDA is a dimensionality reduction method that is quite similar with PCA. "Whereas PCA attempts to find the orthogonal component axes of maximum variance in a dataset, the goal in LDA is to find the feature subspace that optimizes class separability." (Raschka & Mirjalili, 2017, p.155). LDA is defined as a supervised method, as it presupposes that a categorical response variable is provided. While it might be intuitive to think LDA thus is superior to PCA, it is important to note that this will not always be the case (Raschka & Mirjalili, 2017).

The assumptions of LDA are as follows:

1. Normality: $X_{data} \sim N(\mu, \sigma)$
2. Equal covariance matrices: $S_{class1} = \cdots = S_{classQ}$
3. Independence between variables (no multicollinearity problems)

However, "even if one or more of those assumptions are (slightly) violated, LDA for dimensionality reduction can still work reasonably well" (Raschka & Mirjalili, 2017, p.156).

Analogous to PCA, the LDA algorithm also starts by centering the data points, but they are then also scaled using the sample standard deviations – i.e., standardization. For each "i" class, a p-dimensional mean vector is then found by summing the p-dimensional observational vectors and dividing these by the number of group-observations – see (2.3.4.1).

$$m_i = \frac{1}{n_i}(X_1 + \cdots + X_N) \qquad (2.3.4.1)$$

Using the above vectors, two scatter matrices are made; One for within-class (S_W) and one for between-class (S_B). The within-class scatter matrix is found by summing the individual scatter matrices of each individual class "i" (Raschka & Miralili, 2017), see (2.3.4.2). The scatter matrices for an "i" group are found by (2.3.4.3) – the formula is also written with the B-matrix notation used in the PCA-chapter.

$$S_W = \sum_{i=1}^{c} S_i \qquad (2.3.4.2)$$

$$S_i = \sum_{i=1}^{c} (X - m_i)(X - m_i)^T = \sum_{i=1}^{c} B_i B_i^T \qquad (2.3.4.3)$$

The two formulas above assume that the class labels are uniformly distributed. This was not the case with the sample 2 dataset, as certain personality types were more abundant than others. To fix this whenever detected, the Si-scatter matrices are scaled by the group-based observations, which creates a covariance matrix much like the one found in (2.3.3.3). See (2.3.4.4).

$$\Sigma_i = \frac{1}{n_i} S_W = \frac{1}{n_1} \sum_{i=1}^{c} B_i B_i^T \qquad (2.3.4.4)$$

Computation of the between-class scatter matrix are a tad bit easier and does not rely on double summation. Instead of basing the calculation on subtracting the group-specific sample mean vectors from the dataset, the full sample mean vector is subtracted from the group-specific vectors. The outer summation is replaced by multiplication with the group sample sizes. See (2.3.4.5).

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \qquad (2.3.4.5)$$

The remaining steps, which involves finding the eigen -values and -vectors, are similar to PCA. But instead of using the covariance matrix (S, in chapter 2.3.3), these values/vectors are solved for the matrix $S_W^T S_B$. Furthermore, while the maximum number of linear discriminants in PCA was p-1, the maximum for LDA is c-1 (where p is the number of original variables, and c is the number of individual class labels) (Raschka & Miralili, 2017).

Linear Discriminant Analysis was carried out on standardized train-data using Scikit-Learn's "LinearDiscriminantAnalysis" function from the discriminant_analysis sub-pack. Three analyses were made: One with 4-letter type as the response, one with Function-pair as the response, and one with Attitude-pair. The data set was transformed to create the equivalent to the scores data frame, by using the (p x C) dimensional data frame (where $c_1, \dots, c_C$ are the eigenvectors and C = c-1 maximum). Heatmaps, scatter plots, and density plots were made to illustrate how different types had answered Utdanningstesten.

2.3.4 – Modelling with Original data, PCA, and LDA

For this part of the analysis, data from the age groups 1-12 and 13-15 were removed. This was done to better represent the target sample of this thesis, i.e., college and near-college age individuals.

SELECTING NUMBER OF COMPONENTS

After performing PCA and LDA, 3 new data sets were made, all with correlation scores for the previously defined test-dataset: One from the PCA (PC1-14), one from the LDA with 4-letter type (LD1-9), and one from the LDA with function-pair-type (LD1-2).

Though there are many ways to choose the number of components to further analyze, there is no consensus about which method is the best (Hartmann & Krois & Waske, 2018). An easy way to select the number of components is by visually inspecting the scree plot: When the plot flattens out, i.e., when the amount of variance explained by the subsequent components are low and about the same as each other, we would say the optimum number of PC's have been reached. This might be called the plot's «elbow». Another way is by setting an arbitrary level of cumulative variance we want our new variables to explain and pick the number of components that surpass this threshold. Values between 70-90% is common, but it depends on the type of data and problem at hand (Hartmann & Krois & Waske, 2018). Kaiser-Guttman's criterion/rule is also widely used. It states that "any principal component with variance less than 1 contains less information than one of the original variables and so is not worth retaining" (Hartmann & Krois & Waske, 2018).

Using the abovementioned principles, the first 14 components from the PCA were chosen; The first five accounted for the scree plot's "elbow", PC1-14 explain 70% of the variability, and all components had eigenvalues >= 1, thus following Kaiser's rule. Furthermore, in the LDA-TYPE analysis, the first 9 components were chosen: The first 4 accounted for the "elbow", while all described ~ 99.6% of the total variability. In the LDA-Function analysis, 2 components were chosen both due to the "elbow"-effect and them cumulatively describing ~99.6% of the variability.

PERCEPTIONS OF MATHEMATICS

Various models were where the dependent variables of interest were segment 5's question about math "anxiety" pertaining to math class (Del5_1), as well as other variables that directly asked the participants about their perceptions towards mathematics. An overview of the models that were made is presented in the table below.

Table 2.3: *Overview of statistical models made on the test data (40% of the original data set, randomly chosen – the same test data set as used in PCA). "ABxx", "xAxB", and similar names are used to describe paired personality letter types; the position of "A" and "B" in the 4-letter names refer to the position of the dimensions used (I/E – S/N – T/F – P/J). A colon between variables means that the interaction term was added alongside the individual variables. "SAB" mean "same (text) as above".*

NO.	Y	X	Analysis
M1	Del5_1 (math class)	"ABxx" : "xxAB" Year : Age group Gender Semester	Multiple linear regression. Discrete y treated as continuous. Cat. expl. variables set as factors.
M2	SAB	SAB	Generalized multiple linear regression. Discrete y with Poisson distribution. Cat. expl. variables set as factors.

In the models, discrete variables (such as "Age group") were encoded as factors, and the most common group within each variable were set as the reference level. Although it is not strictly necessary to set the common groups as reference, it is good practice to do so, as stated by my main supervisor. This way, the intercept of the model better reflects the average sample population. Reference levels used were: "ABxx" = IS, "xxAB" = FP, "xAxB" = SJ, "AxBx" = IF, "xABx" = SF, "AxxB" = IJ, Age group = 19-30, Gender = Female, Year = 2016, IE = I, SN = S, TF = F, PJ = P. This way, R converted the variables into dummy variables during the modelling process, treating each variable as nominal and not ordinal. The reference level was modelled into the intercept of each respective model.

The two models used were linear regression and a generalized linear regression model with a Poisson distribution. The formula for linear regression is shown in (2.3.5.1). The model assumes that the response is continuous numerical, that the residuals have a mean of 0, a constant variance, and are independently and identically distributed.

$$y_i = \mu + \sum_{j=1}^{p} \beta_j x_{i,j} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma^2)\ iid \qquad (2.3.5.1)$$

The Poisson regression model shares some similarities with linear regression, though it introduces the lambda-variable, and it does not have a residuals-variable in its formula (2.3.5.2). Model assumptions are: The response is a count variable, the variables are independent (i.e., constant variance), the mean and variances are equal, and there is linearity.

$$log(\lambda_i) = \mu + \sum_{j=1}^{p} \beta_j x_{i,j} \qquad \lambda = E(X) = Var(X) \qquad (2.3.5.2)$$

The models' prediction accuracy was compared using both mean absolute error (MAE, 2.3.5.1) and mean squared error (MSE, 2.3.5.2). MAE evaluates the absolute difference between observations and their predictions. This causes negative errors to be weighted the same as positive ones. MSE instead squares the distances, which means that higher errors (i.e., distances) weigh more than the lower ones.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| y_i^{true} - y_i^{pred} \right| \qquad (2.3.5.3)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(y_i^{true} - y_i^{pred} \right)^2 \qquad (2.3.5.4)$$

Furthermore, GLM-Poisson models were also made to assess the relationship between math-related variables and personality types – both singular dimension types and complementary letter pairs. Parameters were estimated using R's summary-function on the model-object, and deviance-scores were found with the anova-function. The deviance represents the error of prediction, e.g., when the true and the predicted is the same, the deviance will be 0. When a new explanatory variable is added to a model, the model's prediction error might be reduced, and the amount of unexplained variance can decrease. A variable deviance score describes to what degree the variable added value to the model; A high variable-deviance score is desirable. Formula (2.3.5.5) show the general method of calculating the deviance of a model.

$$Deviance = 2\sum_{i=1}^{N}\left(y_i \times log\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right) \qquad (2.3.5.5)$$

The following table lists the models that were made.

Table 2.4: *Overview of GLM-Poisson models made, which variables were used and description of the response-variables.*

Response (Y)		Model variables (X)
Name	Description	
Del5_1	Perception of math; Math "anxiety" in relation to math class	A) TF-PJ & IE-SN pairs B) SN-TF & IE-PJ pairs C) SN & PJ & TF & IE D) SN-PJ & IE-TF pairs
Del2_5d	Would want to find a job where they can work with math and formulas	
Del4_4	From 1-6, how they rate math	
Del5_2	Perception of math; Math "anxiety" in relation to solving tasks + Thoughts about own math skills	

LEARNING STRATEGIES

Generalized linear models with Poisson distributions were also constructed with a selection of segment 3 questions as the responses. This segment contained questions about learning preferences. Different combinations of personality type variables, either singular or complementary paired types, were added as explanatory variables (see table 2.4 for an overview of the models). The latter were encoded as factors in which the most common group per variable were set as the reference level.

Parameters were estimated using R's summary-function, and deviance-scores were found with the anova-function

Table 2.5: *Overview of GLM Poisson-models made to assess parameter estimates and deviance scores in relation to learning strategies.*

Name	Description	Model variables
Del3_1d	Working in groups	A) TF-PJ & IE-SN pairs B) SN-TF & IE-PJ pairs C) SN & PJ & TF & IE D) SN-PJ & IE-TF pairs
Del3_1b	Teacher presenting methods and examples before individual work	
Del3_2c	Like problems where there are multiple solutions & creativity is encouraged	
Del3_2a	Teacher starting with examples from everyday life before divulging into theory	
Del3_3b	Like teachers who teach about methods and rules	
Del3_1a	Like when teacher starts with giving tasks, rather than a theory lecture	

TYPE-SEPARATION; PCA vs LDA

To check whether the PC- or the LD-components did a better job at separating the different types, classification models were made for each analysis (PCA 1-14, LDA-TYPE 1-9, and LDA-Function-pair 1-3). The correlation scores (made from the PCA and the LDA) that were calculated from the (original) test-dataset were used as predictors, in models with type (4-letter or function pair) as the categorical response. To properly evaluate the models, the test-data from previous were split into two subsets, wherein all 4-letter types were equally represented. These two data sets will henceforth be called the train and the test data. See the appendix's Code E2.3

The logistic regression method – with standard scaling prior to fitting – was selected for the PCA-models. The same method was used for the LDA-models to ensure the models were otherwise as similar as possible. This selection was done by performing a K-fold cross validation on a temporary sample set of n=1000 individuals from the (correlation scores-) PCA dataset. In the CV-process, 4-letter type was set as the response, 3 folds were done, and the f1-micro score was used to evaluate the splits. Logistic regression, Support Vector Classifier (SVC) and XGBoost were analyzed, and the best model was selected based on the metric's mean and standard deviation.

Despite its name, the linear Logistic regression method is used for classification tasks. It "tries to maximize the conditional likelihoods of the training data" (Raschka & Miralili, 2017, p.81). Though this means the model is somewhat prone to outliers, one huge advantage is its easy implementation. While it is intended for binary classification, it can be extended to multiclass classification via e.g., the OvR-method (one versus rest).

In short, Logistic regression uses the odds ratio (OR), which is the probability of the "positive" event happening divided by it not happening. The logit function is then defined, i.e., the natural logarithm of the OR, which is the linear combination of the model-assigned weights and the observational values for the individual observational rows. Function 2.3.5.1 show the formula written for one such observational row.

$$logit(p) = log(OR) = log\frac{p}{(1-p)} = \sum_{i=1}^{p} w_i x_i = w^T x \qquad (2.3.5.1)$$

The probability that a sample belongs to a certain class is then the inverse form of the logit-function (Raschka & Miralili, 2017), called the sigmoid function (2.3.5.2).

$$\phi(z) = \frac{1}{1+e^{-z}} \qquad (2.3.5.2)$$

Where:

$$z = w^T x \qquad (2.3.5.3)$$

$$\hat{y} = \begin{cases} 1 & if \phi(z) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \qquad (2.3.5.4)$$

In (2.3.5.4), the first condition is also true if z is larger than or equal to zero.

To find the weights (w), the aim of the Logistic regression method is to maximize the likelihood. This can be rewritten to focus on minimizing a cost function defined as the negative log likelihood (2.3.5.5).

$$J(w) = \sum_{i=1}^{N} \left(-y^{(i)} log\left(\phi(z^{(i)})\right) - (1 - y^{(i)}) log\left(1 - \phi(z^{(i)})\right)\right) \qquad (2.3.5.5)$$

Extending the Logistic regression to support multiclass problems is done by default via the OvR-technique (“One versus Rest”, also called the OvA, “Ove versus all”). In it, one model is trained per “c” class, iteratively treating the class of focus as the “positive” label. This yields “c” number of classifiers, which individually predicts each label probability; the class label with the highest confidence is thus chosen (Raschka & Mirjalili, 2017).

The models that were made were:

- 4-letter type ~ PC1-14
- Function type ~ PC1-14
- 4-letter type ~ LD1-9
- Function type ~ LD1-2

Whereas the response, Y, is before the tilde, and the X-variables are after.

To see how the K-fold cross validation was done, or how the models were tuned, see Code 2.4 in the appendix.

3. Results

Chapter 3 is divided into three sections: First general descriptive statistics are presented for the two samples' datasets. Secondly, data from sample 1 is analyzed to assess the validity and reliability of Utdanningstesten. In the third chapter, the historical data from sample 2 is explored.

3.1 – Demographics & Descriptive statistics

3.1.1 – Sample 1

A table was made with various descriptive statistics for all variables in the data set. See table B1 in the appendix.

Apart from the initial Utdanningstesten (Utd1), the number of individuals who took each test varied, as did the distribution of each personality type letter. Various statistics relating to demographics, personality type distributions, and the number of additional tests taken are shown in table 3.1 below. The distributions of Big-Four letters and Big-Five scores are further illustrated in Figure 3.1.

Table 3.1: *Various descriptive statistics divided by the Jungian personality type letters. Lists the observed number of individuals within each letter for each of the Big-Four related tests, as well as how many tests the participants took in addition to Utd.1. The Big-Five rows, with their own header, show the score-distributions for each of the four focus dimensions. For easier interpretability, these are grouped right below the letter-columns that "matches" up with their respective dimensions (Furnham, 1996).*

	TOT									
Gender; n (%)	K M O TOT	42 (77%) 11 (20%) 1 (1%) 54								
Age; Mean (sd)	23.8 (4.1)									
	TOT		I	E	S	N	F	T	J	P
Utd.1; n (%)	54		37 (68%)	17 (31%)	40 (74%)	14 (25%)	33 (61%)	21 (38%)	33 (61%)	21 (38%)
Utd.2; n (%)	13		6 (46%)	7 (53%)	8 (61%)	5 (38%)	6 (46%)	7 (53%)	7 (53%)	6 (46%)
Uroboros; n (%)	17		10 (58%)	7 (41%)	10 (58%)	7 (41%)	6 (35%)	11 (64%)	13 (76%)	4 (23%)
Truity; n (%)	37		22 (59%)	15 (40%)	13 (35%)	24 (64%)	29 (78%)	8 (21%)	23 (62%)	14 (37%)
Tests taken (Mean)	1 2 3 4 Mean	6 28 14 6 (2.4)	4 19 10 4 (2.4)	2 9 4 2 (2.4)	4 22 11 3 (2.3)	2 6 3 3 (2.5)	5 17 8 3 (2.3)	1 11 6 3 (2.5)	5 16 9 3 (2.3)	1 12 5 3 (2.5)
	TOT		Extraversion		Openness		Agreeableness		Conscientiousness	
Big-Five; Mean (sd)	44		72.9 (16.3)		80.4 (11.4)		92.5 (11.4)		84.8 (16.7)	

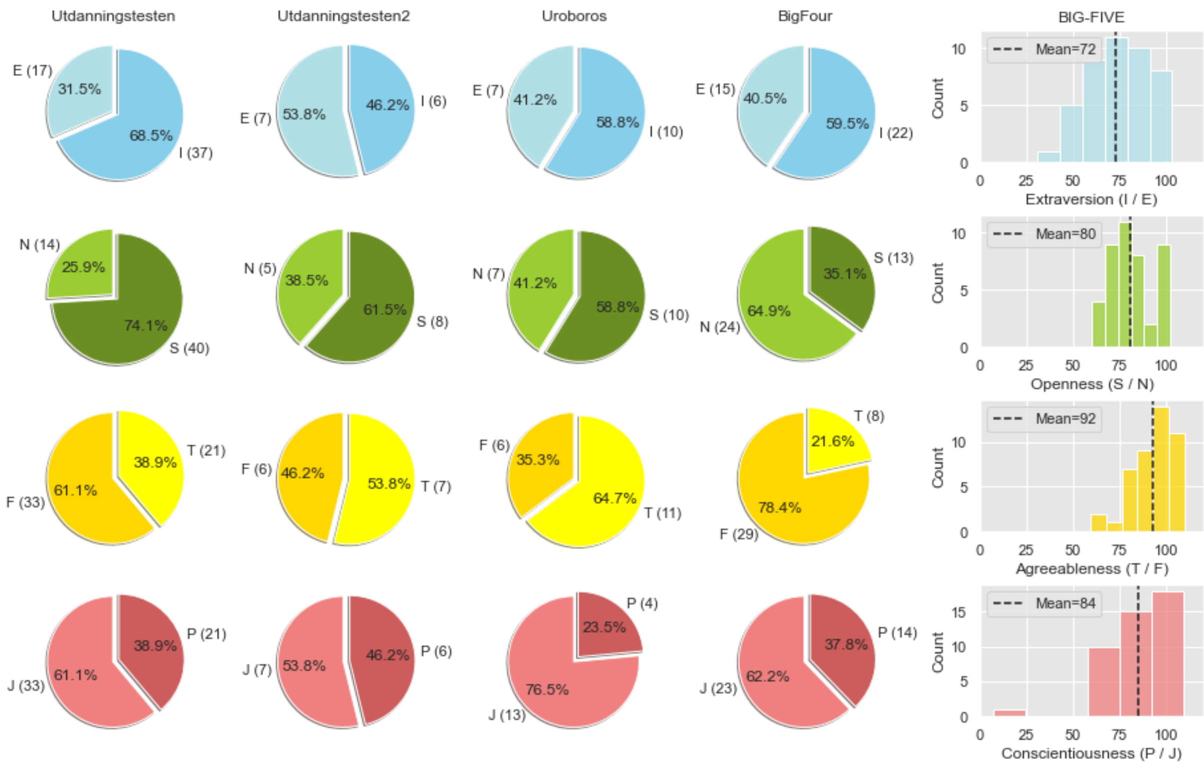


Figure 3.1: *Pie charts showing the distribution of each letter-pair for all tests, with the count in parentheses and the percentwise distributions on the pies. "Utdanningstesten2" is the subset of individuals that took the Utdanningstesten a second time (Utd.2), and "BigFour" is the subset that took the Truity Typefinder test. The pies are colored by dimension, using Big-Five terminology, in order: Extraversion, Openness, Agreeableness and Conscientiousness. In the Big-Five histograms, the mean (black dotted line) is also plotted; the median was approximately the same as the mean for each respective dimension.*

Excluding the Big-Five test and comparing the others, introverts and judgers were consistently found to be more abundant in our sample as opposed to their respective counterparts. The remaining two dichotomies had more variations between tests. Most notable are the differences between the S/N for Utdanningstesten (Utd.1) when compared to the Truity Typefinder Big-Four test; The unbalanced distributions flipped from 71/29 % (S/N) in Utd.1 to 28/72 % (S/N) in the other Big-Four test. The Big-Five extraversion plot showed an approximate normal distribution with a mean of 69 points. The other Big-Five dimensions did also mostly show normally distributed data and mean values on the higher half-end of the scale. Focusing on the Utd.1 test (first column), the most abundant types were: I, S, F and J.

To see the distribution of each 4-letter personality type, see the bar chart in Figure 3.2.

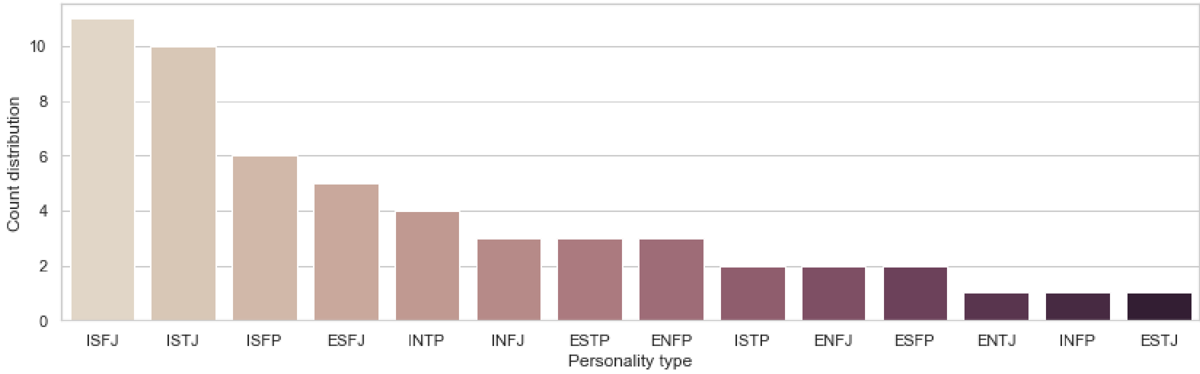


Figure 3.2: *Distribution of personality types as sampled from Utd.1, i.e., the total pool of participants. 13 of the 16 Big-Four personalities are represented in the sample. The missing types are: INTJ, ESTJ, ENTJ.*

The most frequent type was ISFJ, with 11 participants, followed by ISTJ at 8 participants. Both types had the Introverted, Sensing and Judging type. Furthermore, the top 6 types were all Introverted, missing but 2 of the 8 total number of introverted personality types. The least popular types were by default INTJ, ESFP and ENTP, with none receiving these typings.

When the voluntary participants filled out their results on the online form, they could choose to answer a question asking how they agreed with their type / type description. The following table show the distribution of age, genders, and personality type letters, divided by how they answered this question.

Table 3.2: *Descriptive statistics for sample 1, both in total and grouped by each letter (as assigned by Utd1). Participants are also grouped by how they answered the question "Did you think the personality type you received was fitting?" (Good, OK, Bad). As only 2 people answered the latter, the "Bad" column was removed. The two in question were both females, with a mean age of 25. One was ISFP and the other were ISTP, i.e., both ISxP. The "Total" and the type-divided rows includes both the count and percent-count.*

		TOT	Good	OK
Total		54	28 (51%)	24 (44%)
Age; Mean (sd)		23.8 (4.1)	24 (4.7)	23.5 (3.3)
Gender	Female Male Won't specify	40 11 1	22 6 0	18 5 1
Extraversion	E	17	11 (64%)	6 (35%)
	I	37	17 (45%)	18 (48%)
Openness	N	14	7 (50%)	7 (50%)
	S	40	21 (52%)	17 (42%)
Agreeableness	F	33	17 (51%)	15 (45%)
	T	21	11 (52%)	9 (42%)
Conscientiousness	J	33	19 (57%)	14 (42%)
	P	21	9 (42%)	10 (47%)

Table 3.2 show that in total, nearly as many people answered neutrally as affirmative. 6 out of 11 (55%) of males answered positively, while the same percent (22 out of 18) of females did the same. The mean-age was somewhat higher for the "Good" group. Table 3.2 also show some differences between types. Of the 17 extraverted students, two thirds were positive, while the same was only the case for less than half of the Introverted. In the same manner, Intuitives, Feelers and Judgers were more positive than their counterparts.

3.1.2 – Sample 2

Sample 2 were sampled in the period 2016 (March) and until the end of 2020 and contain individuals from across the country. It consisted of N=175287 individuals. Table 3.3 show the distribution of each letter-type and age group, divided by gender. Furthermore, Figure 3.3 show a map of Norway, highlighting the percentwise number of people from each county as well as the relative number of participants (per 1000 inhabitant).

Table 3.3: *Table show the distribution of personality types and age groups, both in total and segmented by gender. The Personality type distributions show the amount of people within each subset (all, females, and males) categorized as each type, as well as the amount of people within each type that was each gender. The latter, placed in parentheses, is shown relative to the total percentwise amount of each gender; E.g., if 70% of the total subset were female (30% male) and 71.5% of all introverts were female, then the result will be +1.5 and -1.5 for females and males respectively, colorized red if negative. To avoid redundancy, these numbers (Δ%) are only shown for the females.*

Personality / Gender	TOT	I	E	S	N	T	F	P	J
All	175287	50.51%	49.49%	62.75%	37.25%	38.55%	61.45%	50.16%	49.84%
Female; % (Δ% type)	121265 (69.18%)	50.96% (0.61)	49.04% (-0.62)	64.47% (1.89)	35.57% (-3.19)	35.12% (-6.15)	64.88% (3.86)	46.16% (-5.52)	53.84% (5.56)
Male; % (Δ% type)	54022 (30.82%)	49.52%	50.48%	58.9%	41.1%	46.24%	53.76%	59.14%	40.86

Age group / Gender	1-12	13-15	16-18	19-30	30+
All	2.16%	23.65%	31.34%	35.47%	7.37%
Female	2.22%	21.6%	33.24%	35.29%	7.65%
Male	2.04%	28.26%	27.07%	35.89%	6.74%

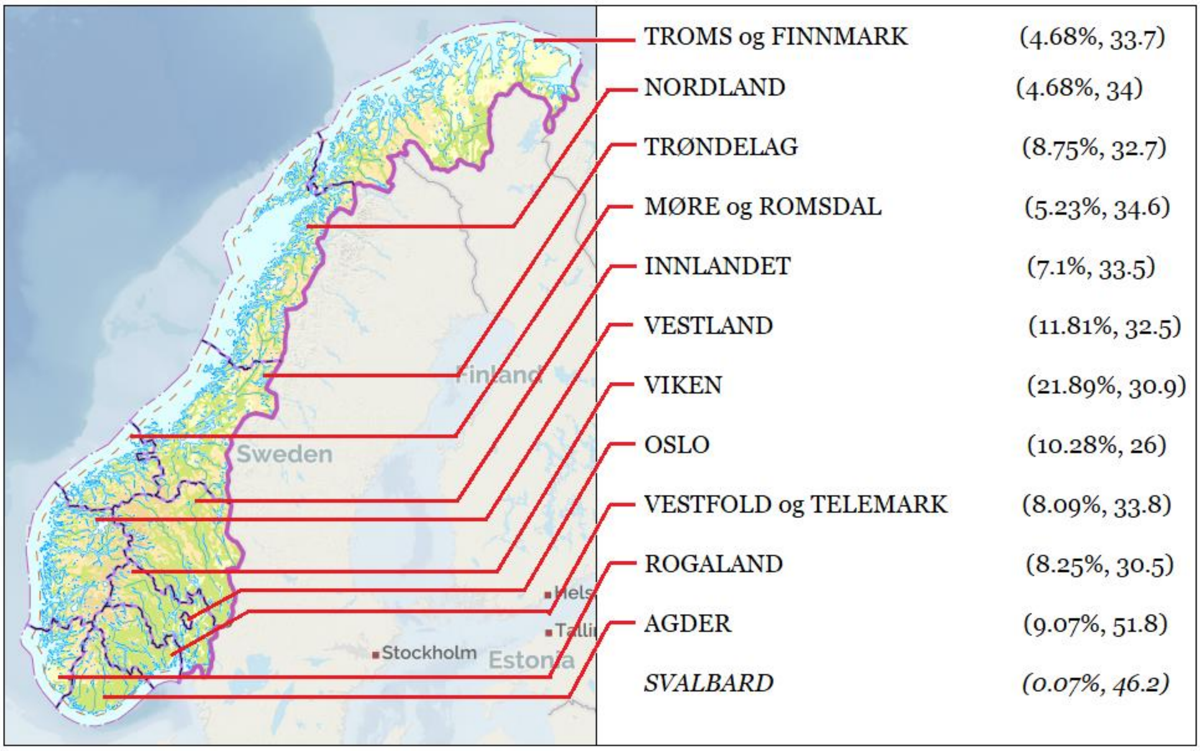


Figure 3.3: *A map of Norway showing the percentwise number of participants from each of the 12 counties (as of late 2020), as well as the relative number of participants from each county (number of people per 1000 inhabitant).*

Every fifth participant lived in Viken county, with an estimated 30 people per 1000 inhabitant (assuming no replicates). Nearly 12% were from Vestland county, while only 0.07% (or n=123) people were from Svalbard.

As shown in Table 3.3 show, more than two thirds of the participants were female. Furthermore, 66.8% were between the ages of 16 and 30, i.e., most were (likely) in the establishing phases of their life (e.g., high school, higher education). Moreover, it showed an approximately equal distribution for the following two dimensions: Introvert/Extrovert and Perceiving/Judging. The former occurred between both genders, but in the latter the males were overrepresented in the Perceiving type.

Approximately 63% were Sensers as opposed to the ~37% Intuitives. Males had a higher frequency of Intuitives (N) than women, with 41.1% compared to 35.5%. Feelers (T/F) dominated in all three groups, but particularly in the female group, where two thirds were Feelers. Men however were closer to an even split, with a slight lean of 3.8% in favor of Feelers. Reversely, this was also the case for women in the P/J-dimension (Conscientiousness), with exactly 3.8% in favor of Judgers. In men the reverse was observed, having a predominance of Perceivers as opposed to Judgers. Thus, the P/J-dimension was the only in which the genders were clearly in disagreement about the most common type. Figure 3.4 show in turn how this affected the frequencies of the 4-letter types.

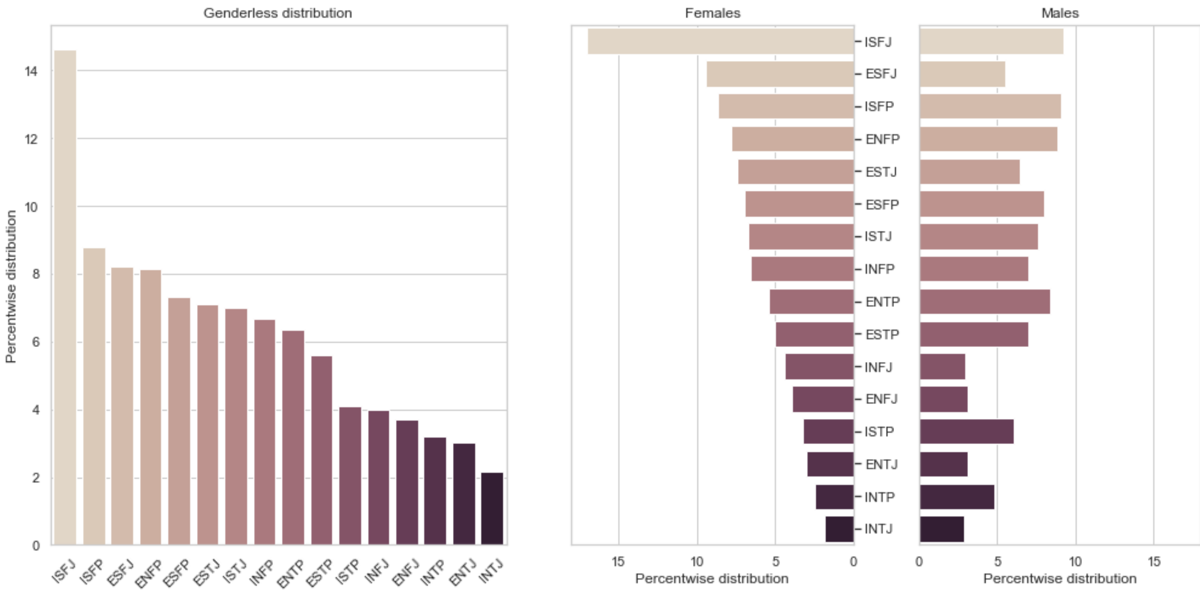


Figure 3.4: *The percentwise distribution of every 4-letter personality type, presented both in total (left), and grouped by gender (right). The gender-divided distributions are ordered decreasingly according to the females-distribution.*

With more than 14% support, the most frequent type was ISFJ; This was more than twice the amount expected if all 16 types had been equally distributed. Following this was ISFP, indicating that ISFx-types were the most recurrent in our sample. The most unique types were xNTJ, with INTJ only having a support of about 2%. When the types were separated by gender, the two most frequent types and the least frequent type were the same as in the total (genderless) histogram. Similarly, the most and the least frequent type for males were the same as the genderless distribution (ISFJ and INTJ). Males were also somewhat more evenly spread across all types compared to the females.

Though ISFJ was number 1 for both males and females, the type was almost twice as common in females. The genders were about equally distributed for ISFP, ENFP, ISTJ, INFP and ENTJ. Besides ISFJ, females had almost twice as many ESFJ and INFJ than men, while men had double the amount of INTP and ISTP (IxTP) than women. The table below show how males, females and the total sample were dispersed according to function pairs and attitude pairs. The biggest differences between the genders were noted for the function “SF” (10.3% more females), and for the attitude “IJ” (7.5% more females). These results are consistent with the observation of the most frequent female type being ISFJ.

Table 3.4 show the number of individuals within each function- and attitude-pair. The most common function pair was the SF, though the sample had 4.5 percentwise points less individuals than the American sample in (Myers et.al., 1998). In comparison, sample 2 had nearly double the amount of NF’s compared to this study. The attitude pairs were more evenly spread, each having between 22.1 and 27.8 percentwise points.

Table 3.4: *The distribution 4-letter types, as well as function and attitude pairs in sample 2 – both in total (first value) and divided by gender (females second, males third). Values in parentheses show the differences in percentwise points between this sample and the American sample examined by (Myers et.al., 1998), colored red/green if the difference is less than -2 or above +2.*

Function						
ST	SF	NT	NF			
23.8% (-6.2) 22.4% 27.1%	38.9% (-4.5) 42.1% 31.8%	14.7% (-1.8) 12.7% 19.2%	22.5% (+12.1) 22.8% 21.9%			
ISTP 4.1% (+1.3) 3.2% 6%	ISFP 8.8% (0) 8.7% 9.1%	INTP 3.2% (-0.1) 2.5% 4.8%	INFP 6.7% (+2.3) 6.6% 7%	22.8% (+0.9) 21% 26.9%	IP	Attitude
ISTJ 7% (-4.6) 6.8% 7.6%	ISFJ 14.61% (+0.81) 17% 9.2%	INTJ 2.2% (+0.1) 1.8% 2.9%	INFJ 4% (+2.5) 4.4% 2.9%	27.8% (-1.2) 30.1% 22.6%	IJ	
ESTP 5.6% (+1.3) 5% 7%	ESFP 7.3% (-1.2) 7% 8%	ENTP 6.3% (+3.1) 5.4% 8.4%	ENFP 8.2% (+0.1) 7.8% 8.9%	27.4% (+3.3) 25.3% 32.2%	EP	
ESTJ 7.1% (-1.6) 7.4% 6.4%	ESFJ 8.2% (-4.1) 9.4% 5.5%	ENTJ 3% (+1.2) 3% 3.1%	ENFJ 3.7% (+1.2) 4% 3.1%	22.1% (-3.2) 23.8% 18.2%	EJ	

The Utdanningstesten questionnaire consisted of 4 segments of non-personality type related questions, with a total of 60 questions (or claims). In table 3.5 below, an abbreviated summary of the results is showcased. Highlighted are the questions/claims wherein the majority of one type (or more) answered either in the negative or the positive. The full table can be found in table B2 in the appendix.

