Performance and diversity in self-managed work teams

MARC LAUNER

PERFORMANCE AND DIVERSITY IN SELF-MANAGED WORK TEAMS

THESIS

DOCTORATE OF BUSINESS ADMINISTRATION (DBA) PROGRAM

2014/15, DURHAM BUSINESS SCHOOL, DURHAM UNIVERSITY, UK

AUTHOR: MARC LAUNER

ANONYMOUS MARKING CODE: Z0958386

DATE: APRIL 23, 2021

NUMBER OF PAGES: 266

WORDCOUNT: 45.600

VERSION: 1.1

I CONFIRM THAT THIS PIECE OF WORK IS THE RESULT OF MY OWN WORK. MATERIALS FROM THE WORK OF OTHERS HAVE BEEN ACKNOWLEDGED, AND QUOTATIONS AND PARAPHRASES SUITABLY INDICATED.

1 DEDICATIONS

To my wife Tanja.

For her love, her advice, her patience, her enthusiasm, and her endless positivity.

Without you this would have never happened.

To all the scholars, advisors, and experts providing time to me over the last six years this project lasted.

Who showed me places and knowledge beyond my horizons.

My world would be smaller, less colorful, and less exciting without having traveled to them.

To Prof. Dr.-Ing. Wolfgang Kubbat and Dr.-Ing. Jens Schiefele.

For inspiring and encouraging me to embark on a journey of life-long learning.

2 CONFIDENTIALITY AND RESTRICTED NOTE

This paper contains Jeppesen (JEPP) and The Boeing Company (TBC) proprietary and confidential information. No part of this paper may be copied, stored, distributed, or shared without prior written permission by Jeppesen/Boeing.

Jeppesen/Boeing grants Durham Business School and Durham University academic staff the right to copy, store and read the paper only for grading and evaluation purposes within the DBA program 2014/15.

3 VERSIONING

Version	Date	Comment
1.0	8th December 2020	Version submitted for evaluation
1.1	23th April 2021	Version with improvements resulting from examination report

4 ABSTRACT

Diversity in Self-Managed Work Teams (SMWTs) performing modern workplace software and service development in Agile multi-national distributed organizations, has become a new norm for many organizations and enterprises. Much research has been done regarding the effects of surface-level and deep-level diversity on team performance and efficiency. In recent years work has been done to conceptualize the different aspects of team performance and its contributing factors, such as the inner/social processes in teams, the output related performance and the matching feedback models. This study investigated multi-national research & development teams in a global aerospace enterprise (The Boeing Company, TBC). It modeled different forms of team performance (self-set goals, external-set goals performance, and member interactions peer performance insights) derived from recently conceptualized models and merges them with surface-level and deep-level diversity to find their contributions to the three levels of team performance. In addition, it investigates the same input diversity factors and the performance levels with team process ratings based on peer evaluations. It strives to reflect the vast variety of existing input parameter based research on team performance and reflecting it over such modern Agile software and services teams. It provides recommendations to practitioners on how to capitalize on these findings. The study confirms many previous findings regarding diversity factors as input parameters. It found additional diversity factors likely contributing to team performance, like native language diversity and education level diversity. It also found contributing factors from the culture dimension models — a field not extensively researched to date. Positive effects from higher levels of risk, directness and independence were found. New knowledge was created by relating the individual effects to the operationalized three levels of team performance. The study found that known input factors and newly found factors are not always contributing to all performance levels at the same level, or even at all to the modern forms of SMWTs. The findings also hint to practitioners that learnings and knowledge from more recent conceptualizations, that focus more on the dynamics and processes within a team, such as the Contingent Configuration Approach (CCA) or the Categorization-Elaboration Model (CEM) need much more attention when using frameworks such as Agile or Pragmatic Marketing. Just focusing on mere input parameters and maximizing there for output may underutilize the concepts existing. Recommendation are developed for managers and practitioners for such

teams based on the empirical findings and the reviewed and conceptualized models of team performance and diversity.

5 TABLE OF CONTENTS

6 TABLE OF FIGURES

7 TABLE OF TABLES

8 INTRODUCTION AND PROBLEM STATEMENT

Teams are a fundamental building block in modern work environments and enterprises. Mathieu et al. (2017) summarize: “Work groups are a vital link between individuals and organizations”. Therefore, it is not surprising that many scholars have been attracted to the problem space of teams since the 1970s, when the concept of work groups and teams started to take over the workspace. Teams are used extensively by industry, as they establish the standard on how creative work is organized and will be the standard for a long time to come. As early as 1997, Cohen and Bailey (1997 p.239) reported: “The use of teams has expanded dramatically in response to competitive challenges. For example, 82% of companies with 100 or more employees reported that they use teams. Sixty-eight percent of Fortune 1000 companies reported that they used self-managing work teams and 91% reported that they used employee participation groups in 1993 compared to 28% and 70% respectively in 1987”. Driven by the efficiency hunts of the automobile industry in the 1980s, teams took over the regime at the conveyor belts and much consideration was given on how to organize such an environment. Substantial work was done by Hackman (1976, 1986) in these years to describe a version of teams that took parts of management and leadership internal to the team and shared responsibility and accountability amongst the team. The term Self-Managed Work Team (SMWT) was introduced. The nature of work has significantly changed and as such, teams is an extensive research field. In recent years the concept of SMWTs has seen a revival, driven by a concept in software development initiated by the Agile Alliance Group (2016): Scaled Agile. Scaled Agile is a concept for how to develop software efficiently through frequent iteration. It also very strongly relies on the concept of the team controlling most processes and being led from within. With the rise of Scaled Agile, the learnings of SMWT are back on the table and more important than ever.

Organizational adaptability and capability to change have become dramatically more important in the aerospace industry in the wake of the COVID 19 crisis, during which this study was concluded. While the author has always been a strong proponent of change and adaptability, many others do not agree with such a viewpoint and favor process stability and structural soundness as the foundation of an organization and its success. Such an identity of an organization is also influenced by how diversity is dealt with in that organization, Brickson (2000) finds. As one learning from the Doctorate in Business Administration

(DBA) studies and involvement with scholars in the macroeconomic research fields, I will take this opportunity to summarize Hannan, Polos, Carroll, and Freeman (2007; 1984), as they provide some insights into this paradox. They claim that regrouping within organizations increases their mortality and risk of failing. At the same time, the environment in which organizations find themselves is constantly changing and that subsequently requires an organization to be able to change. Change ability requires innovation, and innovation requires regrouping. Hence change is not optional. It is risky but necessary. This study will hopefully provide insights into how to increase change ability through diverse teams and organizational forms that allow for stability and changeability at the same time.

Given the much-increased diversity due to e.g. globalization, generational differences and also the much needed and appreciated entrance of more female workers into engineering and software related workplaces, a more specific challenge, first described by Yves Guillaume et al. (2013), moves into the center of attention for the modern software development manager: "Getting diversity at work to work …". Addressing these problem spaces is the new challenge for managers and companies in the software related industry. Not surprisingly, the question of: "What makes teams work?" is back in focus. That is true for how to deal with diversity and manage it for SMWT teams as well as teams in general in organizations. Lawler (1978), within his wide variety of publications on work teams, comes to the conclusion that a significant proportion of teams fail because there is not enough attention to team composition and design. The same view is shared by Coutu (2009) in dialog with Hackman (1998), claiming the "mix" in teams is important to structure, which, when missing, is one of the top five reasons why teams don't work.

Most existing theory and literature around teamwork and team performance does not differentiate between specific types of teams, like SMWTs. It treats teams as a general structure and compares teamwork operations and types with each other. While this possible and certainly leads to results, a more desirable and perhaps more valid way is to validate the potential of SMWTs in a software product development context. This is the method of the current study. It analyses team performance theory and provides insights on SMWT teams through data analysis. Some argue that team design in the field of SMWT has a strong influence on their success. Wagemann (1997) reasons a relationship exists between team design and team effectiveness. Erez et al. (2002) assessed performance of SWMT teams

through peer evaluations and found that more than half of the differences in team performance can be explained by team design and composition.

Most research in many years before that dwell on team diversity and related performance is ambiguous and shows mixed, and in some cases controversial, results. As for example Millikens and Martin (1996) show and elaborate on. Following the well-established and widely used Input-Process-Outcome model attributed to McGarth (1964) one of the three major inputs into the equation of team and performance are Organizational, Team and Individual aspects. The study here focuses mainly on Team and Individual aspects while trying to reflect its results later to practical aspects of the organization. Diversity in studies defining Team and Individual input parameters often use the two most widely used aspects of team composition which are comprised of surface-level diversity and deep-level diversity aspects. Like LePine, Buckmann et al.(2011), LePine, Hanson et al. (2000), Driskell, Hogan and Sallas (1987), or Bell (2007) and many more. Surface-level diversity is commonly referred to as demographic differences. Demographic differences include age, gender, residence and spoken language. Deep-level diversity amongst team members can be divided into two types: psychological or character traits and culture traits. Culture traits like defined by Hofstede et al. (1984). In recent years, scholars have started to investigate the intrapersonal and social mechanisms within teams which potentially contribute to team performance. This includes the categorization-elaboration model (CEM) (Van Knippenberg, De Dreu & Homan 2004) providing an integrated approach to combine mechanisms within teams and groups based on input diversity to determine the performance of teams. More specific related to the personality aspects for team members and their integration and mediation towards team performance is covered in the contingent configuration approach from (Moynihan & Peterson 2001). In recent years, more detailed models of team performance and diversity as direct improvements to the IPO model have emerged, such as the Input-Moderation-Output-Input (IMOI) model developed by Ilgen et al. (2005). This study will investigate those frameworks and locate the theoretical approach taken for it. The methods for measuring team performance varies greatly between studies, and include indirect measures, subjective measures, self-declared measures, quality of inner processes and many kinds of direct measures. Some studies use a combination of methods. The lack of a standardized measurement of team performance makes it difficult to compare results across studies. Most research work follows the fundament approach postulated by Hackman (1987)

when looking at team performance. Hackman uses Task performance, Member Satisfaction with group, and Team viability. While using the same three aspects of team performance the study investigates into how in particular Task performance is operationalized in many studies and how the studies own version relates to these.

The goals of the study is to investigate a practical use case of diversity, using the wide variety of IPO based research, in self-managed software product development teams and analyzed diversity factors for their contribution to team performance in modern terms. The study's hypothesis is that certain aspects of team diversity in SMWTs contribute to team performance in positive or negative ways.

By combining the recent developments in models for team performance and the existing research in team diversity factors, this study contributes new insights for SMWTs their diversity input parameters for teams their composition and individual forming these teams. A matching unique data set reflecting three levels of performance measures, a full set of surface-level and deep-level diversify data and a comprehensive metadata set were analyzed to conduct an empirical analysis. In doing so, this study provides new learnings about team diversity factors and team. This thesis also derives new practical implications and recommendations for managers and leaders working with SMWTs every day.

Once these aspects are identified, the study will provide guidance on how to manage and implement practical implication to them. The team composition (SMWT in software) is important when interpreting findings and recommendations, because, as Joshi and Roh (2009) state: "context matters in team diversity research". The same conclusion is drawn by Hollenbeck et al. (2012) when investigating 42 different types of teams in their meta study. This study provides practical recommendation for the given environment and general recommendations for how to compose diverse teams successfully. This is especially important when reflecting on the nature of the problems that managers deal with in teams and their performance. Managers in a real-world organization are limited in choices and degrees of freedom when composing teams. This is in contrast to many basic assumptions in theory and research. Managers usually are responsible for a specific pool of team members, representing a subset of a larger organization. Adding or subtracting to that pool involves time-consuming processes, which in practice usually comes down to hiring and firing. As such, common theory's definition of what the ideal diverse team looks like can be difficult or

impossible to implement. This study seeks investigating additional, relative options, such as maximizing diversity in one dimension or minimizing in another.

The paper is structured into four major sections. The first section contains an analysis of existing literature on group diversity and team performance and its various aspects, such as performance measures, deep-level and surface-level diversity, and self-managed work teams and diversity. The second section describes the methods used to collect, measure, manage and process data. It also includes the statistical analysis of diversity factors related to team performance. The results and findings from the previous chapters are discussed in the third section. The study concludes with practical recommendations from the findings and feedback to how such findings can influence management of SMWT software teams.

9 DIVERSITY AND ITS INFLUENCE ON TEAM EFFECTIVITY

This chapter is divided into two major sections. First sections will investigate and review into existing models and approaches on how diversity in teams is linked to performance of teams. This review fulfills two purposes. One is the defining the structure and theoretical framing for the empirical analysis of the case at hand. Further it will provide the framework for the second section of the chapter which reviews the existing literature regarding diversity aspects and team performance and its interactions to derive hypotheses to be tested later in empirical analysis.

9.1 CONCEPTUALIZING OF DIVERSITY IN TEAMS AND THEIR PERFORMANCE

This first section is reviewing existing work on theoretical frameworks regarding teams and their diversity in regards to their performance. As van Knippenberg et al. (2004) summarize on basis of Triandis (1994), Jackson (1992), and William & O'Reilly (1998) : "Diversity [in teams] refers to differences between individuals and any attribute that may lead to the perception that another person is different from self". Diversity in teams can be a broad topic consisting of a wealth of aspects. Earlier we explored the reasons why diversity increases and plays a bigger role in regards to Agile software teams in today's world. But in general scholars agree that teams " … will continue to become more diverse in years to come." Van Knippenberg & Schippers (2007). Van Knippenberg et al. (2004) also conclude that the main focus till the mid 2000s has been on attributes which comprise surface-level diversity. Such attributes being things like gender, age, race/ethnicity, education, functional background etc. In the 1990s and 2000s a growing number of studies also brought on the personality of team members and tried to link performance of teams to the individuals and group composition of personality. E.g. Barrick (1998), Mohammed & Angell (2003), or Neuman & Wright (1999). Attributes that according to van Knippenberg & Schippers (2007): " …may not be readily visible but are not always job-related either, such as differences in personality/attitude and values". Such attributes are normally summarized as deep-level diversity. Some scholars also started to integrate surface-level and deep-level diversity and linking them to team performance. E.g. Mohammed & Angell (2004). Most of such research is based on a well-known, and somewhat computational science based, model attributed to McGarth (1964) for

its application to social science. The model of Input – Process – Output (IPO). One of the first researchers investigating what to measure when evaluating team performance was Hackman (1987) in the 1980s. Hackman focused on the three areas postulated by McGarth. The first is the "productive outcome" and how that matches the expectations from the people "who receive and/or review the output". The second is the "group experience" and that it should "satisfy rather than frustrate". Third, he put learning experience and feedback to improve future work into focus. These "should maintain or enhance the capabilities of members to work together on subsequent team tasks". The three areas became the foundation of several concepts of team performance for the following decades

Daniel Ilgen and his co-authors (2005) review this subject in more depth. They concluded that most of the research had moved away from simple Input-Process-Output (IPO) models on to more mathematically defined, model-driven methods and concepts. This models the three performance measures into a structured model linking the processes and interactions. They introduced the IMOI (Input-Mediator-Output-Input) model. Mathieu et al. (2008) contributed a comprehensive set of graphics for the two models, which are depicted in Figure 1 and Figure 2.