Table 3.5: *A summary of Table B2 (appendix), highlighting only questions where one or more personality type answered particularly high/low (ignoring the neutral answers). "Negative; Neutral; Positive" answer labels are {0; 1; 2} (segment 2 and 3), {1,2; 3,4; 5,6} (segment 4), and {1,2; 3; 4,5} (segment 5). Highlighted in the table below are questions which received scores that were at least 20% in favor of negativity/positivity, grouped by personality letter. If no favoritism, the expected distribution of negative / positive scores would be 33.3% (segment 2, 3 and 4) and 40% (segment 5).*

Segment	Part	Claim	Extraversion		Openness		Agreeable		Consc.	
			I	E	S	N	T	F	P	J
2: Work	1A	Animals, Customer service, Customer follow-up			41.3			42.7		40
	1B	Carpenter, Mechanic, Electrician		37.1						37.5
	1C	Design, Entrepreneurship				42				
	2C	Discussion, Ingenuity				45.9				
	3A	Communication, Education		44	42.2			42.5		40.3
	3B	Technical facility, Oil platform	All > 43%, J at nearly 50%							
	3C	Develop & plan new ideas				40.8				
	4C	Renew & Improve; Designer, Architect, Inventor				46.7				
	5A	Practical work helping others	41.9	46.3	47.4			50.5	41.8	46.4
	5D	Formulas & Math	All > 52%							
3: Education style	1B	Examples & Methods first, then exercises	51		52.6		43.2	43.4		50.8
	1C	Tasks without specific final answers	42.9	40.2	52.3		40.8	42.1		47.1
	2B	Theory, example, then tasks	45.6		48.9		43.2	40.3		48.7
	2C	Likes teacher encouraging to find new ways to solve tasks			43.4			42.1		40.2
	3B	Rules & Methods, learning what is right & wrong				43.1				
4: Courses / Topics	1	Toxicology, Lab-work (CHEMISTRY)	All > 40.1%							
	2	Plants, Humans & Animals (BIOLOGY)						36*		
	3	Rocks, Maps, Natural formations (GEOLOGY)	All > 64.1%							
	4	MATHEMATICS	44.8	47	41.2	53.5	39.5	49.9	51.6	41.1
	5	Space, Laws of nature (PHYSICS)		41.1	38.9	36.9		40.7	37.4	38.8
	6	Structural (e.g., human body, computers, and other systems)					36.1*			
5: Math	1	Math class			45.1*		45.3 *			48.1
	2	Solving math problems					43.9 *			44.1 *

* Below threshold, but added to show some hidden tendencies of favoritism (negative/positive)

In segment 2 of the questionnaire, work and career-related questions were asked. The majority of S, F and J-types answered they wanted to help animals or humans in customer-service related work (1a), working with communication and/or education (3a), and preferring more practically skewed work helping others (5a). In fact, apart from N's and T's, all personality types tended to be positive for 5a, the remaining two considered net neutral. The least popular claim was 5d, which stated that the individual would want to work with formulas and mathematics: More than 52% of participants actively rated this claim the lowest among all four claims in segment 2's part 5, regardless of cognitive type. However, the expanded table in the appendix show that among the types, Thinkers reacted the most positively, with 13.1% positive and 52.8% negative. Contrastingly, the type that answered the most negatively was the Intuitives, with 6% positive and 68.1% negative. All types answered net negative for 3b's "Working in a technical facility, e.g., an oil platform", with an average negativity of 46.5%. Of the other three claims presented together with this statement (i.e., segment2-part3), the most liked were 3a's "I would like to work in education or communication"; Within all personality type dimension, one personality type were positively biased, these being: Extraverts, Sensors, Feelers and Judgers.

Segment 3 discussed the topic of preferred educational style. In part 1, the participants almost unanimously answered that they liked best having the teacher present examples and methods first and then offer exercises (1b). Contrastingly they liked open-ended tasks – where they were encouraged to test different methods, and where multiple solutions were possible – the least. Besides the Extraverts, who were neutral for 1b but positive for 1c (thus agreeing with the majority for the latter), the only types deviating from the common perception was the Intuitives and the Perceptive individuals, both answering neutrally. Like 1b, most types also agreed with statement 2b's "Likes theory being presented first, then examples, and then tasks", again excluding the E-, N-, and P-types. The types that showed positive bias towards statement 2b, tended to show negative bias towards 2c's "Prefer multi-road focus" – the exceptions being Thinkers and Introverts, though the latter did have a negativity response of 39.8%, which was nearly at the threshold. In the last part of segment 3, the least divisive statements were 3d's "Does not like being instructed", with an average of 64.1% neutrality (see the full table in the appendix). The only statement in part 3 who received biased results was 3b's "Prefer learning about methods and rules": 43.1% of all Intuitives ranked this as the worst fitting. Perceivers did show slight bias, with 37.9% scoring it the lowest.

The fourth segment discussed various STEM-subjects. These would be translated into: Chemistry, Biology, Geology, Mathematics, and Physics. All types were negative towards chemistry and geology. The latter particularly so, with an average of 67.4%

people rating this topic 1 or 2, i.e., low. The biggest intra-dimensional difference was between Introverts and Extraverts, with 64.2% negative Introverts versus 70.5% negative Extraverts – the latter was the total highest observed negativity in our sample. Math was also a subject that most disliked, though not as bad as geology. Thinkers were the only deviators, with slightly less than 40% negativity versus 25.4% positivity. No significant bias was observed for Biology and IT-/Physical-work – although Feelers were the most positive towards the former, and Thinkers were the most positive towards the latter.

The last segment listed in Descriptive Table 2-S2 found that the only cognitive type that felt actively confident in math class was the Judging-type. Thinkers and Sensors were also confident; however, these results did not surpass the +20% bias-limit, which was set at 48% for the questions in this segment. The second statement considered whether the individuals liked solving math exercises and felt confident in their abilities. In general, there was a higher percentage of participants answering positively compared to negative, but none were above the limit. The most confident types were the Thinkers and Judgers.

Conclusively, I want to provide a summary of some of the claims and questions that showed interesting results yet had to be considered net neutral due to scoring below the chosen threshold values for bias consideration. On the stance "I like fixing things" in segment 2, Extraverts and Judgers did seem to show an aversion when compared to the others, with 36.3% and 35% answering negatively respectively. The claim that received the least negative response was segment 2-5b's "Common sense and practical experience", however about two thirds of the remaining participants subsequently answered neutrally regardless of type. In segment 3, part 3c's "I prefer multimedia learning", 36% of N-type participants answered positively; In comparison their counterpart, the S-type, had only 23.1% positives and even double the number of negatives (30.1% as opposed to 15.9% for N). In segment 3-2d, "Prefer teacher offering guidance in group settings", Introverts and Thinkers answered negatively more often than what Extraverts and Feelers did.

In general, the type that most often answered differently than the others were the Intuitives and the Perceptives. N-types were the only favoring S2-1C ("want to work with design / be an entrepreneur"), S2-2C ("like a workplace where I can discuss, and where ingenuity is praised"), S2-3C ("I like developing and planning new ideas"), and S2-4C ("I would like to work as a designer or inventor, where I can renew and improve things"). As previously mentioned, Intuitives were also the only cognitive type who answered negatively noticeably often on S3-3B.

The following table show how the personality types were distributed in our sample. The data was divided by gender, and the percentwise distribution of the "dominant" type are shown – in this case, dominant means the type who corresponds to a high score within their related Big-Five dimension, e.g., high Extraversion would signify an Extraverted individual, thus Extraverts are the focal point of this dimension. Although the data was not grouped, and individuals only appear in one of the five defined age groups, the data was also grouped by age. This was done to address differences between generations and/or tendencies within personality type development.

Table 3.6: *The distribution of personality type divided by age group and gender. Table shows the percentwise distribution of each age group (header), and the percentwise count of the positive labelled letter for each dichotomous dimension. Intra-cell differences of less than ~5% between the genders are highlighted in blue, and differences above ~10% are highlighted in red*

Personality type \ Age group (%)		1-12 (2.2%)	13-15 (23.7%)	16-18 (31.3%)	19-30 (35.5%)	30+ (7.4%)
E Extraversion	F M TOT	49.7 57.6 52.0	50.6 57.6 53.2	49.4 50.5 49.7	47.6 45.7 47.0	49.3 44.1 47.8
N Openness	F M TOT	43.7 46.1 44.4	41.2 45.3 42.7	31.5 36.0 32.7	34.5 40.4 36.3	39.8 46.4 41.7
F Agreeableness	F M TOT	64.4 54.8 61.6	63.8 54.0 60.2	62.2 48.7 58.6	67.9 57.2 64.5	66.1 54.3 62.8
J Conscientiousness	F M TOT	59.0 47.8 55.7	56.0 42.9 51.2	54.5 41.1 50.9	50.9 37.5 46.7	57.1 47.0 54.2

The distribution of the cognitive types changed across the age groups. Most noticeable was this for Extraversion, where the differences between the genders became less and less noticeable with increased age, but slightly increasing for the 30+ age group. This was due to males' extraversion decreasing with age, while females mostly remained consistent across the ages. The opposite was observed within the

S/N (Openness) dimension, as the differences between the genders became more noticeable with increased age. The amount of Intuitives decreased with ~10% for both genders between the 13-15 and the 16-18 age groups, before increasing again within the 19–30-year-olds. The remaining two dimensions were the most polarizing regarding gender-differences. Females had on average almost 10% more Feeler-types than males, and the largest difference was observed in the 16-18 group. Females also had an abundance of Judgers when compared to males, who consistently had below 48%. In general, age group 16-18 was the most unique, with the fewest number of Intuitives (both genders) and Feelers (only males).

3.2 – Validating Utdanningstesten – Analysis of data from sample 1

The focal point when comparing the various questionnaires to Utdanningstesten (Utd1) was the resulting 4-letter type. This could in turn be split into four dimensions. "Comparison table 0" was thus made to show the number of individuals within each Utd1 type who scored the same letter in each respective test. The type-based Big-Five distributions is also shown.

Table 3.7: *The number of individuals within each type (as determined by Utd1) who received the same letter when taking the other Big-Four tests (row 1-3), and how the same types scored in the Big-Five test (row 4-7). Type matches are represented by count and percent. The count shows the number of individuals who was "rated" a certain type in Utd1 and who received the same type in each respective test as well. The percent-score indicates how many of the original Utd1-rated individuals who received the same type. E.g., in Utd2, six individuals were rated "I" in both Utd1 and Utd2 – this is all of the original I's, meaning no Introverts were "wrongly" rated in Utd2.*

	TOT	I	E	S	N	T	F	P	J
Utd2; n (%)	13	6 (100%)	5 (71%)	7 (87%)	3 (60%)	6 (85%)	5 (83%)	6 (100%)	6 (85%)
Uroboros; n (%)	17	10 (100%)	5 (71%)	7 (70%)	3 (42%)	4 (36%)	4 (66%)	4 (100%)	10 (76%)
Truity; n (%)	37	20 (90%)	10 (66%)	13 (100%)	10 (41%)	7 (87%)	22 (75%)	8 (57%)	18 (78%)
Extraversion; Mean (sd)	43	66.3 (13.6)	88.8 (11)	71.1 (16.2)	78.4 (16.7)	65.5 (13.9)	76.8 (16.3)	76.1 (15.1)	70.9 (17.2)
Openness; Mean (sd)		80 (11.3)	81.6 (12.8)	77.2 (10.8)	89.7 (8.4)	79 (12.6)	81.5 (10.9)	82.9 (13.1)	78.8 (10.5)
Agreeableness; Mean (sd)		90.1 (11.8)	98.2 (9)	92.4 (12.3)	92.9 (9.5)	85.3 (11.9)	98 (7.8)	91.2 (9.4)	93.3 (12.8)
Conscientiousness; Mean (sd)		85.5 (11.9)	82.9 (25.8)	86.6 (12.1)	79.3 (26.9)	86.8 (12)	83.2 (20)	76.5 (20.4)	90 (12.1)

The number of mismatches between 4-letter types were also counted for all pairwise comparisons between the Utd1 and the other Big-Four tests, as shown in “Comparison table 0.1”.

Table 3.8: *The number of letters that was mismatched between Utd1 and the other dichotomous tests. Counts are shown for all possible numbers of mismatches, except 4 (i.e., all letters were rated different between tests) as none were found.*

		TOT	I	E	S	N	T	F	P	J
Letters Mismatch Utd2	0	28	3	4	6	1	3	4	2	5
	1	16	3	1	2	2	3	1	3	1
	2	8	2	0	1	1	1	1	2	0
	3	0	0	0	0	0	0	0	0	0
	TOT	(0.6)	(0.9)	(0.2)	(0.4)	(1)	(0.7)	(0.5)	(1)	(0.2)
Letters Mismatch Uroboros	0	12	2	1	2	1	2	1	1	2
	1	32	5	3	6	2	2	6	4	4
	2	20	4	1	3	2	2	3	2	3
	3	4	1	0	0	1	0	1	0	1
	TOT	(1.2)	(1.3)	(1)	(1.1)	(1.5)	(1)	(1.4)	(1.1)	(1.3)
Letters Mismatch Truity	0	44	5	6	6	5	0	11	3	8
	1	56	11	3	11	3	8	6	5	9
	2	40	8	2	8	2	5	5	4	6
	3	8	1	1	2	0	1	1	1	1
	TOT	(1.1)	(1.2)	(0.8)	(1.2)	(0.7)	(1.5)	(0.8)	(1.2)	(1)

A shown by both table 3.7 and 3.8, not all individuals who received a certain type in Utd1 also received the same type in the other tests. In fact, most were at least 1 letter off. In Utd2 and Uroboros, all who was Utd1-Introverted also received I, whereas only 71% of the Utd1-Extraverted got E in Utd2 and Uroboros respectively. Lowest dimensional numbers were found for S/N. This was in particular the Intuitives, as the majority of Utd1-N’s got the opposite type in both Uroboros and the Big-Four Truity questionnaire. On average, those who received E, N, F and J in Utd1 received higher scores in the Extraversion, Openness, Agreeableness, and the Conscientiousness dimensions than their counterparts.

The average and counted number of mismatches between Utd.1 and the other Big-Four tests varied, as shown in table 3.8. Comparing the Truity Typefinder to Utd.1, the lowest number of mismatches was observed among those classified in Utd1 as Extraverted (0.8), Intuitive (0.7), Feeling (0.8), and Judging (1). However, the latter type had almost as many mismatches on average as their counterpart, the Utd1-Perceivers. On average, fewest mismatches were found between Utd1 and Utd2, with a total average of 0.6 mismatches per individual.

In total each participant could take 4 different personality tests besides the first iteration of Utdanningstesten. 3 of these questionnaires were based on the theories of Myers and Briggs, while the last was a Big-Five test and fundamentally different than the others. The subsequent chapter will thus be divided into two segments, that focus on comparisons related to Big-Four models and the Big-Five model separately.

3.2.1 – Utdanningstesten VS Big-Four tests

Treating the questionnaires as raters, Utd.1 was compared to the other tests. Utd.1's consistency was assessed through estimating Cohen's Kappa values for both total, 4-letter personality types and for each dimension separately. These results are presented in Table 3.9 and 3.10 below.

Table 3.9: *Results from Utdanningstesten compared to itself (Utd1 vs. Utd2) to assess its consistency. The predictions from Utdanningstesten (Utd.1) is presented across the columns, and the second iteration of Utdanningstesten (Utd.2) across the rows. When comparing the 4-letter personality type directly, a Kappa value of 0.497 was obtained.*

<table>
<tr><td></td><td>E</td><td>I</td><td>TOT</td><td rowspan="4">Dim: E / I
Kappa = 0.698</td></tr>
<tr><td>E</td><td>5</td><td>2</td><td>7</td></tr>
<tr><td>I</td><td>0</td><td>6</td><td>6</td></tr>
<tr><td>TOT</td><td>5</td><td>8</td><td>13</td></tr>
<tr><td></td><td>N</td><td>S</td><td>TOT</td><td rowspan="4">Dim: N / S
Kappa: 0.494</td></tr>
<tr><td>N</td><td>3</td><td>2</td><td>5</td></tr>
<tr><td>S</td><td>1</td><td>7</td><td>8</td></tr>
<tr><td>TOT</td><td>4</td><td>9</td><td>13</td></tr>
<tr><td></td><td>F</td><td>T</td><td>TOT</td><td rowspan="4">Dim: F / T
Kappa: 0.69</td></tr>
<tr><td>F</td><td>5</td><td>1</td><td>6</td></tr>
<tr><td>T</td><td>1</td><td>6</td><td>7</td></tr>
<tr><td>TOT</td><td>6</td><td>7</td><td>13</td></tr>
<tr><td></td><td>J</td><td>P</td><td>TOT</td><td rowspan="4">Dim: J / P
Kappa: 0.847</td></tr>
<tr><td>J</td><td>6</td><td>1</td><td>7</td></tr>
<tr><td>P</td><td>0</td><td>6</td><td>6</td></tr>
<tr><td>TOT</td><td>6</td><td>7</td><td>13</td></tr>
</table>

Table 3.10: *Kappa values are shown for each test compared with Utdanningstesten (Utd.1). The "K" column shows the estimated kappa, as well as its 95% confidence interval, when comparing the 4-letter personality types. The subsequent columns show the estimated kappa for each dimension separately.*

	n	K total	I / E	S / N	T / F	P / J
Utd1 VS Utd2	13	0.5 [0.342, 0.652]	0.70	0.49	0.69	0.85
Utd1 VS Uroboros	17	0.08 [-0.064, 0.226]	0.75	0.13	0.03	0.61
Utd1 VS Truity	37	0.24 [0.145, 0.325]	0.59	0.33	0.50	0.36

The highest Kappa-values were found for the comparison of Utd1 versus Utd2, i.e., between two iterations of the same test. Kappa was estimated to 0.5, with a 95% confidence interval covering values between 0.34 and 0.65. Apart from the S/N dimension Utd1 vs. Utd2, the single-dimensional kappa values exceeded that of the 4-letter type comparison Kappa. The lowest Kappa occurred when Utd1 was compared with Uroboros, wherein the K total was 0.08, with a confidence interval that included zero. However, this comparison yielded the highest I/E-dimensional Kappa, of 0.75. Averaging across the three comparisons, the highest dimensional-Kappa was found for the I/E-dimension, while the lowest was for the S/N. The highest deviation between Kappa values, largely in part due to the Uroboros-comparison, were within the T/F-dimension. For the intra-specific test (Utd1 vs. Utd2), the COD was estimated as: I/E (0.49), S/N (0.24), T/F (0.48), and P/J (0.72).

3.2.2 – Utdanningstesten compared to Big-Five

To compare the continuous output of the Big-Five test to the dichotomous output of Utdanningstesten, the former had to be binarized. Two values were chosen as binarization thresholds: 60, as the scale was [0, 120], as well as the median for each dimension. Individuals who scored above these thresholds were labelled E, N, F, and P. The kappa values from these two comparisons are shown in table 3.11.

Table 3.11: Kappa values are shown for the Big-Five test compared with Utdanningstesten (Utd.1). The "K" column shows the estimated kappa, as well as its 95% confidence interval, when comparing the 4-letter personality types. The subsequent columns show the estimated kappa for each dimension separately. To glean the personality type letters from the continuous Big-Five data, all scores were divided at a threshold value; Participants scoring > "thresh" points received (E, N, F or J), and participants scoring <= "thresh" points received (I, S, T or P). In the second row, median values were used. The median values were 72 (Extraversion), 77 (Openness), 93 (Agreeableness), and 89 (Conscientiousness).

	n	K total	I / E	S / N	T / F	P / J
Utd1 VS Big-Five Thresh = 60	43	0.03 [-0.019, 0.071]	0.21	0.02	0.12	0.07
Utd1 VS Big-Five Thresh = Median		0.275 [0.197, 0.353]	0.48	0.47	0.40	0.36

When the Big-Five test was binarized at score = 60, the Kappa was low and the confidence interval – though not wide – covered 0. The best Kappa occurred for the I/E dimension, at 0.21. When binarization was done using the individual median values, the estimates became higher. Kappa for 4-letter type became 0.28, and the dimensional Kappa values were all between 0.35 and 0.50.

To further illustrate any differences between the types in how they scored across the Big-Five dimensions, boxplots were made. Figure 3.5 show one figure for each Big-Four dimension (letter-type determined by Utd1), and how the respective types scored across the four chosen Big-Five dimensions.

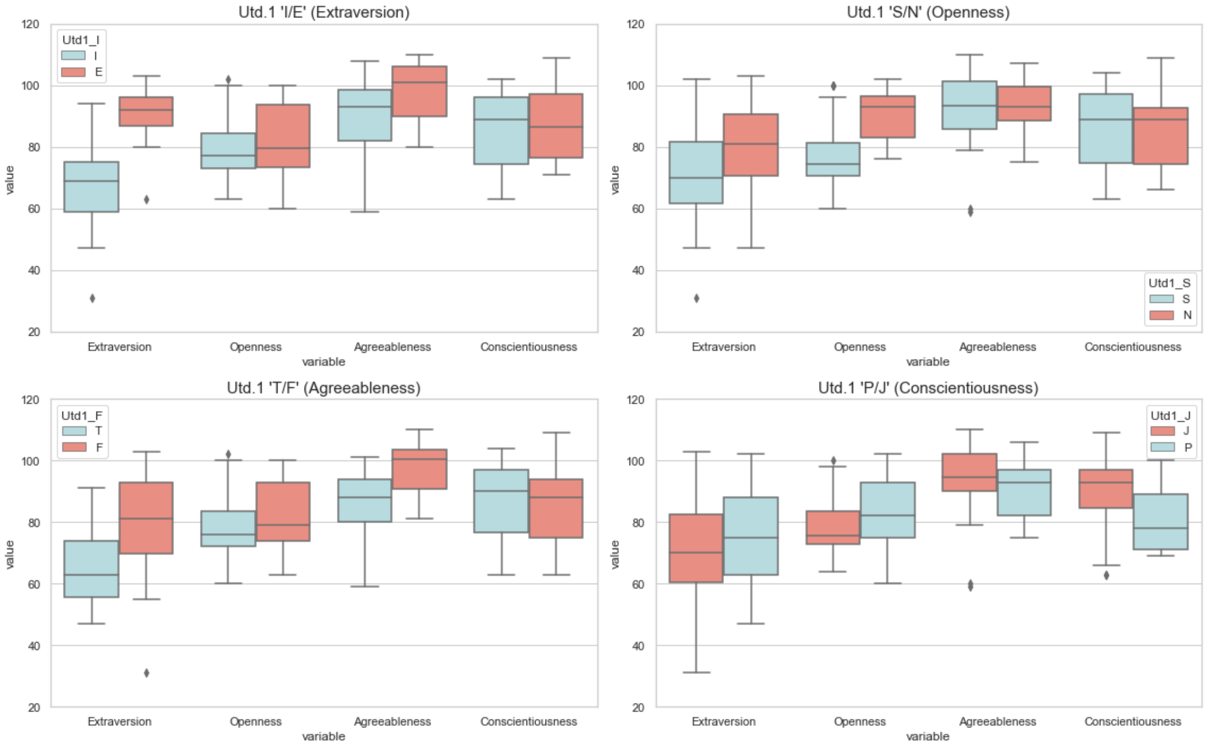


Figure 3.5: The distribution in score for all personality type letters (as determined by Utd.1) across the four chosen Big-Five dimensions. Each subfigure focuses on one dichotomous dimension, with the type-boxplots pivoted against each other. The blue colored types represent the type that could be considered a “low-scorer”, and red a “high-scorer”, in their respective Big-Five counterpart (e.g., E’s are high-Extraversion scorers, and F’s are high-Agreeableness scorers). Ideally, the largest discrimination between paired boxplots should be observable for the Big-Five dimension that is the most correlated with the pairs’ dimension (e.g., between S & N within Openness).

In the four respective subplots above, most Utd1-type boxplots’ interquartile range (IQR) overlapped with their intra-dimensional counterpart. E.g., I and E for Openness, and T and F for Conscientiousness. Extraverts scored higher in Extraversion than Introverts. These individuals also tended to score a bit higher in Agreeableness than I’s, though with some IQR-overlap. Intuitives also scored higher than Sensors for Openness. For T/F, all paired boxplots showed overlapping within their interquartile range; The largest inequality between these types were found in Agreeableness, wherein Feelers tended to score highest. Lastly, all boxplots also overlapped for P/J, but the lowest amount of interquartile range overlap was recorded within the Openness dimension, followed by Conscientiousness. To assess whether the differences were noteworthy, two-sample T-tests and Wilcoxon Rank Sum tests was performed on the data. The results are summarized in Comparison Table 3.

Table 3.12: *The result of performing a two-sample T-test between each paired Big-Four letter-group, across the four chosen Big-Five dimensions. All comparisons assume normality and equal variances within each letter-group. The T-value is shown in conjunction with the p-value in parentheses. Two-sided hypothesis testing. Highlighted in dark green are results where the p-value is below a significance level of 0.025, highlighted in lighter green are p-values below 0.05.*

	I	E	S	N	T	F	P	J
n = 44	n = 31	n = 13	n = 33	n = 11	n = 19	n = 25	n = 17	n = 27
Extraversion	T = -5.27 (0.0)		T = -1.28 (0.208)		T = 2.81 (0.007)		T = -1.01 (0.32)	
Openness	T = -0.41 (0.687)		T = -3.47 (0.001)		T = 0.71 (0.482)		T = -1.13 (0.264)	
Agreeableness	T = -2.23 (0.031)		T = -0.13 (0.894)		T = 4.25 (0.0)		T = 0.57 (0.57)	
Conscientiousness	T = 0.46 (0.644)		T = 1.25 (0.218)		T = -0.68 (0.498)		T = 2.75 (0.009)	

Comparing the Utd1-types across all Big-Five dimensions found 6 cases of significance at p <= 0.05, four of which had p-values lower than 0.01. These were between: E / I in Extraversion and Agreeableness, S / N in Openness, T / F in Agreeableness and Extraversion, as well as between P / J in Conscientiousness. Four of the comparisons who had p-values lower than 0.1 were between types within the ”corresponding” Big-Five dimension.

3.3 – Analysis of historical data from sample 2

The next chapters will explore the results obtained from Principal component analysis on a train subset of the data (3.3.1); Linear discriminant analysis on the same train dataset (3.3.2); And T-tests and modelling using the PCA/LDA transformed data (3.3.3.).

3.3.1 – PCA and Multicollinearity

To assess whether data would become more separable using alternate techniques and identify any trends using fewer axes, PCA was performed. This also mediated the multicollinearity problems that was found in the untransformed (total) dataset, as measured by e.g., the variance influence factor (VIF): After visual inspection of a heatmap featuring the correlations between all variables (data not shown), VIF was calculated using both "Del5" questions ("Del5_1" and "Del5_2") and the "Del4_4" question. The resulting VIFs were: 18.7, 14.9 and 8.5 respectively. When all scales were converted to be on the same format, i.e., only [0, 1, 2], the VIFs were: 6.1, 5.2 and 3.6 respectively. In both cases, at least one VIF value exceeded the commonly accepted threshold of 5.

The predictor-dataset that was used to fit the PCA only included the questions from segment 2, 3 and 4; Segment 5's "math anxiety" questions, as well as the demographic- and personality-related questions, were kept as part of separate matrices. X_train (105172 * 38) thus only consisted of numerical variables with no missing data, as the only variables with missing data was "county" and the two "Del5" questions. This meant no one-hot encoding or imputation was performed. The data set was then standard scaled with the mean values and standard deviations – this preprocessing was done by creating a preprocessor-object, to allow the X_test dataset (70 115 * 38), to be scaled in the same manner with the same values as X_train. Figure 3.6 presents the resulting scree plots.

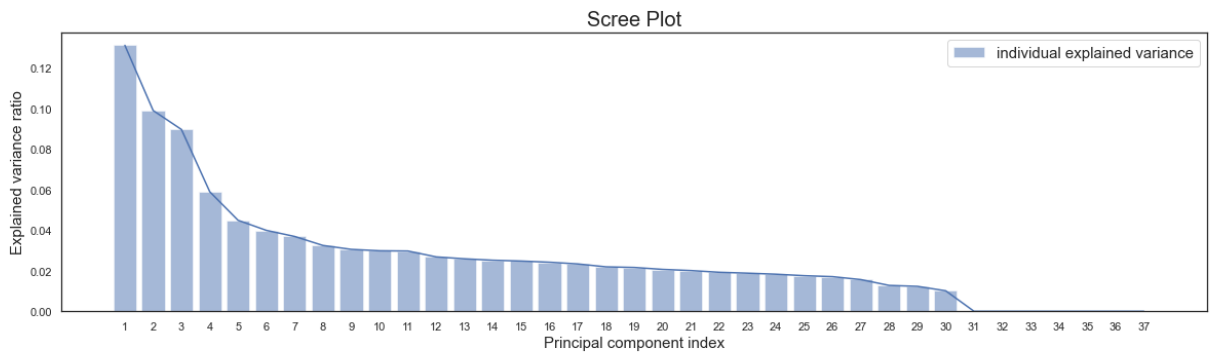


Figure 3.6: *Scree plot that resulted from performing PCA on a train dataset (60% of original data, random sampling using a set random state), that was standard scaled. The variables used were all the questions relating to work, school-related topics, and education preferences.*

The Scree plot show that the variability explained by the individual principal components drastically reduces after the 4th or 5th component. Print outs (data not shown) show that the first five PCs explain 42.4% of the variability. 19 PCs are needed to exceed 80% (81.7%), 24 PCs exceeds 90% (91.4%), while 30 principal components are needed to achieve above 99% (99.99%). Furthermore, inspecting the variance within each component (rounded to 1 decimal) found that the first four PC's had variances above 2 (5.0, 3.8, 3.4, and 2.2), PC1-11 exceeded 1.0, and PC12-14 each had a variance of 1.0.

Circling back to the figure at hand, we see that the last 7 components, 31-37, does not contribute at all. This phenomenon is due to Utdanningstesten's design: Segments 2 and 3 include 8 parts in total, each consisting of 4 claims grouped. As the test-taker must give a score to two of the claims, leaving the rest neutral, this means that 8 of the 38 variables are not linearly independent. The eight non-independent variable was excluded by the algorithm at the process' start, as a maximum of $k = p - 1$ PC-variables can be made.

The correlation between PC-loadings (a vector of p weights for each PC) and original data for each observation (a vector of p values for each observation/row) was then calculated. The resulting correlation scores were then plotted for PC1 against PC2, colorized by various letter-pairs. For visualization purposes, the sign of the PC1 coordinates were changed for all observations. The plots are shown in the three subsequent figures.

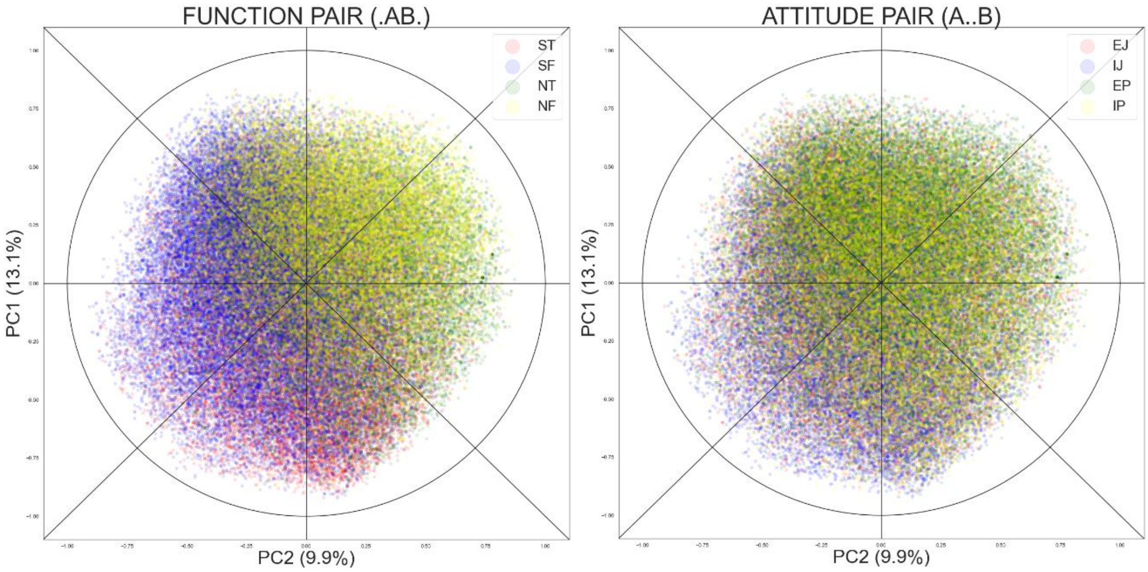


Figure 3.7: *Correlation scores plot with PC1 against PC2, that resulted from performing PCA on the X_train dataset. The points are colorized by function pair (left) and attitude pair (right).*

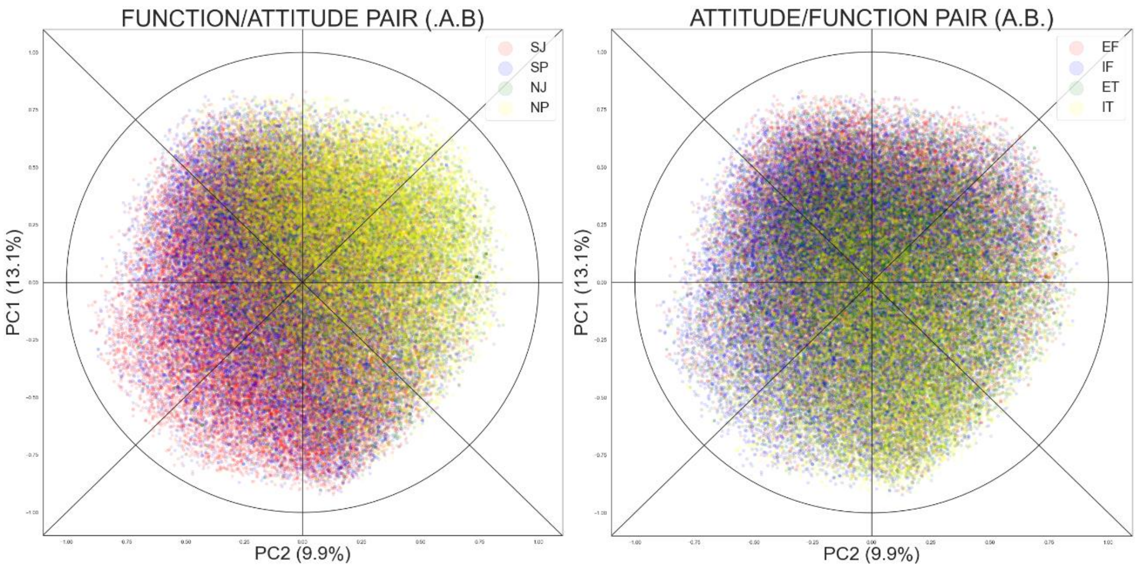


Figure 3.8: *Correlation scores plot, PC1 against PC2, that resulted from performing PCA on the train dataset. The points are colorized by the function/attitude pair consisting of openness (S/N) and conscientiousness (P/J) to the left, and the type consisting of extraversion (I/E) and agreeableness (T/F) to the right.*

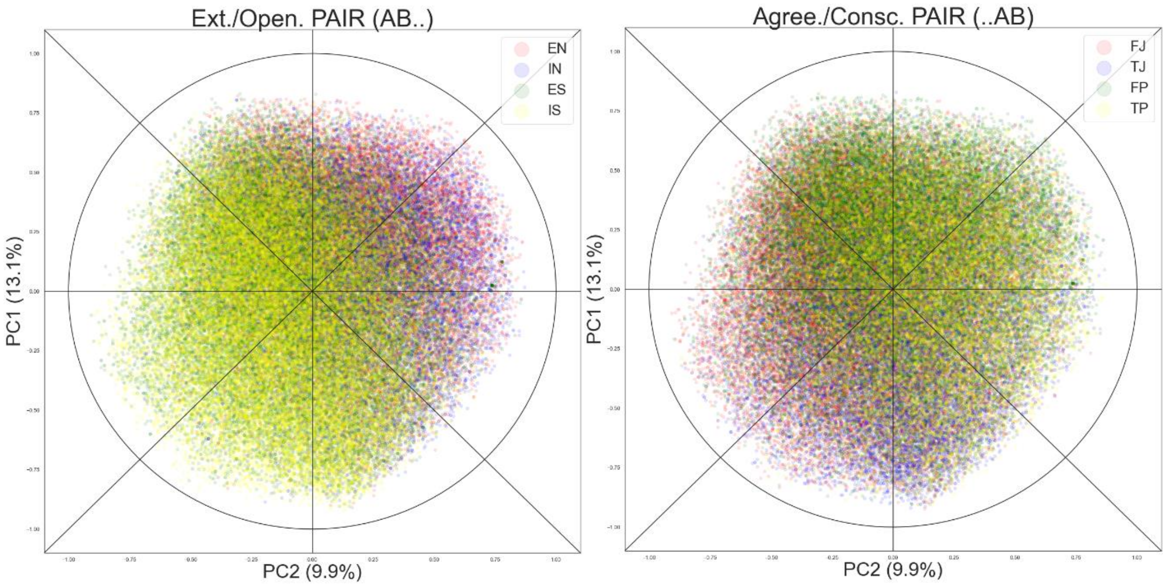


Figure 3.9: *Correlation scores plot, PC1 against PC2, that resulted from performing PCA on the train dataset. The points are colorized by the extraversion (I/E) and openness (S/N) pair to the left, and the agreeableness (T/F) and conscientiousness (P/J) pair on the right.*

Of the 6 respective sub-figures above, the four chosen types were most differentiable in figure 3.7's Function pair subplot. ST's were grouped in the south, SF's in the west, NF's in the north, and NT's in the east. Figure 3.8's first plot also had some separation between types, particularly between the S-types (SJ and SP) and the N-types (NJ and NP). As the cognitive types were most distinguishable when using function pairs, this grouping was prioritized for the following analyzes.

In addition to the correlation scores, the correlation weights were also calculated for a variety of variables, such as the questions, the personality letters and -pairs, and age groups, among others. This was done by using the scores dataset and the regular (scaled) data. The corr.scores figure, colorized by function pair and with these corr.weights added, are shown in figure 3.10.

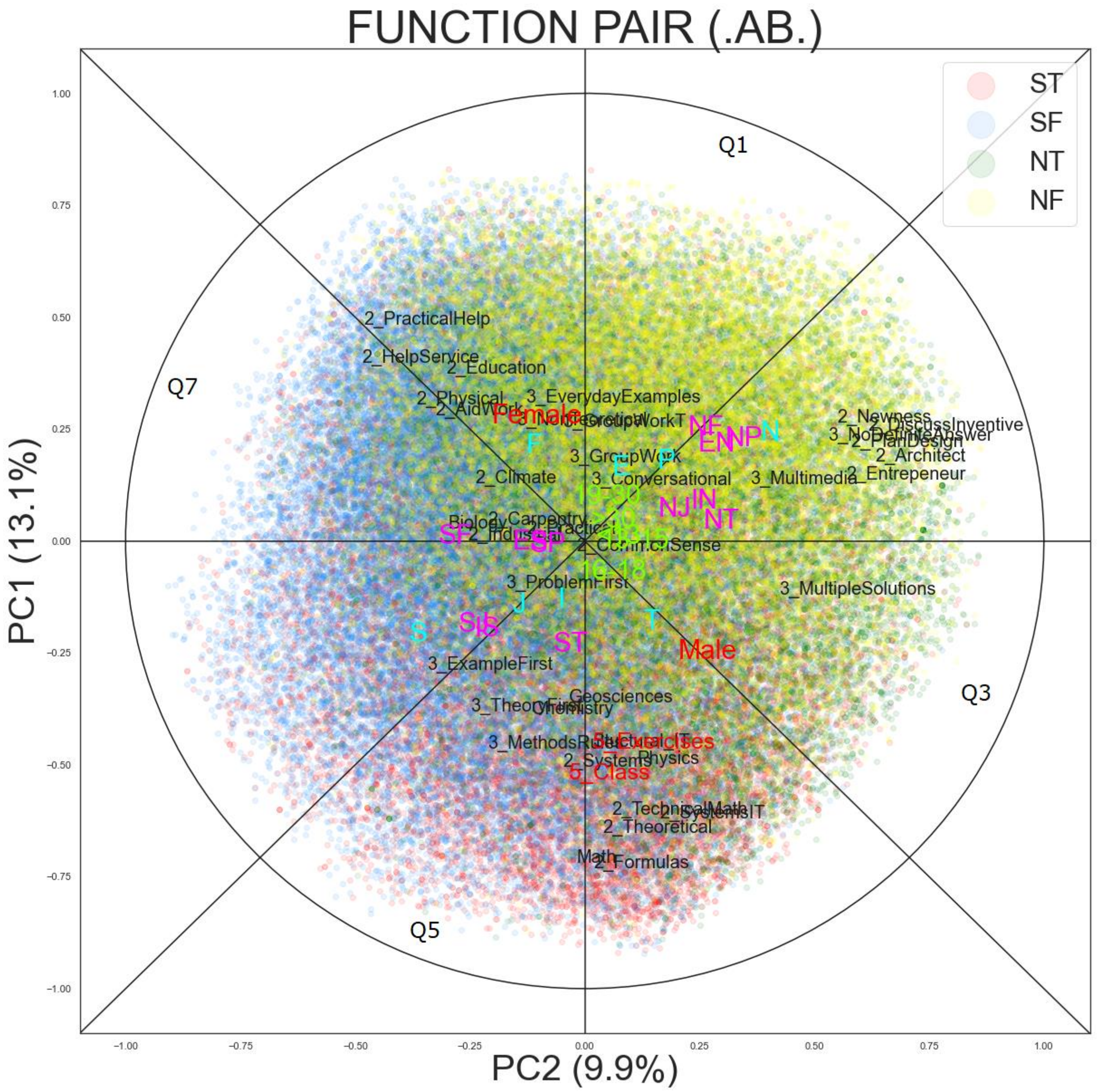


Figure 3.10: *Correlation scores plotted for PC2 against PC1, with the correlation loading weights (for all the non-demographic and non-personality related questions in the questionnaire) plotted as black text points. Other correlation loading weights plotted as different-colored text: Age-group (limegreen), single-letter type (cyan), gender (red), segment 5's math "anxiety" questions (red, smaller font than gender), letter pairs "xABx" & "ABxx" & "xAxB" (purple). Due to visualization, PC1 was plotted on the secondary (y) axis, and all observations (correlation loadings or weights) had their sign changed for PC1.*

In the Q1-pie above, we see that the more theoretical questions have clustered, e.g., segment 2's statement about wanting a job where they can use formulas, or segment 4's Mathematics. The two segment 5 questions, about math class and solving math exercises, were also clustered here. Among all the Q-pies, Q1 were perhaps the most grey-coded, i.e., many individuals from all four function pairs fell within this area – although the T's seemed more numerable.

To assess whether the test-data had similar distributions as the train-data, correlation scores were made for the test-data using the loadings dataframe – that resulted from the original PCA – and the scaled original test-data. The questions and single-letter types were also plotted to make the figure's skeleton otherwise like the train-data based correlation scores plot. The resulting figure is presented figure 3.11, and it shows that the function types are about identically distributed compared to the train data. The only noticeable difference lies in the strengths of the colors, which is due to the differences in dataset size: At 60:40, the train data was 50% larger than the test data.

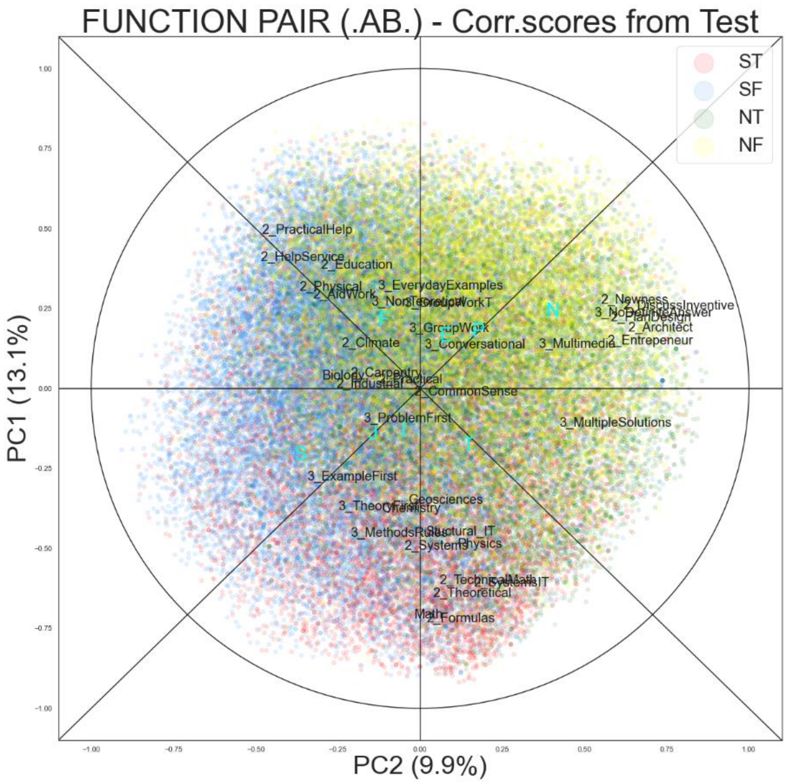


Figure 3.11: *Correlation scores plotted for PC2 against PC1 based on the test-data. The sign before the PC1-coordinates were changed for all correlation scores and -weights. The questions and single-letter types were plotted like previously ("PCA figure 3"), as black and cyan text respectively.*

To further showcase how different types answered, correlation score plots were made highlighting different cognitive types. Below are two plots, showcasing the PC1-PC2 correlation scores of all individuals in the train sample 2 data having the same personality types of myself (ISTJ) and my partner (INFP). These types were opposite in all dimensions save for I/E, and despite missing the latter, the points were almost perfectly weighted in opposite segments; ISTJ in the north-northwest, and INFP in the southeast. The types were equally abundant, but ISTJ's were less localized than

INFP, with a sizeable number of individuals in all segments. INFP's correlated most with the "architect" job, wanting "no definite answers" and "newness". ISTJ's correlated most with most of the courses ("math", "physics", "chemistry" and "IT / Structural sciences"), in addition to the segment 2's "formulas" claim.

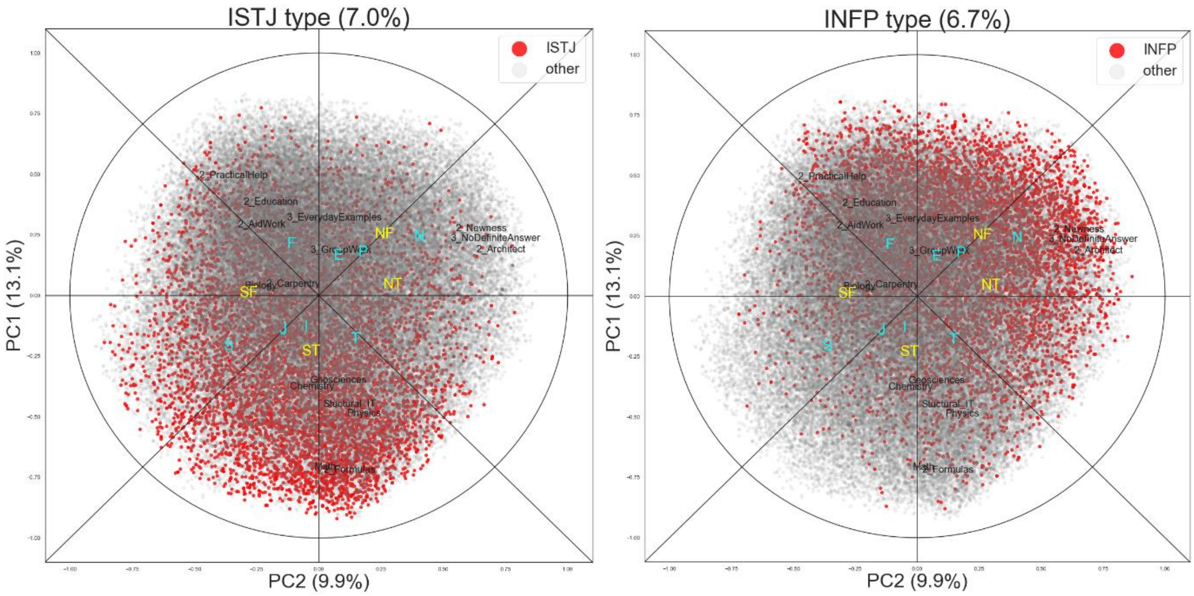


Figure 3.12: *Correlation scores plotted for PC2 against PC1 based on the train data. The sign before the PC1 coordinates were changed. The individual personality letters are plotted as cyan text, and the function pairs as yellow text. The observations are colorized by the respective type in focus: ISTJ (myself, left), and INFP (my creative partner, right). Though not shown above, the extraverted counterparts of these types were almost identically distributed. A handful of the correlation weights for the questions were added as black text points as well. The percent of individuals (in the train sample2) who fell within each type are printed in the titles.*

3.3.2 – LDA

The same mean-scaled train-dataset, previously used for the principal component analysis, was then used to perform a Linear discriminant analysis. As before, this data set contained all questions pertaining to work, education, and topic-preferences, with segment 5's "math anxiety" questions omitted due to missingness caused by their late inclusion. In the first iteration of LDA, the 4-letter type was set as response and the LDA was fitted using the maximum number of components $k = 15$ (due to $k = c - 1$). The ratio of explained variances are shown in Figure 3.13. Furthermore, a heatmap displaying the coefficient table is shown in Figure 3.14.

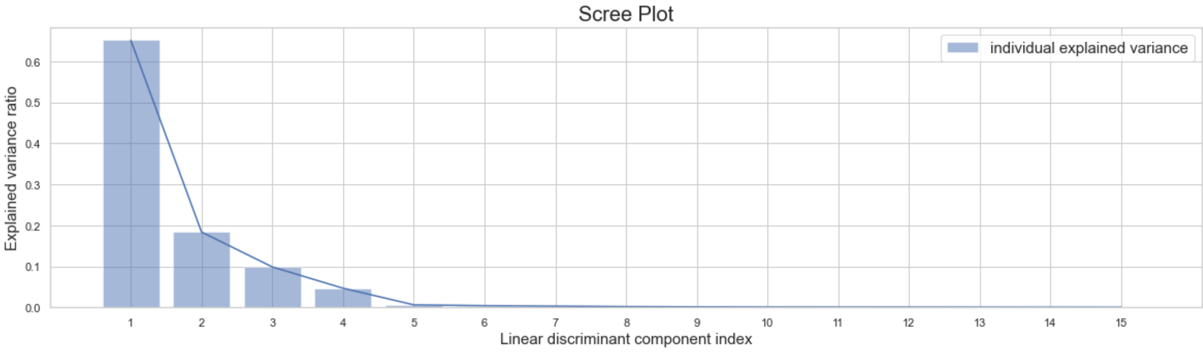


Figure 3.13: *The explained variance ratio by each linear discriminant component, resulting from performing LDA on a mean-scaled train-data set with k=15 components. The first two cumulatively explains 83.5% of the variability, and the first four explains 98%. Component 5 onwards each explain less than 1%.*

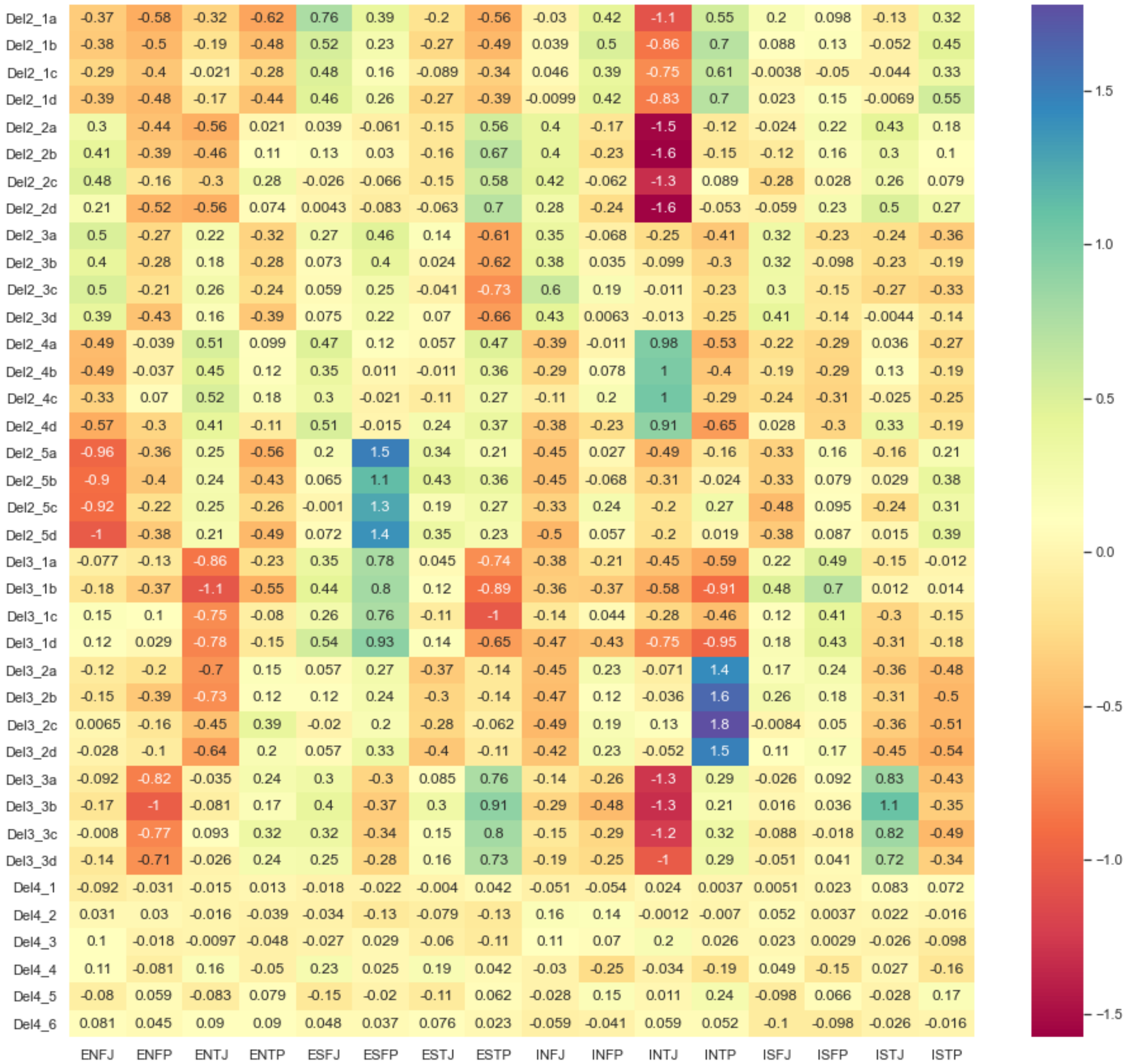


Figure 3.14: *Heatmap plot of the resulting coefficient table from performing LDA analysis with k=15 components. The row-names (y-axis) correspond to the variables / question-statements, and the column-names (x-axis) correspond to the response labels, i.e., 4-letter personality type. Red/orange and purple/blue color coding respectively represents low and*

high coefficients, i.e., the associated personality type tended to answer quite archetypical on these.

In the coefficient table above, we see that the types who received the highest (absolute) values, scoring quite uniquely, were INTJ (segment 2_2 and 3_3), INTP (segment 3_2), and ESFP (segment 2_5). None received particularly high or low coefficients for segment 4's questions about courses / topics. Segment 4 was also the only in which the grouped questions were scored completely independently from each other. The following figure show the sum of all absolute coefficients for each personality type label.

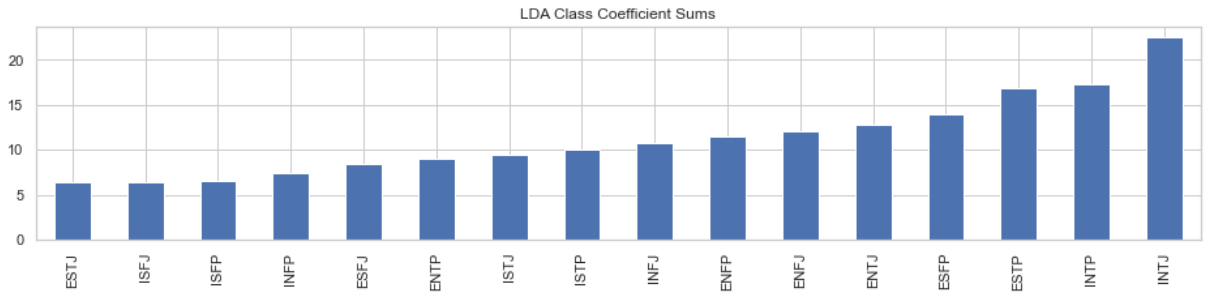


Figure 3.15: *Barplot showing the cumulative (absolute) coefficient scores, across all selected variables, for each of the 16 four-letter personality types that were represented in our dataset. Higher values/bars indicate that more questions in the questionnaire were marked as receiving archetypical answers for the types.*

Figure 3.16 further show that INTJ, INTP and ESFP had some of the (cumulatively) highest absolute coefficient values. The lowest values were observed for ESTJ and ISFJ. In total, the two highest scorers were INT's, while the two lowest – and least differentiable – scorers were SJ's.

The first two LDA components were then used to plot the transformed data. First, a randomly selected subset of the train-data was used, row-selection bounded by the requirement of equal class (i.e., type) representation. Only a small degree of segmentation was shown between observations of different types (Data not shown). A correlation scores plot was then constructed, following the same logic and structure used for the PCA data. The signs for both the LD1 and LD2 was changed for cosmetic reasons. See figure 3.16.

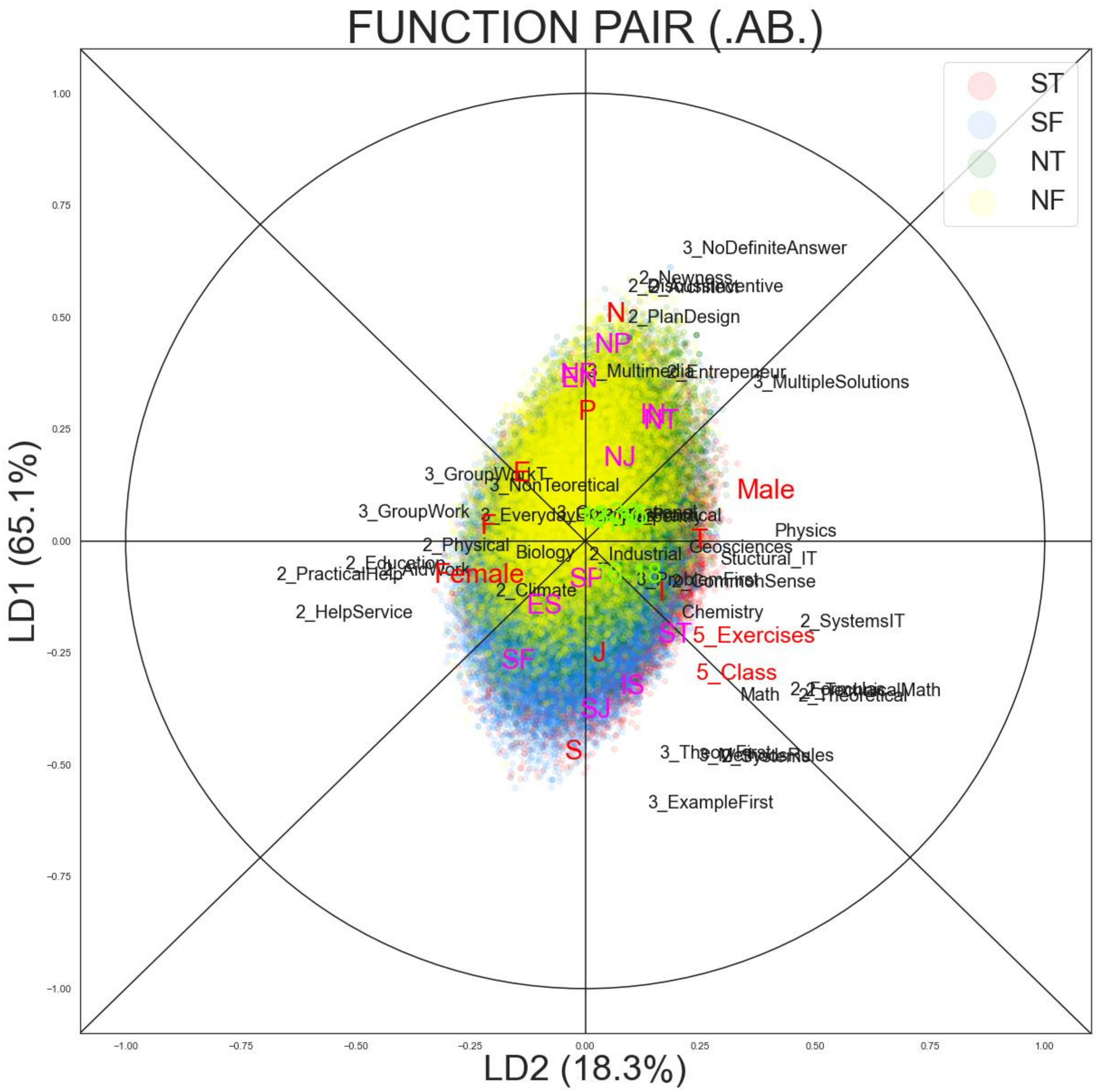


__Figure 3.16:__ Correlation scores and -weights which resulted from performing LDA with 4-letter type as the response. Scores are colorized by function pair. The questions, gender, single & paired letter-types, and age group are plotted as text points. The LD1 and LD2 signs were changed (i.e., the coordinates were flipped) for visualization purposes. The paired letter types shown are: "xAxB", "xABx" (function type), and "ABxx".

The single letter types, when grouped by dimension, dispersed into opposite quadrants. The largest gap between these were within the S/N dimension, followed by P/J. The I/E and the T/F dichotomies each had about the same distance between each other. Intuitive types were grouped with segment 2's "PlanDesign", "Newness", "Entrepeneur", "Discussion" and "Inventive", as well as segment 3's "MultipleSolutions", "NoDefiniteAnswer" and "Multimedia". Perceivers also tended towards the same, albeit less. Perceivers were draw in the direction of "GroupWork" and "NonTeoretical" – the same as Extraverts and Feelers.

Sensing and Judging individuals answered quite similar. Both types leaned towards segment 3's "ExampleFirst", learning about methods and rules, and "Theory". One exception however was the Judging Intuitive (NJ), which grouped with the other N-types.

Differences were also observed between age groups and the genders. Males tended towards "Physics", "geosciences", "IT", "Analytical". Males also showed stronger tendencies towards the direction of Intuitives than Sensers. Furthermore, all ages except the 16-18 group grouped together. While the latter tended in the same direction of the Introverted, Sensing and Judging types, the remaining age groups leaned towards the opposite.

Further grouping was then done, and two iterations of LDA was performed; one with Attitude pair ("AxxB") as the response, and one with Function pairs ("xABx"). The maximum number of components (k = 3) was used in both instances. The amount of variability explained by the first component was found to be 79.5% and 67.2% by the "function pair" and the "attitude pair" methods respectively. The first two components both had a cumulative amount of 99.6% when rounded to 1 decimal. Coefficient tables from each analysis, together with the cumulative coefficient bar plot, was then made. These are shown in figure 3.17.

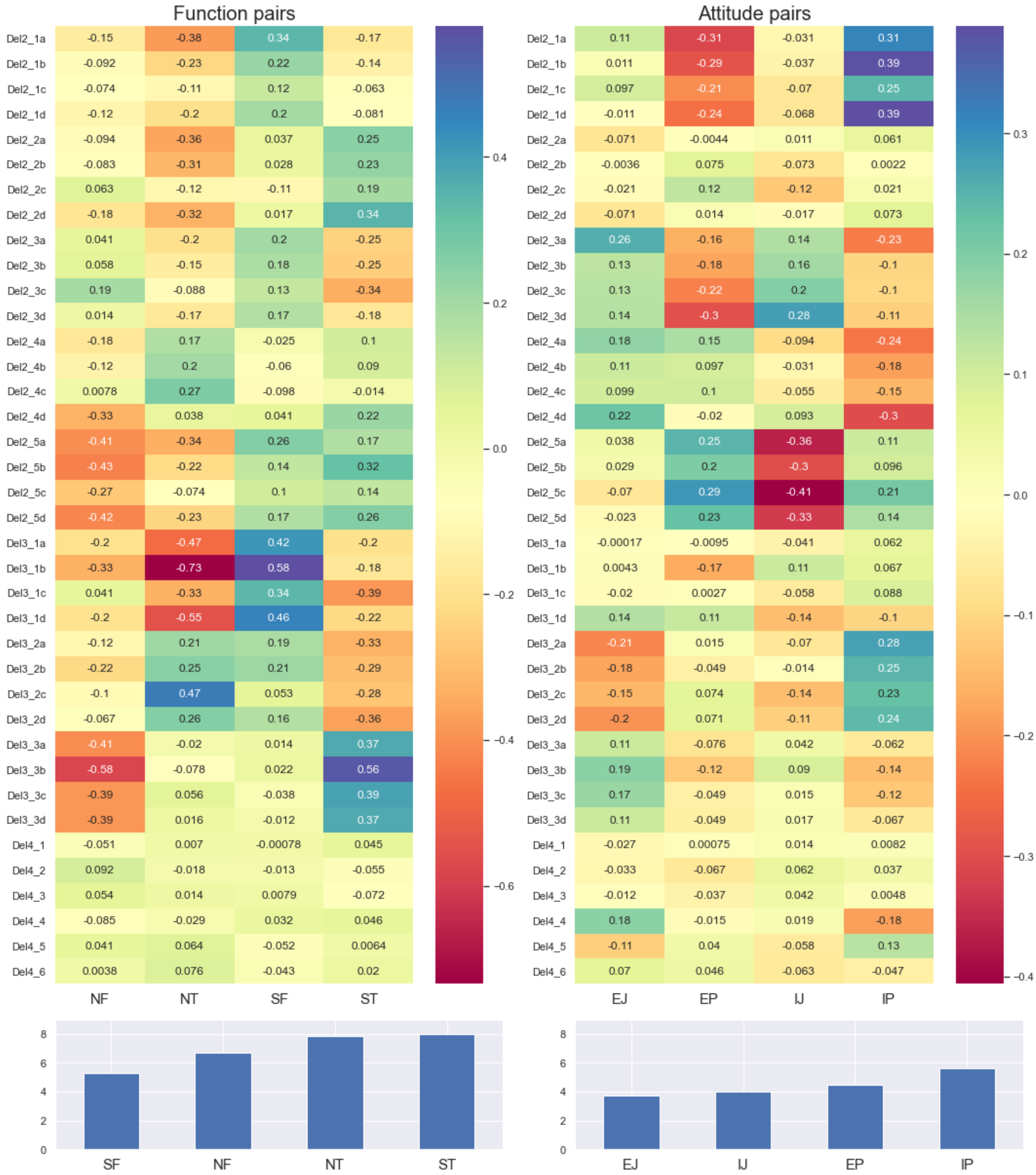


Figure 3.17: *Coefficient tables plotted as heatmaps, as well as barplots of the cumulative absolute class coefficients (below heatmaps). Figures were made after running LDA analysis with the response set as Function pair (left) or Attitude pair (right), for all non-demographic and non-personality type related questions in the Utdanningstesten questionnaire. Dark red or blue colors in the heatmaps indicate that the subset of individuals within each type often answered similarly to each other – but different to other types – for the variable in question.*

Overall, the highest absolute coefficients were observed for the function pairs; The highest Function pair coefficient score (~ 8) were almost +50% the size of that of the highest Attitude pair coefficient score (~ 5.5). The most notable types in terms of cumulative scores were the Thinkers, though they differed some in terms of variables weighted. ST's answered most homogenic (intraspecific) in Del3_3, whereas NT's had the most notably high (negative) coefficients for Del3_1. These two sections were respectively concerned about which educational techniques they prefer the teacher to use, and how they liked exercises to be presented and structured.

For the attitude pairs, the cumulative coefficients were highest for the perceivers, particularly the introverted. Highest coefficients were found in Del2_1 and Del3_2 for the IP's. The former was also the segment in which received highest absolute values for the EP's. Regarding function pairs, the highest coefficients were on average observable for the Del3 questions. These pertained to the participant's preferred educational style. On the other hand, the lowest coefficients were observed for the Del4 questions, which asked the participants to rank six different fields of study/work on a scale of 1-6. The scale used for these questions were the most diverse, with 6 possible answers compared to Del2 and Del3's three possible answers.

A correlation scores plot, identical in ground structure as figure 3.16, is shown in figure 3.18 for the analysis with function pairs as the response.

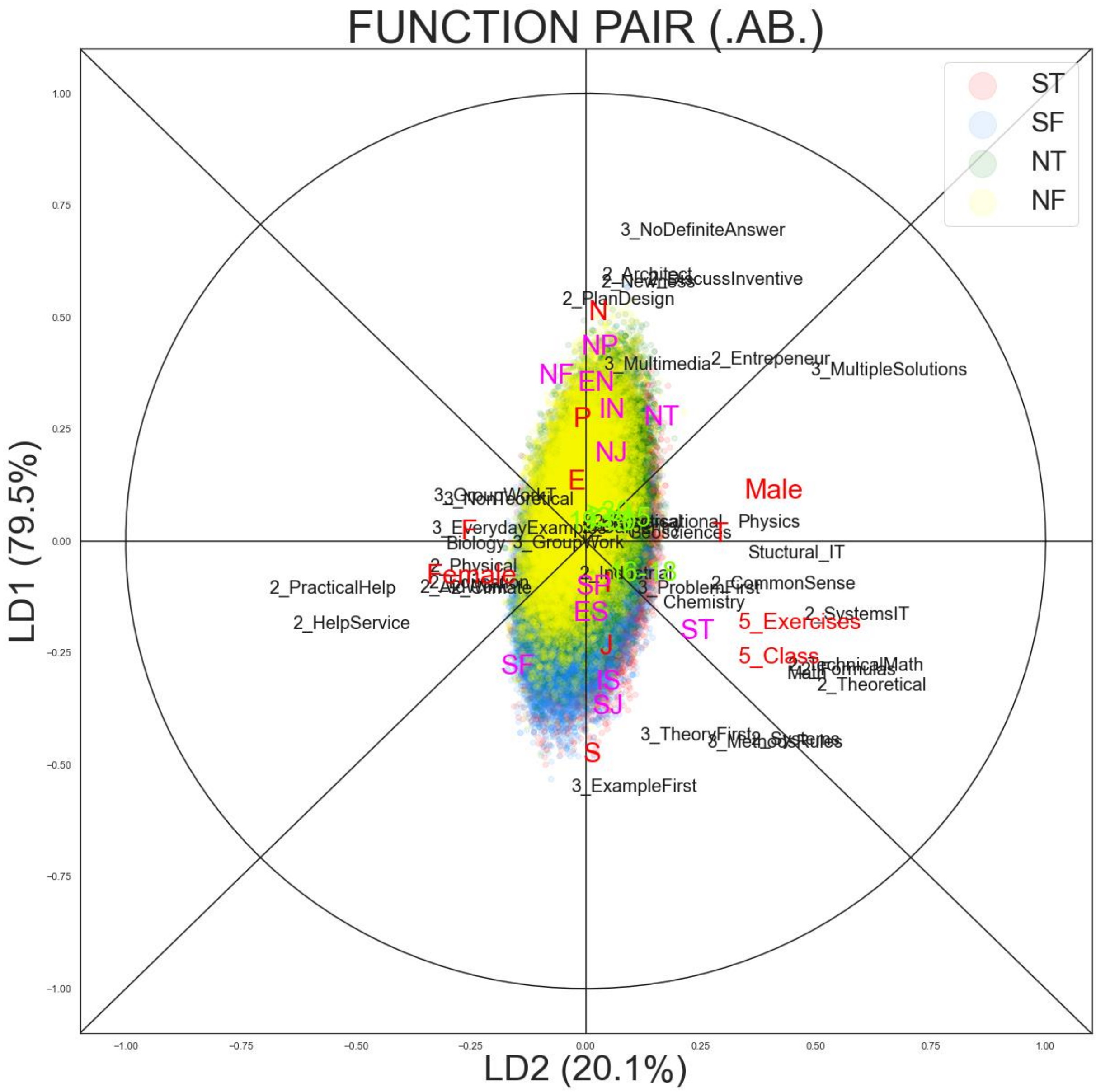


Figure 3.18: *Correlation scores and -weights which resulted from performing LDA with function pair type as the response. Scores are colorized by function pair. The questions, gender, single & paired letter-types, and age group are plotted as text points. The LD1 signs were changed (i.e., the coordinates were flipped) for visualization purposes. The paired letter types shown are: "xAxB", "xABx" (function type), and "ABxx".*

Figure 3.18 show that Intuitives group towards variables such as "Architect", "PlanDesign", "Newness" and "NoDefiniteAnswer". The largest distance between same-dimensional letter-types were found for the S/N dichotomy. Females grouped close to variables such as "Biology", "Climate" and "GroupWork", while males wound up closer to "Physics", "IT", and a more conversational type of lecturing. ST's were the function pair closest to the segment 5 questions, and Thinkers were the letter type closest to most of the course-variables (e.g., "Geosciences") – an exception to this is "Biology", which grouped close to Feelers.