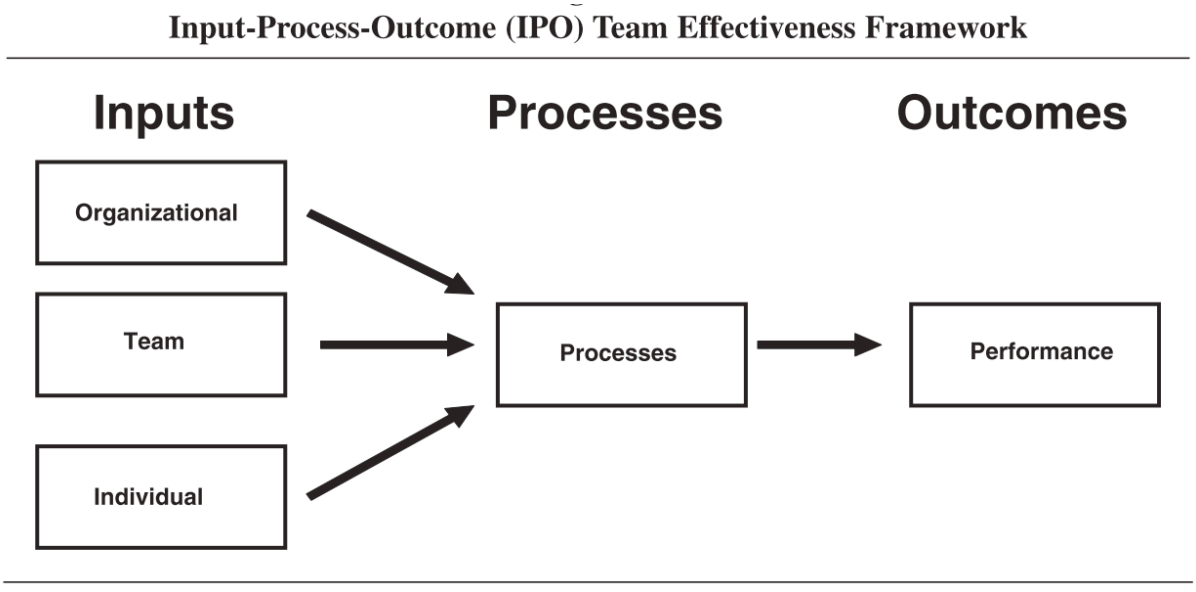


Figure 1 IPO concept.- Mathieu et al. (2008 p.413)

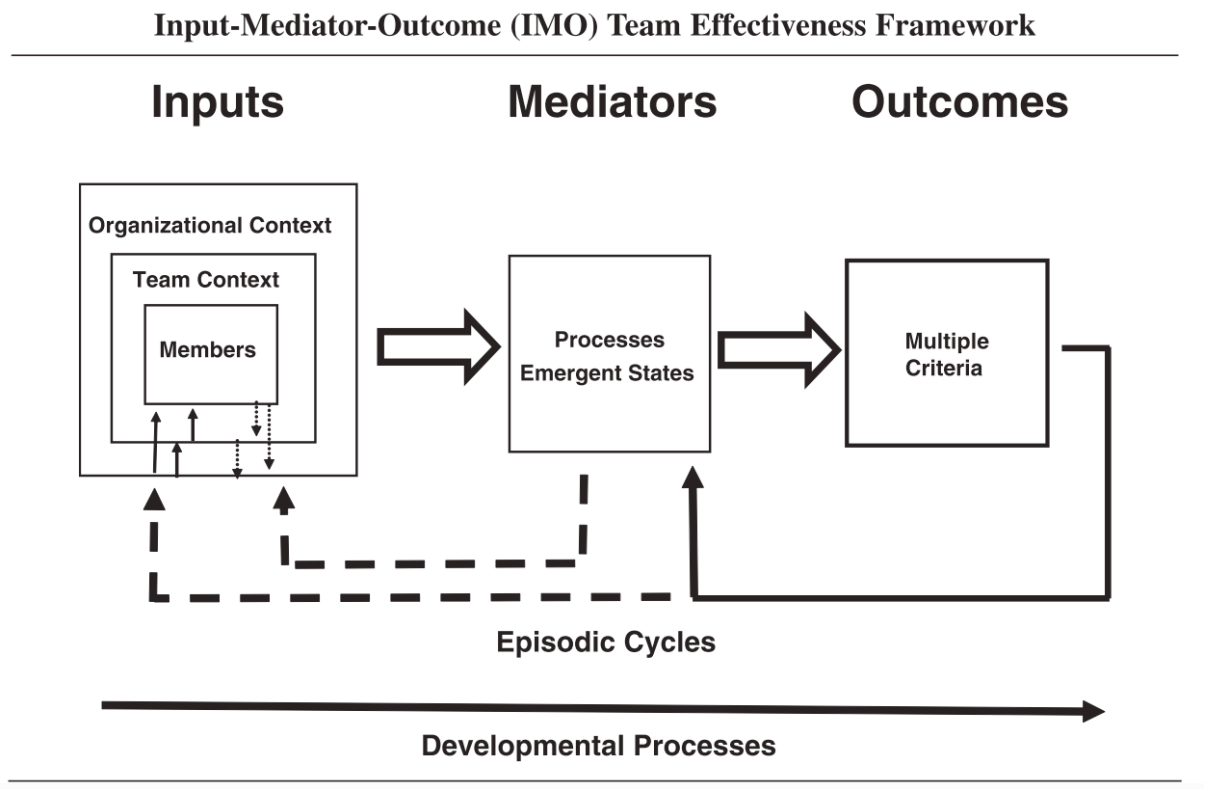


Figure 2 IMOI concept. Mathieu et al. (2008 p.413)

The major three differences between the models are that the IMOI model reflects the internal processes and their effects, assumes a feedback cycle from output to internal processes of a team and to the inputs of a team, and includes the concept of time and longevity of a team process. All three performance measure concepts described earlier are also represented here systematically. Generally, the model from Ilgen et al. is referred to as the IMOI 3x3 model as it considers three internal phases/processes of teams: forming (I-M), functioning (M-O), and finishing (O-I), as well as three aspects within each phase affect, behavior, cognition. The IMOI 3x3 concept also merges simple IPO models with temporal focused work on teams, such as that from Tuckman et al. (1965), who developed the Forming, Storming, Norming, Performing concept. Later, Tuckman together with Jensen (1977), added a fifth phase of Adjourning which fits the temporal assumption in Ilgen's work. A similar concept is the GROW model developed by Fine (2009), which is widely used at TBC. It defines phases as Goals, Reality, Options, and Way forward and assumes the Way Forward as being connected to a revised Goals phase. This later model also fits the IMOI structure conceptually.

As we can see both the basic IPO and the IMOI models include “inner mechanism” (processes or mediators) that are not widely covered by the diversity research summarized earlier. Over the last three decades a branch of research developed that started to look into these “inner mechanisms”. Yet again William & O’Reilly (1998) define two major streams of research efforts to investigate the mechanisms. One being the social categorization perspective and the other the informational/functional perspective. Van Knippenberg et al. (2004) and again in collaboration with Schippers (2007) provide comprehensive summaries of the characteristics and conclusions so far in the two fields. The social categorization perspective can be summarized into the concept that more readily available attributes, as the ones described above as surface-level, are used to categorize team members themselves into groups. That leads to the concept of in-groups and out-group perceptions of self’s and groups themselves. Hence categorization. The concept generally concludes that a more homogenous group, along the attributes at hand, is better performing than a heterogeneous group. The other school of thought, the information/decision making perspective, conversely concludes the contrary, such that heterogeneous groups outperform the homogeneous groups. The thinking is that based on the broader knowledge and skills, members contribute better to the problem of solving the task at hand. It is generally thought that such process leads to better reviewed and reflected solutions and avoids consensus group bias that is more likely to be present in homogenous groups. Compare e.g. Ancona & Caldwell (1992) or Bantel & Jackson (1989). Overall the situation appeared inconclusive and contradictory to van Knippenberg et al. (2004). Therefore they proposed a model dubbed the Categorization-Elaboration Model (CEM) to relate the two streams of research and integrate both into one conclusive model. Among other proposed effects, the model focused to include interaction/moderation between the social categorization and the information/decision approach. The main mechanism introduced here is elaboration. Elaboration being the process of how groups interact and process information and perspectives and ultimately integrate many group aspects and information into a consolidated outcome. As such elaboration defines, in layman’s terms, the “super-process” that combines the two other concepts. Further the processes are segmented into more detailed steps and effects. An overview can be found in Figure 3.

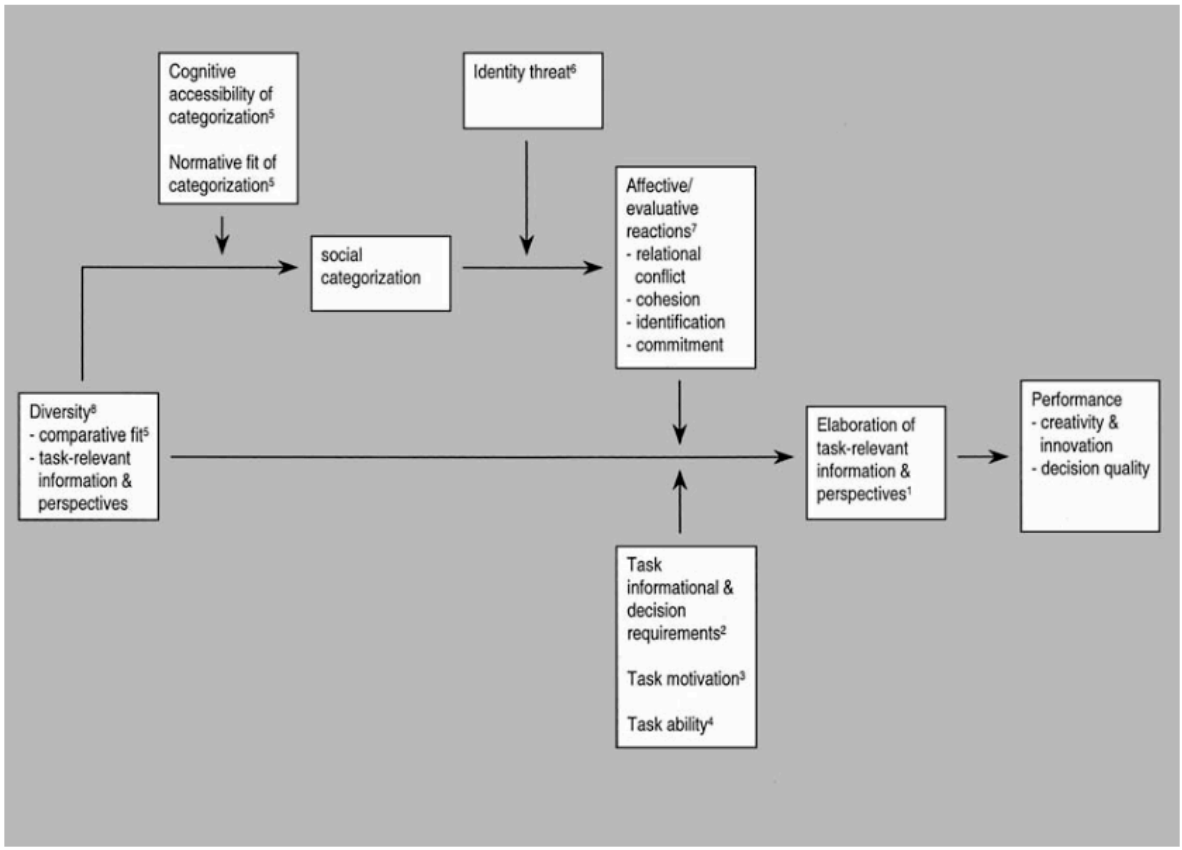


Figure 3 Categorization-elaboration model. Van Knippenberg et al. (2004)

Postulating several implications to future research van Knippenberg et al. in summary indicate that it is advisable to include measures of elaboration in study designs. Also it is implied that task motivation and task ability are playing significant roles in affecting elaboration. Hence such design elements are desirable. Inter group biases and social categorization processes are also influencing the abilities of teams to output performance. The last proposition is to include as many dimensions of diversity into study designs as possible since according to one proposition by van Knippenberg et al. all dimensions possibly affect elaboration as well as social categorization effects.

Reflecting on the study here several aspects need to be evaluated and acknowledged before deciding on a conceptual framework guiding the further work of this study. Being part of an educational part time program lasting several years, and supported by and integrated with the authors work place and enterprise, this study is mainly based on problems located within the organization at question. As we will see in chapter 10.1 data and access to data is predefined for this study since about 2011. While the organization functioning as the case to be analyzed

here and subsequently the data that has been collected and being available being designed by the author, it must be recognized that design of data setup and collection dates back to 2009. The nature of ever changing team constellations (compare Figure 9) does not allow to add any major data collections or insights specific to unique projects. Nor is it practically possible to change or add to data about individual team members due to the nature of a commercial organization and its natural attrition rates over the course of nearly 10 years. As such, the study is bound to the data and subsequent insights available and collected over the last years. Logically, any theoretical framework and conceptualizations chosen to analyze must fit these constraints. A constraint that any other study dealing with historically grown data frames and data designs has to deal with. Thus a specific and common problem of real world case studies and real world organizations insights in contrast to experiments or other environments, where context and setup can be influenced, designed and experimentally manipulated. Specifically the studies data is sharing a limitation that it according to van Knippenberg and Schippers (2007) shares with many studies performed in the past and in the future: ".. due in part to the fact that many studies did not include process measures.". We will see in chapter 10.1 that data available and collected for this study is not including many process measures, with one small none systematic exception as we will discuss in the performance measure section. For that reason the conceptual framework of this study needs to be a more classic one and more focused on theoretical frameworks mainly based on input parameters in the IPO and IMOI frameworks. Still a very active filed with quite recent additions like Curral et al. (2001) and West & Hirst (2003). Analysis will be based on an extensive list of surface-level diversity parameters and deep-level diversity parameters covering personality and culture background aspects. These parameters will build the Input parameters (compare Figure 1) to the IPO/IMOI models (McGarth 1964). More specifically the attributes about *Individuals* as Input parameters. Plus, the relation of such parameters amongst each other, their variety/diversity and the configuration of teams. *Team* factors are conceptually described as Self-Managed work teams (SMWT). Further details about specific designs used and general aspects of diversity in SMWT can be found in chapter 9.2.1. *Organizational* factors such as design features and environmental complexity (Mathieu, Maynard, Rapp, *et al.* 2008) can be found in chapter 10.1.1.1. Also the study systematically collects externally rated performance for teams and well as team set goals as performance ratings. Both measures reflect the *Outcome*. Also assessed in the form of round robin data amongst team members is certain aspects of how interaction within the team is perceived. Mathieu et al. (2000) summarize into

two section for outcome for IPO. One being the *performance* (as such quality and quantity) being reflected in this study as the external and team based performance and second the *affective reactions* (such as satisfaction, commitment, viability) being reflected to some extent in this study by results from Peer reviews. Detailed discussions of the data can be found in chapter 10.1.2.1. As mentioned earlier no extensive data about *Process* is collected to fit conceptual models as this is true with many other studies. Nevertheless the further analysis and results discussion will make use of the Peer review data to reflect models like CEM for the impact and practical recommendations. Figure 4**Fehler! Verweisquelle konnte nicht gefunden werden.** summarizes the studies design and data in relation to the IPO framework. Further in chapter 9.2.4.1 it is detailed how the study operationalizes and conceptualizes in respect to the field of personality and team performance. In particular the contingent configuration approach by Moynihan & Randall (2001) is reviewed. The data available in this study is allowing for testing of moderation in the concept of configuration regarding personality in teams by using the Peer review data. Peer review data comes close to the desired measurements of cohesion, communication, and motivation to work together according to Moynihan & Randall (2004 p. 346). Mathisen, et al. (2008) also use a moderation construct following the IPO structure when measuring relations between teams personality traits and creativity/innovation. They use the Team Climate Inventory (TCI) as a basis for operationalizing Process aspects. TCI was developed by Anderson & West (1994, 1998). The Peer review data in this study is partially reflecting the TCI dimensions. In particular some aspects of participant safety and task orientation. Compare Brodbeck (2000). While the Peer data does not relate formally to the TCI, nor can it replace such construct, it is further suggested to use the Peer data in two ways. First as an statistical Output (Performance measure) and further on as a possible Mediator (Process measure). Same is suggested by Homan et al. (2008) when dealing with personality aspects of team performance. This is at least in part addressing common critique around IPO study designs neglecting inner processes of teams. Compare LePine (2011 p.313).