Contour decomposition plots, for both the "function pair" LDA and the "attitude pair" LDA can be viewed in figure 3.19 The figure illustrates where the various types accumulated along the LD1-LD2 dimensional space.

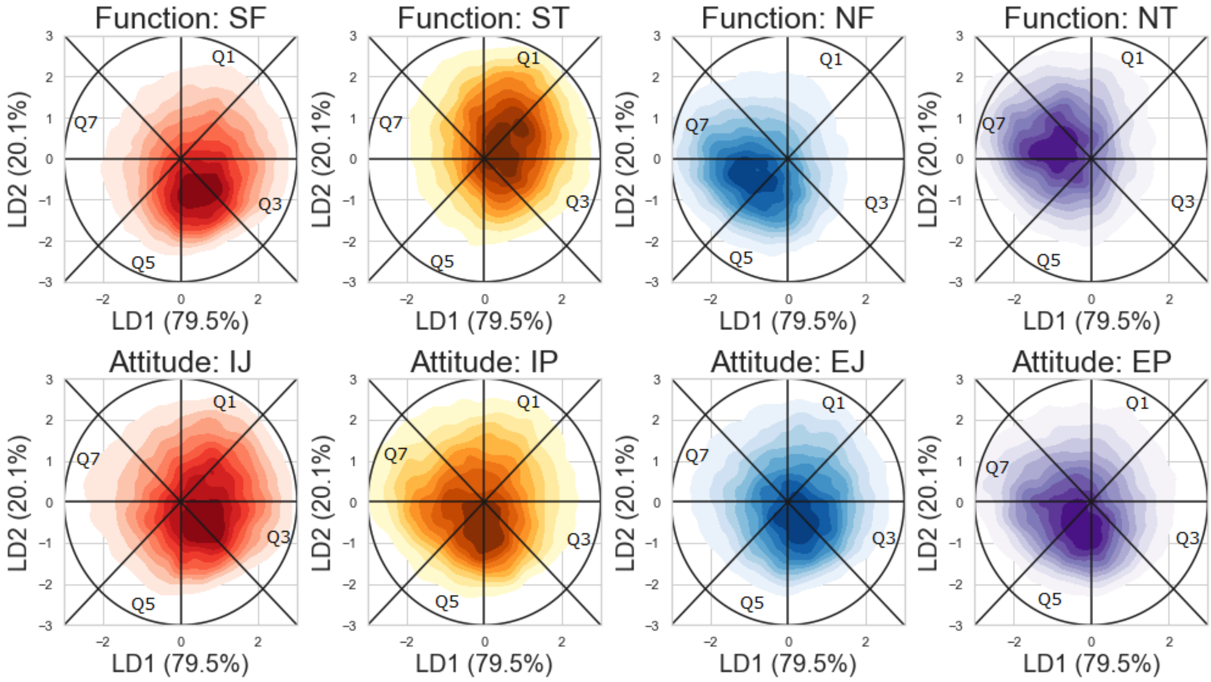


Figure 3.19: *2D LDA decomposition contour plots for component 1 plotted against 2, for the two types of LDA's run (y = function pair and y = attitude pair). All data samples were used to create the contour plots, which are separated by their respective attitudes / functions.*

Through focusing on the center of accumulation of each sub-plot above, it seems that the four function pairs had the least overlap compared to attitude pairs. SF-types accumulated in Q4, NF's in Q6, NT's in Q7, and ST's were somewhat spread between Q1-Q3. Comparatively both IJ and EJ had their centers in Q3-Q4, and IP and EP were both in Q4-Q5 – this means all four attitudes were centered in the lower half in their respective plots, in Q3-Q5.

3.3.3 – Modelling with original data, PCA, and LDA

Using the assumption that the test data was representative of the train data, a set of multivariable linear models was fitted on the test data.

PERCEPTIONS ON MATHEMATICS

Linear regression and Poisson regression models were then fit on the full test data from previously. Results from the analyses follow below.

Table 3.13: *An overview over the various models fitted, significant variables (based on p-value <= 0.05), together with the MAE & MSE values. The model named in the "No." column refer to the names given in the methods-chapter(i.e., linear regression, poisson with). The second column lists significant variables obtained from the R anova() function, together with notable mentions from the R summary() function.*

<table>
<tr><th>Model</th><th>Summary & ANOVA</th><th>MAE</th><th>MSE</th></tr>
<tr><td>M1
REG</td><td><pre>
Coefficients:
                              Estimate Pr(>|t|)
(Intercept)   2.886524 < 2e-16 ***
DecEnergy_pairEN   -0.246618 < 2e-16 ***
DecEnergy_pairES   -0.040601 0.14465
DecEnergy_pairIN   -0.240964 < 2e-16 ***
StrFeel_pairFJ   0.300113 < 2e-16 ***
StrFeel_pairTJ   0.518364 < 2e-16 ***
StrFeel_pairTP   0.210389 3.02e-10 ***
YYYY2018   0.014616 0.91701
YYYY2019   -0.054348 0.69780
YYYY2020   -0.121520 0.38469
Aldersgruppe16-18   -0.599060 0.00555 **
Aldersgruppe30+   0.252131 0.52722
KjonnM   0.198595 < 2e-16 ***
Semester   -0.026481 0.02199 *
DecEnergy_pairEN:StrFeel_pairFJ -0.043232 0.33134
DecEnergy_pairES:StrFeel_pairFJ 0.028280 0.44675
DecEnergy_pairIN:StrFeel_pairFJ -0.050090 0.25177
DecEnergy_pairEN:StrFeel_pairTJ -0.097932 0.04097 *
DecEnergy_pairES:StrFeel_pairTJ -0.069794 0.08832 .
DecEnergy_pairIN:StrFeel_pairTJ -0.079433 0.15178
DecEnergy_pairEN:StrFeel_pairTP -0.066017 0.14593
DecEnergy_pairES:StrFeel_pairTP 0.017383 0.70638
DecEnergy_pairIN:StrFeel_pairTP 0.008456 0.86979
YYYY2018:Aldersgruppe16-18   0.836673 0.00012 ***
YYYY2019:Aldersgruppe16-18   0.852579 8.47e-05 ***
YYYY2020:Aldersgruppe16-18   0.873092 5.67e-05 ***
YYYY2018:Aldersgruppe30+   -0.155659 0.69795
YYYY2019:Aldersgruppe30+   -0.306300 0.44426
YYYY2020:Aldersgruppe30+   -0.292600 0.46423
</pre>
$R^2 = 0.06128$

ANOVA:
All significant except "ABxx" : "xxAB"</td><td>0.9229</td><td>1.2554</td></tr>
</table>

M2 POISSON	Coefficients: Estimate Pr(>|z|) (Intercept) 1.060852 < 2e-16 *** DecEnergy_pairEN -0.086997 1.85e-09 *** DecEnergy_pairES -0.013569 0.34810 DecEnergy_pairIN -0.085014 2.32e-08 *** StrFeel_pairFJ 0.096797 3.22e-16 *** StrFeel_pairTJ 0.159809 < 2e-16 *** StrFeel_pairTP 0.067971 5.21e-05 *** YYYY2018 0.005304 0.94103 YYYY2019 -0.017532 0.80645 YYYY2020 -0.040354 0.57234 Aldersgruppe16-18 -0.219604 0.06327 . Aldersgruppe30+ 0.078633 0.68826 KjonnM 0.063620 < 2e-16 *** Semester -0.008543 0.14502 DecEnergy_pairEN:StrFeel_pairFJ -0.006219 0.78816 DecEnergy_pairES:StrFeel_pairFJ 0.009543 0.61395 DecEnergy_pairIN:StrFeel_pairFJ -0.008807 0.69938 DecEnergy_pairEN:StrFeel_pairTJ -0.015890 0.51312 DecEnergy_pairES:StrFeel_pairTJ -0.017833 0.38258 DecEnergy_pairIN:StrFeel_pairTJ -0.010289 0.71325 DecEnergy_pairEN:StrFeel_pairTP -0.015241 0.51566 DecEnergy_pairES:StrFeel_pairTP 0.006482 0.78115 DecEnergy_pairIN:StrFeel_pairTP 0.010157 0.70258 YYYY2018:Aldersgruppe16-18 0.293558 0.01356 * YYYY2019:Aldersgruppe16-18 0.300674 0.01128 * YYYY2020:Aldersgruppe16-18 0.309269 0.00914 ** YYYY2018:Aldersgruppe30+ -0.047744 0.80866 YYYY2019:Aldersgruppe30+ -0.097205 0.62147 YYYY2020:Aldersgruppe30+ -0.092527 0.63781 AIC = 132 056	0.9219	1.2547

Of the models above, the lowest MAE and MSE was found for the GLM-Poisson model. Variables that were significant for this model at p <= 0.05, were the paired Intuitive-types (EN & IN), all agreeableness-conscientiousness paired letters, gender, and the combined effect of year and the 16-18 age group. Similar findings were also found in the regression model, except that semester, the 16-18 age group, and the EN-TJ combined effect were significant as well. Of the significant Poisson-model variables, these had positive estimates: FJ, TJ, TP, Gender male, and the combined effect of year and the 16-18 age group. Negative estimates were found for the EN and IN types.

The importance of various personality type variables was then assessed in Poisson-models, by evaluating their deviances. Del5_1, and other variables relating to perception of math, were set as the response. A quick summary is found in table 3.14.

Table 3.14: *A summary of various simple glm-Poisson models with a "Response" variable, the deviances of the X-variables (obtained using the anova-function) and whether their estimated (if p-value <= 0.05) were positive or negative.*

Response	Models					
	Variable	Deviance	Estimate	Variable	Deviance	Estimate
Del5_1 *"Math class"*	xxAB ABxx	363.2 321.4	FJ,TJ,TP > 0 EN,ES,IN < 0	xABx AxxB	467.2 215.4	NT,SF,ST > 0 EP,IP < 0
	SN PJ TF IE	307.1 205.2 156.1 14.0	T > 0 N,J,E < 0	xAxB AxBx	513.7 172.2	ET,IT,SJ > 0 NP < 0
Del5_2 *"Math confidence"*	xxAB ABxx	340.0 181.3	ES,FJ,TJ,TP > 0 EN,IN < 0	xABx AxxB	408.5 108.1	NT,ST,EJ > 0 NF,EP,IP < 0
	TF SN PJ IE	254.8 145.3 83.9 31.3	E,T,J > 0 N < 0	AxBx xAxB	300.0 233.7	EF,ET,IT > 0 NJ,NP,SP < 0
Del2_5d *"Math in workplace"*	ABxx xxAB	1740.4 1145.6	FJ,TJ,TP > 0 EN,IN,ES < 0	xABx AxxB	2100.7 805.1	ST > 0 NF,EJ,EP,IP < 0
	SN TF IE PJ	1184.2 874.4 556.2 261.0	T,J > 0 E,N < 0	xAxB AxBx	1457.1 1448.6	ET,IT > 0 EF,NJ,NP,SP < 0
Del4_4 *"Math"*	ABxx xxAB	1196.0 1129.3	FJ,TJ,TP > 0 EN,IN,ES < 0	xABx AxxB	1835.0 491.1	ST > 0 NF,EJ,EP,IP < 0
	SN TF PJ IE	1034.8 793.9 334.6 159.9	T,J > 0 E,N < 0	xAxB AxBx	1373.1 970.5	ET,IT > 0 EF,NJ,NP,SP < 0

For Del5_1, the highest deviances were found for the S/N-dimension, function pair (xABx), and the combination of S/N and P/J (xAxB). T-types tended to have positive estimates (e.g., TJ, TP, NT, ET), while E- and N-types tended to have negative estimates – except for the NT and ET types.

For Del5_2, pertaining to confidence solving math exercises and own capabilities in math, the highest deviances were found for T/F, the combination of T/F and P/J (xxAB), function pair, and the I/E – T/F pair. Extraverted types tended to have positive parameter estimates (except for EN and EP), while N- and P- types mostly had negative estimates (except for TP and NT).

For Del2_5d, of whether the participants would like a job that uses math actively, T- and J- types tended to get positive estimates, whereas Intuitives received negative estimates. Highest deviances were found for I/E-S/N (ABxx), S/N, and function pair. In general, all personality type variables received high deviance scores with Del2_5d as the response variable.

Lastly, in Del4_4, high deviance scores were again observed. Highest were for function pair, S/N, and S/N-P/J (xAxB). Nearly all E- and P- types had negative estimates (except ET and TP), and all Thinkers had positive estimates.

LEARNING STRATEGIES

Similar modelling as was done in table 3.14 above was also done for various educational preference variables as the response. As before, reference levels used were "ABxx" = IS, "xxAB" = FP, "xAxB" = SJ, "AxBx" = IF, "xABx" = SF, "AxxB" = IJ, IE = I, SN = S, TF = F, and PJ = P. The results are shown below.

Table 3.15: *A summary of various simple glm-Poisson models with a "Response", variables, their deviances (obtained using the anova() function), and whether their estimates (if p-value <= 0.05) were positive or negative.*

Response	Models					
	Variable	Deviance	Estimate	Variable	Deviance	Estimate
Del3_1d *"Group work"*	ABxx xxAB	782.9 65.7	EN,ES > 0 IN,TJ,TP < 0	AxxB xABx	808.9 23.6	SF,EP > 0 NT,IJ,IP < 0
	IE TF SN PJ	719.6 56.3 40.6 7.9	E > 0 N,T,J < 0	AxBx xAxB	780.6 48.8	SJ,SP > 0 ET,IT,IF < 0
Del3_1b *"Methods first"*	ABxx xxAB	2192.5 169.7	FJ,TJ > 0 EN,IN,ES < 0	xABx AxxB	1779.1 585.1	SF,ST,IJ,IP > 0 EP < 0
	SN IE PJ TF	1555.1 632.2 167.9 0.8	J > 0 E,N < 0	xAxB AxBx	1731.7 633.5	IF,IT,SJ,SP > 0 NP < 0
Del3_2c *"Multiple solutions"*	ABxx xxAB	1461.4 537.8	EN,IN,ES,TJ,TP > 0	xABx AxxB	1975.0 37.8	NT > 0 SF,ST,IJ < 0
	SN TF IE PJ	1309.4 517.0 144.4 14.7	E,N,T > 0 J < 0	xAxB AxBx	1322.9 677.6	ET,IT,NP > 0 IF,SJ,SP < 0
Del3_2a *"Everyday examples"*	xxAB ABxx	205.0 10.6	--- FJ,TJ,TP < 0	xABx AxxB	165.9 50.6	EP,IP > 0 NT,ST < 0
	TF PJ SN IE	155.8 48.9 6.5 3.9	--- T,J < 0	AxBx xAxB	160.2 55.6	NP,SP > 0 ET,IT < 0
Del3_3b *"Methods rules"*	ABxx xxAB	2169.1 522.6	FJ,TJ,TP > 0 EN,IN,ES < 0	xABx AxxB	2162.7 531.8	NT,SF,ST,IJ > 0 EP < 0
	SN IE TF PJ	1787.9 381.2 266.1 256.1	T,J > 0 E,N < 0	xAxB AxBx	2046.1 647.6	ET,IT,IF,SJ,SP > 0 NP < 0
Del3_1a *"Problem first"*	ABxx xxAB	62.6 33.7	TP > 0 EN,IN,ES,FJ < 0	xABx AxxB	54.2 40.2	NT,SF,ST,EP,IP,IJ > 0
	SN IE TF PJ	34.1 25.6 17.3 16.2	T > 0 E,N,J < 0	xAxB AxBx	51.0 44.8	ET,IT,IF,SJ,SP > 0

Del3_1d asked whether the participant preferred group work. I/E, attitude pair (AxxB), I/E-T/F, and I/E-S/N had the highest deviance values. Sensing and Extraverted types (except ET) received positive parameter estimates. Conversely, Thinkers and Intuitives got negative values.

When asked in Del3_1b whether they liked the teacher to present an example (i.e., go through methods) before divulging into the theory, Judgers and T-types received positive estimates. Extraverts got in comparison negative estimates, alongside Intuitives. Highest were the deviances for function pair, S/N-P/J, the I/E-S/N pair, and the S/N-dimension. T/F had a deviance of near 0.

With Del3_2c as the response, deviances were highest for I/E-S/N, the S/N-dimension, function pair, and S/N-P/J. N-types tended to get positive estimates, while their counterpart the S-type tended to get the opposite. There was also some inclination for Thinkers to get positive numbers, and Judgers to get negative.

Del3_2a asked if the test-taker preferred their teacher to use examples from everyday life. Fewer types received significant parameter estimates at alpha = 0.05 than the previous section. Of these, Perceivers tended to get positive numbers (except TP), and Thinkers got negative estimates.

The most notable variables in the Del3_3b-model was the S/N-dimension, function pair, I/E-S/N, and S/N-P/J. Intuitive and Extraverted individuals got negative estimates, unless paired with the Thinking-letter, and T- and P- types often received positive estimates.

Lastly were the Del3_1a-models, in which received the lowest deviance values for all variables in all four models. Thinkers and S-types (apart from ES) got positive estimates. Extraverts, when not paired with either P or T, received negative estimates.

PCA vs. LDA

Logistic regression models were fitted for all four LDA / PCA models. Scikit-Learn's "GridSearchCV" was run on all four, using either 4-letter type or function pair as the response and PCA or LDA components as the explanatory variables. The best combinations of hyperparameters found were:

- PCA-TYPE & -Function: C = 0.01, penalty = L2
- LDA-TYPE & -Function: C = 1, penalty = L2

The resulting precision and F1-scores, obtained after fitting the models on their respective data sets and predicting on the test-data, is presented in table 3.16.

Table 3.16: *Precision and F1-scores for the various models having either 4-letter type or function pair type as response, and correlation scores (from either PCA or LDA) as the explanatory variables.*

TYPE	Precision		F1-score		Function pair	Precision		F1-score	
	PCA	LDA-Type	PCA	LDA-TYPE		PCA	LDA-function	PCA	LDA-function
ESTJ	0.15	0.15	0.11	0.12	ST	0.44	0.44	0.45	0.45
ESTP	0.09	0.11	0.06	0.10					
ISTJ	0.19	0.19	0.25	0.24					
ISTP	0.15	0.14	0.16	0.14					
ESFJ	0.13	0.16	0.14	0.16	SF	0.44	0.42	0.44	0.43
ESFP	0.14	0.14	0.14	0.14					
ISFJ	0.17	0.16	0.20	0.18					
ISFP	0.12	0.10	0.12	0.10					
ENTJ	0.11	0.09	0.07	0.07	NT	0.43	0.43	0.41	0.42
ENTP	0.12	0.14	0.11	0.13					
INTJ	0.14	0.15	0.10	0.12					
INTP	0.16	0.19	0.19	0.22					
ENFJ	0.14	0.11	0.11	0.08	NF	0.43	0.43	0.43	0.43
ENFP	0.17	0.16	0.20	0.18					
INFJ	0.12	0.16	0.13	0.16					
INFP	0.19	0.19	0.18	0.19					
Accuracy	PCA: 0.15 LDA: 0.15				Accuracy	PCA: 0.44 LDA: 0.43			

Table 3.16 show that there were not many differences between LDA and PCA. However, when modelling and prediction was done using function pair types rather than the four-dimensional types, precision and F1-scores of nearly 4 times the amount were obtained. ISTJ, INTP, INFP and ENFP received the highest scores, and were most easily modelled. Differences between function pairs were smaller, but highest scores were obtained for the S-types.

4. – Discussion

4.1 – Answers to my hypotheses

Associations were found between personality types and perceived difficulties in math (A). Specifically, Sensing and Judging individuals were more positive towards math class and had higher self-esteem regarding their mathematical capabilities. This was partly observed for the Thinking type too, causing the xSTJ to be the most math-positive type. Likewise, other types were in disfavor of the topic. Intuitives and Perceivers consistently tended to dislike math class, felt their skills were inadequate, and avoided math in general. Feelers that had the Judging-attitude tended to be positive, contradicting my theory, while the remaining F-types were mostly neutral. Lastly, the S/N and the T/F were found to be the most important dimensions in predicting whether the individual would like or dislike math. However, the biggest differences were not observable for the T/F dimension, as hypothesized, but rather the S/N.

Analysis further showed that there was a certain segmentation between function-pairs for work preferences and learning styles (B). NF's preferred creative pursuits and open-ended tasks. NT's shared many similarities with NF, but were more inclined towards the technical-mathematical fields than the creative. ST's liked technical-mathematical work and traditional formula-based education, while SF's tended to prefer service-oriented work and group work. Considering the latter, group work was indeed preferred by both Extraverts and Feelers, and somewhat by Sensing individuals as well. The E- and F-types liked working in environments where they could discuss and be part of a group, and disliked authoritative and formula-heavy teachers or courses. Furthermore, Feeling people were found to like service-oriented and altruistic work, more so than any other type. Finally, Intuitives were predisposed towards creative methods of education and open-ended assignments, also in line with my working hypothesis (B).

Analyzing sample 1 found that there was a relationship between Big-Four and Big-Five dimensions. Agreement scores between Utd1 and the Big-Five test were surprisingly high when binarized using median values. Models also found that four of the six T-tests that were significant were between the hypothesized Big-Four and Big-Five dimensions, further supporting my hypothesis (C).

Comparative analyses between Utdanningstesten (Utd1) and the other tests supported the theory that agreement would be the highest between Utd1 and Utd2 (a). This was true for all dimensions except for the I/E dimension, of which the Utd1 – Uroboros comparison were the highest. Besides this dimension, the agreement with Uroboros was low – particularly for the 4-letter- and S/N-type comparisons. This shows that while Utdanningstesten does well in temporally separated studies (Utd1 vs. Utd2), Utdanningstesten might not be representative of the Uroboros questionnaire.

As further theorized, single-dimensional comparisons resulted in the highest agreement. This was only natural, as the full type is dependent on the prediction of four separate dimensions, increasing the chances of "mislabeling". It is unknown why exactly the I/E dimension tended to give the best agreement and S/N the lowest. The former is concerned with the tangible world of socialization and people, while the latter is more internalized. Perhaps that is why Introverts and Extroverts are easier to separate than Sensing and Intuitives, in either a questionnaires wording or the test-taker.

A comparison of the unsupervised PCA to the supervised LDA found that no method appeared to be significantly better than the other. While PCA did a better job at spreading the function types, the Logistic Regression models generally gave similar results for both PCA and LDA in terms of performance metrics. The same variables were also found to group together in the respective figures. However, LD-components appeared to do a slightly better job at predicting 4-letter personality types (b).

Comparison of the two models with Del5_1 (math class "anxiety") as the response found that the GLM Poisson model were slightly better than the standard linear regression model. As expected, gender-, age-, and paired-type-variables were found to be significant predictors – however, age was only significant in the regression model (c). In both models though, the high-school aged participants were negative, while perception became more positive with age. Deviance analysis on the "perception on mathematics" models found that the S/N dimension and function pair tended to be the most important variables. One exception to this was the Del5_2 (math confidence) model, where I/E and T/F appeared to be the most influential.

4.2 – My study; Strengths and Weaknesses

Before I address where my findings stand in comparison to other studies within the field, I will address any strengths, weaknesses, and other particulars of my study.

Sample1 had some weaknesses. For one, it was quite small, and not even all 4-letter types were represented. Among those who did partake, a higher percentage might have been Judgers (i.e., high in conscientiousness) compared to the average student population, as conscientious people are more likely to volunteer (Mike & Jackson & Oltmanns, 2014). Another problem is due to the way sample1 was gathered. First, it was specified that participants would have to wait 1 week between Utd1 and Utd2. This was later decreased to 1-2 days – or 1-2 hours, as in the case with the Vitenparken sampling process. Although reliability tends to be higher when temporally separated studies do not take too long a time between the tests (Myers et.al., 1998), separation should not be too short due to question memorization. It is therefore possible that the agreement between Utd1 and Utd2 got somewhat artificially inflated. Furthermore, agreement between Utd1 and Uroboros may appear lower than it should be due to the long format of the latter. This is supported by the idea of careless responding, which tends to increase when participants must take longer tests (Eisele et.al., 2020).

There was also some discrepancy between sample1 and sample2 which needs to be addressed. While the individuals in sample1 took the post-expansion (/new) version of Utdanningstesten, sample2 only consisted of older data – before the number of personalities discriminating questions were increased from 4 to 12. Thus, the version of Utdanningstesten that went through the validation process is not the same as the one analyzed in sample2. Although the predictive power of a questionnaire can increase if more relevant items are added (Johnson, 2014), other studies have found no significant effects of questionnaire length on accuracy rates (Kato & Miura, 2021). Emphasizing the former, a hypothetical post-expansion sample2 data set could prove to give better and more accurate results in the future. ‘

Sample2 also had very little additional information, besides age, gender and county. This means there was no way for me to control for hidden duplicates caused by people taking the test multiple times but changing their answers for fun. The sample also had no information about socioeconomic, cultural, or ethnic background. These are all factors that affects behavior (Myers et.al., 1998).

The descriptive results obtained from Utd1 in sample 1 generally mirrored (Aspheim, 2020)'s, as most of Aspheim's respondents were also female, Sensing, Feeling, and Judging. The only differences lay in the I/E dimension, where I found more Introverts compared to her equal distribution. The two most common 4-letter types in both sample 1 and sample 2 were also identical to that of (Myers et.al., 1998).

Compared with the American sample in (Myers et.al., 1998), sample2 had an overabundance of Intuitive Feelers (NF). In addition to the Norwegian sample2 having 40% more Intuitives, Perceivers were also more common. This can either be due to socioeconomic and cultural differences, or it is an effect of how the data were sampled: Sample2 was gathered on a voluntary basis, while the sample in (Myers et.al., 1998) was actively recruited. The numbers were however flipped in sample1; While N and P dominated in sample2, S and J dominated in sample1. This may be due to how sample1 mainly consisted of university students, as S- and J-types tend to be the most numerous types in universities (Jang, 2018)(Vinje et.al., 2021).

I found that the balance of type for the Big-Five dichotomies seemed to change with age. (Kainz, 1985) stated that "younger extraverts did show a decrease in reported extraversion over time", supporting my observations that the number of extraverted males decreased with age. The author also described how changes in P/J can be expected during the developmental stages. (Kainz, 1985) further noted that the number of Intuitives usually declines slowly with age, before dropping in among 18–20-year-olds, the same of which I observed for the 16-18 and 19-30 age groups. However, be careful to overinterpret these results, as my data was not paired, i.e., the data did not contain information about the same individual across different ages. As the environment affects how your personality is shaped growing up, and different generations may end up with different worldviews, we should be careful to not overanalyze what could "just" be generational differences.

In the methodical sections of my work, I was careful to use techniques that would ensure reproducibility. When data was shuffled or randomly sampled, either with or without a criteria of equal label representation, one specific random state (or seed, as in the case of R) was used to ensure the same sampling was done each time. Furthermore, data was shuffled before modelling to avoid that any models fitted would wrongly weight any artificial patterns (individuals taking the test multiple times in succession, or data presented in order of type due to the iterative label-ratio-specific sampling). When constructing the Poisson models in RStudio, the categorical X-variables were set as factors to be defined by R as nominal instead of ordinal. The most common group were set as the reference levels, so that the intercept of the model better reflects the average sample population.

The PCA and LDA analyses were performed on the same train dataset, of which was scaled beforehand to ensure that any variables with significantly higher variances than the other was not weighted more as a result (Raschka & Mirjalili, 2017). The Grid Search cross validation found the optimal hyperparameter dimension for the models. On one hand, this increased the models' chances at performing their best for their respective data sets. On the other hand, as all four models were not compared using the exact same model, this came at the cost of comparison. I do however support the choice I made to tune the models differently, as I wanted to compare the LDA- and PCA-models at their relative peak – equity over equality.

Relating to the validity and reliability analyses, it could have been advantageous to present additional inter-rater reliability scores in conjunction with Cohen's Kappa. The Kappa does not factor in "near-misses", e.g., it would label a comparison that were just 1 letter off ("ISTJ vs. ISFJ") the same as it would a comparison that were 3-4 letters off ("ISTJ vs. ENFP"). Weighted Kappa's could be one way to word around this, as it would be able to differentiate between relatively proximal and distant ordinal categories (DeVellis, 2005). Wongpakaran et.al. (2013) also proposed that a statistic called "Gwet's AC1" should be used in place of Cohen's Kappa, as it was found to be more stable and "less affected by (...) marginal probabilities" (p.7). Gwet's AC1 also considers that there may be natural differences between the "raters", which it mediates (Lydersen, 2016). An example of this would be how Uroboros was much longer than Utdanningstesten, and thus a greater cognitive burden (Eisele et.al., 2020).

When a response variable is discrete, a choice must be made as to whether it should be treated as numeric or categorical during modelling. In the case of the variable Del5_1, responses had been coded to integer values between 1-5, but each number corresponded to a specific statement of different wording. E.g., “1” was not translated from “Dislike the most”, just as “3” was not originally “Neither”. Instead, the values 1 through 5 represented statements using different adjectives, with only a general sense of Likert-scale order. For example, “1” describes dreading and being anxious towards math, while “2” describes being bored and not interested in math. When I made the linear regression and GLM-Poisson models with Del5_1 as the response, I treated it as a numeric variable. Classification models should additionally have been made. This could potentially have yielded even better MAE and MSE results, as the 1-5 coded statements were not on a direct Likert-format.

4.3 – Relation to previous studies

Other studies have also found a significance between math “anxiety” and Big-Four personality types. (Hadfield & McNeil, 1994) noted that, among American elementary school teachers, the Feeling-type of the T/F-dimension were predictive of negative feelings towards mathematics. The same finding was mentioned by (Sæbø & Almøy & Brovold, 2015) and (Brovold, 2014). (Hinkle, 1987) specified that types who were defined by reflective observation tended to disfavor math, while types who were more grounded in “concrete experience” had a negative correlation with math “anxiety”. The Intuitive type is considered a reflective and inwards-thinking type (Brovold, 2014), and thus this supports my findings that Intuitives disfavor math.

However, other studies disagree with which dimensions or letter-pairs are prone to negativity. Using Bayesian networks, (Smail, 2016) found that the personality types which had the highest conditional probability of math “anxiety” were ISFJ, INTJ, ENTP, ISFP and INTP. Of these, all had SF or NT as their function type. These results do not entirely support my findings, as I found evidence that having a SF-type was indicative of little anxiety in relation to math class. Furthermore, despite Intuitives’ negative attitudes towards math class and low math-self-esteem, NT’s were found to be net positive. Conversely, ST- and NF-types scored the lowest (i.e., positivity towards math) in the study by Smail, only the first of which my data supports. One dissertation even found that Extraverts, Sensing and Thinking types had higher levels of math “anxiety” than their counterparts (Mancini, 1993), all of which is in direct contract with my findings.

Learning styles used by educators affects the connotations in which students are left with afterwards (Brady & Bowd, 2005). Negative associations towards math can negatively affect student’s learning outcome (Onwuegbuzie, 2010), but offering alternative teaching methods have been shown to alleviate student’s anxiety (Brady & Bowd, 2005). Therefore, it is important that the school system offer some degree of alternative or personalized education. Previous studies have noted that different types are predisposed to like certain learning styles (Vinje et.al., 2021), and have suggested specific learning styles that is best suited / preferred by the types. Before I compare their findings to mine, I want to further synthesize and interpret my results. Table C4 in the appendix contains my suggestions as to how lectures and tasks “should” be formulated for each letter type, as decided by their preferences. A short summary is presented in table 4.1.

Table 4.1: *Overview over learning strategies and -structure as preferred by different Big-Four types, synthesized from my findings.*

Educational style		Types
Presentation style	Instructive, Theory and Details	S, T ISxJ
	Everyday examples / Try first	(S), N, F, P ExFP
Working style	Individual	I, N, T
	Group	E, S, F ENxx
Task formulation	Open-ended, Creative, Contextual	N, P INxP, xNTx
	Close-ended, Concrete	S, J STxx, IxxJ, IxTx
Formulas	Yes	I, S, T, J IxTJ
	No, General concepts	E, N, F, P ExFP

As discussed, type-dependent preferences towards learning styles were found in my study. Extraverts’ preferences for group work and a contextual-centered presentational style, as well as introverts’ individualistic-reflective and more traditional sequential learning style, did not come as a surprise (Vinje et.al., 2021)(Myers et.al., 1998)(Melvin, 2013). For the S/N dimension, Myers et.al. (1998) defined Intuitives' to prefer reflective judgement. This can be represented through the Intuitives in sample2, who had an apparent aversion towards group work. Both (Myers et.al., 1998) and (Vinje et.al., 2021) support my findings that Intuitives also likes holistic / contextual learning with clear goals, as well as open-ended questions. (Myers et.al., 1998) noted that Feeling types often liked to be taught using experimental methods and a holistic view of the subject at hand. I did not find significant results to support the former, but the F-types in my sample did prefer it when lectures would start with general examples to put the themes into context. Myers et.al. also found Thinkers and Judgers to enjoy fact-retentive tasks, in the same way as to how I found them to like courses teaching about methods and rules. Furthermore, both (Myers et.al., 1998) and (Vinje et.al., 2021) also found that Perceivers tend to prefer a certain level of learning-by-doing, and that the type had an innovative streak to them.

The trends that emerged in my correlation scores plots regarding work preferences reflected the results obtained in Brovold's 2014 doctoral dissertation. In his PLS PCA correlation scores plot (p.356), F were grouped with service- and humanitarian-oriented work, while Thinkers and Introverted were angled towards theoretical work. Furthermore, "his" Perceivers and Intuitives often chose creative and abstract work as well, while Judgers and Sensers often chose the opposite – technical and practical fields of work. One difference I found between our studies, however, was that the Extraverted in my sample seemed to be drawn between the F-direction of public service and practical work, and the N-direction of newness and creative work. In comparison, the Extraverted individuals in (Brovold, 2014)'s sample were far removed from abstract work. One important note though is that Brovold had used the Uroboros questionnaire while my results stem from Utdanningstesten. However, when Utdanningstesten was compared to Urorobos in my sample1, the Kappa score for the I/E dimension was particularly high. To that extent, I believe my results holds some merit. The remaining explanation in differences between our studies may therefore be due to methods (and perhaps preprocessing) used, and sample size.

As stated above, other studies were also in favor of group-centered work for Extraverted types (Vinje et.al., 2021)(Myers et.al., 1998)(Melvin, 2013). The study by

Vinje et.al. also specified that Feeling types like to work with others and “study in dialog” (p.187). S-types were furthermore also defined as collaborative workers (Myers et.al., 1998). In the same way, I, T, and N types tend to prefer working alone (Vinje et.al., 2021). When talking about personality types, it can be natural to define each dichotomous type in relation to which way they are turned – towards the outer world of people and objects, or the inner worlds of thoughts and possibilities. E, S, and F have been defined as outward-oriented (Puji & Ahmad, 2016)(Brovold, 2014). It is then no coincidence that these were the types I found to like group-work the best.

When formulating hypothesis (C), I based it on (Furnham, 1996)’s analyses on the correlations of Big-Four and Big-Five dimensions. Furnham found that the biggest significance lay between I/E in Extraversion, S/N in Openness, T/F in Agreeableness, and J/P in Conscientiousness. My findings mirrored his. Furthermore, the Big-Five results (grouped by Utd1-letter-type) provided similar distribution tendencies to that of McCrae and Costa; In their 1989 paper, they found that Extraverted scored higher in Extraversion than Introverted, N-types were higher in Openness, Feelers in Agreeableness, and Judgers scored higher in Conscientiousness than their counterpart. (McCrae & Costa, 1989) did however also find additional significant paired differences. At alpha=0.05, Feelers scored significantly higher in Extraversion, as I too found. They did however not find differences between I/E for Agreeableness, as I did. Nor did (Furnham, 1996). On the other hand, contradictory to McCrae and Costa, I did not find any significant differences between T/F in Conscientiousness.

In my hypothesis (a), I defined the expectation that the test-retest reliability of Utdanningstesten (Utd1 vs. Utd2) would be the highest among all comparisons. Test-retest (intra-rater) reliabilities are often found to be higher than inter-test reliabilities (Carlson, 1985)(Schlager et.al., 2018). This is due to inter-rater sensitivity created by differences in scales (Cohen, 2017). The standard for what is considered a “good” or “acceptable” kappa value is arbitrary (Xia, 2020)(McHugh, 2012). Kappa values between 0.4 and 0.75 are often considered to be moderate to good, while >= 0.75 represents excellent agreement (Queen’s University, 2022)(Xia, 2020). As such, only the Utd1-Utd2 comparison yielded sufficiently good Kappa values for all five type comparisons (4-letter type, and the four dimensions). In addition to this, the P/J-dimension in “vs Utd2” and the I/E-dimension in “vs Uroboros” were the only that received excellent agreement.

In (Myers et.al., 1998), the test-retest correlations of the MBTI were: I/E (0.79), S/N (0.83), T/F (0.62), and P/J (0.82). While the test-retest correlations of Utdanningstesten (estimated by squaring the Kappa values (McHugh, 2012)) were both quite high and close to the MBTI's for the I/E-, T/F- and P/J-dimension, the S/N barely surpassed a score that was ¼ of the MBTI's correlation.

A 2019 paper by Almanza-Ojeda et.al. sought to predict Big-Four personality traits from social media data that was transformed using either PCA or LDA. (Almanza-Ojeda et.al., 2019) made three models: SVC, Logistic Regression, and Random forest – the latter of which is what the XGBoost algorithm I briefly tested is based on. Unlike me they 1) classified directly on the single-letter dimensions (as opposed to function pair and 4-letter type), and 2) they used (non-correlation) scores for *all* components for the analysis. In SVC and Random Forest, the prediction accuracy was on average the same between methods, resembling the results I found for function pair classification. In the Logistic Regression model however, PCA performed the best. This was particularly true for the attitude dimensions I/E and P/J. These results oppose both my original theory, as well as my findings (somewhat).

However, it is worth noting that the PCA models I made contained only the first 14 components, totaling ~70% of the original variance. In comparison, the LDA-models I made (both the function pair- and type-response versions) had ~99.6% of their variability explained by the components used. If the number of PCA-components used in the models had been increased, and the total data variance explained by the components were closer to that of the LDA, the PCA-models could potentially have performed better (Howley et.al., 2005). Alternatively, the PCA components selected could have been subset differently to improve results (Sutter & Kalivas & Lang, 1992). This effect would have been more noticeable for the PCA than LDA due to the higher number of components in the former.

There have been found some contradictory results as to which factors are predicative of math perception. Using a sample of N=134 American undergraduate students, (Tapia & Marsh, 2004) noted that gender had no such effect. However, consensus seem to be that there is an effect of gender (Lazarides & Rubach & Ittel, 2016).

(Woodard, 2002) and (Akey, 2001) both found that female students were significantly more math anxious than males, and (Dønnestad, 2019) noted that females were less motivated by math than males. This I too found, as represented by the factor-level "male" receiving a significantly high (positive) model estimate. Despite this, girls at lower-level education in Norway tend to receive math grades that are similar (or higher) than males and struggle less in math class (Dahle, 2009)(Teigmo, 2019), possibly because they tend to better at self-regulating than males (Dønnestad, 2019).

(Woodard, 2002) further analyzed the effect of age, by grouping their participants into two categories: below and above 25 years old. As opposed to my study, they found no significance of age in predicting math "anxiety". On the other hand, it has been noted that young children "tend to hold positive academic self-concepts that are not strongly correlated with their actual achievement" (Lazarides & Rubach & Ittel, 2016, p.127), indicating that themes such as math-confidence and general interest/enjoyment are considered age-specific.

Other studies have also found psychological types to be associated with preferred learning styles. My findings, if we focus on the single-letter types, mirrored that of (Myers et.al., 1998) and (Vinje et.al., 2021). The former presented preferences based on 4-letter types as well, specifying that function pair is the most important. In fact, while the four scales separately have found widespread approval to predict behavior and preferences, support behind 16 different types have not been as strong (Murray, 1990). On the other hand, support have been found for the combined S/N – T/F (function) types (Myers et.al., 1998), of which my findings seemed to support.

4.5 – Conclusion

In summary, my study found clear associations between types and educational preferences. It could then be wise to incorporate personality type theory into how education is practiced. For math-heavy courses, such as statistics, the most care should be put into the S/N dichotomy (for content definition and theoretical angling), T/F (for implementation and feedback), and P/J (for course structure).

However, Utdanningstesten might not be the optimal tool to assess personality on a Big-Four scale – at least not currently. Comparison with both the Uroboros and the Truity Typefinder yielded promising results in the Extraversion dimensions, but its creator could do wise to further expand or reword the questions used for the remaining dimensions. As (Myers et.al., 1998) notes, the paired contradicting statements should be so "stereotyped" that the participant will be resolute in their forced choices.

Although it is important as an educator to challenge your students with tasks and methods that do not come naturally, it would be unwise to ignore their natural cognitive predispositions. For example, although an individual might be Introverted, working with other people in groups is an imperative skill necessary to enter the work market. Or regarding the use of repetition tasks in the classroom; Though some types find this tiresome, an adequate emphasis on the drill of flow patterns is needed to properly synthesize the material, and to avoid the excessive cognitive load of consistently having to "reinvent" mathematics caused by poor integration (Brovold, 2014). A balance of adversity and ease is suggested, with some structure of choice and some mandatory, to help the students prosper in an environment that is new yet not entirely unfamiliar.

Lastly, I will present you this quote by (Myers et.al, 1998):

"Higher educational professionals who are concerned about retention of students may consider type theory as one way of understanding the students' diversity of needs (…). It is wise, however, to be careful not to stereotype students by (…) psychological type. Educators must ultimately deal with individual students, who often vary from others of their (…) gender or type. Furthermore, students must be provided not only support for their preferred learning styles but also challenge to learn skills that do not come easily and naturally. An appreciation of both differences and developmental needs can help educators to seek an optimum balance" (p.280).

4.6 – Suggestions for future work

Chapter 4.2 presented some weaknesses in my study, as well as what I may have done differently looking back. This chapter will shortly present ideas for further studies and may be seen in context with chapter 4.2.

FOR RESEARCHERS / STUDENTS:

Although the inter-reliability between Utdanningstesten and the Big-Four test showed promising results, I suggest that the inter-reliability be analyzed one more time. If available, funding should be allocated to allow for the use of the official MBTI. In the case of Uroboros, a non-random sample of highly motivated individuals may be necessary to sample, as this could help control for the dangers of careless responding (Eisele et.al., 2020). Gwet's AC1 could also be interesting to analyze alongside Cohen's Kappa. Also, the relationship between the question-variables for the same dimensions may need to be analyzed, particularly for the S/N dichotomy, to see if these describe the same phenomenon and/or are easy to understand by the test-takers.

Should another explorative study be carried out, I suggest implementing additional data besides the variables I have analyzed. This could be data about mental health, Big-Five Neuroticism, choice of studies, ethnic and cultural background, performance (exam scores) in courses, and more. It could also be interesting to link personality to genetics or the microbiome.

The methods I presented regarding how lectures or tasks should be designed could be applied to educational material, and its type-dependencies tested. Perhaps in another study like that of (Aspheim, 2020), who used eye-tracking and galvanic sensors to quantify cognitive processes during learning. Brain-computer interface technology could also be an alternative to the eye-tracking. BCI's directly measures the participant's brain waves and may be a better tool to analyze cognitive processes real-time than steady-state eye-tracking technology (Shih & Krusienski & Wolpaw, 2012).

With reference to what I wrote in chapter 4.2, I would also like to suggest that the wording of Utdanningstesten's segment 5 questions be changed, so that obtained data may be easier to model and understand. The wording on the 5 statements from each sub-section may be changed to better reflect a Likert-scale system. Alternatively, these should be split up. After all, you can be good at math (Del5_1, D/4) while not finding it interesting (Del5_1,B/2), and you can also be afraid to stand in front of the class and solve questions (Del5_2, A/1) while still loving math (Del5_2, F/5) – the first of which may be more linked to mental health problems than math "anxiety".

FOR EDUCATORS:

Educators may review the learning styles and design formats I suggested in table 4.1 and in table C4 in the appendix. Would it be possible to introduce some of these methods into your course while still maintaining a certain infrastructure and not overwhelming the students, and if so, do the students' learning outcomes and overall course satisfaction increase after they obtain the possibilities to choose?

5 – Sources

- 16personalities B (2021). Our Framework. 16personalities.com. https://www.16personalities.com/articles/our-theory
- 16personalities C (2022). Personality types. 16personalities.com. https://www.16personalities.com/personality-types
- Akey, Wayne L. (1991). Personality type and mathematics anxiety factors affecting remedial college freshman. Ohio State University. https://www.proquest.com/openview/0df56bd021b1c4d1c271fb86235fd895/1?pq-origsite=gscholar&cbl=18750&diss=y
- Alemdag, Ecenaz & Cagiltay, Kursat (2018). A systematic review of eye tracking research on multimedia learning. Computers & Education, 125, 413-428. https://doi.org/10.1016/j.compedu.2018.06.023
- Allport, Gordon W. & Odbert, Henry S. (1936). Trait-Names: A psycho-lexical study. Psychological Review Company, Princeton and Albany. http://psych.colorado.edu/~carey/Courses/PSYC5112/Readings/psnTraitNames_Allport.pdf
- Almanza-Ojeda, Dora-Luz & Gomez, Juan C. & Ibarra-Manzano, Mario A. & Moreno, Daniel R.J. (2019). Prediction of Personality Traits in Twitter Users with latent features. The 2019 International Conference on Electronics, Communications and Computers (CONIELECOMP).. https://doi.org/10.1109/CONIELECOMP.2019.8673242
- Anglim, Jeromy & Horwood, Sharon & Smillie, Luke D. & Marrero, Rosario J. & Wood, Joshua K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. Psychological Bulletin, 146(4), 279-323.. https://doi.org/10.1037/bul0000226
- Annesley, T.M. (2010). "It was a cold and rainy night": Set the Scene with a Good Introduction. Clinical Chemistry, 56(5), 708-713. https://doi.org/10.1373/clinchem.2010.143628
- Annesley, T.M. (2010). The Discussion Section: Your closing argument. Clinical Chemistry, 56(11), 1671-1674.. https://doi.org/10.1373/clinchem.2010.155358
- APA 1 (2020). Reciprocal determinism. APA Dictionary of Psychology, American Psychological Association. https://dictionary.apa.org/reciprocal-determinism
- Aspheim, Janne H. (2020). Kva innsikt kan biometriske målingar gje om studentar si oppleving av statistikkundervisning?. NMBU, fakultet for Kjemi, Bioteknologi og Matvitenskap. https://hdl.handle.net/11250/2716937
- Atzil, Shir & Hendler, Talma & Feldman, Ruth. (2011). Specifying the Neurological basis of human attachment: Brain, Hormones, adnd Behavior in Synchronous and Intrusive Mothers. Neuropsychopharmacology, 36, 2603-2615. https://www.nature.com/articles/npp2011172
- Baldwin, Kylee & Moua, Selena & Perryman, Teylor & Hayden, Alyssa (2020). Let's get personal: The influence of personality type assessments on team communication and structure. Concordia Journal of Communication Research. https://digitalcommons.csp.edu/comjournal/vol7/iss1/1/
- Barnhart, Huiman X. & Haber, Michael J. & Lin, Lawrence I. (2007). An Overview on Assessing Agreement with Continuous Measurements. Journal of Biopharmaceutical Statistics, (17:4), 529-569. https://doi.org/10.1080/10543400701376480
- Beck, Roger. (2007). A Brief history of Ancient Astrology. Wiley-Blackwell, 1st ed,. ISBN: 9781405110877
- Boyle, Gregory J. (1995). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. Australian Psychologist, 30(1).. https://doi.org/10.1111/j.1742-9544.1995.tb01750.x
- Brady, Patrick & Bowd, Alan. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among pre-service education students. Teachers and Teaching; Theory and practice, 11(1), 37-46.. https://doi.org/10.1080/1354060042000337084

- Britannica. (2022). Education in the earliest civilizations. Britannica.com (Retrieved 11.04.2022), Ch.: "The old world civilizations of Egypt, Mesopotamia, and China". https://www.britannica.com/topic/education/Education-in-the-earliest-civilizations
- Brovold, H. (2014). Invarians drøftet i et nevropsykologisk perspektiv med spesiell referanse til realfag kognisjon. "Fire veier inn i matematikken". Thesis, Norwegian University of Science and Technology. http://hdl.handle.net/11250/271186
- Bruni, Renato & Catalano, Giuseppe & Daraio, Cinzia & Gregori, Martina & Moed, Henk F. (2020). Studying the geterogenity of European higher education institutions. Scientometrics, 125, 1117-1144.. https://link.springer.com/article/10.1007/s11192-020-03717-w
- Carlson, John G. (1985). Recent assessments of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 49(4), 356-365.. https://doi.org/10.1207/s15327752jpa4904_3
- Christoffersen, Line & Johannessen, Asbjørn (2018). Forskningsmetode for lærerutdanningene. Abstrakt forlag, Oslo. (1st issue, 2nd ed). ISBN: 978-82-7935-328-7.
- Cohen, Yoav. (2017). Estimating the Intra-Rater reliability of Essay Raters. Frontiers in Education, 2(49). https://doi.org/10.3389/feduc.2017.00049
- Corr, Philip J. & Matthews, Gerald (edited by). (2009). The Cambridge Handbook of Personality Psychology. Cambridge University Press. ISBN: 978-0-511-59654-4
- Costa, P. (2008). The revised NEO personality inventory (NEO-PI-R). Research Gate. https://www.researchgate.net/publication/285086638_The_revised_NEO_personality_inventory_NEO-PI-R
- Dahle, Dag Y. (2009). Guyyer sliter mest med matte. Teknisk Ukeblad (retrieved 15.5.2022). https://www.tu.no/artikler/gutter-sliter-mest-med-matte/242440
- DAT200. (2022). DAT200 Anvendt Maskinlæring. NMBU.no (retrieved 26.04.2022).. https://www.nmbu.no/emne/dat200
- Datatilsynet (2022). Databehandleravtale. datatilsynet.no (Retrieved 29.03.2022).. https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/databehandleravtale/
- Dean, A. & Voss, D. & Draquljic, D. (2017). Design and Analysis of Experiments. Springer International Publishing AG 2017, 2nd ed, ISBN: 978-3319522500.
- Dean, Michelle A. & Conte, Jeffrey M. & Blakenhorn, Tom R. (2006). Examination of the predictive validity of Big Five personality dimensions across training performance criteria. Personality and Individual Differences, 41(7), 1229-1239.. https://doi.org/10.1016/j.paid.2006.04.020
- Degnan, J. (2017). Measures of Agreement. University of New Mexico. https://math.unm.edu/~james/week14-kappa.pdf
- DeVellis, Robert F. (2005). Inter-rater reliability. Encyclopedia of Social Measurement, 317-322. . https://doi.org/10.1016/B0-12-369398-5/00095-5
- Dønnestad, Ingrid K. (2019). Kjønnsforskjeller i matematikk. Universitetet i Agder.. https://uia.brage.unit.no/uia-xmlui/bitstream/handle/11250/2623671/D%C3%B8nnestad%2C%20Ingrid%20Karoline.pdf?sequence=1
- Edsys (2019). 20 Best Education System in the World. Edsys Pvt. Ltd.. https://www.edsys.in/best-education-system-in-the-world/
- Eisele, Gudrun & Vachon, Hugo & Latif, Ginette & Kuppens, Peter & Houben, Marlies & Myin-Germeys, Inez & Viechtbauer, Wolfgang. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. SAGE journals, 29(2), 136-151.. https://doi.org/10.1177/1073191120957102
- Enachea, Rodia Gabriela & Matei, Raluca Silvia (2017). Study on Self-Awareness and Vocational Counseling of High School students. New trends and issues proceedings on Humanities and Social Sciences, (3:3), 372-378. https://doi.org/10.18844/gjhss.v3i3

- Fairhurst, A.M. & Fairhurst, L.L. (1995). Effective teaching, effective learning; Making the personality connection in your classroom. Nicholas Brealey, ISBN: 978-0891060789.
- Felder, R.M. & Felder, G.N. & Dietz, E.J. (2013). The effects of personality type on engineering student performance and attitudes. Journal of engineering education, 91, 3-17. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2002.tb00667.x
- Felder, R.M. & Felder, G.N. (2013). The effects of personality type on engineering student performance and attitudes Journal of Engineering Education, 91(1), 3-17.. https://doi.org/10.1002/j.2168-9830.2002.tb00667.x
- Felder, R.M. & Silverman, L.K. (1988). Learning and teaching styles in engineering education. Engineering education, 78(7), 674-681. https://www.researchgate.net/publication/257431200_Learning_and_Teaching_Styles_in_Engineering_Education
- Fellmann, Megan (2018). Scientists determine four personality types based on new data. Northwestern University. https://news.northwestern.edu/stories/2018/september/are-you-average-reserved-self-centered-or-a-role-model/
- Fleeson, William & Gallagher, M. Patrick. (2009). The Implications of Big-Five Standing for the Distribution of Trait manifestation in Behavior: Fifteen Experience-sampling studies and a meta-analysis. J Pers Soc Psychol., 97(6), 1097-1114.. https://doi.org/10.1037/a0016786
- Forbes (2022). What are durable skills and why is there a shortage?. Forbes.com (Retrieved 11.03.22, 4:37 pm). https://www.forbes.com/sites/forbeshumanresourcescouncil/2022/03/11/what-are-durable-skills-and-why-is-there-a-shortage/?sh=4aaa794d56d4
- Fordham, Frieda (2021). Carl Jung . Britannica.com (Retrieved 12.03.22, 7:55 pm). https://www.britannica.com/biography/Carl-Jung
- Furnham, Adrian (1996). The big five versus the Big Four; The relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. Personality and Individual Differences, 21(2), 303-307. https://doi.org/10.1016/0191-8869(96)00033-5
- Gavrilescu, Mihai & Vizireanu, Nicolae. (2018). Predicting the Big-Five personality traits from handwriting. EURASIP Journal on Image and Video Processing, 57.. https://jivp-eurasipjournals.springeropen.com/articles/10.1186/s13640-018-0297-3
- Gerlach, Martin & Farb, Beatrice & Revelle, William & Amaral, Luís A. Nunes (2018). A robust data-driven approach identifies four personality types across four large data sets. Nature Human Behavior, (2), 735-742. https://doi.org/10.1038/s41562-018-0419-z
- Geyer, Peter (1995). Quantifying Jung; The origin and development of the Myers-Briggs Type indicator. ResearchGate. https://doi.org/10.13140/2.1.2591.4240
- Gholipour, Bahar (2019). How accurate is the Myers-Briggs Personality Test?. Livescience. https://www.livescience.com/65513-does-myers-briggs-personality-test-work.html
- Gjefle, Mina T. (2021). Student's learning preferences and perception of mathematics; How different cognitive types view contrasting pproaches to statistics education. NMBU, (semesteroppgave i regi ECOL300 "Naturvitenskapelig metode"), fakultet for Miljøvitenskap og Naturforvaltning.
- Gliem, Joseph A. & Gliem, Rosemary R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Midwest Research-to-Practice Conference, Ohio State University. Adult, Continuing and Community Education.. https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem%20&%20Gliem.pdf?s..
- Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. Personality psychology in Europe, 7-28.
- Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. Personality Psychology in Europe (vol7, 7-28). Tilburg, The Netherlands: Tilburg University Press.. https://ipip.ori.org/A%20broad-bandwidth%20inventory.pdf

- Goldberg, L.R. (2008). The Eugene-Springfield community sample: Information available from the research participants. ORI Technical Report, 48(1). (Retrieved 29.03.2022).. https://ipip.ori.org/ORI_TechnicalReport_ESCS_Mar08.pdf
- Haanæs, Øystein R. (2021). 3D-teknologi kan gi mer effektiv læring. Forskning.no (retrieved 08.04.2022).. https://forskning.no/de-regionale-forskningsfondene-informasjonsteknologi-partner/3d-teknologi-kan-gi-mer-effektiv-laering/1846865
- Hadfield, Oakley D. & McNeil, Keith. (1994). The relationship between Myers-Briggs personality type and Mathematics anxiety among preservice elementary teachers. Journal of Instructional Psychology, 21(4).. https://www.proquest.com/openview/bd6623148f39be306d327b792502794d/1?pq-origsite=gscholar&cbl=2029838
- Hamm, S.C. (2018). The Myers-Briggs Type Indicator and a Student's College Major. The University of Texas, San Antonio. https://hdl.handle.net/20.500.12588/65
- Hammond, Zaretta. (2014). Culturally responsive teaching & The brain; Promoting authentic engagement and rigor among culturally and linguistically diverse students. Excerpt from Facinghistory.org (retrieved 28.04.2022).. tinyurl.com/FacingHistoryDependentLearners
- Harrington, Rick & Loffredo, Donald A. (2009). MBTI personality type and other factors that relate to preference for online versus face-to-face instruction. The Internet and Higher Education, 13(1-2), 89-95.. https://doi.org/10.1016/j.iheduc.2009.11.006
- Hartmann, K. & Krois, J. & Waske, B. (2018). E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Ch.: "Choose principal components". Departmenf of Earth Sciences, Freie Universitaet Berlin.. https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/Principal-Component-Analysis/principal-components-basics/Choose-principal-components/index.html
- Hinkle, K.S. (1987). An investigation of the relationships among learning style preferences, personality types, and mathematics anxiety of college students. The University of Maryland College Park.. https://www.elibrary.ru/item.asp?id=7491406
- Hogg, Robert V. & McKean, Joseph W. & Craig, Allen T. (2020). Introduction to Mathematical Statistics. Pearson Education Limited, Essex, 8th ed.. ISBN: 978-1-292-26476-9.
- Howley, Tom & Madden, Michael G. & O'Connell, Marie-Louise & Ryder, Alan G. (2005). The effect of Principal component analysis on Machine learning Accuracy with High dimensional spectral data. Applications and Innovations in Intelligent Systems XIII, Proceedings of AI-2005, the Twelth-fifth SGAI International Conference on Innovative Techniques and Applications of Artificia Intelligence [conference], Cambridge, UK.. https://doi.org/10.1007/1-84628-224-1_16
- Hull, Lynnette R. (2007). Using personality preferences to predict persisters and dropouts among residential students at a two-year technical college. Capella University ProQuest Dissertations Publishing.. https://www.proquest.com/openview/3c3739775b6d5c6ad9cb0869b18162ac/1?pq-origsite=gscholar&cbl=18750
- IGI Global. (2022). What is Massification of Higher Education. IGI Global; Publisher of Timely Knowledge. (Retrieved 25.04.2022).. https://www.igi-global.com/dictionary/massification-of-higher-education/69992#:~:text=1.,for%20these%20numbers%20of%20students.
- IMRT100. (2022). IMRT100 Innføringsemne - fagorientert prosekt. NMBU.no (retrieved 26.04.2022).. https://www.nmbu.no/emne/imrt100
- INF200. (2022). INF200 Videregående programmering. NMBU.no (retrieved 26.04.2022).. https://www.nmbu.no/emne/INF200
- IPIP 1. . https://ipip.ori.org/InterpretingIndividualIPIPScaleScores.htm
- James, Gareth & Witten, Daniela & Hastie, Trevor & Tibshirani, Robert. (2017). An introduction to statistical learning: With applications in R. Springer-Verlag New York Inc., 2nd ed.. ISBN: 9781071614174

- Jang, Hyun-Jung. (2018). MBTI Personality Types of the University Students in an Area. he Journal of the Korea Contents Association, 18(3), 486-498.. https://doi.org/10.5392/JKCA.2018.18.03.486
- Johnson, J.A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. Journal of Research in Personality, 51, 78-89. https://doi.org/10.1016/j.jrp.2014.05.003
- Jung, Carl (1910). The association method. American Journal of Psychology, 31, 219-269. https://psychclassics.yorku.ca/Jung/Association/lecture1.htm
- Jung, Carl (1921). Psychological types (Collected Works of C.G. Jung). H.G. Baynes, Trans. & Sir H., Read, Ed., Volume 14. Routledge.. ISBN: 978-0-415-04559-9
- K. Katahira, Yoshihiko Kunisato & Shinsuke, Suzuki (2020). Commentary: A robust data-driven approach identifies four personality types across four large data sets. Frontiers in Big Data. https://doi.org/10.3389/fdata.2020.00008
- Kainz, Richard I. (1985). Myers-Briggs Type Indicator Preference scores and age. The University of Florida.. https://ufdcimages.uflib.ufl.edu/AA/00/05/29/86/00001/AA00052986_00001.pdf
- Kappe, Rutger & Flier, Henk van der. (2010). Using multiple and specific criteria to assess the predictive validity of the Big Five personality factors on academic performance. Journal of Research in Personality, 44(1), 142-145.. https://doi.org/10.1016/j.jrp.2009.11.002
- Kato, Takumi & Miura, Taro. (2021). The impact of questionnaire length on the accuracy rate of online surveys. Journal of Marketing Analytics, 9, 83-98.. https://link.springer.com/article/10.1057/s41270-021-00105-y
- Kennair, Leif E.O. (2018). Personlighet. Store Norske Leksikon (Retrieved 11.04.2022). . https://snl.no/personlighet#:~:text=Personlighet%20er%20definert%20som%20de,rekke%20teorier%20innen%20fagomr%C3%A5det%20personlighetspsykologi.
- KJB310 (2021). Proteinkjemi. NMBU.no (Retrieved 12.03.22, 12 am). https://www.nmbu.no/emne/KJB310?studieaar=2021
- KJB310. (2022). KJB310 Proteinkjemi (emnebeskrivelse). NMBU.no (retrieved (08.04.2022).. https://www.nmbu.no/emne/KJB310
- Lahey, B.B. (2009). Public health significance of neuroticism. American Psychologist, 64(4), 241-256. (Retrieved 29.03.2022).. https://doi.org/10.1037/a0015309
- Lay, David C. & Lay, Steven R. & McDonald, Judi J. (2016). Linear Algebra and its Applications. Pearson Education Limited, Essex, 5th ed.. ISBN: 978-1-292-09223-2.
- Lazarides, Rebecca & Rubach, Charlott & Ittel, Angela. (2016). Motivational profiles in mathematics: What role do gender, age and parents' valuing of mathematics play? International Journal of Gender Science and Technology, 8(1).. http://genderandset.open.ac.uk/index.php/genderandset/article/view/406
- Lim, Annabelle Q.Y. (2020). The Big Five Personality Traits. (Retrieved April 24, 2021).. https://www.simplypsychology.org/big-five-personality.html
- Lindstrom, S. (2021). What does different personality test and behavioral test measure? And do we overuse them?. Journal of Psychology and Neuroscience, 3(2), 1-3.. https://unisciencepub.com/storage/2021/06/What-Does-Different-Personality-Test-and-Behavioral-Test-Measure-And-Do-We-Overuse-Them.pdf
- Løvås, Gunnar G. (2018). Statistikk for universiteter og høgskoler. Universitetsforslaget, ISBN: 9788215031040..
- Lydersen, Stian. (2016). Measuring agreement between raters. Presentation at RBUP, NTNU.. https://folk.ntnu.no/slyderse/medstat/Interrater_fullpage_9March2016.pdf
- MacLeoud, Saul A. (2021). Id, Ego, and Superego. Simply psychology. https://www.simplypsychology.org/psyche.html
- Mancini, Teresa M. (1993). The relationship between Mathematics anxiety and personality type. The University of Connecticut.. https://www.elibrary.ru/item.asp?id=5788424

- Mayer, D.G. & Butler, D.P. (1993). Statistical Validation. Ecological Modelling, 68, 21-32. https://www.sciencedirect.com/science/article/pii/0304380093901052
- McAdams, D.P. (1997). Handbook of Personality Psychology. . https://doi.org/10.1016/B978-012134645-4/50002-0
- McCaully, M. & Godleski, E. & Yokomoto, C. & Harrisberger, L. & Sloan, E. (1983). Applications of psycholocial type in engineering education. Engineering education, 73(5), 394-400.. tinyurl.com/GoogleBooks1983
- McCrae, R.R. & Costa, P.T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the Perspective of the Five-Factor Model of Personality. Journal of Personality, (57:1). tinyurl.com/ResearchGateMBTIBigFive
- McCrae, R.R. & Costa, P.T. (1989). The Structure of Interpersonal Traits: Wiggin's Circumplex and The Five-Factor Model. Journal of Personality and Social Psychology, 56(4), 586-595. https://doi.org/10.1037/0022-3514.56.4.586
- McCrae, Robert & Costa, Paul (1989). Reinterpreting the Myers-Briggs Type Indicator from the Perspective of the Five-Factor Model of Personality. Journal of Personality, 57(1), 17-40.. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x
- McCrae, Robert & Costa, Paul (1992). Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional Manual. Odessa, FL: Psychological Assessment Resources..
- McCrae, Robert & Terracciano, Antonio & Sanchéz-Bernardos, M.L. & Djuric Jocic, Dragana & Halim, Magdalena S. (2005). Universal Features of Personality Traits from the Observer's Perspective; Data from 50 cultures. Journal of Personality and Social Psychology, 88(3), 547-561. https://doi.org/10.1037/0022-3514.88.3.547
- McEwen, Bruce S. (2020). Hormones and behavior and the integration of brain-body science. Hormones and Behavior, 119, 104619.. https://doi.org/10.1016/j.yhbeh.2019.104619
- McHugh, M.L. (2012). Interrater Reliability; The Kappa Statistic. Biochemica Medica, 22(3), 276-282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/
- McLeod, Saul A. (2007). What is reliability?. Simply Psychology.. https://www.simplypsychology.org/reliability.html
- MedicineNet (2019). Nature vs. Nurture; Is it in our genes or Our Environment?. Retrieved April 16, 2021. https://www.medicinenet.com/nature_vs_nurture_theory_genes_or_environment/article.htm
- Melvin, Jenna. (2013). Personality type as an indicator of learning style. University of Rochester, Center fr Excellence in teachng and learning.. http://hdl.handle.net/1802/28102
- Menard, Scott. (2001). Applied Logistic Regression Analysis. SAGE publications, 2nd ed.. ISBN: 978-0761922087
- Mendenhall, William & Sincich, Terry. (2014). A Second Course in Statistics: Regression Analysis. Pearson Education Limited, Essex, 7th ed. . ISBN: 978-1-292-04290-9.
- Mike, Anissa & Jackson, Joshua J. & Oltmanns, Thomas F. (2014). The conscious retiree: The relationship between conscienciousness, retirement, and volunteering. Journal of Research in Personality, 52, 68-77.. https://doi.org/10.1016/j.jrp.2014.07.002
- Mitchell, Rachel L.C. (2016). Hans Eysenck's interface between the brain and personality: Modern evidence on the cognitive neuroscience of personality. Personality and Individual Differences, 103, 74-81. (Retrieved 29.03.2022). https://doi.org/10.1016/j.paid.2016.04.009
- Murray, John B. (1990). Review of research on the Myers-Briggs Type Indicator. Perceptual and Motor Skills, 70, 1187-1202. . https://doi.org/10.2466/pms.1990.70.3c.1187
- Myers & Briggs foundation (2022). How frequent is my type. myersbriggs.org. https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm

- Myers & Briggs Foundation (2022). MBTI® basics. myersbriggs.org (Retrieved 12.03.22, 8:20 pm). https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/
- Myers & Briggs Foundation 1 (2022). Type Dynamics. The Myers and Briggs foundation (retrieved 04.02.2022).. https://www.myersbriggs.org/my-mbti-personality-type/understanding-mbti-type-dynamics/type-dynamics.htm
- Myers & Briggs Foundation 2 (2022). The Eight function attitudes. The Myers and Briggs foundation (retrieved 04.02.2022).. https://www.myersbriggs.org/my-mbti-personality-type/understanding-mbti-type-dynamics/the-eight-function-attitudes.htm
- Myers Briggs Company (2022). The history of the MBTI assessment. themyersbriggs.com (Retrieved 12.03.22, 7.31 pm). https://eu.themyersbriggs.com/en/tools/MBTI/Myers-Briggs-history
- Myers, I.B. & McCaulley, M.H. & Quenk, N.L. & Hammer, A.L. (1998). MBTI Manual. A Guide to the Development and Use of the Myers-Briggs Type Indicator. Palo Alto, California. Consulting Psychologists Press. https://archive.org/details/mbtimanualguidet00myer
- Myers-Briggs Company (2021). The history of the MBTI Assessment. . https://eu.themyersbriggs.com/en/tools/MBTI/Myers-Briggs-history
- Myers-Briggs foundation (2001). How Frequent is my type. . https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm
- NIH (2012). The World of Shakespeare's Humors. (Retrieved April 16, 2021).. https://www.nlm.nih.gov/exhibition/shakespeare-and-the-four-humors/index.html#section2
- Norgeskart (2022). Norgeskart. Kartverket.. https://www.norgeskart.no/#!?project=norgeskart&layers=1002&zoom=3&lat=7197864.00&lon=396722.00
- Norman, Warren T. (1967). 2800 Personality Trait Descriptors: Normative Operating Characteristics for a University Population. Michigan University, Ann Arbor College of Lit., Sci., Arts.. https://files.eric.ed.gov/fulltext/ED014738.pdf
- Onwuegbuzie, Anthony J. (2010). Academic procrastination and statistics anxiety. Assessment & Evaluation in higher education, 29(1), 3-19.. https://doi.org/10.1080/0260293042000160384
- Otero, Inmaculada & Cuadrado, Dámaris & Martínez, Alexandra. (2020). Convergent and Predictive validity of the Big-Five factors assessed with Single-stimulus and Quasi-Ipsative Questionnaires. Journal of Work and Organizational Psychology, 36(3), 215-222.. https://doi.org/10.5093/jwop2020a17
- Owens, Molly & Carson, Andrew D. (2021). Research summary for the Typefinder Personality Assessment. Truity Psychometrics LLC. tinyurl.com/TypefinderResearchSummary
- Owens, Molly (2021). The Typefinder Personality Assessment. Truity Psychometrics LLC. https://www.truity.com/sites/default/files/fillpdf/typefinder-technicaldoc.pdf
- Patten, B.C. (1977). Systems Analysis and Simulation in Ecology. Academic Press Inc, ISBN: 978-0125472043.
- Pervin, Lawrence A. (1994). A Critical Analysis of Current Trait Theory. Psychological Inquiry, 5(2), 103-113. Taylor & Francis, Ltd.. https://doi.org/10.1207/s15327965pli0502_1
- Pervin, Lawrence A. (ed.) & John, Oliver P. (ed.) (1999). Handbook of Personality: Theory and Research. The Guildford Press, New York/London. 2nd ed, edited by Lawrence A. Pervin & Oliver P. John.. ISBN: 1-57230-483-9.
- Pittenger, David. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. Consulting Psychology Journal Practice and Research, 57(3), 210-221.. https://doi.org/10.1037/1065-9293.57.3.210
- Potters, Charles. (2021). Variance Inflation Factor (VIF). Investopedia.com (retrieved 07.04.2022).. tinyurl.com/InvestopediaVIF

- Psychologia (2022). Four temperaments: Sanguine, Phlegmatic, Choleric, and Melancholic Personality types. Psychologia.co. https://psychologia.co/four-temperaments/
- Puji, Rully P.N. & Ahmad, Abdul R. (2016). Learning style of MBTI personality types in history learning at higher education. Scientific Journal of PPI-UKM, 3(6), 289-295.. https://doi.org/10.27512/sjppi-ukm/ses/a13122016
- Pykes, Kurtis. (2020). Cohen's Kappa; Understanding Cohen's Kappa coefficient. Towardsdatascience.com (retrieved 28.04.2022).. https://towardsdatascience.com/cohens-kappa-9786ceceab58#:~:text=Lastly%2C%20the%20formula%20for%20Cohen%27s,the%20probability%20of%20random%20agreement.
- Queen's University. (2022). Reproducibility; Kappa Values. Queen's University School of Medicine, (retrieved 12.05.2022). tinyurl.com/QueensUniversityKappa
- Raschka, Sebastian & Mirjalili, Vahid. (2017). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt Publishing Limited, 2nd ed. . ISBN: 9781787125933.
- Regjeringen (2005). Kunnskapsløftet - reformen i grunnskole og videregående opplæring. . https://www.regjeringen.no/globalassets/upload/kilde/ufd/prm/2005/0081/ddd/pdfv/256458-kunnskap_bokmaal_low.pdf
- Rosati, P. (1998). Academic progress of canadian engineering students in terms of MBTI personality type. International journal of engineering education, 14, 322-327. https://www.ijee.ie/articles/Vol14-5/ijee1037.pdf
- Rostad, Jens Kristian - NMBU. (2022). Utforming av kontrakter og avtaler. nmbu.no (Retrieved 29.03.2022). https://www.nmbu.no/forskning/forskere/juridiskbistand/kontrakter/node/7759
- Sæbø, Solve & Almøy, Trygve & Brovold, Helge (2015). Does academia disfavor contextual and extraverted students?. Uniped, (4:38), 274-283. https://www.idunn.no/doi/10.18261/ISSN1893-8981-2015-04-03
- Sarmento, David. (2022). Chapter 22: Correlation types and when to use them. Ademos (retrieved 10.04.2022).. https://ademos.people.uic.edu/Chapter22.html#2_a_brief_overview_of_correlations
- Schlager, Angela & Ahlqvist, Kerstin &Rasmussen-Barr, Eva & Bjelland, Elisabeth K. & Pingel, Ronnie & Olsson, Christina & Nilsson-Wikmar, Lena & Kristiansson, Per. (2018). Inter- and Intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods. BMC Musculoskeletal Disorders, 19, 376.. https://doi.org/10.1186/s12891-018-2290-5
- Schmidt, M.E. & Steindorf, K. (2006). Statistical Methods for the Validation of Questionnaires; Disrepancy between Theory and Practice. Methods of Information in Medicine, (45:04), 409-413. https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0038-1634096
- Schmidt, M.E. & Steindorf, K. (2006). Statistical methods for the Validation of questionnaires; Disrepancy between Theory and Practice. Methods Inf Med, 45(4), 409-413.. https://pubmed.ncbi.nlm.nih.gov/16964357/
- Sevik, Kristine. (2016). Programmering i skolen: Notat fra Senter for IKT i Utdanningen. Senter for IKT i Utdanningen, UDIR.. https://www.udir.no/globalassets/filer/programmering_i_skolen.pdf
- Shih, Jerry J. & Krusienski, Dean J. & Wolpaw, Jonathan R. (2012). Brain-Computer Interfaces in Medicine. Mayo Clinic Proceedings, 87(3), 268-279.. https://doi.org/10.1016/j.mayocp.2011.12.008
- Shipley, Beverly & Weiss, Alexander & Der, Geoff & Taylor, Michelle & Deary, Ian J. (2007). Neuroticism, Extraversion, and Mortality in the UK Health and Lifestyle Survey: A 21-Year