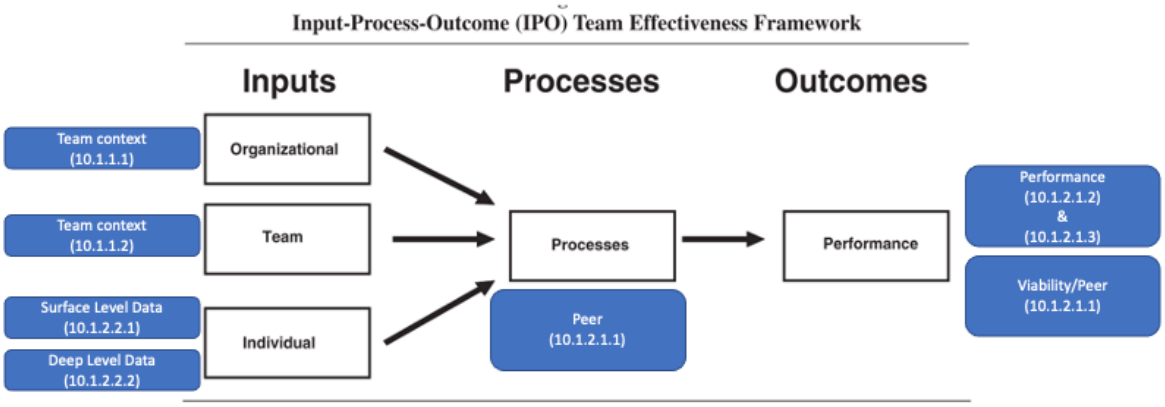


Figure 4 Data in relation to conceptual framework

While the conceptual framing of the study is rather traditional, the study offers a very broad variety of diversity dimension. This might be perceived as a question too broad to be addressed but as other researcher quite recently attempted to integrate a large variety of surface-level and deep-level attributes, like Mohammed and Angell (2004), there is great value in analyzing a broad and rather complete set of dimension such as this study has available. Also the modern aspects of Agile and more software and multi-location teams in combination with a more classic concept will improve understanding of influences of diversity attributes and provide valuable input in similar studies using equally or more narrow sets of attributes while using methods and concepts like CEM.

9.2 DIVERSITY IN TEAMS

At the onset of team diversity research, most studies focused on the direct influence of individual diversity factors on team performance. Diversity, as described by Josh et al. (2003 p.802), "… refer[s] to the distribution of personal attributes among interdependent members of a work unit." This, as Iglen (1999) shows in a meta-analysis, led to a focus on demography and other more readily available variables, such as group composition such as in Levine et al. (2006). As the research field expanded, researchers started exploring the mechanisms that describe and model the effects inside of teams and how team members interact and why they do what and in what way (social aspects). The goal here, as described by Guillaume et al. (2012; 2013), is to model the effects of diversity in teams so as to actively manage it and predict and maximize team output and efficiency. The hunt for individual factors and the model/interaction driven approach continues to attract scholars to this day, as evidenced by

Mathieu et al. (2017), who found more than 100 pieces of team diversity research published in the Journal of Applied Psychology in only ten years 2007-2017.

Ilgen (1999) asserts that no general assumption can be made about whether or not diversity is good or bad for teams and their performance. As we have seen earlier this is also reflected by the controversial finding around social categorization and Information/Decision approaches. In his article he describes how too much diversity can make it difficult to communicate, while too little diversity can hinder creativity when only a homogeneous skill set dominates the team. Later, van Kippenberg et al. (2004) develop the CEM model as we have discussed earlier. Highlighting on the mediating effects and interaction between the mediating effects. Focusing more on how to manage and work with diversity Guillaume et al. (2013 p.126) investigated more detailed classification of factors to better understand their influence. They summarize the different factors into four categories:

- *Simple demographics*: How does the demographic background of a person affect their work outcomes?
- *Relational demographics*: How does being demographically dissimilar from peers affect a person's work outcomes?
- *Work group diversity*: How does work group diversity affect work outcomes?
- *Diversity management*: How does diversity management affect work outcomes?

This study contributes to the work group diversity category of research.

Mathieu et al. (2017) provide a comprehensive overview of how the complex aspects of human relationships and interactions influence the way teams work. Their abstract schematic

is presented in Figure 5.

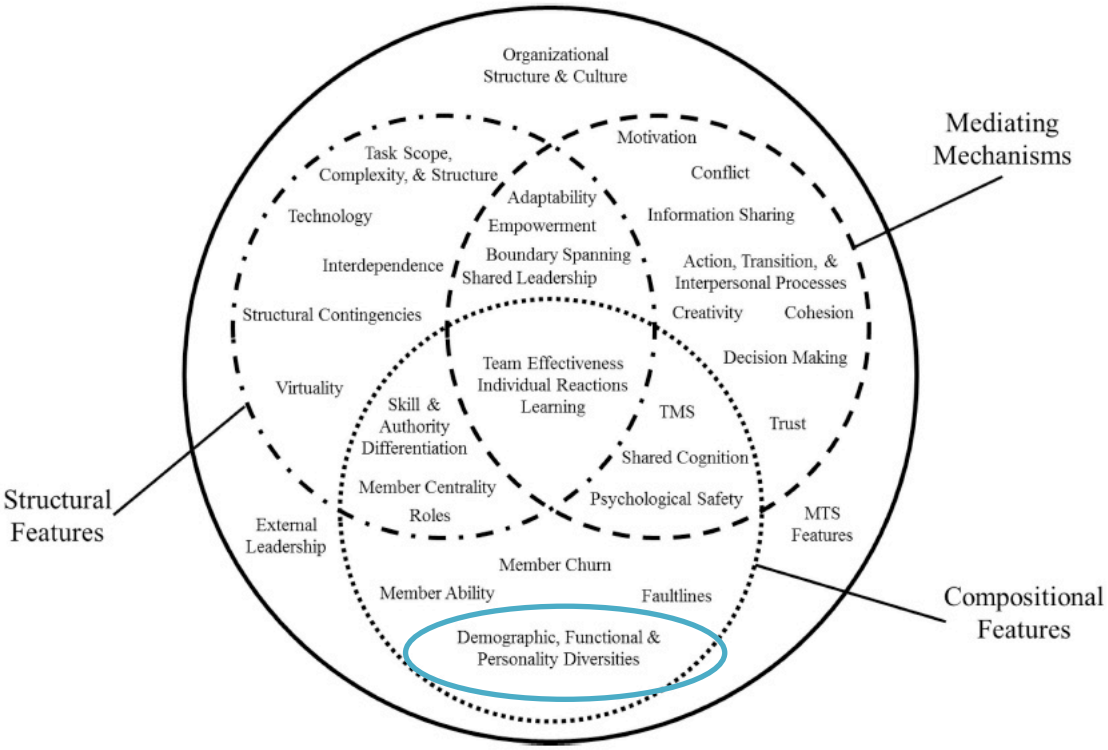


Figure 5 Model of contributing factors to team research by Mathieu et al (2017)

The data collected in this study, highlighted by the blue oval in Figure 5, covers the compositional features of demographic, functional & personality diversity. As Mathieu et al. (2008) and also Bell (2007) claim and document in their meta-studies, the aspects for personality and culture of team related research a is a relatively new field. Publications relating to this field started in the mid-1990s. The bulk of the literature can be found in the early 2010s, and in recent years, the many scholars have shifted focus to structural features, modeling approaches and mediating mechanisms.

Diversity in teams describes the differences along a variety of attributes that individual team members bring to a team. Typically, these diversity aspects are grouped into demographic aspects like age, gender, etc. described as surface-level diversity, aspect like personality described as deep-level diversity, and functional aspects like education and subject matter expertise. Sometimes the combination of these aspects is referred to as team composition. When designing a study targeting team performance and diversity, certain questions need to be answered. First, how are existing studies measuring team performance, and how do these methods enable a cross-study comparison of results? Second, how do results differ across

studies using different performance measurements and performance concepts? Additionally, what samples of teams are used by previous studies and how do these samples influence a comparison of results?

The following section also includes reviews and discussions of four more topics important to designing a study on team diversity and performance. First, a short review is provided on diversity measures outside of the classic team composition described earlier. This section covers concepts like General Mental Ability (GMA). Second, the influence of time to team diversity and performance is discussed as this impacts the study's design and results. Next, a review of how previous studies have measured and operationalized diversity is provided. Lastly, a literature review for the special field of SMWT and diversity is given.

9.2.1 SELF-MANAGED WORK TEAMS (SMWT) AND DIVERSITY

Hackman (1986) classifies teams as "units." In his authority matrix (see Figure 6), Hackman distinguishes between the responsibilities of management and the responsibilities of the team (unit). He sets four layers of responsibility levels when it comes to management of work in teams:

- Setting overall direction
- Designing the performing unit and its context
- Monitoring and managing work process
- Executing the task

Figure 6 Authority Matrix adapted from Hackman (1986)

Based on who is responsible for the above listed responsibilities, management, or the team (unit), he identifies four classifications for teams:

- Manager-led unit
- Self-managing unit
- Self-designing unit
- Self-governing unit

According to the Hackman's (1986 p.92) definition, this study investigates a self-managing unit, in which "…members have responsibility not only for executing the task but also for monitoring and managing their own performance."

Another differentiation of types of teams is offered by Katzenbach and Smith (2011) .They try to answer the question of the difference between working groups and teams. They assume that not all groups are teams, counter to the assumption found in most other team research. They differentiate the two using a variety of characteristics, summarized in Figure 7. Two major differences described by Katzenbach and Smith and applied in the context of the team in this study are:

1. Collective work products in teams vs. individual work products for working groups.

Concluding for the study here the desired outcome from projects performed by teams is to get a collective work product back. This work product should use all the skillsets and various viewpoints that exist within the group of people comprising a team. Individual results produced by working groups would not fulfill the challenge of defining and inspiring new products and services within the business units. Even when the individual results are to some extent loosely coupled as coordinated individuals.

2. Encourages open-ended discussions and active problem-solving meetings for teams vs. running effective meetings for working groups.

Along the lines of the first argument for the work performed here the focus of the organization must be on unique new solutions and not primarily on being most efficient while potentially hampering with the optimal joint innovative solution.

Working group	Team
Strong, clearly focused leader	Shared leadership
Individual accountability	Individual and mutual accountability
The group's purpose is the same as the broader organizational mission	Specific team purpose that the team itself delivers
Individual work products	Collective work products
Runs efficient meetings	Encourages open-ended discussion and active problem-solving meetings
Measures its effectiveness indirectly by its influence on others (such as financial performance of the business)	Measures performance directly by assessing collective work products
Discusses, decides, and delegates	Discusses, decides, and does real work together

Figure 7 Group characteristics adapted from Katzenbach and Smith (2011)

Therefore, the type of team used in this study can be described as a self-managing unit and team, otherwise known in the research as a self-managed work team (SMWT). To see rational and design on the operational model used see Launer (2011).

SMWT were first described in the late 1970s and early 1980s by Hackman (1976, 1982, 1986) in the context of manufacturing, primarily in the automobile industry. Today, as Yeats et al. summarize (1994; 1998), they are also widely used in high tech creative companies. Recently, one more addition was made to the dimensions of Hackman, Katzenbach and Smith. Eseryel et al. (2020) expanded the definition to include Self-Managed Virtual Work Teams to address the complexity of remote work.

Most research on team performance, and more specifically, performance of SMWT teams, uses a general definition of teams, such as the one from Baker and Salas (1992), namely a "…distinguishable set of two or more individuals who interact interdependently and adaptively to achieve specified, shared, and valued objectives." Klaus and Glaser (1968) do the inverse and define what teams are not. Seldom does research reflect or specify the type of team or the form of operation when dealing with their performance. Yeatts and Hyten (1998), on the other hand, provide a comprehensive background on types of teams and their operation. More recent meta studies on the subject of team performance do recognize a difference in team organization, but struggle to sort through literature while keeping that aspect relevant and visible (Bell 2007). Hence some caution must be used when deriving hypotheses and conclusions on special cases to be analyzed.

9.2.2 SAMPLE SELECTION AND COMPOSITION

The reviewed research describes many types and compositions of teams. Almost all studies describe the population they use and the resulting limitations. To compare hypotheses and findings in regards to types of team, three important aspects are described in this section.

First, the majority of studies use students and study groups as their subjects in teams' research. In over 100 studies reviewed for the current dissertation, this study design represents more than a third of all papers. Examples include Seeber et al. (2014), Mathieu et al. (2015), Barry and Stewart (1997), Harrison et al. (2000), Bouchard et al.(1974), Watson et al. (1993), Mohammed and Angell (2003), Chatman and Flynn (2001), Price et al. (2006),

Brodbeck et al.(2010) and Schippers et al. (2013). This is unsurprising as most scholars reside in a university context, and study groups are readily accessible. Experiments and data collection can be set up relatively easy and free of constraints. But it presents problems when transferring the findings into a practical context, as is the intention of this study. Student groups tend to be younger and in early stages of career and life. Therefore, this team model is often less diverse in terms of age, education and experience. The homogeneous age distribution of the student groups makes it unusable for control on age-based effects. Many researchers (compare 9.2.5.3) suggest age and its diversity are an important factor in team performance.

The second aspect to consider is how the way teams are composed might influence results and should be carefully reflected when comparing studies.. For example, some studies explicitly investigate Top Management Teams (TMT) as a research subject, such as the study by Elron (1997). van Knippenberg et al. (2011) state that small differences in team composition can result in big effects. TMT groups are usually characterized as self-selecting, meaning the group decides who will join or not through selection mechanisms such as top down promotion decisions to control group membership. This is in contrast to most industry and other work-related group compositions and how teams are formed. Usually an outside entity e.g., a manager, decides on group composition without strong influence from within the group. This aspect can influence inner processes within a group. Therefore, some attention should be paid to this difference in team types when analyzing and comparing results.