Prospective Cohort Study. Psychosomatic Medicine, 69(9), 923-931. (Retrieved 29.03.2022).. https://doi.org/10.1097/PSY.0b013e31815abf83

- Silverman, Bernie I. (1971). Studies of Astrology. The Journal of Psychology, 77(2), 141-149. . https://doi.org/10.1080/00223980.1971.9916861
- Smail, Linda. (2016). Using Bayesian networks to understand relationships among math anxiety, genders, personality types, and study habits at a university in Jordan. Journal on Mathematics Education, 8(1), 17-34.. https://doi.org/10.22342/jme.8.1.3405.17-34
- Statistics.com (2021). Test-Retest Reliability. . https://www.statistics.com/glossary/test-retest-reliability/
- STIN100. (2022). STIN100 Biologisk Data-Analyse. NMBU.no (retrieved 08.04.2022). https://www.nmbu.no/emne/STIN100
- Sutter, Jon M. & Kalivas, John H. & Lang, Patrick M. (1992). Which principal components to utilize for principal component regression. Journal of Chemometrics, 6(4), 217-225.. https://doi.org/10.1002/cem.1180060406
- Tabery, James (2014). Nature vs. Nurture. (Retrieved April 16, 2021).. https://eugenicsarchive.ca/discover/tree/535eed0d7095aa0000000241
- Tapia, Martha & Marsh, George E. (2004). The relationship of math anxiety and gender. American Exchange Quarterly, 8(2), 130-134.. http://www.rapidintellect.com/AEQweb/5may2690l4.htm
- Teigmo, Anne-Mai. (2019). Hvorfor velger mange jenter bort matematikk? Forskning.no (retrieved 15.05.2022).. https://forskning.no/kjonn-og-samfunn-matematikk-oslomet/hvorfor-velger-mange-jenter-bort-matematikk/1606300
- Thomsen, Jens P. (2012). Exploring the heterogenity of class in higher education: social and cultural differentiation in Danish university programmes. British Journal of Sociology of Education, 33(4), 565-585.. https://doi.org/10.1080/01425692.2012.659458
- tinyurl.com/GoogleBooks1995"
- UDIR (MAT1-04). Læreplan i matematikk fellesfag (MAT1-04). UDIR. https://www.udir.no/kl06/MAT1-04/Hele/Kompetansemaal/kompetansemal-etter-2.-arssteget
- UDIR (MAT1-05). Kompetansemål og vurdering. UDIR. https://www.udir.no/lk20/mat01-05/kompetansemaal-og-vurdering/kv20
- Uhl, Norman P. (1981). Personality type and Congruence with Environmet: Their relationship to College Attrition and Changing of major. ERIC Journal, ED205130.. https://eric.ed.gov/?id=ED205130
- Ursachi, George & Horodnic, Ioana A. & Zait, Adriana. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. Procedia Economics and Finance, 20, 679-686.. https://www.sciencedirect.com/science/article/pii/S2212567115001239?ref=pdf_download&fr=RR-2&rr=70b89ceb08ea1c02
- Uslainer, Barry M. (1990). The combined use of the Myers-Briggs Type Indicator and Demos Dropout Scale as predictors of risk for dropout status. Pace University ProQuest Dissertations Publishing.. https://www.proquest.com/openview/0d3ac317e3b2f0cc12204de7fbcfce24/1?pq-origsite=gscholar&cbl=18750&diss=y
- Vinje, H. & Brovold, H. & Almøy, T. & Frøslie, K.F. & Sæbø, S. (2021). Adapting statistics education to a cognitively heterogeneous student population. Journal of statistics and data science education, 29(2), 183-191.. https://doi.org/10.1080/26939169.2021.1928573
- Vittinghoff, E & Glidden, D.V. & Shiboski, S.C. & McCulloch, C.E. (2011). Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Springer-Verlag New York Inc., 2nd ed.. ISBN: 9781461413523

- Wankat, P.C. & Oreovicz, F.S. (1992). Teaching engineering. McGraw-Hill College, ISBN: 978-0070681545, 244-281. https://engineering.purdue.edu/ChE/aboutus/publications/teaching_eng
- WatchWorldTest (2021). Jung's Theory of Psychological Types. (Retrieved April 19, 2021).. http://watchwordtest.com/types.aspx
- Wiggins, J.S. & Pincus, A.L. (1989). Conceptions of personality disorders and dimensions of personality. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1(4), 305-316.. https://doi.org/10.1037/1040-3590.1.4.305
- Wongpakaran, Nahathai & Wongpakaran, Tinakon & Wedding, Danny & Gwet, Kilem L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Medical Research Methodoloy, 13(61).. https://doi.org/10.1186/1471-2288-13-61
- Woodard, Teresa S.H. (2002). The effects of math anxiety on post-secondary developmental students as related to achievement, gender, and age. Argosy University.. https://www.proquest.com/openview/9b79491c5effc7742c56f9352a445537/1?pq-origsite=gscholar&cbl=18750&diss=y
- Xia, Yinglin. (2020). Chapter Eleven - Correlation and association analyses in microbiome study integrating multiomics in health and disease. Progress in Molecular biology and Translational Science, 171, 309-491.. https://doi.org/10.1016/bs.pmbts.2020.04.003

6 – Appendix

6.1 – Appendix A – General

.

Table A1: *Courses at NMBU specifically targeted for student recruitment. Lists course codes, name, examples of which studies has the course as part of their mandatory curricula, and which faculties the mentioned study-programs belong to.*

Course code	Course name	Example of studies (NMBU) with this course	Faculty of studies
STAT100	Intro. statistics	**Most studies**	
MATH100	Intro. mathematics	**Most studies**	
MATH111*	Calculus 1	Civil Engineering**	REALTEK, KBM, MINA, Biovit
STIN100	Biological data analysis	Biotechnology, Food Sciences, Animal Science, Biology	KBM, Biovit
ECN110	Intro. to Macroeconomics	Economy and Business, Social economics, Renewable Energy, Ecology and Nature Management, Forestry	HH, MINA, REALTEK
EDS115	Intro. to Research Methods	International Environment and Development Studies	Landsam
LAD103	Cartography and Geodatabases in Planning	Real Estate Jurisprudence, Urban- and Regional Planning	Landsam
STAT351	Applied Statistics	Bioinformatics and Applied Statistics	KBM, REALTEK

* Often offered as a substitute for MATH100 in non-engineering studies

** e.g.: Chemistry and Biotechnology, Industrial Economics, Applied robotics

6.1.2 – Project announcement (sample 1)

Tittel:

Invitasjon til å delta i forskning på undervisning

Kunngjøring:

Kunne **du** tenke deg i å delta med informasjon som kan bidra til å forbedre undervisningen?

Det er en godt etablert teori at ulike personer tenker og lærer ulikt, og at personlighetstyper eller «kognitive profiler» kan brukes til å studere dette. Ved NMBU har vi lenge forsket på hvordan dette påvirker eksamensresultater i statistikkurset STAT100. Målet er å tilby ulike undervisningsformer og læremidler, så studenter med ulike læringsstrategier kan finne noe som passer for seg.

For å samle informasjon om studenters personlighetstype og læringsstrategier har vi brukt et spørreskjema kalt *Utdanningstesten*. **Nå trenger vi deres hjelp til å analysere hvor god denne testen er.**

I forbindelse med masterprosjektet til student Mina Therese Gjefle ved KBM, søker vi derfor etter studenter – uavhengig alder og studieretning – som kan delta i studien ved å ta minst to personlighetstype-undersøkelser.

Deltakelse vil i praksis være anonymt, og innebærer at du svarer på 2 eller flere personlighets-undersøkelser over nett. Du rapporterer deretter personlighetstypene du fikk via et nettskjema.

For å lese mer om prosjektet og hvordan man kan delta, gå inn på denne linken:

- [PLACEHOLDER FOR ARKEN-LINK]

Vi er avhengige av at så mange som mulig velger å delta i denne studien. Din deltakelse er uvurderlig, og vil hjelpe ikke bare masterstudenten Mina, men også NMBU som helhet!

Spre gjerne budskapet (linken ovenfor) videre til andre studenter, også utenfor NMBU!

Med vennlig hilsen,

Mina Therese Gjefle (M-BIAS) og Kathrine Frey Frøslie (prosjektansvarlig)

Tidligere forskning ved NMBU (interessant lesning for de ekstra interesserte):

Fra 2021: Ulike typer studenters resultater i statistikkurs etter innføring av «flipped classroom»

- Full article: Adapting Statistics Education to a Cognitively Heterogeneous Student Population (tandfonline.com)

Fra 2015: Gjør ekstroverte og kontekstuelle personer det dårligere på universitet og høgskolene?

- Does academia disfavor contextual and extraverted students? - Nr 04 - 2015 - Uniped - Idunn

6.1.3 – Website for recruitment (sample 1)

Below are screenshots from the website made by Mina Therese Gjefle, in collaboration with the supervisors, used to easily recruit students to part 1 of this thesis. Screenshots were taken January 7th, 2022, at 4:15 pm. NB, the website is maintained by Mina and was created using www.Simplesite.com, but will go offline after the master's completion.

Address: www.personlighet-forskning-nmbu.com

6.1.3.1 – Home Page

A simple green background with the text "Vil du delta I forskningsprosjektet: 'Sammenligning av ulike spørreskjemaer som gir informasjon om læringsstrategier og personlighetstyper'".

6.1.3.2 – Information page

Forskning på læringsstrategier og personlighetstyper

Hjem Informasjon Personlighetstester Kontakt Oss

Norges miljø- og biovitenskapelige universitet

Vil du delta i forskningsprosjektet "Sammenligning av ulike spørreskjemaer som gir informasjon om læringsstrategier og personlighetstyper"

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor det overordnede formålet er å forbedre statistikkundervisningen. Da er det viktig å ha en kvalitetssikring av tallene vi samler inn. I dette skrivet gir vi deg informasjon om prosjektet og hva deltakelse vil innebære for deg.

Formål

I flere år har vi sett på hvorfor det innledende ststiastikk-kurset STAT100 ved NMBU kan virke utfordrende for mange. Vi ønsker å tilrettelegge undervisningen for en mer variert studentmasse som vi møter her på NMBU. Prosjektet startet i 2014, og til dette har vi brukt et spørreskjema kalt Utdanningstesten, for å samle informasjon om studenters personlighetstype og læringsstrategier.

Utdanningstesten er utviklet for en yngre aldersgruppe og et annet refleksjonsnivå enn hva voksne studenter er på. Spørreskjemaet inneholder spørsmål om yrkesinteresser, hvordan du liker å lære på skolen, og hvilke naturvitenskapelige fag du liker best. Likevel tror vi resultatene fra Utdanningstesten gir en grei indikasjon på det vi er interessert i, og det er derfor vi har brukt den i tidligere studier. Dette er det vi skal undersøke i dette prosjektet.

En masterstudent problematiserte i 2020 hvorvidt Utdanningstesten gir den informasjonen vi ønsker. Vi har dermed behov for å analysere og validere selve testen, før vi kan gå i dybden på hva svarene forteller oss. Dette prosjektet er relevant for studenter på tvers av alle fag og disipliner. Vi er derfor ikke bare interesserte i studenter som har tatt eller tar STAT100. Deltakelse i prosjektet kan dermed bidra til å forbedre forskningen på undervisning ved NMBU, og på lang sikt forbedre selve undervisningen.

Hvem er ansvarlig for forskningsprosjektet?

NMBU v/ Kathrine Frey Frøslie, førsteamanuensis i statistikk, er ansvarlig for dette prosjektet. Prosjektet er en del av masteroppgaven til student Mina Therese Gjefle.

Hvorfor får du spørsmål om å delta?

Vi ønsker å samle så mange studenter som mulig til å delta på dette prosjektet, slik at vi kan få et stort nok utvalg til å trekke konklusjoner som er statistisk holdbare.

Hva innebærer det for deg å delta?

Alle som samtykker til å delta må ta Utdanningstesten fra Nasjonalt senter for realfagsrekruttering. Det vil ta deg ca. 10 minutter. Alle som samtykker til å delta må også ta minst én av de 4 følgende testene - GJERNE alle:

1) Utdanningstesten, én gang til (med noen dagers mellomrom). *Ta samme digitale Utdanningstesten et par dager etter at du tok den første gang, slik at vi kan se om svarene (personlighetstypen) du fikk hver gang samsvarer. Dette vil gi oss informasjon om testens reliabilitet.*

2) En lengre versjon av Utdanningstesten (Uroboros) - dette vil ta ca. 30-45 minutter. *Ta en lengre versjon av Utdanningstesten, for å se hvor like resultat du får. Dette vil gi oss informasjon om påliteligheten til den kortere testen.*

3 / 4) En Big-Four / Big-Five test - dette vil ta ca. 15 minutter. *Ta en validert Big-Four / Big Five test. Dette vil gi oss informasjon om hva Utdanningstesten forteller oss.*

Du må deretter rapportere dine resultater fra disse testene i dette digitale spørreskjemaet. Der spør vi også om annen informasjon, som kjønn og alder. Undersøkelsen er derfor ikke helt anonym, men gitt utvalgets størrelse så vil det i praksis ikke være mulig å identifisere enkeltpersoner i datasettet ut ifra denne begrensede personinformasjonen.

Alt i alt er det forventet 20-60 minutters bidrag ifra de som velger å delta.

NB! Lenke til testene & Nettskjemaet finner du under "Personlighetstester" på nettsiden

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du ikke trekke samtykket tilbake. Dette er fordi vi ikke kommer til å samle inn direkte identifiserende informasjon, og dataene vil dermed ikke kunne spores tilbake til deg.

Ditt personvern - hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

1 - Forelesere tilknyttet STAT100, og PHD-studenter og masterstudenter tilknyttet biostatistikkgruppen, vil ha tilgang til opplysningene.

2 - Spørreskjema som skal benyttes tilbys i form av utdanningstesten til Nasjonalt senter for realfagsrekruttering (www.utdanningstesten.no). Dataene blir av Nasjonalt senter for realfagsrekrutering automatisk slettet etter 6 måneder.

3 - Samtykke innhentes gjennom et elektronisk nettskjema. Deltakerne rapporterer selv inn resultatene de fikk fra testene gjennom det samme skjemaet. Informasjon som rapporteres fra testene er: Hvilke 4 bokstaver deltakeren fikk som svar (dersom deltaker tok Utdanningstesten / en Big-Four-test), eller hvilke tallverdier deltakeren fikk på de 4 kategoriene «Ekstroversjon», «Åpenhet», «Medmenneskelighet» og «Planmessighet» (dersom deltaker tok Big Five test).

4 - Resultatene som deltakerne melder inn via det elektroniske nettskjemaet vil bli lagret på NMBU sin server. Ingen informasjon som kan spores tilbake til deltakerne vil bli lagret. Prosjektansvarlig Kathrine Frey Frøslie og personer tilknyttet forskningsgruppen BIAS vil ha tilgang til disse anonyme dataene.

Deltakerne vil ikke kunne gjenkjennes gjennom publikasjoner eller datamaterialet senere.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes innen 31.desember 2024, men de fullt anonymiserte data vil forbli lagret på NMBU sine servere. Det anonyme datamaterialet blir oppbevart videre kun for eventuell etterprøvbarhet, oppfølgingsstudie og arkivering for senere forskning.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene
- å få rettet personopplysninger om deg
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninge

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke. På oppdrag fra Norges miljø- og biovitenskapelige universitet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

For kontaktinformasjon, gå til "Kontakt Oss" på denne nettsiden

Med vennlig hilsen,

Kathrine Frey Frøslie (Prosjektansvarlig)

Mina Therese Gjefle (Masterstudent og databehandlingsansvarlig 2021-2021)

6.1.3.3 – Personality test page

Forskning på læringsstrategier og personlighetstyper

Hjem   Informasjon   **Personlighetstester**   Kontakt Oss

Personlighetstester & Nettskjema

Velkommen til nettsiden for forskningsprosjektet *"Sammenligning av ulike spørreskjemaer som gir informasjon om læringsstrategier og personlighetstyper"*. Du kan lese mer om prosjektet under "Informasjon". På denne siden finner du oversikt over de ulike personlighetstestene, og link til nettskjema hvor du anonymt kan melde inn resultatene (personlighetsypene) dine

KORT OPPSUMMERT: Alle som velger å delta må ta minst 2 persontester: **Nr1 =** Utdanningstesten, **Nr2 =** Minst én (men gjerne flere) av de følgende 4: Utdanningstesten *(igjen, etter et par dager)* / Utdanningstesten Uroboros / Bigfive-test / Truity's Typefinder test. Du sender deretter inn resultatet (persontypene) du fikk via nettskjemaet som er linket til nederst på denne siden.

STEG 1: Ta Personlighetstestene

Personlighetstester

TEST1: Utdanningstesten.

Alle som velger å delta i prosjektet må ta **Utdanningstesten** én gang (10 min). Disse resultatene må du skrive ned: Under "1) Din personlighetsstil", noter deg de 4 ordene/setningene du fikk (f.eks. "Ekstrovert" - "Sansebasert" - "Rasjonell" - "Se i sammenheng")

LINK: Utdanningstesten

TEST 2: Velg minst én

a) Utdanningstesten (Trenger spesielt flere deltakere)

Ta Utdanningstesten igjen, noen dager (eller lenger) etter at du tok den første gangen.

Disse resultatene må du skrive ned: Under "1) Din personlighetsstil" noter de 4 ordene/setningene du fikk *(f.eks. på bildet så er resultatet: "Introvert" - "Sansebasert" - "Rasjonell" - "Se i detaljer")*

LINK: Utdanningstesten

b) Big-five test

<u>Disse resultatene må du skrive ned:</u> Tallene du fikk på disse dimensjonene: 1) Extraversion, 2) Openness, 3) Agreeableness, 4) Conscientiousness.

LINK: [bigfive-test]

c) Big-Four test

<u>Disse resultatene må du skrive ned:</u> De 4 bokstavene du fikk *(f.eks. om du fikk typen "INTJ", som på bildet, så er bokstavene dine: "I", "N", "T" og "J")*

LINK: [Typefinder-Test]

d) Utdanningstesten Uroboros (Trenger fler deltakere)

Dette er en lengre versjon av Utdanningstesten .

<u>Disse resultatene må du skrive ned:</u> De 4 bokstavene du fikk (f.eks. om du fikk typen "ENFP", så er bokstavene "E", "N", "F" og "P")

LINK: [Utdanningstesten Uroboros]

Denne testen krever innlogging. For å få tildelt en bruker, skriv inn din epostadresse i feltet under og trykk send. Innloggingsdetaljer og eventuelle testresultater vil bli sendt til deg på denne epostadressen. Dette er ikke en bindende påmelding. *NB! Dersom du velger å ikke rapportere resultatet du fikk i nettskjemaet, vil ikke dataene dine bli brukt i prosjektet.*

NBNB! Dersom du enda ikke har mottat mailen etter et par timer, sjekk spamfolderen for avsenderen "Uroboros Utdanningstest". Det tar normalt 10-30 min for mailen å ankomme.

E-postadresse*

Send

STEG 2: Rapporter resultatene

Nettskjema

Link til nettskjema.no, hvor du anonymt kan melde inn resultatene dine, finner du på denne linken

NB, innen du begynner fylle ut skjemaet, skal du ha tatt minst 2 personlighetstester: Utdanningstesten, og en til test (les "TEST2" ovenfor)

6.1.3.4 – "Contact Us" page

Forskning på læringsstrategier og personlighetstyper

Hjem Informasjon Personlighetstester **Kontakt Oss**

Om du har noen spørsmål knyttet til forskningsprosjektet, ta kontakt med:

Kathrine Frey Frøslie (førsteamanuensis NMBU, Prosjektansvarlig)

kathrine.frey.frøslie@nmbu.no, MOBIL: +47 672 32 591

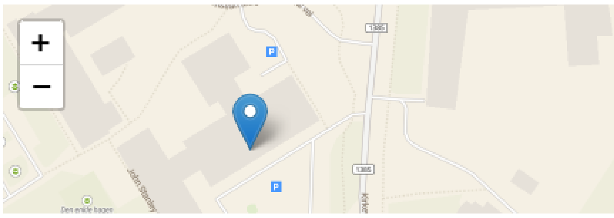
Mina Therese Gjefle (masterstudent NMBU, Databehandler)

mina.therese.gjefle@nmbu.no, MOBIL: +47 941 64 647

Dersom du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

NSD - Norsk senter for forskningsdata AS

EPOST: personverntjenester@nsd.no, TELEFON: 55 58 21 17



Kontoradresse

Kathrines kontor er lokalisert i 2.etg i Meieribygningen, avdeling Bioinformatikk og anvendt statistikk (BIAS)

6.1.4 – Nettskjema web form – an overview

Spørreskjemaet lages i den elektroniske løsningen Nettskjema som er utviklet av UiO/USIT, og som er tilgjengelig for ansatte ved NMBU via databehandleravtale med UiO/USIT.

Samtykkeerklæring (Obligatoriske spørsmål)

Jeg har mottatt og forstått informasjon om prosjektet «Sammenligning av ulike spørreskjemaer som gir informasjon om læringsstrategier og personlighetstyper» og har fått anledning til å stille spørsmål . Jeg samtykker til å delta i studien slik den er forklart i informasjonsskrivet jeg fikk på e-post.

Ja, Nei

Jeg samtykker med dette også til at mine opplysninger (resultater ifra testene) behandles frem til prosjektet er avsluttet, 31.desember 2024, da vil all persondata som kan knyttes tilbake til meg være slettet.

Ja, Nei

Kjønn:

Kvinne, Mann, Ønsker ikke å svare

Hva er din alder?

Tall mellom 1 og 100

TEKST:

Alle deltakere skal ha tatt den korte versjonen av Utdanningstesten én gang. Hver deltaker fikk også valget mellom å i tillegg ta 1) Den store versjonen av Utdanningstesten, 2) Den samme utdanningstesten én gang til med 1 ukes mellomrom, 3) Testen bigfive-test.com, og/eller 4) En annen BigFour test

Resultater fra Utdanningstesten – liten (10 minutters) versjon [OBLIGATORISK]

Jeg har tatt Utdanningstesten som ligger på utdanningstesten.no, og fikk følgende resultater på 1 Din personlighetsstil:

1
Ekstrovert (E) eller Introvert (I)
2
Fantasibasert (N) eller Sansebasert (S)
3
Rasjonell (T) eller Følsom (F)
4
Se i sammenheng (P) eller Se detaljer (J)

Hvor korrekt var Utdanningstestens beskrivelse? [VALGFRITT]

FRIVILLIG SPØRSMÅL: Under «1 Din personlighetsstil», hvor bra syntes du at beskrivelsen passet deg?

Veldig bra, Sånn passe, Veldig dårlig

Tilleggstest **[OBLIGATORISK, mer enn 1 (ett) svar er mulig]

I tillegg til å ta Utdanningstesten én gang, tok jeg også denne/disse testene:

Utdanningstesten igjen, 1 uke etter den første gangen

Den store versjonen av Utdanningstesten

BigFive testen

Ingen

Resultater fra Utdanningstesten – liten [VISES KUN OM «utdanningstesten igjen ...»]

Jeg fikk også i oppdrag å vente 1 uke og ta den samme Utdanningstesten én gang til, og fikk følgende resultater:

1
Ekstrovert (E) eller Introvert (I)
2
Fantasibasert (N) eller Sansebasert (S)
3
Rasjonell (T) eller Følsom (F)
4
Se i sammenheng (P) eller Se detaljer (J)

Resultater fra Utdanningstesten – stor [VISES KUN OM «den store versjonen ...»]

Jeg fikk også i oppdrag å ta den utvidede versjonen av Utdanningstesten, og fikk følgende resultater:

1
Ekstrovert (E) eller Introvert (I)
2
Fantasibasert (N) eller Sansebasert (S)
3
Rasjonell (T) eller Følsom (F)
4
Se i sammenheng (P) eller Se detaljer (J)

Resultater fra bigfive-test [VISES KUN OM «big five testen»]

Jeg fikk også i oppdrag å ta big five testen, og fikk følgende resultater:

Ekstroversjon
Tall mellom 0 og 120

Åpenhet for erfaringer
Tall mellom 0 og 120

Medmenneskelighet
Tall mellom 0 og 120

Planmessighet
Tall mellom 0 og 120

6.2 – Appendix B – Descriptive analysis

Table B1: Descriptive statistics across every variable of interest in the sample 1 dataset. The amount of uniques (n) and the most common group/value (Common) is also printed.

	Column	Dtype	mean (std)	median	[min, max]	n	Common -- (n, %)
0	Gender	Nominal	x	x	x	3	K -- (42, 77.78 %)
1	Age	Integer	23.81 (4.05)	23.0	[19, 44]	14	23 -- (11, 20.37 %)
2	UTD1 letter1 (0=E, 1=I)	Binary	0.69 (0.47)	1.0	[, 1]	2	1 -- (37, 68.52 %)
3	UTD1 letter2 (0=N, 1=S)	Binary	0.74 (0.44)	1.0	[, 1]	2	1 -- (40, 74.07 %)
4	UTD1 letter3 (0=T, 1=F)	Binary	0.61 (0.49)	1.0	[, 1]	2	1 -- (33, 61.11 %)
5	UTD1 letter4 (0=P, 1=J)	Binary	0.61 (0.49)	1.0	[, 1]	2	1 -- (33, 61.11 %)
6	'Was the UTD1 type fitting?'	Ordinal	x	x	x	3	Good -- (28, 51.85 %)
7	UTD2 letter1 (0=E, 1=I)	Binary	0.46 (0.52)	0.0	[0, 1]	43	0 -- (7, 53.85 %)
8	UTD2 letter2 (0=N, 1=S)	Binary	0.62 (0.51)	1.0	[0, 1]	43	1 -- (8, 61.54 %)
9	UTD2 letter3 (0=T, 1=F)	Binary	0.46 (0.52)	0.0	[0, 1]	43	0 -- (7, 53.85 %)
10	UTD2 letter4 (0=P, 1=J)	Binary	0.54 (0.52)	1.0	[0, 1]	43	1 -- (7, 53.85 %)
11	Uroboros letter1 (0=E, 1=I)	Binary	0.59 (0.51)	1.0	[0, 1]	39	1 -- (10, 58.82 %)
12	Uroboros letter2 (0=N, 1=S)	Binary	0.59 (0.51)	1.0	[0, 1]	39	1 -- (10, 58.82 %)
13	Uroboros letter3 (0=T, 1=F)	Binary	0.35 (0.49)	0.0	[0, 1]	39	0 -- (11, 64.71 %)
14	Uroboros letter4 (0=P, 1=J)	Binary	0.76 (0.44)	1.0	[0, 1]	39	1 -- (13, 76.47 %)
15	Extraversion	Continuous	72.91 (16.47)	72.5	[31, 103]	42	70 -- (4, 9.09 %)
16	Openness	Continuous	80.42 (11.58)	77.0	[60, 102]	34	73 -- (3, 6.98 %)
17	Agreeableness	Continuous	92.5 (11.55)	93.5	[59, 110]	35	90 -- (5, 11.36 %)
18	Conscientiousness	Continuous	84.77 (16.94)	89.0	[7, 109]	41	97 -- (5, 11.36 %)
19	Truity letter1 (0=E, 1=I)	Binary	0.59 (0.5)	1.0	[0, 1]	19	1 -- (22, 59.46 %)
20	Truity letter2 (0=N, 1=S)	Binary	0.35 (0.48)	0.0	[0, 1]	19	0 -- (24, 64.86 %)
21	Truity letter3 (0=T, 1=F)	Binary	0.78 (0.42)	1.0	[0, 1]	19	1 -- (29, 78.38 %)
22	Truity letter4 (0=P, 1=J)	Binary	0.62 (0.49)	1.0	[0, 1]	19	1 -- (23, 62.16 %)
23	Number of tests taken	Ordinal	2.37 (0.83)	2.0	[1, 4]	4	2 -- (28, 51.85 %)
24	Personality type (UTD1)	Nominal	x	x	x	14	ISFJ -- (11, 20.37 %)
25	Personality type (UTD2)	Nominal	x	x	x	12	ISTJ -- (3, 23.08 %)
26	Personality type (Uroboros)	Nominal	x	x	x	11	ISTJ -- (5, 29.41 %)
27	Personality type (Truity)	Nominal	x	x	x	13	INFJ -- (8, 21.62 %)
28	Personality type (Big-Five)	Nominal	x	x	x	17	ENFJ -- (30, 69.77 %)
29	% mismatch; UTD1-UTD2	Continuous	0.15 (0.19)	0.0	[0, 0.5]	44	0 -- (7, 53.85 %)
30	all matched; UTD1-UTD2 (0=no, 1=yes)	Binary	0.54 (0.52)	1.0	[0, 1]	43	1 -- (7, 53.85 %)
31	% mismatch; UTD1-URO	Continuous	0.31 (0.21)	0.25	[0, 0.75]	41	0.25 -- (8, 47.06 %)
32	all matched; UTD1-URO (0=no, 1=yes)	Binary	0.18 (0.39)	0.0	[0, 1]	39	0 -- (14, 82.35 %)
33	% mismatch; UTD1-Tr	Continuous	0.27 (0.22)	0.25	[0, 0.75]	21	0.25 -- (14, 37.84 %)

	Column	Dtype	mean (std)	median	[min, max]	n	Common -- (n, %)
34	all matched; UTD1-Tr (0=no, 1=yes)	Binary	0.3 (0.46)	0.0	[0, 1]	19	0 -- (26, 70.27 %)
35	% mismatch; UTD1-BF	Continuous	0.49 (0.24)	0.5	[0, 1]	16	0.5 -- (19, 44.19 %)
36	all matched; UTD1-BF (0=no, 1=yes)	Binary	0.07 (0.26)	0.0	[0, 1]	13	0 -- (40, 93.02 %)

Table B2: *Descriptive statistics for the most important variables in the sample 2 dataset; Gender, Age, County, Type, as well as the variables relating to education, school subjects, and math anxiety. For each question-related row, the percentwise distribution across all possible answers is shown. If odd number of possible answers (m), negative/positive coded answers are the bottom/top (m-1)/2 number of answers. If even numbers, m/2. For segment 4- and 5-questions, the cells list the cumulative negative/positive scores. Cells of interest have been colored red/green to signify a negative/positive biased percentwise count-score of at least +10%, e.g., 10% biased positivity for a question with 3 possible score-types = 33.3 * 1.1 = 36.6%. Bold lettering marks particularly high percentwise count-scores (at least 35% biased). Underlined scores marks observations that are otherwise interesting. (This is a shortened version and does not account for every question / variable in the dataset. For the full table, see the appendix)*

		TOT							
Gender; n (%)	F M	67.9% 32.1%							
Age; Group (%)		19-30 (33.72%) 16-18 (31.18%)							
County; Group (%, pr1000)		Viken (21.63%, 41) Svalbard (0.08%, 70)							
Personality type (4-letter)		ISFJ (14.4%) ENTJ (3.1%)							

	TOT	I	E	S	N	F	T	J	P
Personality type	236240	50.42%	49.58%	61.93%	38.07%	61.51%	38.49%	49.58%	50.42%
2_1a; Help humans/animals	0 = 24.0% 1 = 39.2% 2 = 36.8%	0 = 26.2% 1 = 38.4% 2 = 35.5%	0 = 21.7% 1 = 40.1% 2 = 38.2%	0 = 22.5% 1 = 36.2% 2 = 41.3%	0 = 26.4% 1 = 44.1% 2 = 29.5%	0 = 18.9% 1 = 38.3% 2 = 42.7%	0 = 32.0% 1 = 40.6% 2 = 27.4%	0 = 22.2% 1 = 37.8% 2 = 40.0%	0 = 25.7% 1 = 40.5% 2 = 33.7%
2_1b; Practical work	0 = 35.2% 1 = 47.4% 2 = 17.4%	0 = 33.3% 1 = 48.6% 2 = 18.1%	0 = 37.1% 1 = 46.2% 2 = 16.6%	0 = 36.1% 1 = 46.4% 2 = 17.5%	0 = 33.7% 1 = 49.1% 2 = 17.2%	0 = 35.5% 1 = 47.2% 2 = 17.3%	0 = 34.7% 1 = 47.8% 2 = 17.5%	0 = 37.5% 1 = 46.1% 2 = 16.5%	0 = 32.9% 1 = 48.8% 2 = 18.3%
2_1c; Creative solutions	0 = 11.3% 1 = 57.8% 2 = 30.9%	0 = 12.6% 1 = 59.3% 2 = 28.2%	0 = 10.0% 1 = 56.4% 2 = 33.7%	0 = 13.7% 1 = 62.2% 2 = 24.1%	0 = 7.3% 1 = 50.6% 2 = 42.0%	0 = 12.3% 1 = 59.5% 2 = 28.2%	0 = 9.7% 1 = 55.1% 2 = 35.2%	0 = 12.0% 1 = 60.3% 2 = 27.7%	0 = 10.6% 1 = 55.4% 2 = 34.0%
2_1d; Structural / systematic	0 = 29.5% 1 = 55.7% 2 = 14.8%	0 = 28.0% 1 = 54.0% 2 = 18.1%	0 = 31.1% 1 = 57.4% 2 = 11.5%	0 = 27.7% 1 = 55.3% 2 = 17.0%	0 = 32.6% 1 = 56.2% 2 = 11.2%	0 = 33.3% 1 = 55.1% 2 = 11.6%	0 = 23.6% 1 = 56.5% 2 = 19.9%	0 = 28.3% 1 = 56.0% 2 = 15.7%	0 = 30.7% 1 = 55.3% 2 = 13.9%
2_2a; Climate, aid work	0 = 34.6% 1 = 48.4% 2 = 17.1%	0 = 33.4% 1 = 48.6% 2 = 18.1%	0 = 35.8% 1 = 48.2% 2 = 16.0%	0 = 33.5% 1 = 47.5% 2 = 19.0%	0 = 36.3% 1 = 49.7% 2 = 13.9%	0 = 31.9% 1 = 49.5% 2 = 18.7%	0 = 38.9% 1 = 46.6% 2 = 14.5%	0 = 32.8% 1 = 48.9% 2 = 18.3%	0 = 36.4% 1 = 47.8% 2 = 15.8%
2_2b; Physical work	0 = 23.8% 1 = 48.1% 2 = 28.0%	0 = 25.2% 1 = 48.3% 2 = 26.4%	0 = 22.4% 1 = 47.9% 2 = 29.7%	0 = 23.5% 1 = 46.7% 2 = 29.7%	0 = 24.3% 1 = 50.4% 2 = 25.3%	0 = 21.7% 1 = 47.2% 2 = 31.2%	0 = 27.3% 1 = 49.7% 2 = 23.1%	0 = 24.8% 1 = 47.7% 2 = 27.5%	0 = 22.9% 1 = 48.5% 2 = 28.6%