The third aspect relates to inner processes and performance of groups, namely, the form of leadership that is designed into, assigned onto or selected within groups. In SMWTs, no formal leadership role is assigned to the team. As such the analysis of processes concerning the inner dynamics of teams should be also carefully reviewed and interpreted accordingly. Leadership forms can influence the perception of processes within teams and the inner relations of teamwork. Diversity and team performance research specifically investigating SMWTs include like Hackman (1976, 1987), Humphrey et al. (2007), Yeatts and Hyten (1998), Armstrong and Priola (2001), Carter et al. (2006), Coutu (2009), Zarraga and Bonache (2005), Yeatts et al. (1994), Eseryel et al. (2020), Barry and Stewart (1997), Nishii and Mayer (2009) and Markova and Perry (2014). These studies have found that the quality

and form of leadership established in a work group has influence on its performance. This is supported by Stewart's (2006) meta-analysis in which he relates small performance effects on the types of teams which are differentiated by forms of leadership. The same is argued by Eseryle et al. (2020), who recently found that SMWTs perform better when assuming functional leadership within their group. Similar findings were described by Carte et al. (2006) a few years earlier.

For this study the described three aspects are taken in consideration whenever possible to determine their impact and effects and when comparing results.

In addition to performance parameters in team research deep-level and surface-level diversity are the two main study fields when it comes to team performance. The next two chapters will review the state of research in these two fields.

9.2.3 GENERAL DIVERSITY

A subset of the literature reviewed does not break down the details and factors contributing to team diversity, but describe diversity as a general construct, and investigate whether diverse teams perform better than less diverse ones. A few studies aggregate all deep-level or surface-level diversity variables and collapse them into one single or highly aggregated measure of diversity. An analysis across teams is then performed to understand how a general concept of diversity influences teams.

Studies which generalize diversity present contradicting findings. In this review, as many studies claimed that higher diversity is good for team performance (like (Ancona and Caldwell (1992), Cohen and Bailey (1997), Campion and Medsker (1993)) as claimed the opposite (like Bettenhausen (1991), Van der Vegt and Janssen (2003), Watson et al. (1993), Mohammed and Harrison (2013), and Fritzsche et al. (2017)). A deeper analysis by the author found large differences in experiment setup and types of teams used. A commonality across some of the studies is how they use the diversity construct as a moderator to team performance, as a contributing factor or as a control variable, while attempting to link other effects such as individual perception and handling of team or interdependence other variables. Results depended on what variables were available to be calculated, collapsed or

aggregated into a diversity measure. Overall effect sizes were generally not high and tended to be lower than in specialized less general studies.

Some studies collapsed diversity factors into two separate, aggregated variables, deep-level and surface-level diversity. Two studies by Guillaume et al. (2012) and Nishii and Mayer (2009) claim positive effects from higher surface-level diversity on team performance. Harrison et al. (2000) claim the opposite effect. This is likely an effect of what was measured in the studies, why diversity was used, how it was operationalized and on how performance measures were constructed. Guillaume et al., for example, analyzed social effects between work groups and individuals. This method does not directly focus on diversity, but instead uses it as a proxy .

A few pieces of work should be mentioned as they use diversity constructs adjacent to diversity and may show related results. While focusing on time effects, surface-level diversity, and deep-level diversity, Harrison et al. (2000) also used the concept of General Mental Ability (GMA). GMA is commonly known as the g-factor, cognitive ability test or general intelligence. Harrison et al. found that higher GMA in a team leads to better performance. Three more meta-analyses also used the concept of GMA. Devine, Philips (2001) and Stewart (2006) confirmed the positive relation. Bell (2007) confirmed it for field studies, but could not confirm it in studies performed in a laboratory setup.

LePine et al. (LePine, Piccolo, Jackson, *et al.* 2008) included the size of teams as a possible factor in a multidimensional model of team performance. They concluded that smaller team sizes are favorable for team performance. Given the findings around the importance of social interaction in teams a coherent picture.

Other team performance enhancing factors were found by Mathieu et al. (2015), matching competence with team performance in a study focusing on cohesion. Certain cognitive styles (intuitive behavior) were identified as positive contributions to team efficiencies by Armstrong and Priola (2001). Chen et al. (2007) found that individual empowerment, in conjunction with team empowerment, as a team behavior in SMWTs contributes to team performance. Low perceived stress levels in teams and their job design can also improve team performance, according to Karasek (1979).

In summary, generalized diversity studies reveal two important findings for the purposes of this study. One is that the intents and constructs used for many studies do not match the schematic approach of the current study. Therefore, it is not possible or advisable to draw hypotheses and conclusions from them as input to the work here. The second finding is that some of the studies support the hypotheses of some researchers in the field of performance models like IMOI around performance relations for teams, such as better social processes and better identification with the problem space leads to better team output performance. Therefore, no additional hypotheses are derived from this section of literature other than the conclusion of confirmed and reinsurance of other derived ones.

9.2.4 DEEP-LEVEL DIVERSITY

Deep-level diversity sometimes is also known as psychological diversity. It summarizes the aspects of diversity that are harder to observe. Attitude, beliefs, subject matter expertise, personality and culture are common examples. The last two attributes are of particular importance to this study. Deep-level diversity, in contrast to surface-level diversity, is not as frequently studied in team diversity research. Personality traits and culture dimension models are not easy to sample and to collect within a subject group, and the underlying models bare higher complexity due to privacy, data confidentiality and access issues. Nevertheless, this chapter reviews the existing research on deep-level diversity and its influence on team performance. Personality and culture related research will be reviewed in separate sub-chapters.

9.2.4.1 DEEP-LEVEL DIVERSITY: PERSONALITY

Before going into details and hypothesis design around team performance and personality traits diversity a few words around personality in general should be spend.

Personality traits are a broad subject in social science and psychology. Over the last decades one model has become the most widely-used method to describe individual psychological traits. The tool named NEO-I was originally developed as a dimension-based psychological trait model by McCrea and Costa in the 1970s. It has evolved over the years, with its authors adding dimensions to the model, and revising the definitions of dimensions and their assessment. In the late 1980s, Costa and McCrea published the model and a questionnaire-

based test under the name NEO-PI. It is commonly known as the Big5 or OCEAN model as it consists of five dimensions: Neuroticism, Extraversion, Openness to experience, Agreeableness and Conscientiousness. Broadly defined, extroversion is characterized by the tendency to be assertive, active, dominant, and sociable. Conscientious- ness is characterized by the tendency to be purposeful, responsible, and determined. Neuroticism is characterized by the tendency to experience negative emotions such as fear, embarrassment, and guilt. Agreeableness is characterized by a desire to get along with and have sympathy for the problems of others. Openness is characterized by intellectual curiosity, active imagination, and preference for variety. McCrea & Costa (1992), Moynihan &Randall (2001). Further versions developed by McCrae and Costa (2007) are NEO-PI-R and NEO-PI-3 revising dimension delineation and definitions/test. The NEO-PI-R/NEO-PI-3 contain 240 aspects to evaluate, mainly in interview form. As a research tool, is a very cumbersome and time-consuming method. Thus, a version called NEO-FFI, which is directly based on NEO-PI-3, was developed. The NEO-PI-3 assessment is only 60 questions long and is self-reported. Most research reviewed here used this form of the model. This study also used NEO-FFI. McCrea and Costa (1990) affirm that the NEO-PI-3/NEO-FFI model is the right framework when working with an adult group age 21 to 91 years.

The area of personality and its influence on team performance marks a somewhat special area as in 2001 Moynihan and Peterson (2001) created a theoretical framework to conceptualize the influence of personality traits on team performance in much detail. They summarize that most studies in this field follow one of three approaches. First the *universal* approach assuming that certain personality traits of a person directly and universally influence the ability to perform as a team. Second the *contingent* approach which assumes that certain traits and levels influence through organizational culture and task type as moderators. Lastly the *configuration* approach is assuming that the mix or variance of certain traits are influencing the ability to perform. Moynihan & Peterson conclude that these three concepts are not mutually exclusive but rather inclusive and persist all three at the same time. They propose to investigate into all three at the same time and call the approach proposed the *contingent configuration* approach. They also conclude that all three approaches are valuable and valid and as such need to be integrated. Further suggestions are made on how to structure and measure personality input into team performance studies. For the study at hand it becomes clear from their work that input to the universal and configuration approach is viable. The

contingent approach almost always requires laboratory experiments setups. (2001 p.344). The data inputs to the universal approach are quite straight forward and described by levels of traits. The configuration approach requires at a minimum also information about the variance of traits within the group (p.346). That data is also available in this study. More details on data structure, operationalization, and preparation in this context can be found in chapter 10.1.2.2.2. They suggest to control for the influence of levels and variance in regression or other statistical analysis to see the effects of the level of diversity and their relation of personality factors. This is also implemented by this study as we will see later.

LePine et al. (2011) come to the concussion that the effect of team members personality and team composition are stronger with internal team processes and behavioral aspects within the team as they are with results or in other words output oriented measures. They argue that the compositional factors (one of two pathways of effect they propose) have greater influence. In addition LePine et al. also conclude that personality traits are of higher importance than other parameters in the field of surface-level diversity. Subsequently they argue, also on the basis of quite some other scholars work, that personality traits have higher influence on the mediation factors and processes in the team dynamics. Same is supported by Homan et al. (2008 p.1205,1217).

In addition to the theoretical and conceptual framing, Moynihan & Peterson also describe some effects to be expected when analyzing personality traits for team performance influence. Along the discussions of heterogeneous or homogenous distribution of personality traits within a team they say that certain traits lead to higher performance of teams. One claim is that more heterogeneous groups in personality should show higher Peer level results (p.346). More specifically on single dimensions a heterogenous group in extraversion should increase team output performance. In contrast a more homogenous group in conscientiousness is prone to higher team output (p.348). At the same time a generally higher level of conscientiousness will lead to a better team performance in their analysis. But with the exception of creative tasks, for which they claim the relation is not true. As such the study here will hypothesize that the effect is not visible as mainly creative work is performed in the groups observed. Interesting conclusions are drawn on the effects of a high diversity in the sense of disparity for conscientiousness. Meaning that when mixing high and low levels together it is negatively influencing performance. It is even going as far as saying that one

single low conscientiousness team member can influence the whole team negatively. Furthermore, a homogeneously high level of agreeability and openness seem to positively influence team performance. Homan et al. (2008 p.1209) support the same claim of higher openness in personality traits to be favorable for team performance due to improved sub group processes.

Humphrey et al. (2007) developed several propositions in a conceptual, non-empirical study suggesting that high group extraversion and a high degree of conscientiousness in the group mix positively influence team performance. In a revised version of the study, Humphrey et al. (2011) again found effects of high conscientiousness and extraversion on team performance. In the revised study, they used a diversity approach with a min-max method (disparity) in their experimental setup. Their findings suggest maximizing a group's extraversion disparity and minimizing a group's conscientiousness disparity. Zarraga and Bonache (2005) suggest similar findings regarding maximizing levels of extraversion. They argue that a "high care atmosphere' and "high transfer and creation of knowledge" is a predictor for higher team performance. They conclude that leadership style and communication, attributed to high extraversion in the team, fosters such an environment.

Similarly, Eseryel et al. (2020) approach team performance from a leadership perspective. They found positive relational behavior and the resulting functional leadership in teams resulting from a high degree of openness, extraversion and agreeableness on the team. All correlated to a higher performance of the team when using inner performance of the team processes as measures for performance. Mathieu et al. (2015) investigated shared leadership and member competencies to understand how they influence team efficiency. They found that high member competence and agreeableness positively influence performance and group cohesion. van Kleef et al. (2010) assert that high levels of agreeableness are related with high performing teams. They also find high average extraversion, similar to Eseryel et al., relates to higher degrees of team efficiency. In comparable ways, Carte, Chidambaram and Becker (2006) link extraversion and the team developing an internal leadership model with higher performance.

Barrick et al. (1998) could not confirm the influence of higher agreeability on team performance when conducting study on 650 employees in an industry/company context. They did, however, find the same positive correlation for high levels of conscientiousness and

extraversion. They also found an additional small correlation of lower levels of neuroticism in teams with higher team performance. Barry and Stewart (1997) found somewhat different results in a separate study on SMWT teams. In their study, extraversion showed curved linear behavior and suggested that neither too low nor too high levels help with team performance. In the 1997 study, they do not find evidence for high conscientiousness influencing team performance. Harrison et al. (2000), however, did see high levels of conscientiousness as the only personality trait predictor of lower team performance. They provide some unique insight with how they compare inner perceived diversity vs. measured diversity. They suggest that how a team perceives diversity, especially deep-level diversity, such as personality, has a higher influence on the outcomes than the actual measured diversity. A similar argument can be found with van Dick (2008). In their view, positive results from diversity rely on a fundamental positive attitude in teams towards diversity itself. Stewart (2006) later contributed through a meta-analysis on team design and team performance and concluded that high conscientiousness and high agreeability in a team relates to higher performance.

In a study with 59 student groups, Mohammed and Angell (2003) concluded that higher levels of diversity in agreeability and extraversion predict higher team performance. Conversely, higher levels of neuroticism diversity predict lower team performance. Markova and Perry (2014) found that higher levels of openness and agreeability lead to higher performance in internal team processes when looking into team and group cohesion effects, as these factors allow for better opinion convergence and compliance. Similar effects are observed by Diekmann and König (2015). Neumann, Wagner and Christiansen (1999) looked into personality traits influencing retail stores' team performance. The findings are in line with other researchers, namely higher agreeableness, openness, and conscientiousness as signs of higher team performance. In addition, they concluded that a higher diversity in extraversion and neuroticism increase team performance. It is somewhat unclear how the calculation and operationalization was done for the diversity measures. This is the second study after Zarraga and Bonache showing this effect for extraversion diversity. The few extraverted members help the leadership effects, while too many extraverted members hinder the process of leadership emergence. The same year, Neuman and Wright (1999) concluded in a job analysis study of four-member teams that higher agreeability and conscientiousness lead to better performance. As others before them, they found that high levels of neuroticism negatively influence performance, though the link was weak.

Hackman (1976), in one of his early publications on SMWTs, was already looking into the design, structure and composition of the group and its performance. He conceptually argued a high degree of extraversion present with the team members would increase team performance. He also found a high group level of agreeableness to contribute to team performance. He based his findings on knowledge of dysfunctions and conflict in teams and his empirical results. In performing a meta study around deep-level compositions of teams in relation to performance, Bell (2007) identified a high degree of agreeableness and conscientiousness in the team as predictors of higher team performance. The strongest direct links to performance were shown by higher degrees of agreeability and extraversion, and lower neuroticism within the team.

The links of agreeableness and openness to direct outcomes from teams are weaker and more complex. Studies, like Philips et al. (2006), show that both factors, when present in teams, lead to more efficient internal processes, such as convergence and task relation, thus, contributing to better team outcomes team internal performance.

Almost all studies reviewed here used a concept of lower or higher average levels of personality dimensions as possible predictors of team performance. Exceptions are the study by Humphrey et al. (2011) who used a min-max disparity method and Mohammed and Angell (2003) who used a diversity measure. Their derived conclusions, however, still follow a high/low average principle as well. This provides two possible ways to form and then operationalize the current study. Based on the literature review, one way of forming hypotheses is to interpret a high level of a personality trait as low diversity for that trait. The second is to use the average levels of a desired personality trait as the basis. This study uses the latter, which is also the most common method in other studies such as meta-analysis dealing with several other studies. This method only partially describes diversity, but allows it to be compared with the existing body of research. In addition the study follows the recommendation by Moynihan and Peterson (2001) and includes variance measures for such dimension in form of standard deviation (SD) levels

In summary to the discussion above the following hypotheses already formulated in reflection of the structure of this study were derived from the literature review on personality aspects of deep-level diversity:

HYPOSTHESIS 1A: Higher average agreeableness in teams will predict higher team performance.

HYPOSTHESIS 1B: Higher average extraversion combined with low diversity (SD) in teams will predict lower team performance.

HYPOSTHESIS 1C: Higher average conscientiousness while minimizing their SD in teams will predict higher team performance except for creative work. Which leads to the hypothesis for this study that the effect will not be present.

HYPOSTHESIS 1D: Higher average openness in teams will predict higher team performance.

HYPOSTHESIS 1E: Lower average neuroticism and/or lower neuroticism SD in teams will predict higher team performance and peer reviews.

HYPOSTHESIS 1F: Lower variance in conscientiousness will predict higher team performance.

9.2.4.2 DEEP-LEVEL DIVERSITY: CULTURAL

The second aspect of deep-level diversity and team performance important to this study is the area of culture diversity. Like before some remarks will be necessary to reflect the literature review in this study.

Most reviewed studies use Hofstede's dimensions to describe culture orientation for teams with formal structure. Most studies assume differences when going across country boundaries or cultural regions, and do not measure differences directly. Measurable culture differences using scales such as Hofstede's do exist when individuals from different countries are part of a team. It is assumed but not measured. While criticized for the methods in designing parameters and dimensions, Hofstede's model is still the most widely used and provides a comprehensive method for culture assessments according to Elron (1997) and Kogut and Singh (1988). A wide variety of reference studies on countries and other entities is available by Smith et al.(1996), Triandis et al. (1990), and Hofstede and Bond (1984).

Hofstede's scale uses following dimensions:

- Power Distance: Egalitarian ← → Embraces Hierarchy
- Collectivism vs. Individualism: Collectivist ← → Individualist
- Uncertainty Avoidance Index: Comfortable with uncertainty ← → Uncomfortable with uncertainty
- Femininity vs. Masculinity: Nurture Important ← → Power important
- Short-Term vs. Long-term Orientation: Traditional and short-term ← → Futuristic and long-term
- Restraint vs. Indulgence: Normative repression ← → Satisfaction is good

Richard Hackman (1987, 1982, 1976) investigated culture dimensions and their influence on team performance. He concluded in 1987, and even earlier in 1982 and 1976, that relationship orientation (collectivism) contributes positively to team performance. He based his findings on conceptualized studies. Elron (1997) found two positive relations of cultural diversity with TMTs. He predicts that higher cultural diversity contributes positively to team internal performance and development as well as to team outcomes. He does not distinguish in dimensions for culture, however, but on different cultural backgrounds assumed by different countries of residence. Another concept used with regards to culture is the notion of origin for team members. Like in van der Vegt and Janssen (2003). Similar to Elron, this concept assumes different origins represents different cultures and behaviors. This is sometimes referred to as ethnic diversity. The term is used in a number of race-based diversity studies on teams, such as Ely et al. (2012). One study by Brodbeck et al. (2010) found a negative relationship between diversity of origin amongst team members and performance. The same is true for the Ely et al. (2012) in the context of racial diversity with aspects of deep-level culture. Both effects weakened significantly over longer time periods, however. The same effect is found by Harrison et al. (2000), who proposed that surface-level effects on performance vanish over time, and deep-level effects increase their influence. The negative effect of origin diversity in teams on efficiency was also found by Chatman and Flynn (2001). Dahlin et al. (2005) argue that origin diversity has a curved linear relation with team efficiency. They found that too low or too high diversity of origin in teams is counterproductive. The effect in their work is related to the quality of information related processes within teams.

In a field study and two lab studies, Earley and Moaskowski (2000) confirmed the same curved linear relation. Seo, Kang and Song (2020) found the same rational again for cross border Research and Development (R&D) teams. They also found that a higher educational diversity in these teams increased the effect. The opposite positive effect on teams with diverse origin and their effectiveness was found by Kearney and Gebert (2009). In their work, it is moderated by transactional leadership. They claimed that if transactional leadership is present, the effect vanishes. Aronoff et al. (1983) claim around personality research in context of conscientiousness that a higher level of egalitarism behavior will help team performance.

Cultural influence on teams has not received as much attention in research as the personality attribute. Few studies directly measure cultural diversity on a parameterized concept such as Hofstede's. Diversity is usually defined as a low average in terms of country of origin or country of residence of individual team members. As with personality aspects of deep-level diversity, the most common concept in the existing research for analyzing average levels of culture is used. Cultural diversity it is usually represented by origin, language or ethnic considerations.

Based on the literature reviewed in summary following hypothesis already reflecting the structure of this study are developed.

HYPOSTHESIS 2A: Higher average collectivism in teams will predict higher team performance in terms of outcome and team internal aspects.

HYPOSTHESIS 2B: Higher levels of egalitarism will predict higher team performance even with lower conscientiousness present.

Cultural diversity is also represented in surface level diversity with regards to native language, country of origin and race. These aspects are reviewed in the next section.

9.2.5 SURFACE LEVEL DIVERSITY

Surface level diversity influence on team performance is one that is studied more often than deep level diversity as the author found. It is believed in part due to the nature of the data, which is generally easier to access and collect. Participants in surface level diversity studies

are less concerned about sharing demographic data when compared to data on personality traits, which can be perceived as more invasive, personal, and revealing as the author experienced first-hand in this study. In other cases, like education, the necessary data is usually shared without hesitation in educational and business contexts, as it is usually considered a performance attribute or distinction. In addition, some parameters, such as gender, can be observed by researchers directly. The findings are grouped according to the attributes investigated.

9.2.5.1 SURFACE-LEVEL DIVERSITY: EDUCATION

Stewart (2006) claims that a higher average level, and hence less diversity, of education within a team increases its efficiency and outcome in a study with business context. Investigating 75 R&D teams, Shin an Zhou (2007) paid exclusive attention to education diversity in regards to field specialization. They found that higher education field diversity in teams leads to higher performance, greater creativity and transformational leadership within the team. For 62 R&D teams, Kearney and Gebert (2009) concluded that educational background diversity elevates team performance. In contrast, Hackman (1976), in his conceptual work on SMWTs, predicted negative effects of diversity in education background on team outcomes. He sees it as a negative influence on team conflict, not as a direct effect to outcome. Hackman (1987) repeated his postulate in his work from 1987.

In line with other researchers, Ancona and Caldwell (1992) saw positive relationships of more diversely educated teams in terms of their efficiency. Cohen and Bailey (1997) postulate that educational background diversity correlates with higher performance for work teams, but could not find conclusive evidence. Bantel and Jackson (1989) successfully showed the relationship for TMTs. Campion and Medsker (1993) demonstrated a positive relationship between educational diversity and team performance amongst a set of financial industry teams. Four more studies by Keller (2001), Dahlin et al. (2005), van Knippenberg et al., and Bunderson and Sutcliffe (2002) contributed the same finding of increased educational levels and elevated team performance.

Overall researchers tend to assume that more “hidden” dimensions, as Van Knippenberg et al. (2004) name them, have less influence than the more readily visible attributes as gender, age or race etc. Van Knippenberg et al. argue further though, that factors like education have far

greater influence on job related aspects in teams and such have greater influence in information/decision-making processes. Hence educational diversity for example should have positive influence on job related processes and performance. Van Knippenberg and Schippers (2007) summarize that there is a general tendency by researchers to claim that surface-level diversity in the sense of demographic diversity seems to have negative influence on team performance while diversity that relates to the mentioned information driven processes within teams fosters positive contribution to team performance.

9.2.5.2 SURFACE-LEVEL DIVERSITY: GENDER

Bouchard et al. (1974) concluded that neither a higher nor lower gender diversity has a performance effect on teams. Hackman (1976) predicted a negative influence of gender diversity on team performance. Two more recent studies showed positive relationships between gender diversity and team results: Mohammed and Angell's (2004) more conventional statistical analysis of 45 student projects and van Knippenberg et al.'s (2011) study using more unusual fault-line analysis on TMTs.

9.2.5.3 SURFACE-LEVEL DIVERSITY: AGE

One of the most investigated diversity factors on team performance is age and age diversity. In line with demographic dissimilarities, Jackson et al. (1991) found that an increased diversity in age within teams lead to degraded performance. They also found higher turnover related to these effects. Concurrent results were produced by Kearney and Gebert (2009), portraying the same negative relationship of age diversity with team efficiency. Hackman (1976) assumed a negative correlation of age diversity on team performance. In contrast, Bettenhausen (1991) found positive relationships of age diversity on group performance. In his 500+ papers meta-study, he included groups in general, not just work-related groups, such as workgroups and teams. He also investigated social constructs such as ethnic groups and clubs. Hence his findings are applicable to more general group dynamics, and less so to specialized work-groups.

Bantel and Jackson (1989) showed negative effects for higher average age and age diversity in TMT groups. They also found higher turnover as a consequence. The same negative correlation effect was found by Ellwart et al. (2013). A negative correlation of team

effectiveness and age diversity in teams was shown by Lieberman et al. (2013). They found negative health effects for 50+ years age groups when dealing with age diverse groups. Schippers et al. (2003) did not find any relationship, not positive or negative, between gender, age or education diversity for teams and their performance. Their main performance focus was a rather narrow construct of reflexivity and interdependence; two strictly team internal factors. It is therefore not surprising to see differences compared to other studies more focused on team outcomes.

Following hypotheses are derived for surface-level diversity based on above literature review. All hypotheses are already reflected and summarized towards the structure of this study.

HYPOSTHESIS 3A: Higher educational diversity in teams will predict higher team performance.

HYPOSTHESIS 3B: Higher gender related diversity in teams will predict higher team performance in terms of outcome.

HYPOSTHESIS 3C: Higher gender related diversity in teams will predict higher team performance in terms of outcome through the moderation effects of information/decision-making in teams.

HYPOSTHESIS 3D: Higher age-based diversity in teams will predict lower team performance in terms of outcome.

HYPOSTHESIS 3E: Gender and age diversity will have higher direct effects on team performance than education diversity.

HYPOSTHESIS 3F: Higher diversity in country of residence of team members will predict lower team performance.

9.3 TEAM PERFORMANCE AND DIVERSITY

Several aspects about team outcomes and their measures regarding diversity studies will be reviewed. The different forms of performance are discussed, as well as the various ways of measuring and operationalizing them.

9.3.1 PERFORMANCE/EFFICENCY IN DIVERSITY RESEARCH OF TEAMS

The apparent clean structure in team diversity research and performance measures disappears when it comes to operationalizing these performance measures. The somewhat diffuse concepts are confirmed by Dionne et al. (2004), who, in a 180 papers spanning meta-study attempting to do a multi-level analysis, found diffuse definitions and constructs of performance. After reviewing more than 100 studies on diversity and team performance, very few use direct measures or clear definitions of the original Hackman work, or the later work from Bell, Beal and Ilgen. As discussed in chapter 9.1 when conceptualizing the theoretical framework for this study and reviewing the frameworks used and developed for teams diversity and their performance, a multitude of definitions and constructs are used to define team performance. Many, in particular the more recent ones, follow basic structures like the already mentioned threefold outcome concept from Hackman. While for many meditation effects and some related categories of performance somewhat defined structures and concepts exist, like the Team Climate Inventory (TCI) or similar, the constructs around what defines outcome of teams in terms of throughput, procured assets etc. are more diffuse. This is where this study will try to provide some overview. The author categorized performance measures in diversity studies into indirect, direct, subjective, self-declared, inner processes, and composites. All six categories will be reviewed in the following sections.

9.3.1.1 INDIRECT MEASURES (PROXIES)

Indirect measures of team performance are the most widely used. They are characterized by the fact that the measures do not directly quantify or qualify anything that would fulfill a definition of e.g., Hackman/Ilgen et al. as an outcome/expectation measure. One example is the very narrow measure of budget accuracy of teams by Keller (2001) to link performance to diversity, communications and job stress. In this case, the budget accuracy did not directly measure any outcome of a team, but the accuracy of budget tracking was used as a proxy to assume higher performance of teams. Customer satisfaction of technical support teams was used by Neumann et al. (1999) to correlate the personality composition of teams to performance. In another study by Neumann (1999), together with Wright, turnover rate was used in the teams as an effectiveness measure to relate to personality and cognitive abilities. Even more indirect and somewhat questionable, Bunderson and Sutcliffe (2002) use the

profitability of the related company as a measure to operationalize team performance. The profitability of the company itself is a proxy that has a very low direct correlation with teams and their performance. Performance of companies depend only in part on performance of teams, even management teams, which was the focus of this study. External factors such as market demand and financial market impacts add significantly to such a proxy and should be controlled. Campion and Medsker (1993) as well as Jackson et al. (1991) use turnover as a performance measure. Humphrey et al. (2011) used the speed of teams in task fulfillment as the sole performance measure. Cost and revenue was used by Rapp et al. (2016). More indirect financial measures were used by Jackson, Joshi and Erhardt (2003), Halfhill (2005), and Brodbeck et al. (2010). Bouchard et al. (1974) used creativity as an indirect measure. Team throughput of calls was used in research by Scheider (1987). Innovation capability of the company (as opposed to the individual teams) was used by Bantel and Jackson (1989), and later by Vander Vegt and Janssen (2003).

In real-world empirical studies, limitations on data collection and its implications on study design and methodology exist. The author questions whether some of these indirect measures are related to team performance or influenced by the teams in a meaningful way. For example, financial performance of projects and companies is generally not only related to team performance. External factors, such as sales activities, budget control, stock market deals and color of money considerations under national accounting rules, can influence results more than the performance of teams.

9.3.1.2 SUBJECTIVE MEASURES

Subjective measures are used almost as widely as indirect measures. They are categorized by the aspect that the performance of teams is rated by an external party along a subjective scale. The scale does not quantify any specific outcome, such as units produced, questions answered or products released, but rather how an observer subjectively feels about how the team performed. Armstrong and Priola (2001) and later by Carte et al. (2006) observed behavior to measure team performance. Decision quality measures were used by Seeber at al. (2014) to determine relationships to diversity. Subjective and structured supervisor ratings (mostly 7 or 5 Scale Likert) are used by many researchers e.g. Barrick et al. (1998) and Kearney at al. (2009), and earlier by Ancona and Caldwell (1992). Another method is the one

used by Elon (1997) in late 1990s in which the performance of organizational units is subjectively rated by top management. All subjective measures involve either hierarchically higher entities like managers, supervisors, and boards, or an external observer (usually the researchers) on their subjective impression of how a team performed while working on a project or execution of a task. In some cases, the observation used structured processes to do so, such as guiding questions focusing the rater's attention on specific aspects like conflict or work quality. Some did have no structure and relied on a single, simple Likert scale measured dimension.