	TOT								
2_2c; Discussion & Ingenuity	0 = 14.3% 1 = 54.9% 2 = 30.8%	0 = 16.8% 1 = 55.7% 2 = 27.5%	0 = 11.7% 1 = 54.1% 2 = 34.2%	0 = 19.3% 1 = 59.2% 2 = 21.5%	0 = 6.1% 1 = 48.0% **2 = 45.9%**	0 = 15.3% 1 = 54.7% 2 = 30.0%	0 = 12.7% 1 = 55.2% 2 = 32.1%	0 = 17.5% 1 = 56.7% 2 = 25.8%	0 = 11.1% 1 = 53.2% 2 = 35.7%
2_2d; Theory + Practice	0 = 27.3% 1 = 48.7% 2 = 24.1%	0 = 24.6% 1 = 47.4% 2 = 28.0%	0 = 30.0% 1 = 50.0% 2 = 20.0%	0 = 23.6% 1 = 46.6% 2 = 29.8%	0 = 33.2% 1 = 52.1% 2 = 14.8%	0 = 31.1% 1 = 48.8% 2 = 20.1%	0 = 21.1% 1 = 48.6% 2 = 30.3%	0 = 24.9% 1 = 46.8% 2 = 28.3%	0 = 29.5% 1 = 50.6% 2 = 19.9%
2_3a; Education / Communication	0 = 17.7% 1 = 43.1% 2 = 39.2%	0 = 21.0% 1 = 44.6% 2 = 34.5%	0 = 14.4% 1 = 41.6% 2 = 44.0%	0 = 17.7% 1 = 40.1% 2 = 42.2%	0 = 17.6% 1 = 48.0% 2 = 34.4%	0 = 15.4% 1 = 42.1% 2 = 42.5%	0 = 21.4% 1 = 44.6% 2 = 34.0%	0 = 16.9% 1 = 42.8% 2 = 40.3%	0 = 18.5% 1 = 43.4% 2 = 38.1%
2_3b; Practical / Technical facility	0 = 46.5% 1 = 39.2% 2 = 14.2%	**0 = 45.3%** 1 = 39.7% 2 = 14.9%	**0 = 47.7%** 1 = 38.7% 2 = 13.5%	**0 = 45.1%** 1 = 39.6% 2 = 15.3%	**0 = 48.8%** 1 = 38.6% 2 = 12.6%	**0 = 46.5%** 1 = 39.2% 2 = 14.3%	**0 = 46.5%** 1 = 39.4% 2 = 14.1%	**0 = 49.3%** 1 = 37.7% 2 = 13.0%	0 = 43.8% 1 = 40.8% 2 = 15.4%
2_3c; Innovation / Creation	0 = 12.9% 1 = 59.8% 2 = 27.2%	0 = 13.8% 1 = 59.2% 2 = 27.1%	0 = 12.1% 1 = 60.6% 2 = 27.3%	0 = 16.8% 1 = 64.4% 2 = 18.8%	0 = 6.7% 1 = 52.5% 2 = 40.8%	0 = 12.7% 1 = 59.6% 2 = 27.7%	0 = 13.4% 1 = 60.2% 2 = 26.4%	0 = 14.9% 1 = 61.9% 2 = 23.2%	0 = 11.1% 1 = 57.8% 2 = 31.1%
2_3d; Mathematical calculation	0 = 22.7% 1 = 58.0% 2 = 19.3%	0 = 19.8% 1 = 56.7% 2 = 23.4%	0 = 25.7% 1 = 59.2% 2 = 15.1%	0 = 20.2% 1 = 56.1% 2 = 23.7%	0 = 26.8% 1 = 61.0% 2 = 12.2%	0 = 25.2% 1 = 59.3% 2 = 15.5%	0 = 18.7% 1 = 55.9% 2 = 25.4%	0 = 18.8% 1 = 57.8% 2 = 23.4%	0 = 26.6% 1 = 58.1% 2 = 15.3%
2_4a; Clean water	0 = 22.8% 1 = 46.0% 2 = 31.3%	0 = 24.5% 1 = 46.9% 2 = 28.6%	0 = 21.0% 1 = 45.0% 2 = 34.0%	0 = 22.0% 1 = 43.9% 2 = 34.1%	0 = 24.0% 1 = 49.4% 2 = 26.6%	0 = 19.5% 1 = 46.5% 2 = 34.0%	0 = 28.0% 1 = 45.0% 2 = 26.9%	0 = 22.0% 1 = 46.4% 2 = 31.6%	0 = 23.5% 1 = 45.5% 2 = 31.0%
2_4b; Fixing things	0 = 33.7% 1 = 48.1% 2 = 18.1%	0 = 32.4% 1 = 49.6% 2 = 18.0%	0 = 35.0% 1 = 46.7% 2 = 18.3%	0 = 34.8% 1 = 46.4% 2 = 18.7%	0 = 31.9% 1 = 50.9% 2 = 17.1%	0 = 33.5% 1 = 48.5% 2 = 18.0%	0 = 34.1% 1 = 47.6% 2 = 18.3%	0 = 36.3% 1 = 46.5% 2 = 17.1%	0 = 31.2% 1 = 49.7% 2 = 19.1%
2_4c; Designer / Architect / Inventor	0 = 15.4% 1 = 53.4% 2 = 31.2%	0 = 17.0% 1 = 52.5% 2 = 30.6%	0 = 13.7% 1 = 54.4% 2 = 31.9%	0 = 20.6% 1 = 57.7% 2 = 21.7%	0 = 6.9% 1 = 46.4% **2 = 46.7%**	0 = 15.5% 1 = 53.6% 2 = 30.9%	0 = 15.2% 1 = 53.2% 2 = 31.7%	0 = 18.5% 1 = 55.3% 2 = 26.2%	0 = 12.3% 1 = 51.6% 2 = 36.1%
2_4d; Technical control systems	0 = 28.1% 1 = 52.5% 2 = 19.4%	0 = 26.1% 1 = 51.1% 2 = 22.9%	0 = 30.2% 1 = 54.0% 2 = 15.8%	0 = 22.6% 1 = 52.0% 2 = 25.4%	0 = 37.2% 1 = 53.3% 2 = 9.5%	0 = 31.5% 1 = 51.4% 2 = 17.1%	0 = 22.7% 1 = 54.2% 2 = 23.1%	0 = 23.2% 1 = 51.8% 2 = 25.0%	0 = 33.0% 1 = 53.2% 2 = 13.8%
2_5a; Practical + Helping	0 = 14.4% 1 = 41.5% 2 = 44.1%	0 = 16.0% 1 = 42.1% 2 = 41.9%	0 = 12.9% 1 = 40.8% **2 = 46.3%**	0 = 14.1% 1 = 38.5% **2 = 47.4%**	0 = 14.9% 1 = 46.4% 2 = 38.7%	0 = 10.3% 1 = 39.1% **2 = 50.5%**	0 = 21.0% 1 = 45.2% 2 = 33.8%	0 = 14.1% 1 = 39.5% **2 = 46.4%**	0 = 14.8% 1 = 43.4% 2 = 41.8%
2_5b; Common sense + Practical experience	0 = 8.6% 1 = 65.3% 2 = 26.0%	0 = 8.5% 1 = 65.2% 2 = 26.4%	0 = 8.8% 1 = 65.5% 2 = 25.7%	0 = 8.5% 1 = 63.3% 2 = 28.2%	0 = 8.9% 1 = 68.7% 2 = 22.4%	0 = 8.7% 1 = 68.7% 2 = 22.7%	0 = 8.6% 1 = 60.0% 2 = 31.4%	0 = 9.0% 1 = 64.0% 2 = 27.0%	0 = 8.3% 1 = 66.6% 2 = 25.1%
2_5c; New & Exciting > Rules & Routines	0 = 16.6% 1 = 63.2% 2 = 20.2%	0 = 18.4% 1 = 61.6% 2 = 19.9%	0 = 14.7% 1 = 64.8% 2 = 20.5%	0 = 21.9% 1 = 65.7% 2 = 12.4%	0 = 8.0% 1 = 59.1% 2 = 32.9%	0 = 16.0% 1 = 64.7% 2 = 19.3%	0 = 17.5% 1 = 60.8% 2 = 21.8%	0 = 21.7% 1 = 63.9% 2 = 14.4%	0 = 11.5% 1 = 62.5% 2 = 26.0%
2_5d; Formulas & calculations	0 = 60.3% 1 = 30.1% 2 = 9.7%	**0 = 57.1%** 1 = 31.1% 2 = 11.8%	**0 = 63.5%** 1 = 29.0% 2 = 7.5%	**0 = 55.4%** 1 = 32.6% 2 = 11.9%	**0 = 68.1%** 1 = 25.8% 2 = 6.0%	**0 = 65.0%** 1 = 27.5% 2 = 7.5%	**0 = 52.8%** 1 = 34.1% 2 = 13.1%	**0 = 55.1%** 1 = 32.7% 2 = 12.2%	**0 = 65.4%** 1 = 27.5% 2 = 7.1%
3_1a; Problem then theory	0 = 27.8% 1 = 57.8% 2 = 14.5%	0 = 25.7% 1 = 60.0% 2 = 14.3%	0 = 29.9% 1 = 55.5% 2 = 14.7%	0 = 24.3% 1 = 61.9% 2 = 13.8%	0 = 33.4% 1 = 51.1% 2 = 15.6%	0 = 28.7% 1 = 57.6% 2 = 13.7%	0 = 26.2% 1 = 58.0% 2 = 15.7%	0 = 26.8% 1 = 60.1% 2 = 13.1%	0 = 28.8% 1 = 55.4% 2 = 15.8%
3_1b; Formula then excercise	0 = 15.4% 1 = 41.2% 2 = 43.3%	0 = 11.5% 1 = 37.5% **2 = 51.0%**	0 = 19.4% 1 = 45.1% 2 = 35.5%	0 = 9.7% 1 = 37.8% **2 = 52.6%**	0 = 24.7% 1 = 46.9% 2 = 28.3%	0 = 14.9% 1 = 41.7% 2 = 43.4%	0 = 16.2% 1 = 40.6% 2 = 43.2%	0 = 11.1% 1 = 38.1% **2 = 50.8%**	0 = 19.6% 1 = 44.3% 2 = 36.0%
3_1c; Free exploration > One solution	0 = 41.6% 1 = 43.1% 2 = 15.3%	0 = 42.9% 1 = 42.5% 2 = 14.6%	0 = 40.2% 1 = 43.6% 2 = 16.1%	**0 = 52.3%** 1 = 39.8% 2 = 7.9%	0 = 24.2% 1 = 48.5% 2 = 27.3%	0 = 42.1% 1 = 42.9% 2 = 15.0%	0 = 40.8% 1 = 43.4% 2 = 15.8%	**0 = 47.1%** 1 = 41.4% 2 = 11.5%	0 = 36.2% 1 = 44.7% 2 = 19.1%
3_1d; Prefer group work	0 = 15.2% 1 = 58.0% 2 = 26.8%	0 = 19.8% 1 = 60.1% 2 = 20.1%	0 = 10.5% 1 = 55.9% 2 = 33.6%	0 = 13.6% 1 = 60.7% 2 = 25.7%	0 = 17.7% 1 = 53.6% 2 = 28.7%	0 = 14.2% 1 = 57.9% 2 = 27.8%	0 = 16.7% 1 = 58.1% 2 = 25.2%	0 = 15.0% 1 = 60.5% 2 = 24.5%	0 = 15.3% 1 = 55.6% 2 = 29.1%
3_2a; General info, then specific	0 = 15.0% 1 = 56.5% 2 = 28.5%	0 = 13.7% 1 = 58.1% 2 = 28.2%	0 = 16.2% 1 = 54.9% 2 = 28.9%	0 = 13.8% 1 = 58.6% 2 = 27.6%	0 = 16.9% 1 = 53.0% 2 = 30.1%	0 = 13.1% 1 = 55.7% 2 = 31.2%	0 = 18.0% 1 = 57.7% 2 = 24.3%	0 = 14.7% 1 = 59.7% 2 = 25.5%	0 = 15.2% 1 = 53.3% 2 = 31.5%

	TOT								
3_2b; Theory, then exercises	0 = 17.4% 1 = 41.2% 2 = 41.4%	0 = 14.5% 1 = 39.8% **2 = 45.6%**	0 = 20.3% 1 = 42.6% 2 = 37.1%	0 = 12.7% 1 = 38.4% **2 = 48.9%**	0 = 25.0% 1 = 45.7% 2 = 29.2%	0 = 18.2% 1 = 41.5% 2 = 40.3%	0 = 16.1% 1 = 40.7% 2 = 43.2%	0 = 13.2% 1 = 38.1% **2 = 48.7%**	0 = 21.5% 1 = 44.3% 2 = 34.3%
3_2c; Prefer multi-road focus	0 = 37.8% 1 = 46.2% 2 = 16.0%	0 = 39.8% 1 = 46.5% 2 = 13.7%	0 = 35.8% 1 = 45.9% 2 = 18.2%	0 = 43.4% 1 = 45.2% 2 = 11.4%	0 = 28.8% 1 = 47.9% 2 = 23.4%	0 = 42.1% 1 = 44.8% 2 = 13.1%	0 = 31.0% 1 = 48.5% 2 = 20.4%	0 = 40.2% 1 = 45.8% 2 = 14.1%	0 = 35.5% 1 = 46.7% 2 = 17.8%
3_2d; Prefer teacher-group-guidance	0 = 29.8% 1 = 56.2% 2 = 14.1%	0 = 31.9% 1 = 55.6% 2 = 12.4%	0 = 27.6% 1 = 56.7% 2 = 15.7%	0 = 30.1% 1 = 57.8% 2 = 12.1%	0 = 29.3% 1 = 53.5% 2 = 17.3%	0 = 26.7% 1 = 58.0% 2 = 15.3%	0 = 34.8% 1 = 53.2% 2 = 12.0%	0 = 31.8% 1 = 56.5% 2 = 11.7%	0 = 27.8% 1 = 55.8% 2 = 16.4%
3_3a; Easy language & practical experience	0 = 26.1% 1 = 47.5% 2 = 26.5%	0 = 26.3% 1 = 47.2% 2 = 26.5%	0 = 25.9% 1 = 47.7% 2 = 26.4%	0 = 26.9% 1 = 47.9% 2 = 25.2%	0 = 24.7% 1 = 46.8% 2 = 28.5%	0 = 24.3% 1 = 46.5% 2 = 29.1%	0 = 28.9% 1 = 49.0% 2 = 22.1%	0 = 28.2% 1 = 47.9% 2 = 24.0%	0 = 24.0% 1 = 47.1% 2 = 28.9%
3_3b; Prefer methods & rules	0 = 32.5% 1 = 41.1% 2 = 26.4%	0 = 29.1% 1 = 40.8% 2 = 30.1%	0 = 35.9% 1 = 41.5% 2 = 22.6%	0 = 26.0% 1 = 41.1% 2 = 32.9%	0 = 43.1% 1 = 41.2% 2 = 15.8%	0 = 34.5% 1 = 41.3% 2 = 24.2%	0 = 29.2% 1 = 40.8% 2 = 30.0%	0 = 27.0% 1 = 40.8% 2 = 32.3%	0 = 37.9% 1 = 41.4% 2 = 20.6%
3_3c; Prefer multimedia learning	0 = 24.7% 1 = 47.3% 2 = 28.0%	0 = 27.1% 1 = 47.7% 2 = 25.2%	0 = 22.2% 1 = 47.0% 2 = 30.9%	0 = 30.1% 1 = 46.8% 2 = 23.1%	0 = 15.9% 1 = 48.1% 2 = 36.0%	0 = 24.4% 1 = 47.9% 2 = 27.7%	0 = 25.1% 1 = 46.4% 2 = 28.5%	0 = 27.4% 1 = 46.6% 2 = 25.9%	0 = 22.0% 1 = 48.0% 2 = 30.0%
3_3d; Does not like being instructed	0 = 16.7% 1 = 64.1% 2 = 19.1%	0 = 17.5% 1 = 64.4% 2 = 18.1%	0 = 16.0% 1 = 63.9% 2 = 20.2%	0 = 17.0% 1 = 64.3% 2 = 18.7%	0 = 16.3% 1 = 63.9% 2 = 19.8%	0 = 16.7% 1 = 64.3% 2 = 19.0%	0 = 16.8% 1 = 63.8% 2 = 19.4%	0 = 17.4% 1 = 64.8% 2 = 17.9%	0 = 16.1% 1 = 63.5% 2 = 20.4%
4_1; Chemistry & Toxicology	1 = 22.8% 2 = 20.2% 3 = 21.1% 4 = 20.0% 5 = 10.7% 6 = 5.3% 43 /15.9	1 = 20.4% 2 = 19.8% 3 = 21.7% 4 = 21.0% 5 = 11.5% 6 = 5.6% 40.2 /17.1	**1 = 25.2%** **2 = 20.6%** 3 = 20.5% 4 = 18.9% 5 = 9.8% 6 = 5.0% 45.8 /13.8	1 = 21.1% 2 = 19.5% 3 = 21.1% 4 = 21.0% 5 = 11.6% 6 = 5.6% 40.6 /17.3	**1 = 25.5%** **2 = 21.2%** 3 = 21.1% 4 = 18.2% 5 = 9.1% 6 = 4.9% 46.7 /14	1 = 23.6% 2 = 20.8% 3 = 21.3% 4 = 19.1% 5 = 10.3% 6 = 4.9% 44.4 /15.2	1 = 21.5% 2 = 19.2% 3 = 20.9% 4 = 21.3% 5 = 11.2% 6 = 5.9% 40.7 /17.1	1 = 21.6% 2 = 19.2% 3 = 20.8% 4 = 20.7% 5 = 11.9% 6 = 5.9% 40.8 /17.7	**1 = 24.0%** **2 = 21.2%** 3 = 21.5% 4 = 19.3% 5 = 9.4% 6 = 4.7% 45.2 /14
4_2; Biology	1 = 9.4% 2 = 14.4% 3 = 20.2% 4 = 22.3% 5 = 19.2% 6 = 14.5% 23.8 /33.7	1 = 8.6% 2 = 13.9% 3 = 20.0% 4 = 22.9% 5 = 19.7% 6 = 14.9% 22.5 /34.6	1 = 10.1% 2 = 14.9% 3 = 20.4% 4 = 21.7% 5 = 18.8% 6 = 14.1% 25 /32.9	1 = 9.2% 2 = 14.3% 3 = 19.9% 4 = 22.5% 5 = 19.6% 6 = 14.5% 23.5 /34.1	1 = 9.6% 2 = 14.6% 3 = 20.6% 4 = 22.1% 5 = 18.6% 6 = 14.6% 24.2 /33.1	1 = 8.0% 2 = 13.1% 3 = 20.1% 4 = 22.8% 5 = 20.5% 6 = 15.6% 21.1 /36	1 = 11.6% 2 = 16.5% 3 = 20.4% 4 = 21.6% 5 = 17.1% 6 = 12.7% 28.1 /29.9	1 = 9.2% 2 = 13.6% 3 = 19.7% 4 = 21.8% 5 = 20.3% 6 = 15.4% 22.8 /35.7	1 = 9.5% 2 = 15.2% 3 = 20.7% 4 = 22.8% 5 = 18.1% 6 = 13.6% 24.7 /31.8
4_3; Geology	1 = 42.7% 2 = 24.7% 3 = 16.3% 4 = 9.8% 5 = 4.5% 6 = 2.0% 67.4 /6.5	**1 = 39.3%** **2 = 24.9%** 3 = 17.6% 4 = 10.9% 5 = 5.2% 6 = 2.1% 64.2 /7.3	**1 = 46.1%** **2 = 24.4%** 3 = 14.9% 4 = 8.7% 5 = 3.9% 6 = 2.0% 70.5 /5.9	**1 = 42.7%** **2 = 24.7%** 3 = 16.4% 4 = 9.7% 5 = 4.5% 6 = 2.0% 67.4 /6.5	**1 = 42.7%** **2 = 24.7%** 3 = 16.0% 4 = 9.9% 5 = 4.6% 6 = 2.1% 67.4 /6.7	**1 = 43.2%** **2 = 25.1%** 3 = 16.0% 4 = 9.4% 5 = 4.3% 6 = 2.0% 68.3 /6.3	**1 = 41.9%** **2 = 24.0%** 3 = 16.7% 4 = 10.3% 5 = 5.0% 6 = 2.1% 65.9 /7.1	**1 = 41.8%** **2 = 24.4%** 3 = 16.7% 4 = 10.3% 5 = 4.8% 6 = 2.1% 66.2 /6.8	**1 = 43.6%** **2 = 25.0%** 3 = 15.9% 4 = 9.3% 5 = 4.3% 6 = 1.9% 68.6 /6.2
4_4; Mathematics	**1 = 30.4%** **2 = 15.5%** 3 = 17.1% 4 = 16.8% 5 = 12.6% 6 = 7.6% 45.9 /20.2	1 = 29.3% 2 = 15.5% 3 = 16.8% 4 = 16.6% 5 = 13.4% 6 = 8.4% 44.8 /21.8	**1 = 31.5%** **2 = 15.5%** 3 = 17.4% 4 = 16.9% 5 = 11.9% 6 = 6.8% 47 /18.7	1 = 26.6% 2 = 14.6% 3 = 17.2% 4 = 18.0% 5 = 14.5% 6 = 9.1% 41.2 /23.6	**1 = 36.6%** **2 = 16.9%** 3 = 17.1% 4 = 14.8% 5 = 9.5% 6 = 5.1% 53.5 /14.6	**1 = 33.3%** **2 = 16.6%** 3 = 17.3% 4 = 15.8% 5 = 10.8% 6 = 6.2% 49.9 /17	1 = 25.8% 2 = 13.7% 3 = 16.8% 4 = 18.3% 5 = 15.5% 6 = 9.9% 39.5 /25.4	1 = 25.5% 2 = 14.6% 3 = 17.0% 4 = 17.8% 5 = 15.0% 6 = 10.2% 41.1 /24.1	**1 = 35.2%** 2 = 16.4% 3 = 17.3% 4 = 15.8% 5 = 10.3% 6 = 5.1% 51.6 /15.3
4_5; Physics & Space	1 = 19.8% 2 = 18.3% 3 = 19.5% 4 = 17.8% 5 = 13.8% 6 = 10.8% 38.1 /24.6	1 = 17.4% 2 = 17.8% 3 = 19.4% 4 = 18.4% 5 = 15.1% 6 = 11.8% 35.2 /27	1 = 22.3% 2 = 18.8% 3 = 19.5% 4 = 17.3% 5 = 12.5% 6 = 9.7% 41.1 /22.1	1 = 20.3% 2 = 18.6% 3 = 19.1% 4 = 17.7% 5 = 13.8% 6 = 10.5% 38.9 /24.3	1 = 19.1% 2 = 17.8% 3 = 20.0% 4 = 18.0% 5 = 13.8% 6 = 11.3% 36.9 /25.1	1 = 21.2% 2 = 19.5% 3 = 20.1% 4 = 17.4% 5 = 12.7% 6 = 9.0% 40.7 /21.8	1 = 17.6% 2 = 16.3% 3 = 18.4% 4 = 18.5% 5 = 15.5% 6 = 13.6% 33.9 /29.2	1 = 20.2% 2 = 18.6% 3 = 19.2% 4 = 17.6% 5 = 13.9% 6 = 10.5% 38.8 /24.4	1 = 19.4% 2 = 18.0% 3 = 19.8% 4 = 18.0% 5 = 13.7% 6 = 11.1% 37.4 /24.8
4_6; IT / Practical /	1 = 9.3%	1 = 8.5%	1 = 10.2%	1 = 8.8%	1 = 10.1%	1 = 10.2%	1 = 7.9%	1 = 9.0%	1 = 9.6%

	TOT								
Physical work	2 = 14.2% 3 = 21.7% 4 = 24.1% 5 = 18.9% 6 = 11.8% 23.5 /30.7	2 = 14.3% 3 = 22.2% 4 = 24.5% 5 = 19.4% 6 = 11.2% 22.8 /30.5	2 = 14.2% 3 = 21.1% 4 = 23.7% 5 = 18.4% 6 = 12.4% 24.4 /30.8	2 = 14.0% 3 = 21.8% 4 = 24.7% 5 = 19.5% 6 = 11.2% 22.8 /30.7	2 = 14.6% 3 = 21.5% 4 = 23.1% 5 = 18.0% 6 = 12.7% 24.7 /30.7	2 = 15.3% 3 = 23.1% 4 = 24.1% 5 = 17.6% 6 = 9.7% 25.5 /27.3	2 = 12.4% 3 = 19.5% 4 = 24.1% 5 = 21.0% 6 = 15.1% 20.3 /36.1	2 = 13.7% 3 = 21.4% 4 = 24.4% 5 = 19.6% 6 = 12.0% 22.7 /31.5	2 = 14.7% 3 = 22.0% 4 = 23.9% 5 = 18.2% 6 = 11.6% 24.3 /29.8
5_1; Math class	1 = 10.9% 2 = 19.4% 3 = 29.3% 4 = 30.6% 5 = 9.9% 30.3//40.5	1 = 11.6% 2 = 17.8% 3 = 29.0% 4 = 30.7% 5 = 11.0% 29.4//41.7	1 = 10.2% 2 = 21.0% 3 = 29.6% 4 = 30.4% 5 = 8.8% 31.2//39.2	1 = 9.7% 2 = 15.7% 3 = 29.5% 4 = 33.3% 5 = 11.8% 25.4//45.1	1 = 12.8% 2 = 25.4% 3 = 29.1% 4 = 26.1% 5 = 6.7% 38.2//32.8	1 = 12.3% 2 = 20.0% 3 = 30.3% 4 = 29.1% 5 = 8.3% 32.3//37.4	1 = 8.6% 2 = 18.4% 3 = 27.7% 4 = 32.8% 5 = 12.5% 27//45.3	1 = 9.5% 2 = 13.3% 3 = 29.1% 4 = 34.9% 5 = 13.2% 27.9//48.1	1 = 12.2% 2 = 25.3% 3 = 29.6% 4 = 26.3% 5 = 6.7% 37.5//33
5_2; Solving math exercises	1 = 12.1% 2 = 13.1% 3 = 37.4% 4 = 28.3% 5 = 9.1% 25.2//37.4	1 = 15.5% 2 = 12.2% 3 = 35.3% 4 = 27.7% 5 = 9.2% 27.7//36.9	1 = 8.6% 2 = 14.0% 3 = 39.4% 4 = 28.9% 5 = 9.0% 22.6//37.9	1 = 11.3% 2 = 11.5% 3 = 36.2% 4 = 30.6% 5 = 10.4% 22.8//41	1 = 13.5% 2 = 15.8% 3 = 39.3% 4 = 24.4% 5 = 7.0% 29.3//31.4	1 = 14.6% 2 = 14.0% 3 = 38.1% 4 = 26.3% 5 = 7.0% 28.6//33.3	1 = 8.2% 2 = 11.8% 3 = 36.1% 4 = 31.4% 5 = 12.5% 20//43.9	1 = 11.5% 2 = 10.1% 3 = 34.3% 4 = 32.3% 5 = 11.8% 21.6//44.1	1 = 12.7% 2 = 16.1% 3 = 40.4% 4 = 24.3% 5 = 6.5% 28.8//30.8

6.3 – Appendix C – Learning strategies & personality

Table C1: *Link between each of the dichotomous Big-Four dimensions and the Big-Five dimensions, together with each letters preferred learning style. Reworked from (Vinje et.al., 2021) and (Myers et.al., 1998).*

MBTI	Big Five	Learning style
E – Extrovert	Extraversion: HIGH	**Participatory and Activity:** • Interaction (group, cooperative behavior). • Talking & Discussing • Concrete & experimental (e.g., projects) • Variety of methods • Dependent learner (needs to be activated and teacher to carry most of the cognitive load (Hammond, 2014)) • Goal-oriented
I – Introvert	Extraversion: LOW	**Individualism and Observation:** • Individualism (quiet reflection, processing at their own pace). • Reflection and observation. • Visual-Auditory. • Abstract sequential style. • Independent learner (self-activating).
N – Intuition	Openness: HIGH	**Abstraction and Possibilities:** • Associations and Meanings. Read between the lines, and general concepts. • Deadlines to avoid procrastination. • Conceptual • Visual-Auditory • Holistic learners • Reflective judgement • Self-directed & Innovative • Goal-oriented
S – Sensing	Openness: LOW	**Practicalities and Facts:** • Stay connected to practical realities. Learn by doing. • Observation and facts. Moving towards abstract concepts and principles. • Concrete & experimental • Variety of methods • Sequential & Collaborative • Dependent learner • Fact-retention
F – Feeling	Agreeableness: HIGH	**Guidance and Qualities:** • Learn by positively worded feedback. • Study in dialog • Experimental • Dependent • Holistic
T – Thinking	Agreeableness: LOW	**Figurative and Data-driven:** • Logical flow to the subject. • Analyze to bring order to confusion. • Abstract • Fact-retention • Goal-oriented

J – Judgment	Conscientiousness: HIGH	**Sequential; Bits and Particulars:** • Clear structure, Formalized instructions that moves in orderly sequences. • Aim towards completions and closure. • Flexible; Abstract conceptual or Concrete sequential • Fact-retention • Independent studies • Goal-oriented
P – Perception	Conscientiousness: LOW	**Global thinking: Patterns and Relations** • Open exploration without preplanned structure; Flexible handling of problems • Spontaneously following their curiosity. • Newness • Experimental & Active • Holistic • Innovative

Table C2: *The Big Five model with its 5 dimensions and 30 facets. Descriptions in cursive explain what a high value in the current dimensions would signify. Reworked from bigfive-test.com*

Dimension	Facets					
Openness *Open to new experiences & possibilities*	Imagination	Adventurous-ness	Emotionality	Intellect	Liberalism	Artistic Interests
Conscientiousness *Careful, disciplined, consciously control and direct impulses*	Dutifulness	Achievement-Striving	Cautiousness	Self-Efficacy	Self-Discipline	Orderliness
Extraversion *Pronounced engagement with the external world*	Friendliness	Gregariousness	Assertiveness	Activity Level	Excitement Seeking	Cheerfulness
Agreeableness *Values cooperation and social harmony*	Sympathy	Cooperation	Modesty	Trust	Morality	Altruism
Neuroticism *Pessimistic, tendency to experience negative feelings*	Depression	Self-consciousness	Immoderation	Anger	Anxiety	Vulnerability

Table C3: *Correlations between the Big-Five dimensions and Big-Four dimensions as registered by (Furnham, 1996) and (McCrae & Costa, 1989). In the table, the two studies are respectively titled "Fur" and "McC Costa".*

Big-Five / Big-Four	Extraversion		Openness		Agreeable.		Neuroticism		Consc.	
	Fur	McC Costa	Fur	McC Costa	Fur	McC Costa	Fur	McC Costa	Fur	McC Costa
Introvert (I)	-0.46		0.22				0.26			
Extravert (E)	0.69		0.21				-0.24			
EI	-0.70	-0.74	-0.22	0.03*			0.25	0.16		
Sensation (S)	-0.18		0.52						0.20	
Intuitive (N)	0.16		0.49						0.24	
SN			0.48	0.72					-0.16*	-0.15
Thinking (T)			0.22		0.40		-0.16		-0.28	
Feeling (F)			0.22		0.40		0.18		-0.28	
TF	0.04*	0.19	-0.24	0.02*	0.47	0.44			-0.23	-0.15
Perceiving (P)			0.24						-0.41	
Judging (J)			-0.24						0.50	
JP	0.02*	0.15	0.17	0.30					0.52	-0.49

* Not significant by their respective study

Table C4: *Preferred learning style per letter type. This table suggests how an educator can formulate their lectures or tasks, to appeal to the natural cognitive inclinations of each type.*

Type	Preferred learning styles
I	Start with theory and an example walkthrough, then give tasks, particularly if paired with S and J. Tasks should be formulated concrete, with a practical angle. Open-ended tasks can be used too, if paired with N but not J. Avoid excessive group work, particularly if paired with N. With J and T, traditional exams using formulas is best suited.
E	Do not be too authoritative or instructive, and avoid using hard technical terms, especially if paired with F or P. Start with general examples before theory. Tasks should have a practical angle, particularly if paired with S, and let them be able to find their way by themselves. Open the classroom for discussion and group work.
S	Initiate the lectures with theory; Instruct the S-types and give them "recipes" for solving problems. Tasks should focus on details and the use of formulas and not be too open-ended. Group work can be a great idea.
N	As with E-types, the relationship to their educator is important to N's. Avoid excessive instruction and authority. Tasks should leave room for interpretation and exploration and have a somewhat holistic focus. Project- or report-oriented teaching style could be preferred, though avoid group work unless paired with E.
T	Lectures and tasks should focus on details, methods and rules. Encourage the student to look for multiple ways to solve a problem (unless ST). Avoid group work.
F	Put the theory and lecture into perspective by using examples from everyday life. Do not be too instructive or overtly theoretical. Allow to work in colloquium groups with you and TA's present.
P	Do not be too instructive – teach using a more conversational tone. Allow for hands-on tasks, where they can learn as they go. If paired with N, give tasks that open for interpretation and exploration, where there is not necessarily one true answer. Unless paired with I or T, group work is encouraged.
J	Initiate the lectures with theory; Instruct and give them "recipes" for solving problems. However, if paired with N, avoid excessive instruction. Open for hands-on tasks, where they can learn as they go, if paired with S.

6.4 – Appendix D – Utdanningstesten

Table D1: *Overview of the personality-determining questions from Utdanningstesten, before the 2021-expansion ("Bf.") and after ("A"). Sample 1 took the test after the expansion, and the individuals in sample 2 took the older version. Questions are divided by dimension and which statements were positively coded for each letter ("Low-scorer"; I/S/T/P, "High-scorer"; E/N/F/J). In part reworked from (Aspheim, 2020).*

Dim		Low-scorer	High-scorer
I / E	Bf.	Jeg er stille og rolig, og liker godt å arbeide konsentrert på egenhånd, f.eks. med leksene mine. Jeg trives best i mindre grupper, og jeg unngår ofte å ta ordet i større forsamlinger eller i klasserommet. Jeg kan bli sjenert hvis jeg får mye oppmerksomhet, og kun mine nærmeste venner vet mye om meg.	Jeg er utadvendt, pratsom og lett å bli kjent med, og jeg trives med mye folk rundt meg – både på skolen, hjemme og i gruppearbeid. Jeg kan godt presentere noe for hele klassen, og jeg gjør det gjerne med intensitet, godt humør og engasjement.
	A	1B: Jeg er nok en litt reservert, innadvendt og en mer stille type, særlig i større forsamlinger. Jeg er vanligvis rolig og avventende, ettertenksom, beskjeden og behersket. 2B: Jeg er en noe privat person, og lytter gjerne heller enn å prate selv. Trenger litt tid til å tenke gjennom noe før jeg uttaler meg. Liker bedre å snakke om en sak eller et emne enn å mene noe om hva de andre mener. 3B: Jeg foretrekker å være heller litt alene eller sammen med de jeg kjenner godt, de jeg er trygge på framfor å hevde meg eller utfolde meg blant ukjente eller i større grupper. Bedre til å lytte, observere og forstå enn å mene eller selge mine meninger.	1A: Jeg er utadvendt, sosial og av natur ganske åpen. Jeg er lett å bli kjent med, pratsom og jeg trives godt med å arbeide i større grupper. Tenker ofte høyt, deler og utprøver gjerne mine tanker med andre. 2A: Jeg er spontan, trives med variasjon, er handlekraftig og ikke spesielt redd for å stille spørsmål eller snakke i forsamlinger. Engasjerer meg gjerne i gruppeaktiviteter. 3A: Jeg tåler godt avbrytelser og er fri og omgjengelig i kommunikasjonen, men blir lett utålmodig med å måtte jobbe lenge alene og med de samme oppgavene. Trives med at ting skjer rundt meg, og da helst sammen med andre.
S / N	Bf.	Jeg er praktisk, ærlig og jordnær, og er opptatt av det jeg faktisk kan oppfatte gjennom sansene mine. Jeg trives godt med oppgaver som krever aktivitet og jeg finner lett løsning på praktiske utfordringer. Jeg liker å se at det jeg gjør har en nytteverdi.	Jeg er kreativ og engasjert, og liker å gjøre ting på min egen måte. Tankene mine vandrer lett av sted, og jeg undrer ofte på både store og små ting i livet. Jeg liker å filosofere, og lar meg lett inspirere av kunst og kultur.
	A	4A: Jeg oppfatter meg selv som realistisk, jordnær, faktaorientert og fornuftig. Å forholde seg til det vi vet virker er bedre enn å måtte tolke, fantasere og tro. 5A: Jeg trives godt med oppgaver som er tydelige og som enkelt beskriver hva som skal gjøres. Jeg er opptatt av nytteverdi og resultat. Lærer best gjennom en praktisk teoretisk tilnærming, og gjennom det å kunne få prøve og feile og å prøve igjen. 6A: Jeg foretrekker faktainformasjon og brukermanualer framfor abstrakt teori og fantasifulle tanker om hva eventuelt kan bli eller er mulig.	4B: Jeg oppfatter meg selv som kreativ, og jeg trives med å finne på, designe eller lete etter nye måter å gjøre ting på. Repetisjon er kjedelig. 5B: Jeg er impulsiv, og i mindre grad opptatt av orden og disiplin. Jeg har et stort behov for frihet slik at jeg kan gjøre tingene litt på min egen måte og slik få bruke min fantasi og oppfinnsomhet i løsningene. 6B: Å få lov til å tenke nytt og på hva som er mulig og kanskje en mer hensiktsmessig løsningsmåte, er for meg mye mer spennende og inspirerende enn å følge læreboken.