9.3.1.3 SELF-DECLARED MEASURES

Some studies, such as the one by Schippers et al. (2003), used the concept of teams self-rating their performance. They used a Likert-scale based model to gather performance measures from the teams. The main difference to the previously described subjective performance measure is that the measure of performance comes from within the team, not an external view. While still subjective, it is expected that a team internal view of its performance can differ from an external view as we found earlier with e.g. Ilgens' IMOI model.

9.3.1.4 INNER PROCESSES

Another common form of evaluating team performance is the measurement of how team members perceive the project or task that they worked on; the internal view of how teams think the process went, not the outcome as above. While in 9.3.1.3 the teams are asked on how good they think they fulfilled the external expectations, in the Hackman/Ilgen definition, when using inner process measures, teams are asked about their view on internal processes and behaviors. The concept of *atmosphere* is used by Zarraga an Bonache (2005). Peer ranking in terms of performance is used by Barry and Stewart (Barry & Stewart 1997) to determine the role of personality within groups. A more narrow evaluation is used by Mohammed an Angell (2004) as they ask the members to evaluate conflict within their team. Markova and Perry (2014) used group cohesion and individual well-being as constructs. Bell and Marentette (2011) used a full team viability construct as a performance measure to try predicting future team performance. Ellwart et al. (2013) performed a similar case study.

9.3.1.5 DIRECT MEASURES

A few studies use measures of team performance that are directly related to something the team is expected to do. Usually, outcome related aspects define an a priori expectation to the work of the team. Such aspects are usually manifested in goals, expectations and statements of work, which describe the intended outcome along measurable scales. One of the often-used methods here is grades in schools and universities of group projects and assignments, which was used by Harrison et al. (2000) in one study. Mohammed and Angell (2003) argue that the use of grades of written results as measurements are the best way to reflect the common efforts towards expected outcomes for teams. Schippers et al. (2013) used an instructor grade method in which they even ensure instructors are not part of the research team. A similar method is used by Watson, Kumar and Michaelsen (1993). They used four formally defined dimensions (range of perspectives, number of potential or existing problems, generation of multiple alternatives and quality of the recommended solution). Overall performance on the task was the average of the ratings of the four criteria. This was found by the author to be the most comprehensive direct measure.

9.3.1.6 COMPOSITES

Some studies refer to performance measures as composites of multiple measures, such as Schippers et al. (2003) and Seeber et al. (2014). Some combine different measures and aggregate them into single measurement for statistical analysis, such as in Rapp et al. (2016). Cohen and Bailey's (1997) meta study performed further analysis by combining measurements into a single variable. While this may be a necessary, and sometimes the only possible, step to reduce the complexity of analytical models, it is not reflective of the previously described threefold distinction of team performance (team outcome, team inner performance and team self-defined goals). More recent meta studies by Mathieu, Joshi and Erhardt (2003) and by Mathieu et al (2008) propose separating out the three areas to get better insights into team mechanics and diversity.

The intent of this study is to maintain the presence of the three different performance measures in the data collected. For team outcome, project expectations are provided to the team which are graded externally after project completion. The inner process view is derived from peer review data. Teams also derive their own goals for a project, which are graded

externally. Data collection in this study is amongst very few showing clear feedback loops and direct connection from goals set to team performance and as such, allows for new insights in diversity and team performance.

9.3.2 TEAM PERFORMANCE DIVERSITY

As discussed in section 8, and on the previous pages here, literature and theory define three types of performance for teams: the outcome focused on set goals, the team behavior to achieve those goals and how the inner mechanisms amongst the team members perform. Some of the reviewed literature identified relationships amongst these performance measures. The current study employs all three measures.

Bell (2007) claimed that inner conflict in teams effects performance negatively for the team. Her main argument was that teamwork quality relied on inner team processes and interpersonal functions. A similar route was taken by Hiller, Day and Vance (2006). Constructs around problem solving, support, power distance and collective leadership were used to investigate team performance at different points. They predicted that teamwork quality correlates with team outcome performance in SMWTs. Beal et al. (2003) saw the same effect in a meta-analysis. They found performance to be influenced by positive team behavior toward the goals being set. In their study focused on reflexive behavior, Schippers et al. (2013) found effective teams' inner processes relate to a higher team outcome performance. Markova and Perry (2014) came to the same conclusions for SMWTs formed by students. In a more conceptual study, Seeber et al. (2014) had the same findings. Another meta-analysis by LePine et al. (LePine, Piccolo, Jackson, *et al.* 2008) described positive relationships of teams' inner processes on team behavior. The same is true for the meta-study by Guzzo and Dickson (1996). Gist et al. (1987) noticed the performance of small groups going up when inner team processes are more positively seen. The same effect was found by Bradley et al. (2012) while studying conflict topics in teams. Kristof-Brown and Stevens (2001) concluded concurrently while researching goal related team effects.

Following this review of literature hypotheses are developed in the frame of this study.

HYPOSTHESIS 4A: Higher rated Peer performance leads to higher levels of team outcome performance (PO_Exp).

HYPOSTHESIS 4B: Higher rated team goals performance (PO_Team) leads to higher levels of team outcome performance (PO_Exp).

9.4 INFLUENCE OF TIME ON DIVERSITY

This study will not investigate longevity effects of diversity in teams. The main reason for this is that by the nature of the teams observed, only cyclic data can be collected, and the way life cycles of teams are designed does not allow for looking into time-based effects. Still, it is advisable to review some literature describing time as it still been found to have effects in the field of team diversity. Reason is to understand and possibly compensate or adjust for effects within the somewhat short-lived teams in this study, should such effects exist. The concepts and constructs around time effects and team diversity are based on the facts and theories that many factors, such as the ones described over the last few pages, underly temporal effects. Two constructs dominate in the field.

The first is that time influences team members individually. One example here is the widely discussed effects of time and age on the Big5 personality traits. Traits can change over the course of a lifetime. The changes are small and take a long time, according to as Funder (2010), spanning years or decades. A meta-study was performed by Roberts et al. (2006b) which showed that over time, personality traits can change. They also reflect that this is very much in line with current state of science in personality development research. They showed that personality changes over time for four of the six categorized traits. Due to the base studies and data available (n=92), some adaptions to the categories were made from the original Big 5. The study also showed that the changes are gradual and cover long time periods. Changes usually take decades to develop into relevant measurable factors. Wu (2016) also concluded that personality change happen over longer periods of time. Two studies by McCrea, Martin and Costa (2005) and Roberts et al. (2006a) indicated that there is a common vector, in strength and direction, for changes during a lifetime. As an example, they found that openness starts generally to increase when approaching 40 and continues to do so when getting older at a slightly higher rate. Based on these findings the study here concluded that taking the NEO-FFI test one time in a study length of five to six years is enough to the predict accurate results when analyzing team diversity along the personality traits and not for longevity effects.

Work-related factors like strong forms of stress, may influence personality traits temporarily or even permanently. Wu (2016) found evidence that higher job stress decreases extraversion and consciousness, and higher job control helps to increase agreeability, consciousness and openness over the course of five years and with small effect sizes.

The second time construct is how the mechanics of how members interact with each other change over time. Watson et al. (1998) research the temporal effects of team-orientation and self-orientation. They concluded that team-orientation grows over time and self-orientation shrinks. Chatman and Flynn (2001) found that the effect of demographic differences in origin/place negatively affecting team performance through lower cooperation levels in teams fades over time. Harrison et al. (1998) concludes that surface-level diversity influence on team performance weakens over time while deep-level influence increases. The effect time scales are multiple years and they compared different groups by tenure, not in a long-term, in-group design. Bell and Marentette (2011) see team viability increasing and a positive influence on the team's ability to perform in long term-teams. Rapp et al. (2016) include time aspects when searching for positive effects of external leaders and coaching for teams and their performance.

In summary it has been found that time effects on processes and factors of diversity in teams are long-term. Typically, the findings see multiyear effects. Some even see effects in decades. As the project teams in this study have a lifecycle of three to four months, it can be concluded that the above effects do not have to be considered for the research at hand.

9.5 MEASUREMENT OF DIVERSITY

A study on diversity bares the question on how to measure and operationalize diversity. One common and agreed upon approach in the literature to measure diversity is provided by Harrison and Klein (2007). Harrison and Klein based their work on the need to define diversity in a "specific, narrow, and usable way". They claim that authors usually use terms like diversity, heterogeneity, dissimilarity and dispersion "casually". They suggest limiting it to "…describe the distribution of differences of a unit with respect to a common attribute X …" This definition is used to analyze the data in this study. Harrison and Klein (2018) define three states of diversity:

a) Separation
b) Variety
c) Disparity

Separation is defined as an opposite opinion or disagreement. Variety is defined as differences in kind or category, such as information or knowledge. Disparity is difference in concentration of e.g., pay or status. Figure 8 provides a visual representation of the definitions and concept.

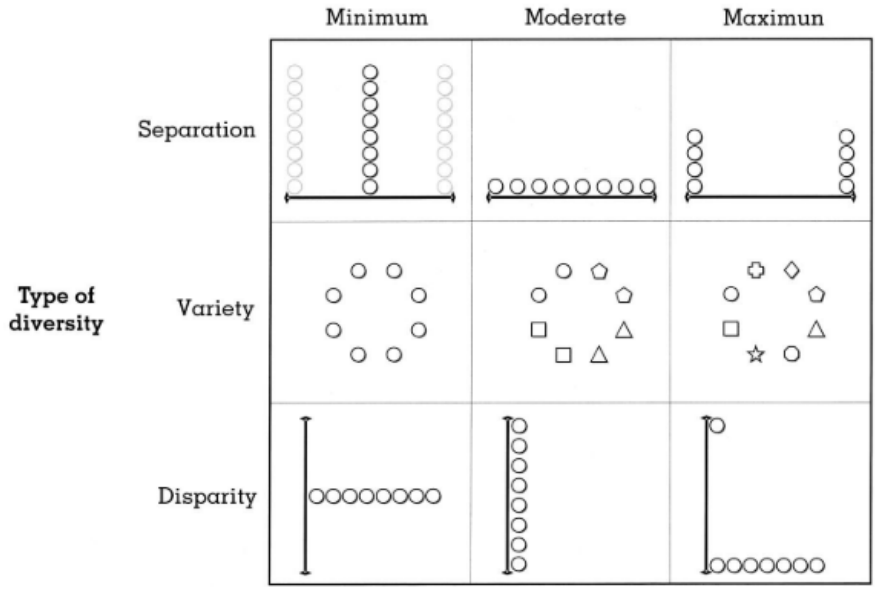


Figure 8 Separation, Variety, Disparity by Harrison and Klein (2007)

Taxonomy and diversity in biology and ecological science, as well as entropy in chemistry and physics, represents as a potential viable solution to operationalize variety. The concept is mathematically similar in many fields whilst different names for the concept are used. In summary, the most used diversity indices in biology are the Simpson Index or the Shannon Index. Both indices are based on the concept of probabilistic/statistical considerations that, for example, one species are of the same origin. The Simpson Index is defined as (Simpson 1949):

$$D = 1 - \sum_{i=1}^{s} \frac{n_i(n_i - 1)}{n(n-1)}$$

where *n* is the number of individuals of a given species, *i*. A higher diversity will converge to 1.

The above indices are similar to the Blau Index, which is a concept often used in social-science-based studies to describe diversity. Originally published by Blau (1977) in the context of sociology, the Blau Index expresses how "equally" some members of a team, *Vk,* are spread over a distinct category, *K*. Another widely used term for such an index is "evenness." As Harrison and Klein (2007) describe, "Having an even spread of members over the richest number of information sources corresponds well to our definition of maximum variety; evenness and richness contribute to a higher Blau index." The highest possible Blau Index is 1. It means that all measures are equally distributed over the scale of measurement. In contrast, a Blau Index of 0 describes the state that all measurements are at the same discrete value. For measuring variety in this study, this measure is necessary to analyze the data at hand.

The Blau Index is defined as:

$$B = 1 - \sum p_k^2;$$

where p is the proportion of group members in the category, k. The study here used an evenness measurement which resembles a ratio conversion of the Blau Index. So Blaus 1 equals 100%. Evenness is defined as:

$$ev = -\sum_i (p_i \log p_i) / \log B;$$

where *pi* is the proportion of entities comprising the category, *i*.

The Blau Space is another method to operationalize variety. The Blau Space is simply a multidimensional space in which coordinates are defined by Blau Indices along different axes defined by distinct categories. This forms a geometrical definition of the variety amongst multiple dimensions. It is mainly used and experimented with in the area of urban planning and analysis, as Matthews (2019) concludes. This allows, for example, for vectorized and parametric analysis. While this seems to be a workable solution applicable to the problem space in the social science context, one aspect prevents its use: the creation of a Blau Space requires a definition of the relationships of the categories (axis) to each other. This implies

that one knows the relationship or angles of the axis in space. This requires an understanding and model the relationships of factors and their interaction/interrelation. Research has not yet defined a universal working model and construct for team performance, as Guillaume et al. (2013) state. Once reached, the concept of Blau Space and vectorized models may improve methods for predicting team performance. Such an understanding is also the prerequisite to enter the field of machine learning with data describing team performance.

Following the discussion from Harrison & Klein (2007), Knippenberg & Schippers (2007) argue that in some cases it is desirable to go beyond the described concepts of dispersion and more multi-dimensional measures like fault-lines and concepts allowing of more detailed description of the particulate relation of attributes and individuals to each should be employed. At the same time they acknowledge that such connects are rather complex and apply more to theoretical research rather than empirical research. Nevertheless the discussion along concepts and model in the field of personality by Moynihan & Peterson (2001) with their contingent configuration approach provide structure to applied some of the deeper measures into empirical research as well. Therefore this study employs such suggestions to the diversity measures and operating this study. Detailed computation can be found in chapter 10.1.2.2.2.

10 METHODS

The previous chapter provided a literature review of surface-level and deep-level diversity and their influence on team performance. This chapter describes the methods used to analyze the data collected to validate the previously defined hypotheses. The results of the analysis provide more conclusive influences.

Three major topics will be covered in this chapter. First, the data sampling and data collection methods are described. Also, the context and description of the teams is given. The second topic discusses the measures used to operationalize the collected data to allow for statistical analysis. Operationalizing diversity measurement is reviewed to specify where this work fits into the existing team performance research. The section will also deal with missing data and data privacy issues. Lastly, the chapter defines the statistical analysis that was conducted, including necessary data conversion, parametric testing, correlation and regression analysis.

The necessity to embed the data given what data was able to be collected while keeping the advantages of the unique consistent framework has proven to be challenging and not trivial. Overall, the author hopes to achieve what Rumsey (2011 p.9) once stated: “Every data set has a story, and if statistics are used properly, they do a good job of uncovering and reporting that story”.

10.1 DATA AND DATA MANAGEMENT

This chapter describes the process of how data was sampled and collected. Context on the teams investigated is given. In a second section measures used will be discussed.

10.1.1 SAMPLING AND DATA COLLECTION (TEAM AND ORGANIZATIONAL CONTEXT (INPUT))

Sampling and data collection will be covered in this section. Some context of the teams sampled are covered as well.

10.1.1.1 CONTEXT AND EVIRONMENT OF SAMPLE

The hypotheses were tested using data that was collected over the course of seven years in an R&D environment at a service and software group within Boeing Global Services. The data includes 115 individual projects, their results, surface-level and deep-level diversity attributes for team members and a range of attributes describing aspects of project related teamwork.

The work presented here is founded on an organizational change rooted in the year 2000. In September of 2000 The Boeing Company (TBC) in Chicago, USA acquired Jeppesen Sanderson Inc. (JEPP), headquartered in Denver, CO, USA. Jeppesen is the industry leader in aviation data management, air-navigation service provider/aggregator and electronic solutions for flight operations solutions. For more context on the air transport system and air-navigation system, the author recommends Hirst (2008 p.223) compendium on the subject. The acquisition has driven various waves of organizational change for JEPP's operation as integration moves on. Now in 2020, the process is at its peak as JEPP, a non-fully integrated subsidiary organization up to now, is being fully integrated into The Boeing Company.

As part of the process of integration, in 2010, a major change was initiated to redefine the strategy, mission and vision for JEPP. Integration into the TBC caused the complexity of JEPP products to increase. The airplane platforms developed by core Boeing drove closer integration of software products for pilots and Airline Operations Centers (AOC). These platforms started to become digitally enabled, and the airlines themselves reached new heights in digitation of their enterprises. Thus, modern and more efficient software development processes were needed internally at JEPP. This led to the wide adoption of Scaled Agile and Pragmatic Marketing. Throughout the integration, the organization needed to answer how it was going to address these changes (see Bowlin (2011)).

The answer by the innovation and research department was to design a radically different organization and operational model than the ones that were common for such organizations. The organizational form chosen was based on Self-Managed Work teams (SWMT). Though out of scope of the current study, the detailed motivations, systematics, discourse and approach that led to the design the organization can be found in previous work by the author (2011). A summary of the key aspects of the organization are:

- Self-Managed Work Teams (SMWT)
- Short periods of teams working together (three to four months only)

- Changing team assignments and respectively composition from project to project
- Internationally mixed teams (Cultural background and multi-location)
- Strict delineation for team tasking on “What” (Management) and “How” (Team)

The operational model has been executed as described above since 2012, for the most part unchanged, providing a great opportunity for deriving insights from that teamwork for a large number of projects. A full 115 projects and over 1000 data points are included in this study from 2012 through 2018.

Through an integrated performance process covering 360-degree feedback and continuous improvement interests by management, a comprehensive set of data is collected for each of the projects and team members involved. The data includes information about a wide variety of parameters for projects, teams and team members. The data is also unique in terms of consistency over time. It contains data that was collected in a stable and controlled environment, and conditions such as consistent questionnaires, population, leadership and performance management. Therefore, the data provides a unique opportunity into team research and diversity. Many other data sources pose deficiencies in time and structural frameworks, as Wu (2016) argues.

The data allows for unique insights in surface-level and deep-level diversity aspects of teamwork. It also provides a stable and consistent approach to link diversity aspects to different types of team performance. This aspect is often a weak spot in previous literature as it is described in later chapters of this work. See chapter 11.

10.1.1.2DESCRIPTION OF TEAMS

The analyzed teams have a typical size of between two and six members, which is a typical size in team research according to Bettenhausen (1991). In this study, 65 members form the team population. Team compositions are not stable over time. Every time a new project starts, teams are reconstructed. The projects usually all end and start at the same time. All team members participated in at least one project, but most participated in many more over time. As no fixed team structure exists and teams are always mixed differently for every project a new constellation of team members was selected and grouped. This leads to unique team constellations for the 115 projects feed by the same 65 people pool.

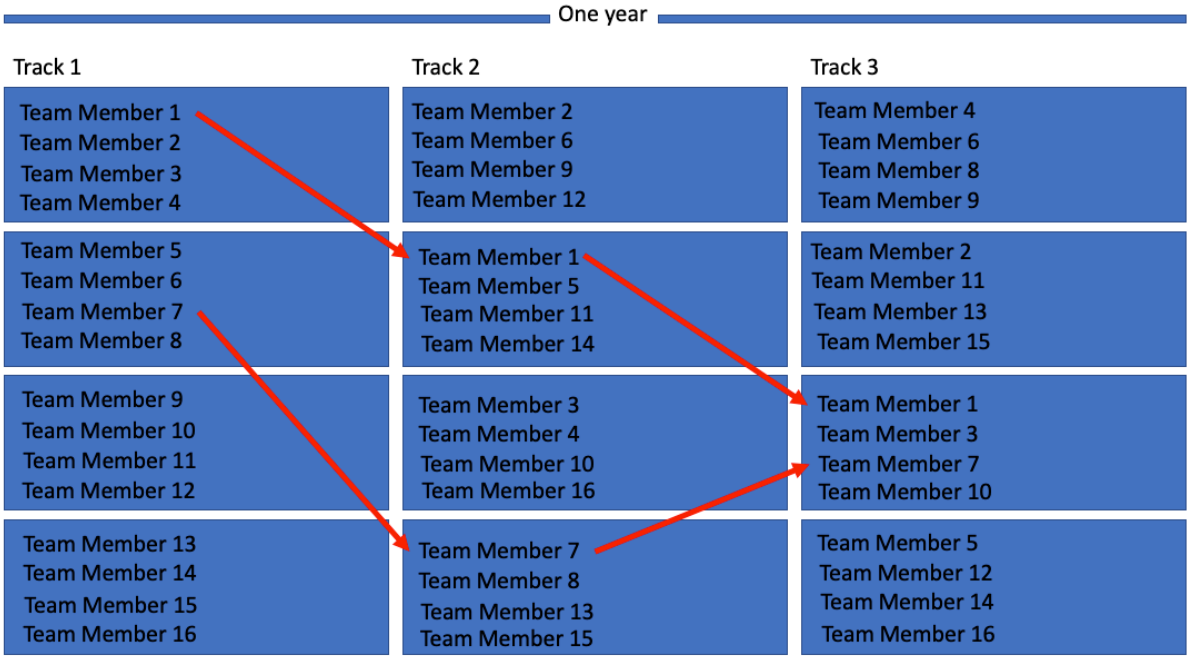


Figure 9 Team members and project tracks

Figure 9 illustrates the movement of two selected team members (Member one and seven) through three different projects over a given year. The groups of projects and their time span of four months are called “Tracks” internally. The organization hosting the teams is embedded and works integrated into a project matrix for their day-to-day work. That means their functional manager acts more as a coach or sideline function to support the individual team member. The manager has no direct influence on tasks and work structure. In larger scale, the organization represents a weak, functional matrix in regards that management is only assigned to cross-project and/or corporate reporting (see Burton et al. (2006), Burns (1989), Sy (2005)). The teams are designed to work as SMWT teams, as defined in Chapter 9.2.1. Work is organized in a combination of Scaled Agile and SMWT structure. In layman’s terms, the management is responsible for the “what” and the teams have full responsibility and latitude for the “how”. A typical team member has an engineering or business background education at a graduate level or above. Team members are mostly located in three location: Denver, CO, USA; Frankfurt, Germany and Gdansk, Poland. Some team members are included from Copenhagen, Denmark, Singapore, and Seattle, WA, USA. The gender distribution is 86% male and 14% female. More recently that ratio shifted towards a more equal distribution due to hiring more female team members into the teams. Age ranges from 25 to 70 years of age. Average employment time for the sample population is 10.2

years. The sample shows a relative equal and homogeneous distribution of cultural traits throughout the team. Six different languages are spoken natively (English, German, Polish, Spanish, Portuguese and Danish). The ability to speak multiple languages varies between one and four. Some further basic demographic statistics are included in Appendix B.

10.1.1.3 DATA COLLECTION

Data in this study is collected two different ways: periodic data collection at the end of each project, and one-time collection of stable surface-level and deep-level data for each team member.

10.1.1.3.1 PERIODIC PROJECT DATA COLLECTION

In reference to Figure 10, at the end of every project a set of data is collected. This set of data is comprised of performance measures covering all three performance aspects described in chapter 9.3.1. External rated team output performance and team goal performance is assessed by the Project Owner (PO) in reference to the pre-negotiated tasks and goals. Data is collected in a team wide distributed spreadsheet and discussed with the team after the project. An example can be found in Appendix A.1and Appendix A.3. The results are uploaded into a calculation and evaluation tool called Employee Surveys Evaluation (ESE). Peer performance is collected after the end of the project by means of online survey tools. Each team member receives an email invitation to access the online survey. That survey guides a given team member through the peer questionnaire for every other team member. The questionnaire can be found in A.2 Figure 14 through Figure 18. This questionnaire also includes team members to provide feedback on the project itself. The participation is strictly voluntary and has been discussed at length with employees, Human Resources (HR) and the works councils, and subsequently approved by all involved parties. Also see chapter 10.1.2.4 for more detail on that topic. Peer data is also uploaded into the ESE tool and available anonymously for evaluation.

10.1.1.3.2 ONE-TIME TEAM MEMBER DATA COLLECTION

For every team member included in the sample population of this study, a comprehensive set of surface-level and deep-level data was collected. This set of data was collected once over

the runtime of this study for each team member. For deep-level diversity two sets of data were collected. One set was the Big5 character traits via a paper/email-based assessment using the NEO-FFI questionnaire. Culture data was collected using the online questionnaire of the Aperian GlobeSmart tool. See Doherty (2016) and Global Aperian (2016) for more information. Both deep-level datasets were converted into a per team member spreadsheet. For the surface-level data, interview-based questionnaires were used with every individual team member. Data was merged into the overall data set per member.

10.1.2 MEASURES

This chapter provides the measures for each individual data set collected and the data model used. The chapter further addresses data privacy aspects, missing data a detailed section on measuring and operationalizing diversity. In Figure 20 in Appendix B.2, the complete list of data fields for the collected raw data is listed in the form they were collected.

10.1.2.1 PERIODIC PERFORMANCE DATA (OUTPUT)

Measures for the periodically collected data are explained. The periodic data collected per project is resembling the output data in the IPO model as defined in chapter 9.1 for this study.

10.1.2.1.1 PEER REVIEW

Peer review is assessed using a concept developed in a former study by the author (2011). The previous work isolated three parameters as relevant in in-between peer behavior: communication, collaboration and contribution. The research in 2011 showed these three factors as critical to team performance and to be the main contributors to team performance when looking into peer related processes for SMWTs and peer interaction. Launer (2011). The data collected here resemble the main drivers for interpersonal behavior monitoring individual and team performance against these critical parameters, a set of per project evaluation questions was developed. Peer review data is coded into fields [Q1] trough [Q8]. All of them are coded with a 5-point Likert scale. Scale is dependent on the question, but generally, 1 is low or negative while 5 is high or positive. Data is available on a project-by-project basis about every individual. The full questionnaire can be found in Appendix A, Figure 14 through Figure 18. The concept is resembling mostly to the construct of

Interpersonal Behavior moderating Member Satisfaction with group outcome for teams as defined by Moynihan & Peterson (2001). The Peer data construct is not a formal representation of such academic measures but resembles the intended insights on cohesion, communication, and motivation to work together. Therefore two possible ways of using the data are thinkable. One is using the Peer results to test for moderation for personality trait effects, as intended by Moynihan & Peterson. Second to use it as a related outcome reflecting team performance or in Hackman definitions reflecting team viability. Both approaches will be performed and reflected in chapter 10.6 by this study.

The Peer data collected here is representing a multilevel theory problem in team research when brought forward to an analysis. In contrast to all other data collected the Peer data sets contain multiple measures per team member per project in a round robin fashion. As such, the data is a second level, nested data structure, to the otherwise first layer data on the project aggregation. Or as Klein et al. (2000) headline it: “Aggregation of lower-level data to higher units of analysis”. Simple means to aggregate are not sufficient in such cases as the dimensions are not necessary bi-directional and continuously defined. Also several answers around multiple topics need to be aggerated and correctly reflecting the state. In this case the state of the team. In statistical and systematics terms for sociological studies the state that needs to be computed here is the Within-Group Agreement per project and team. Bliese in Klein & Kozlowski (2000) addresses the systematic problem with such data sets. Such measures describe to what degree the ratings are interchangeable between team members and how strong team members agree in their perception of the situation. Summarizing the recommendations from multiple pieces of work in that area by Lawrence et al. (1993), LeBreton & Senter (2008), Koo & Li (2016), Smith-Crowe et al. (2013) it was decided to look into the within group agreement by means of the r_{wg} Index and into the across group agreement reliability by means of the ICC based reliability. Detailed tables of results can be found in chapter C.6.. Summary is that all project aggregations show r_{wg} values close to 1 (>0.96). That suggests a high level of within group agreement on the situation described by Q1 thru Q8. (Klein et al. (2000), Bliese (2000), Lawrence et al. (1993)). In addition the ICC has been analyzed and it can be concluded that also across group agreeability reliability calculations show a moderate reliability of above 0.5 for the given sample. 0.5 – 0.7 ICC are considered moderate reliability by Koo & Li (2016). Therefore, it can be assumed that the

data can be aggregated and grouped into a first level data structured as it will be described later in chapter 10.2.1. The data is further on referenced as “Peer results” in this study.

10.1.2.1.2 PROJECT RESULTS PERFORMANCE (PO EXPECTATIONS)

In addition, the performance results of the project are recorded once the project ends. These measures represent the external view on project performance as pre-negotiated with the PO. Expectations are communicated and negotiated with the team before the project works starts. An example of a documented expectation matrix can be found in Figure 19 in Appendix A.3. Measurements are percent values (0%-100%) on how the team fulfilled the expectation in each of the negotiated goals. Goals are weighted and cumulated to a project level single percent measurement as a single number project result. Results are transfer into the ESE tool for further processing. The data is described as “PO_Exp” in the cause of the following chapters.

10.1.2.1.3 TEAM GOALS PERFORMANCE (PO TEAM)

Team self-set goals are also measured. The teams are expected to set self-defined goals in addition to the externally PO_Exp driven goals in the expectation matrix. These goals are also graded by the PO after the project has ended using a percentage complete scale. Analogue to the project results, individual goals are rated using a percentage scale (0-100%). Goals are weighted and consolidated into a single team goals performance value on a percent scale. An example is in Figure 19 in Appendix A.3. Team goals are documented and included in the Expectation Matrix and rated there. Results are also transferred into the ESE tool for further processing. The data is referenced as “PO_Team” further on.

10.1.2.2 ONE-TIME TEAM MEMBER DATA (INDIVIDUAL INPUT)

Data about every team member on surface-level and deep-level diversity factors were measured and collected one time per member during the time frame of this project.

10.1.2.2.1 SURFACE-LEVEL DIVERSITY DATA

A set of surface-level data marks the first block of data related to individuals. Data consists of birthdate coded by full year compliant with ISO 8601. See ISO (2019). Age when project was

performed is also recorded and coded according to ISO 8601. In addition, the start date with the organization is measured and coded in accordance with ISO 8601. That data element allows to calculate the tenure with the organization at the point in time when the project was performed. Gender is part of the data collection as well. While the common coding of ISO 5218. See ISO (2004) is used, only male and female coding appear for the team at question here. Education is coded in categories as: no formal [0], College [1], Bachelor [2], Master [3], and Doctoral [4] level, or equivalent. This section is following the recommendation from the UNESCO ISCED-97. See UNESCO (1997) education classification system. During literature review for this study, it was found that most researchers use the above mentioned categories or very similar ones to classify educational levels. Job level grade is also recorded. The organization classifies jobs of the same category into six levels: Researcher Level 1, or Junior Researcher, to Researcher Level 5, or Principal Researcher. Levels are coded by integers. No project management or management level is involved directly in any of the projects due to the nature of self-managed teams. Thus, no other coding is needed. Country of origin is assessed by collecting the country codes that the team member originated from. The data is coded according to ISO 3166-1 Alpha-3. See ISO (1999). For further processing the data into diversity measures, country codes are converted into integer values. The same is true for the current country of residence for every team member. Coding and integer conversion follow the same method. The native language spoken is coded the same way. Decoding tables can be found in Appendix B.4.

10.1.2.2.2 DEEP-LEVEL DIVERSITY: PERSONALITY

For individual personality measures, information about the character traits along the NEO-PI-3 concept were recorded. The data fields neuroticism [N], extraversion [E], openness to experience [O], agreeableness [A], and conscientiousness [C] are the summary of the NEO-FFI test that most team members provided to the research (see chapter 9.2.4.1). All dimensions were coded according to the NEO-FFI test version used by the study using a scale of 0 to 50. For example, 0 refers to low/no agreeability and 50 refers to high agreeability. Based on country level work by Bartram (2013), a comparison of descriptive statistics is shown in Table 1 Descriptive statistics comparison for Big5 for the collected sample. The data was rescaled to 1-10 dimensions to match data frames used by Bartram. The country

level findings are compared to the data from this study which are listed as Sample in the table.

	N		E		O		A		C		
	M	SD	M	SD	M	SD	M	SD	M	SD	M SD
Germany	6,08	2,00	5,96	1,88	6,08	1,86	5,87	1,80	5,32	1,84	1,88
United States	5,70	2,05	5,84	2,09	5,29	2,05	5,34	1,97	5,72	2,03	2,04
Poland	4,99	2,01	4,64	2,03	5,84	1,81	4,75	1,69	5,35	2,07	1,92
Denmark	6,38	1,82	6,37	1,98	5,60	2,01	5,76	1,86	5,51	2,06	1,95
Sample	3,90	0,70	4,90	1,20	5,10	1,20	5,70	1,10	5,50	1,30	1,10

Table 1 Descriptive statistics comparison for Big5

The team members sample is a rather homogeneous group compared to the country level samples as indicated by the lower SD values. They also show a relatively low neuroticism mean. That might be rooted in the recruiting practices or relatively high education levels that act as filters compared to average population. The comparatively low extraversion can be attributed to the approximately 25% Polish researcher group being lower in country mean and influencing the group's mean, as well as being exclusively software engineers, which have an average lower extraversion according to research by Aqeel et al. (2019).

Based on the findings from Moynihan and Peterson (2001) that we discussed earlier, two decisions for the study here are derived. One is to include a diversity measure for the individual traits per team. Or in our case, per project, which defines an unique team constellation. As such the study converted all team level data to the recommended mean level and variance variables as input into the further analysis. For variance the calculation the standard deviation (SD) was chosen as this also the most widely used methods so far in related research as recommended by Moynihan & Peterson. This will allow a certain level of compatibility going forward. This inclusion of SD data (variance) will also allow for the recommended control for effects between the mean level and the SD for personality attributes in regression analysis later. Second is related to the within group Peer level data that is available to this study here. As we have seen in chapter 10.1.2.1.1 the data collected here describes and models a limited section on communication, collaboration, and contribution. Moynihan and Petersen (p. 346 & p. 357) mark such data as measures defining process quality, hence mediators/moderators, influencing member satisfaction with group. Therefore the study here will perform a mediation analysis based on the results of the regressions analysis in chapter 10.6. This step will also allow to test for a few more hypotheses found earlier.

10.1.2.2.3 DEEP-LEVEL DIVERSITY: CULTURE

In the field of deep level-diversity, cultural background for every individual is included in the data. Culture is measured using a formal assessment through culture dimensions. TBC has been using a commercially available tool to help deal with culture and culture differences within and outside of its organization for over a decade now. The tool is sold, managed and designed by Aperian Global (2020) and is called GlobeSmart. The tool itself is designed around culture dimensions, such as the very commonly used construct from Hofstede (2011) in its various states and the latest version from 2011. This results in a GlobeSmart Profile along five dimensions for a given team member:

- Independent ← → Interdependent
- Egalitarianism ← → Status
- Risk ← → Certainty
- Direct ← → Indirect
- Task ← → Relationship

As GlobeSmart has large amounts of data available to reference, most of its gained value does not arise from self-reflection on a personal profile. The value is in accessing profiles of whole organizations, teams and countries, and directly comparing them. The tools also provide suggestions on what the challenges may be when faced with culture differences by e.g., traveling abroad and gives the user advice on what to do or not and what to look out for. As almost all team members in this study have an existing GlobeSmart profile and shared it with the author, the profiles were used as the basis of analysis for culture diversity in this study. GlobeSmart does give academic reference to their work and validity. Doherty (2016) shows the reliability, scale composition and construct validity for the psychometric properties of the GlobeSmart concept. Some company internal work has also be done to evaluate its applicability to the organization by Schiefele (2011). All values are coded in a 1 through 5 scale, 1 being the value for low and 5 being the value for high. Figure 21 in Appendix B.5 shows a country level comparison extracted from GlobeSmart for the sample population collected. Following the arguments risen by scholars around the level vs variance of parametrized data like personality and their influence on group internal mechanisms the author decided here to introduce the same measures as we have seen in the previous chapter

for personality. Moynihan and Peterson (2001). Standard deviation values describing variance are calculated as well as mean values for all five parameter of culture. Compare 10.1.2.2.3.

10.1.2.3DATA MODEL

Figure 10 provides a Unified Markup Language (UML) model of the data structure. The interdependencies described in the diagram also help to understand the data consolidation and conversion described in 10.2.1. The data in raw form shows a one level nested structure and therefore requires further data handling and conversion to adjust for the intended analysis. Some rational was already explained in previous chapters.

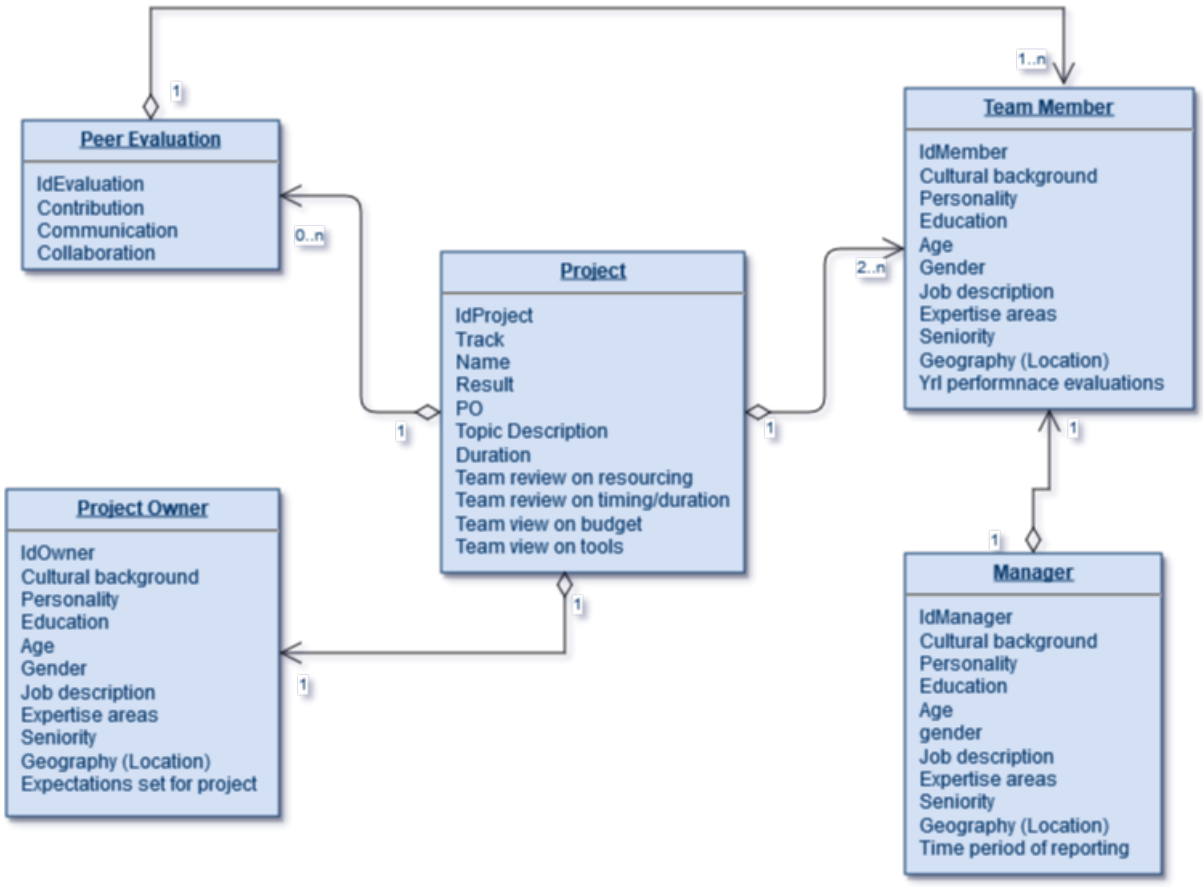


Figure 10 Data structure UML nesting

10.1.2.4DATA PRIVACY, ETHICS AND CONSENT

Due to the nature of the data collected, data privacy, ethics and consent precautions needed to be taken.

Two aspects are important here. First, the legal aspects of handling and processing data within the enterprise and its subsidiaries. Second, the aspects of transparent handling and processing of data according to policies of the institution on the academic side of this work, the Durham University (DU).

The aspect of policies of the institution (DU) will be discussed first, starting with the ethical considerations of the University (2020) and the Durham Business School (2015). The program handbook references the Durham University rules. When reviewing the Ethics considerations, it becomes clear that the author needs to address two aspects. One is the "collection or processing of sensitive personal data (including from secondary sources) without explicit consent" and the other is "sensitive personal data transfer to partners outside the EEA" (Durham University 2020).

The first aspect of sensitive personal data is partially covered by the fact that the data is anonymized. The reason for that is twofold. First, it is to protect sensitive personal information for the sake of the University policies. Second, it is to satisfy General Data Protection Regulations (GDPR) requirements and work contract related legal reasons imposed by the organization in which the data is collected. The employer is partially located in Germany and Poland and as such subject to the GDPR of the European Union (EU) (2016) and local German and Polish labor law. A small portion of the test subjects are in the United States of America (USA). To ensure compliance with the policies, numerous steps were taken. The full process of data collection, handling and storing was reviewed and disclosed with the relevant bodies listed below:

- Polish entity's Human Resource department
- Polish entity's Workers Council
- USA entity's Human Resource department
- German entity's Human Resource department
- German entity's Workers Council (WOC)
- European Data Privacy Officer of the Boeing Company (DPO)

Several decisions were taken to gain approval from the above parties, which is the second aspect of the DU regulations about consent:

- All data is anonymized during the collection and conversion process. This step is taken as early as possible in the process to ensure minimal risk of data breaches. The data is transformed into an ID based system that only allows for matching one ID based record to another one. The coding-key to that process is a strict "need to know" documented process. As additional measures, any data that contains names is deleted and destroyed after that conversion step. This process is reviewed and documented with the DPO and the workers councils. Dependent on legislation, it is done e.g., by a 'Verfahrensmeldung' (procedure report to the authorities) or 'Betriebsvereinbarung' (company agreement) in Germany. After that step, it is no longer considered Personal Identifiable Information (PII) by EU regulations.
- Consent on collection for all data was discussed with all participants/team members. As agreed with the works councils and HR departments, all participation in any of the data collection is strictly voluntary. Every member can freely decide to provide data and/or opt in or out at any given point in time. All these aspects have been discussed and disclosed with all team members in meetings (including WOC and HR), briefings and documentation.
- The goals of the study were fully disclosed to every participant and each was ensured that full anonymity is built into the data. Results are shared at request of any of the participants.

Following the above logic, process, documentation and decisions, it can be concluded that the aspects of sensitive data handling are covered and satisfied, and thus not subject to further review.

The handling of the export compliance requirement is also a twofold aspect. One point is the potential threat to export or transfer data out or in the European Economic Area (EEA). Through fully complying, documenting, auditing and following the procedures of GDPR as an organization, that aspect was also reviewed with the DPO. The conclusion was that since the data is anonymous and the process is documented, the export aspects of GDPR are covered. The second aspect which was reviewed is the subject of International Traffic in

Arms Regulations (ITAR) respectively the EU Export Administration Regulations (EAR). While that aspect is not explicitly mentioned in the context of the University regulations, they are fully applicable to the enterprise in which the data for this study was collected. A review took place with the enterprise assigned Export Compliance Office to review the data. It was concluded that it does not fall under ITAR or EAR regulations.

Overall, the author states that any necessary ethical and regulative concerns have been addressed to ensure full compliance with the University's regulations and the legal and regulative aspects imposed in general and by the enterprise in which the study was conducted.

10.1.2.5HANDLING OF MISSING DATA

Due to the schematic approach of how data is collected for this study, and due to the voluntary contribution of data, it is overt that the set contains missing data values. Three missing data problems needed to be addressed.

The first missing data problem is when a team member, per project or per peer evaluation, does not provide data/feedback at all. While that case seems to be trivial in terms of mathematical complexity (no data structure, no problem), it bears some fundamental validity questions for an analysis. In cases of using diversity driven indices to measure and define the nature of a project team and to then correlate theses indices to project performance, the risk that a team member does not provide data at all could be a statistically driving factor. In other words, if a rather "unique" team member, like the only speaker of the Danish language, does not enter data, it would have significant effect to a diversity factor of, for example, a team including this person and three English native speakers. It will make the team appear as homogeneous while in reality being skewed in native language. There is no systematic and simple way of dealing with such cases other than eliminating samples. There is some aspect in the current study to cope with the issue. The management team has a full overview over the full true composition of the teams since 2011. Also known are the responses and input of each individual team member. After consolidating and managing all raw data records, two cases existed where a team member did not give any responses at all. Hence these two cases where removed from the data and subsequently from the study. The method used here is

usually described as case deletion or list-wise deletion. Compare with Schafer and Graham (2002).

The second case is by far the most common problem in dealing with survey-based data sets and voluntary collections: incomplete or partial records. Reasons for incomplete records vary widely from personal preference to not disclose, to technical problems while dealing with online surveys. The data set collected in this study is no exception. Multiple cases in the data needed to be assessed and handled due to this cause.

The third case for missing data is mostly a problem in long-term studies. As this dataset is spanning seven years of data collection, team members joined or left the enterprise during that time frame. As with many long-lasting studies, during the time of data collection, insights and