T / F	Bf.	Jeg oppleves som seriøs, fornuftig, faktaorientert og rettferdig, og jeg stoler oftere på min logiske sans enn på magefølelsen. Jeg uttrykker meg gjerne på en ærlig og direkte måte.	Mange opplever meg som mild, snill og tillitsfull, og jeg er ofte opptatt av å skape god stemning og harmoni både i klasserommet og hjemme. Jeg er sympatisk, og ofte svært medfølende og lyttende når folk rundt meg har problemer.
	A	7B: Jeg er fornuftig og saksorientert, seriøs, direkte og ærlig. Jeg er ikke redd for å si hva jeg mener. Utsagn jeg oppfatter som lite logiske eller følelsesbaserte, kan jeg ha problemer med å godta. 8B: Jeg trives godt når jeg kan finne forklaringer på hvorfor innviklede ting er som de er. Jeg har en meget sterk rettferdighetssans og sier ofte det jeg mener selv om ikke alle nødvendigvis liker det. 9B: Er kanskje en mer analytisk logisk type som forholder meg sterkere til hva folk tenker og sier mer enn hva de føler. Er kanskje ikke alltid så varsom med andres følelser, og kan da lett oppfattes som noe kritisk og argumenterende.	7A: De fleste oppfatter meg nok som snill, føyelig, mild, hjelpsom og tolerant. Jeg er opptatt av å skape harmoni, smil og trivsel rundt meg, og jeg inngår lett kompromisser for å få dette til. 8A: Jeg er følsom, og bryr meg om og tilpasser meg til hvordan de rundt meg reagerer og tenker. Gruppearbeid er bra og nyttig, men jeg mistrives ekstra der det er personlige motsetninger, krangel og uenighet. 9A: Jeg er en person som lett roser andre, leter etter å finne det vi egentlig er ganske enige om og har likheter. Er en rimelig varm person som bryr meg om andre, er tjenestevillig, men også til tider for vennlig og mild på bekostning av egen selvhevdelse og egennytte.
P / J	Bf.	Jeg er impulsiv og fleksibel, liker best variasjon og er mer opptatt av å se og forstå sammenhengene og helheten enn å fokusere på detaljene. Jeg kan ofte utsette ting f.eks. skolearbeidet, men jobber godt når jeg først kommer i gang og det er travelt og hektisk rundt meg.	Jeg oppfattes ofte som nøyaktig og strukturert, og trives best når jeg har kontroll også på detaljene. Jeg liker å legge planer, som jeg gjennomfører i jevnt, godt tempo – uten stress eller forstyrrelser underveis
	A	10A: Jeg begynner som regel å jobbe med ting i siste liten, men da hardt og effektivt. Jeg liker godt å jobbe med flere ting samtidig, og leter gjerne etter alternative løsninger selv etter at noe er endelig bestemt. Kan oppfattes som en person med noe nedprioritert orden og struktur. 11A: Jeg har ofte min egen rekkefølge og måte å arbeide på, og jeg liker derfor dårlig å måtte følge helt bestemte instrukser. Jeg er tankemessig åpen og fleksibel, opptatt av å undres eller la tankene vandre fritt om det finnes en orden, mønstre og sammenhenger i livets mer komplekse forhold. 12A: Liker ikke å repetere kjente løsninger, men heller prøve å finne nye. Flink til å tilpasse meg endrede situasjoner, er fleksibel og romslig. Har en tendens til å utsette ubehagelige oppgaver lengst mulig, og klarer heller ikke alltid gjøre noe helt ferdig før jeg starter med noe interessant nytt. Mer nysgjerrig og søkende enn beslutningsvillig og handlingsorientert.	10B: Jeg er seriøs, nøyaktig og ordentlig. Jeg må ha kontroll, og liker å gjøre ting på riktig måte og i riktig rekkefølge. Jeg kommer kjapt i gang med oppgavene og arbeider jevnt og stødig. Noen vil kanskje oppfatte meg som litt mye administrativ og detaljfokusert. 11B: Jeg oppleves å ha en god orden og en viss systematisk legning, men blir av noen kanskje også oppfattet som litt for utålmodig når andre dveler, nøler og tviler. Kan slik til tider være for villig og for rask til å konkludere og administrere en løsning. 12B: Liker å ha og kunne følge planer, er organisert og kontrollert og har glede i å få ting gjort ferdig. Mer beslutningsvillig, strategisk og handlingsorientert enn metodisk nysgjerrig, utprøvende og nyskapende.

Table D2: *Description of every question in Utdanningstesten, for the version used during the period 2010-2020. The format of Utdanningstesten was changed in 2021 to include more visual effects such as emojis, as well as expanding the personality-determining questions from 1 to 3 questions per dichotomous dimensions. These additional questions are not covered here. In red are examples of shortened descriptors / key words for each statement or question, used e.g., in the PCA plots.*

<table>
<tr><th>Segment</th><th>Scoring</th><th colspan="2">Question / Claim</th></tr>
<tr><td rowspan="4">Part 1; Personality type</td><td rowspan="4">Test-taker is presented with four questions, each consisting of two opposing claims.

Test-taker must choose one.
The claims represent one side of each dichotomous dimension</td><td>Introvert (I): Jeg er stille og rolig, …</td><td>Ekstrovert (E): Jeg er utadvendt, …</td></tr>
<tr><td>Fantasibasert (N): Jeg er kreativ og engasjert, …</td><td>Sansebasert (S): Jeg er praktisk, ærlig og jordnær, …</td></tr>
<tr><td>Rasjonell (T): Jeg oppleves som seriøs, …</td><td>Følsom (F): Mange opplever meg som mild, …</td></tr>
<tr><td>Se i sammenheng (P): Jeg er impulsiv og fleksibel, …</td><td>Se detaljer (J): Jeg oppfattes ofte som nøyaktig og strukturert, …</td></tr>
<tr><td rowspan="2">Part 2; Work</td><td rowspan="2">{0, 1, 2}
5 segments of questions, each consisting of 4 claims, are presented to the test-taker.

The test-taker must assign a score to two of the four claims; Which claim was the most / best fitting (score = 2) and which was the least fitting (score = 0).

The remaining two claims are then given a score of 1 (neutral)</td><td colspan="2">• 1A: Jeg trives med å hjelpe/pleie både mennesker og dyr slik at de holder seg friske og har det godt. Yrker innen kundeoppfølging kan passe bra for meg. HelpService
• 1B: Jeg liker godt å arbeide med praktiske oppgaver og annet arbeid der jeg raskt ser resultater av det jeg gjør, f.eks. som snekker, mekaniker, elektriker eller lignende. Practical.
• 1C: Jeg liker å tenke ut nye løsninger på ulike problemstillinger, prøve meg frem og utforske ulike muligheter, f.eks. innenfor design eller det å starte opp en bedrift. Entrepeneur.
• 1D: Jeg lurer ofte på hvordan ting eller systemer er bygd opp rent logisk, teknisk eller matematisk. Yrker innenfor utvikling av apper og datasystemer eller koding kan passe meg godt. SystemsIT.</td></tr>
<tr><td colspan="2">• 2A: Jeg tenker mye på miljø- og klimaspørsmål, og er opptatt av at de som kommer etter oss, skal ha noe å leve av og en jord å leve på. Jeg kunne derfor gjerne jobbet med informasjon og kommunikasjon knyttet til nødhjelp eller bistandsarbeid. Climate.
• 2B: Jeg kan tenke meg en jobb innenfor et praktisk yrke. Jeg liker fysisk arbeid godt. Physical.
• 2C: Jeg liker godt å diskutere nye ideer, og kan godt tenke meg et yrke hvor jeg får bruke min fantasi og oppfinnsomhet. DiscussInventive.
• 2D: Jeg er opptatt av hvordan ting fungerer teknisk. Jo mer innviklet og teoretisk, desto morsommere synes jeg det er. Jeg kunne passe godt i yrker som er en blanding av teori og praksis. Theoretical.</td></tr>
</table>

		• 3A: Jeg liker å dele ting jeg har lært med andre. Yrker innen undervisning, formidling eller kommunikasjon er interessant for meg. Education. • 3B: Jeg kunne godt tenke meg å jobbe med praktiske oppgaver, for eksempel på en oljeplattform i Nordsjøen eller på et teknisk anlegg. Industrial. • 3C: Jeg er glad i å utvikle, planlegge og designe nye ideer. Jeg kan trives i alle slags yrker, bare jeg får arbeide med nytenking og ideskaping. PlanDesign. • 3D: Jeg liker å tenke systematisk, beregne løsninger og finne svar på kompliserte problemer. Jeg kunne tenkte meg å jobbe i et miljø som arbeider med tekniske spørsmål. TechnicalMath.
		• 4A: Jeg ønsker å jobbe med oppgaver og utfordringer som må løses for at barn i fattige land skal få nok mat og rent drikkevann. AidWork. • 4B: Jeg liker å snekre, fikse og ordne ting, og å løse problemer ved å kombinere teori og praksis. Carpentry. • 4C: Jeg trives med å bruke fantasien min til å fornye og forbedre ting. Et yrke som oppfinner, designer eller arkitekt kunne derfor passe meg godt. Architect. • 4D: Jeg er glad i å systematisere og strukturere, slik at det blir orden og system i det jeg jobber med, f.eks. i arbeid med systemer for teknisk kontroll, slik som alarm-, lys- og klima-regulering etc. Systems.
		• 5A: Jeg ønsker meg et yrke hvor jeg kan kombinere en praktisk jobb med det å gjøre noe for andre mennesker. PracticalHelp. • 5B: Jeg ønsker meg en jobb hvor jeg kan bruke mitt sunne vett og min praktiske erfaring til å finne effektive eller funksjonelle løsninger. CommonSense. • 5C: Jeg ønsker meg et yrke hvor det å skape noe nytt og spennende er viktigere enn krav om å følge regler og rutiner. Newness. • 5D: Jeg ønsker meg en jobb hvor jeg kan bruke formler og matematiske beregninger som verktøy til å løse ulike problemer. Formulas.
Part 3; Education style	{0, 1, 2} Same scoring system as Part 2	• 1A: Jeg foretrekker oppgaver der jeg først skal løse et praktisk problem, og så lære teorien etter hvert. ProblemFirst. • 1B: Jeg liker at læreren forklarer løsningsmetoden nøyaktig, før jeg deretter skal løse lignende oppgaver selv. ExampleFirst. • 1C: De oppgavene jeg liker best, har ikke fasitsvar eller en løsningsoppskrift, men lar min fantasi og kreativitet finne svaret. NoDefiniteAnswer. • 1D: Jeg foretrekker oppgaver hvor jeg får jobbe sammen med andre, der vi kan diskutere oss fram til en løsning som vi sammen står ansvarlig for. GroupWork.

		• 2A: Jeg liker undervisningen best når læreren først presenterer en rekke eksempler fra min hverdag som får meg til å skjønne selve faget. EverydayExamples. • 2B: Det beste er når læreren først går igjennom teorien, deretter eksempler og til slutt gir oss oppgaver som blir vanskeligere etter hvert. TheoryFirst. • 2C: Jeg foretrekker lærere som utfordrer meg til å lete etter flere metoder eller fremgangsmåter når en oppgave skal løses. MultipleSolutions. • 2D: Jeg lærer best når læreren sitter sammen med oss i gruppearbeid, og veileder oss frem til gode svar. GroupWorkT.
		• 3A: Lærere, lærebøker og oppgaver bør ikke bruke for vanskelige faguttrykk. Jeg lærer bedre når jeg først får prøve ut ting i praksis enn hvis jeg bare må lese i læreboka. NonTheoretical. • 3B: Jeg foretrekker lærere som underviser i metoder og regler, og som gjerne forteller hva som er riktig og galt før oppgavene skal løses. MethodsRules. • 3C: Min favorittlærer bruker inspirerende bilder og andre ord enn læreboka når emner forklares. Han/hun stiller gjerne spørsmål som gir rom for undring og egne ideer. Multimedia. • 3D: Det viktigste for meg er at læreren bruker et forståelig og hverdagslig språk, og fokuserer på samtale og dialog. Jeg trives ikke med å bli instruert. Conversational.
Part 4; Work topics	{1,2,3,4,5,6} The test-taker is given 6 claims they must each rate from 1-6 1 = "Likes the least" 6 = "Likes the most"	1. Jeg kunne godt tenke meg å finne ut hva slags giftige stoffer som finnes i ulike produkter. CHEMISTRY 2. Jeg synes det er artig med planter, mennesker og dyr. BIOLOGY 3. Jeg liker tanken på å kunne undersøke bergarter, studere naturformasjoner og lese kart. GEOSCIENCES 4. Jeg liker å jobbe med matematikk. MATH 5. Jeg synes alt som har med verdensrommet, planeter og naturlover å gjøre, er spennende. PHYSICS 6. Jeg liker å finne ut hvordan saker og ting er satt sammen og virker, uansett om det dreier seg om matretter, motorer, menneskekropper eller datamaskiner. STRUCTURAL_IT
Part 5; Thoughts around math	{1,2,3,4,5} The test-taker is given 2 questions with 5 claims each and are tasked with choosing one. A = 1 ("worst")	Class • A: Jeg gruer meg veldig til mattetimene, og har angst for faget. Jeg får lett jernteppe. • B: Jeg kjeder meg skikkelig i mattetimen, og tankene flyr fort av gårde til noe annet og mer interessant. • C: Matte er et nødvendig fag, som jeg bare må komme meg gjennom. Jeg verken liker eller misliker det. • D: Mattetimene er interessante, og jeg mestrer matematikken helt greit. • E: Jeg gleder meg til mattetimene! Matematikk er mitt favorittfag!

	E = 5 ("best")	Exercises • A: Jeg er livredd når jeg må stå foran klassen og løse matteoppgaver. • B: Jeg kommer nok aldri til å forstå matematikk, og jeg liker ikke faget spesielt godt. • C: Jeg glemmer ofte noen av detaljene, og det hender derfor at jeg får feil svar. Ellers går det greit i mattetimene. • D: Jeg mener selv at jeg er ganske god i matematikk. • E: Jeg er glad i matte, og fikser matematikken enkelt og greit!

6.5 – Appendix E – Codes

6.5.1 – Sample 1

Code E1.1: Make rater's matrix

```
def make_type_matrix(df, test1_idx, test2_idx):
    """
    Rater1/test1 is set as rows, rater2/test2 is set as columns
    """
    import pandas as pd
    import numpy as np

    #Unique personality types
    values, counts = np.unique(list(df.iloc[:, test1_idx]) + list(df.iloc[:, test2_idx]),
                               return_counts = True)
    values = list(values)
    counts = list(counts)

    value_count = {}
    for value, count in zip(values, counts):
        value_count[value] = count

    raters_df = pd.DataFrame(np.zeros(shape = (len(values), len(values))),
                             index = values,
                             columns = values)

    for test1_val, test2_val in zip(df.iloc[:, test1_idx], df.iloc[:, test2_idx]):
        raters_df.loc[test1_val, test2_val] += 1

    raters_df = raters_df / df.shape[0]
    return raters_df
```

Code E1.2: Given a rater's matrix, estimate Cohen's Kappa and standard error of the Kappa estimate. The following function then uses this information as input to make confidence intervals and perform hypothesis tests

```
def agreement_scores(raters_df, n, score = "kappa", decimals = 3):
    """
    The raters_df needs to be of shape (n,n), i.e. equally dimensioned
    Rater1/test1 is rows, Rater2/test2 is columns
    """
    if score == "kappa":
        ########################
        p_obs = 0
        for i in range(raters_df.shape[0]):
            #Sum all the diagonal elements
            p_obs += raters_df.iloc[i, i]

        #########################
        #rowsums_rater1 = Sum across all cols for a specific row (row, all cols = i.)
        #colsums_rater2 = Sum across all rows for a specific col (col, all rows = .i)
        rowsums_rater1 = []
        colsums_rater2 = []
        for i in range(raters_df.shape[0]):
            the_row = raters_df.iloc[i, :] #Row across all cols
            the_col = raters_df.iloc[:, i] #Col across all rows
            rowsums_rater1.append(sum(the_row))
            colsums_rater2.append(sum(the_col))

        #########################
        p_exp = 0
        for i in range(len(rowsums_rater1)):
            p_exp += rowsums_rater1[i]*colsums_rater2[i]

        kappa = (p_obs - p_exp) / (1 - p_exp)

        if kappa < 0.21: message = "None"
        elif 0.21 <= kappa < 0.40: message = "Minimal"
        elif 0.40 <= kappa < 0.60: message = "Weak"
        elif 0.60 <= kappa < 0.80: message = "Moderate"
        elif 0.80 <= kappa < 0.91: message = "Strong"
        elif kappa >= 0.91: message = "Almost perfect"
        else: message = "An error occurred"

        ##########################
        p_sum = 0
        for i in range(raters_df.shape[0]):
            r_rater1 = rowsums_rater1[i]
            c_rater2 = colsums_rater2[i]

            p_sum += r_rater1 * c_rater2 * (r_rater1 + c_rater2)

        se_k = 1/((1-p_exp) * sqrt(n)) * sqrt(p_exp + (p_exp)**2 - p_sum)

        Z_score = kappa / se_k

        return {"observed_agreement": round(p_obs, decimals),
                "expected_agreement": round(p_exp, decimals),
                "kappa": round(kappa, decimals),
                "level_of_agreement": message,
                "stderror": round(se_k, decimals),
                "Z": round(Z_score, decimals)}
```

```python
def hypothesis_test(parameter, SE, H0 = 0, param_name = "Kappa",
                    test_type = "Z", alpha = 0.05, decimals = 3):
    """
    parameter (float, estimated value of parameter. This could be a mean, or maybe a Kappa)
    statistic (float, estimated value of test type statistic, e.g. Z or T)
    SE (float, standarderror of parameter that was used to calculate the statistic)
    test_type (str, can be "Z" or "T")
    """

    if test_type.lower() == "z":
        table_statistic = stats.norm.ppf(1-alpha)
        table_statistic_half = stats.norm.ppf(1-alpha/2)

    statistic = (parameter - H0)/SE

    ############################## Two sided test creating a confidence interval
    a = parameter - table_statistic_half * SE
    b = parameter + table_statistic_half * SE

    ##############################
    if a <= H0 <= b:
        conclusion = f"Keep H0 as {param_name} ~= {H0}"
    else:
        conclusion = f"Accept H1 as {param_name} != {H0}"

    ############################### One-sided test (larger than)
    if statistic > table_statistic:
        conclusion_one = f"Accept H1 Kappa > {H0}"
    else:
        conclusion_one = f"Keep H0 Kappa <= {H0}"



    return {"Parameter": parameter,
            "Parameter type": param_name,
            "Param_interval": [round(a, decimals), round(b, decimals)],
            "conclusion": conclusion,
            "conclusion_onesided": conclusion_one}
```

Code E1.3: Find Kappa values for TYPE- and single-letter comparisons

```python
def print_kappa(the_df, decimals = 3, silent = False):
    df = the_df.copy()
    kappa_all = cohen_kappa_score(y1 = df.iloc[:, -4],
                                  y2 = df.iloc[:, -3])

    if not silent:  # = if silent is not True, but False
        print(f"Direct type comparison = {round(kappa_all, decimals)}")

    dict_res = {"Kappa_all": round(kappa_all, decimals)}

    for i, dimension in zip(range(4), ["I/E", "S/N", "T/F", "J/P"]):
        kappa = round(cohen_kappa_score(y1 = df.iloc[:, 2+i],
                                        y2 = df.iloc[:, 6+i]), decimals)

        if not silent:
            print(f"{dimension} = {kappa}")

        dict_res[dimension] = kappa

    return dict_res
```

Code E1.4: Examples of code-usage

```
print_kappa(df_US, decimals = 3)
```

```
Direct type comparison = 0.081
I/E = 0.746
S/N = 0.131
T/F = 0.025
J/P = 0.611
```

```
#Comparing the full personality type

summary = pd.DataFrame({".": [], "n": [], "K tot": [],
                        "K (E)": [], "K (N)": [], "K (F)": [], "K (J)": []
                        })
H0 = 0

for data, name in zip([df_UU, df_US, df_UB, df_UT],
                      ["Utdanningstesten", "Uroboros", "BigFive", "Truity"]):
    raters_df = make_type_matrix(data, -4, -3)
        #.style.set_caption("Matrix table of rater agreement")
    agree_scores = agreement_scores(raters_df, n = data.shape[0])
    H0_test = hypothesis_test(agree_scores["kappa"], agree_scores["stderror"], H0 = H0)

    a = H0_test["Parameter"]
    b = H0_test["Param_interval"]
    c = H0_test["conclusion"]
    d = H0_test["conclusion_onesided"]

    print(f"{name} >>> Kappa = {a}, Interval = {b}, TWOSIDED: {c}, ONESIDED: {d}")

    all_kappas = print_kappa(data, decimals = 2, silent = True)
    #^ gir dict med nøklene: "Kappa_all", "I/E", "S/N", "T/F", "J/P"

    if name == "Utdanningstesten":
        row = ["Utd.test1 VS Utd.test2"]# , "nval"] #, f"{a} {b}", "", "", "", ""]
    if name == "Uroboros":
        row = ["Utd.test1 VS Uroboros"] #, "nval"] #, f"{a} {b}", "", "", "", ""]
    if name == "Truity":
        row = ["Utd.test1 VS Truity"] #, "nval"] #, f"{a} {b}", "", "", "", ""]
    if name == "BigFive":
        all_kappas = print_kappa(bincode_continuous(data.copy()),
                                 decimals = 5, silent = True)
        row = ["Utd.test1 VS BigFive"] #, "nval"] #, f"{a} {b}", "", "", "", ""]

    row += [str(data.shape[0]).rstrip("0").rstrip("."),
            f"{round(a, 2)} {b}",
            round(all_kappas["I/E"], 2),
            round(all_kappas["S/N"], 2),
            round(all_kappas["T/F"], 2),
            round(all_kappas["J/P"], 2)
           ]

    summary.loc[len(summary)] = row
```

```
Utdanningstesten >>> Kappa = 0.497, Interval = [0.342, 0.652], TWOSIDED: Accept H1 as Kappa != 0, ONESIDED: Accept H1 Kappa > 0
Uroboros >>> Kappa = 0.081, Interval = [-0.064, 0.226], TWOSIDED: Keep H0 as Kappa ~= 0, ONESIDED: Keep H0 Kappa <= 0
BigFive >>> Kappa = 0.026, Interval = [-0.019, 0.071], TWOSIDED: Keep H0 as Kappa ~= 0, ONESIDED: Keep H0 Kappa <= 0
Truity >>> Kappa = 0.235, Interval = [0.145, 0.325], TWOSIDED: Accept H1 as Kappa != 0, ONESIDED: Accept H1 Kappa > 0
```

6.5.2 – Sample 2

Code E2.1: Function randomly splitting a data set into train and test, given a train size (fraction), random state, missing-data limit for removing columns (fraction), whether to one-hot encode, whether to impute missing data, whether to take a group (response; a columns in X) into consideration when imputing, whether to scale X (only if all numerical with no missing, or if onehot = True and imputing = mean).

```
def train_test_split_MTG(X, y, trainsize = 0.8,
                         random = True, random_state = 8841, sort_by_col = "Dato",
                         rmv_rows = None, col_rmv_lim = 0.9,
                         onehot = True, imputing = "mean", bygroup = "TYPE",
                         Xscale = "minmax",
                         int_num = True, onehot_ignore = "", do_not_split = False):

    the_df = X.copy()
    the_y = y.copy()



    if bygroup in the_df.columns:
        the_bygroup = the_df.loc[:, bygroup]
    elif bygroup.lower() == the_y.name.lower():
        the_bygroup = the_y.copy()
    else:
        raise ValueError("Variable 'bygroup' must be either the name of a col in X or the name of the y array")


    if isinstance(onehot_ignore, str):
        if onehot_ignore in the_df.columns:
            #Fjerne kolonnene fra X_train og X_test
            the_df = the_df.drop(onehot_ignore, axis = 1)
    elif isinstance(onehot_ignore, list):
        cols_to_remove = []
        for val in onehot_ignore:
            if val in the_df.columns:
                cols_to_remove.append(val)
        the_df = the_df.drop(cols_to_remove, axis = 1)


    #REMOVE COLS WITH TOO MANY MISSING
    col_all = []
    for col in list(the_df.columns):
        thecol = X[col]
        n_nans = thecol.isna().sum()

        if (n_nans / len(thecol)) > col_rmv_lim:
            col_all.append(col)

    for colname in col_all:
        print(f"Column '{colname}' was removed due to missing above {col_rmv_lim * 100}%")

    the_df = the_df.drop(col_all, axis = 1)


    #SPLIT DATA IN TEST AND TRAIN
    n_rows = int(the_df.shape[0] * trainsize)

    if random:
        X_train = the_df.sample(n = n_rows, random_state = random_state)
        X_test = the_df.drop(list(X_train.index), axis = 0)
```

```
else:
    i_train = list(range(n_rows))
    i_test = list(np.arange(n_rows, the_df.shape[0]))

    #Sort the dataframe first
    the_df = the_df.sort_values(by = sort_by_col)

    X_train = the_df.iloc[i_train, :]
    X_test  = the_df.iloc[i_test, :]

y_train = the_y.loc[X_train.index]
y_test = the_y.loc[X_test.index]


#ENSURE THE BYGROUP ALSO IS CORRECTLY SPLIT
bygroup_train = the_bygroup.loc[X_train.index]
bygroup_test  = the_bygroup.loc[X_test.index]


#ONE-HOT ENCODE THE DATA
index_train = X_train.index
index_test = X_test.index

if onehot:
    numerics = ["float", float, "float64"]
    categorics = ["object", object]

    if int_num:
        numerics += ["int", int, "int64", "int32"]
    else:
        categorics += ["int", int, "int64", "int32"]

    numeric_feat = X_train.select_dtypes(include = numerics).columns
    numeric_trans = Pipeline(steps = [("nothing_numeric", None)])

    categ_feat = X_train.select_dtypes(include = categorics).columns
    categ_trans = Pipeline(steps = [("one_hot_encoder",
                                 OneHotEncoder(handle_unknown="ignore", sparse = False))])

    colnames_after_onehot = make_colnames(X = the_df,
                                      num_types = numerics,
                                      categ_types = categorics)

    preprocessor = ColumnTransformer(
         transformers = [
             ("numeric", numeric_trans, numeric_feat),
             #("nothing", ignore_trans, ignore_feat),
             ("categoric", categ_trans, categ_feat)
         ])

    fitted_preprocessor = preprocessor.fit(the_df)

    X_train = pd.DataFrame(fitted_preprocessor.transform(X_train))
    X_train.columns = colnames_after_onehot.copy()
    X_train.index = index_train

    X_test = pd.DataFrame(fitted_preprocessor.transform(X_test))
    X_test.columns = colnames_after_onehot.copy()
    X_test.index = index_test
```

```python
#IMPUTING THE DATA
if isinstance(imputing, str):
    numerical_features = list(X_train.columns)
    X_train_index = list(X_train.index)
    X_test_index = list(X_test.index)

    if str(imputing).lower() in ["mean", "median"]:

        numerical_trans = Pipeline(steps = [
            ("imputerstr", SimpleImputer(missing_values = float("nan"), strategy = imputing)),
            ("imputernan", SimpleImputer(missing_values = np.NaN, strategy = imputing))
        ])

        preprocessor = ColumnTransformer(
            transformers = [
                ("num", numerical_trans, numerical_features)
        ])

        fitted_preprocessor = preprocessor.fit(X_train)

        X_train = pd.DataFrame(fitted_preprocessor.transform(X_train))
        X_test = pd.DataFrame(fitted_preprocessor.transform(X_test))

        X_train.columns = numerical_features.copy()
        X_test.columns = numerical_features.copy()

        X_train.index = X_train_index.copy()
        X_test.index = X_test_index.copy()

    elif str(imputing).lower() == "bygroup":

        X_train_interlude = X_train.copy()
        X_test_interlude  = X_test.copy()

        X_train_interlude["group_col"] = list(bygroup_train)
        X_test_interlude["group_col"] = list(bygroup_test)


        X_train = X_train_interlude.groupby("group_col").transform(
            lambda x: x.fillna(x.mean())
        )
        X_test = X_test_interlude.groupby("group_col").transform(
            lambda x: x.fillna(x.mean())
        )
    else:
        raise ValueError("Variable 'imputing' must be one of: ['mean', 'median', 'bygroup']")



#SCALING THE DATA
if isinstance(Xscale, str):
    numerical_features = list(X_train.columns)
    X_train_index = list(X_train.index)
    X_test_index = list(X_test.index)

    if Xscale.lower() == "minmax":
        numerical_trans = Pipeline(steps = [
            ("scaler", MinMaxScaler())
        ])

    elif Xscale.lower() == "maxmabs":
        numerical_trans = Pipeline(steps = [
            ("scaler", MaxAbsScaler())
        ])

    elif Xscale.lower() == "mean":
        numerical_trans = Pipeline(steps = [
            ("scaler", StandardScaler())
        ])
```

```
        else:
            raise ValueError("Xscale variable must be minmax, maxabs or mean")

        preprocessor = ColumnTransformer(
           transformers = [
               ("num", numerical_trans, numerical_features)
        ])

        fitted_preprocessor = preprocessor.fit(X_train)

        X_train = pd.DataFrame(fitted_preprocessor.transform(X_train))
        X_test = pd.DataFrame(fitted_preprocessor.transform(X_test))

        X_train.columns = numerical_features.copy()
        X_test.columns = numerical_features.copy()

        X_train.index = X_train_index.copy()
        X_test.index = X_test_index.copy()



    return X_train, y_train, X_test, y_test
```

Code E2.2: Variables used when splitting data with the function from Code E1. This resulted in the train and test data used during the PCA and LDA processes, of which the test-data was transformed and used to model on later. The classification models having PCA/LDA components as explanatory variables were based on this transformed test-data (that had been randomly reduced so that there was an equal amount of rows per 4-letter type).

```
X_train, y_train, X_test, y_test = train_test_split_MTG(X = df_onehigh.drop(["Del5_1", "Del5_2", "Fylke"], axis = 1),
                                                        y = df_onehigh.TYPE,
                                                        trainsize = 0.6, random = True,
                                                        onehot = True, onehot_ignore = "MM-YYYY",
                                                        #do_not_split = True,
                                                        imputing = "bygroup", bygroup = "TYPE",
                                                        Xscale = False)  #can be None or False
```

Code E2.3: TYPE-dependent splitting on data that are otherwise preprocessed. Also shows how it was used. The “df_PCA” here is the same as the “X_test” from above (after correlation-scores transformed using PCA with component 1-14)

```
def train_test_bygroup(df, groupname, random_state, testsize = 0.33):
    the_df = df.copy() #To ensure no overwrites

    smallest_group_n  = the_df.loc[:, groupname].value_counts()[-1]
    n_per_group_test  = int( testsize * smallest_group_n )
    n_per_group_train = smallest_group_n - n_per_group_test

    unique_groups = list(set(the_df.loc[:, groupname]))

    test_indexer = []
    for group in unique_groups:
        df_subset = the_df.loc[the_df[groupname] == group, :]

        testsample = df_subset.sample(n = n_per_group_test,
                                      random_state = random_state)

        test_indexer += list(testsample.index)  #Add indexer to total index

    #Define test and, by proxy, the train
    X_test = the_df.loc[test_indexer, :]
    X_train = the_df.drop(test_indexer, axis = 0)

    #But, train is not yet scaled to have similar amount of each group
    train_indexer = []
    for group in unique_groups:
        df_subset = X_train.loc[X_train[groupname] == group, :]

        trainsample = df_subset.sample(n = n_per_group_train,
                                       random_state = random_state)

        train_indexer += list(trainsample.index)

    X_train = X_train.loc[train_indexer, :]

    #Shuffle the rows using sklearn.utils.shuffle
    X_train = shuffle(X_train, random_state = random_state)
    X_test  = shuffle(X_test, random_state = random_state)

    return X_train, X_test
```

```
train_PCA, test_PCA = train_test_bygroup(df_PCA, "TYPE",
                                         random_state = rs,
                                         testsize = 0.33)

#########
train_LDA_TYPE = df_LDA_TYPE.loc[list(train_PCA.iloc[:, 1]), :]
test_LDA_TYPE = df_LDA_TYPE.loc[list(test_PCA.iloc[:, 1]), :]

#########
train_LDA_Func = df_LDA_Func.loc[list(train_PCA.iloc[:, 1]), :]
test_LDA_Func = df_LDA_Func.loc[list(test_PCA.iloc[:, 1]), :]
```

Code E2.4: Machine Learning K-Fold Cross-Validation, hyperparameter-tuning using grid search, and subsequent model fitting

```
clfs = []

clfs.append(("LogReg",
             Pipeline([("Scaler", StandardScaler()),
                       ("LogReg", LogisticRegression(n_jobs = -1,
                                                     random_state = rs,
                                                     multi_class = "ovr"))
                     ])))
clfs.append(("LogReg_unsc",
             Pipeline([("LogReg", LogisticRegression(n_jobs = -1,
                                                     random_state = rs,
                                                     multi_class = "ovr"))
                     ])))

clfs.append(("SVCrbf",
             Pipeline([("Scaler", StandardScaler()),
                       ("SVC", SVC(kernel = "rbf"))
                     ])))

clfs.append(("SVCrbf_unsc",
             Pipeline([("SVC", SVC(kernel = "rbf"))
                     ])))

clfs.append(("XGBoost",
             Pipeline([("Scaler", StandardScaler()),
                       ("XGB", xgb.XGBClassifier())
                     ])))

clfs.append(("XGBoost_unsc",
             Pipeline([("XGB", xgb.XGBClassifier())
                     ])))
```

```
the_X_train = df_PCA.iloc[:, [1,2,3,4,5,6,7,8,9,10,11,12,13,14]]
the_y_train = df_PCA.loc[list(the_X_train.index), "TYPE"]

scoring = "f1_micro"
n_folds = 3
results, names = [], []

for name, model in clfs:
    kfold = KFold(n_splits = n_folds)
    cv_results = cross_val_score(model,
                                 the_X_train, the_y_train,
                                 cv = n_folds, n_jobs = -1,
                                 scoring = scoring)
    names.append(name)
    results.append(cv_results)
```

```
y_PCA_Function = train_PCA.loc[:, "Function_pair"]
y_PCA_TYPE     = train_PCA.loc[:, "TYPE"]

x_PCA          = train_PCA.loc[:, ["PC1", "PC2", "PC3", "PC4", "PC5",
                                   "PC6", "PC7", "PC8", "PC9", "PC10",
                                   "PC11", "PC12", "PC13", "PC14"]
                              ]
```

```
pipe_logreg_PCA = make_pipeline(StandardScaler(),
                                LogisticRegression(n_jobs = -1,
                                                   random_state = rs,
                                                   multi_class = "ovr")
                               )

param_penalty = ["l1", "l2"]
param_C = [0.001, 0.01, 0.1, 1, 10]

param_grid = [{"logisticregression__C": param_C,
               "logisticregression__penalty": param_penalty
              }]

gs = GridSearchCV(estimator = pipe_logreg_PCA,
                  param_grid = param_grid,
                  scoring = "f1_micro",
                  cv = 5, refit = True, n_jobs = -1
                 )
```

```
gsFunc_PCA = gs.fit(x_PCA, y_PCA_Function)
```

```
clf_logreg_PCA_Func = make_pipeline(StandardScaler(),
                               LogisticRegression(random_state = rs,
                                                  n_jobs = -1,
                                                  multi_class = "ovr",
                                                  C = 0.01,
                                                  penalty = "l2"
                                                 )
                              )
```

```
clf_logreg_PCA_Func.fit(x_PCA, y_PCA_Function)
```

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway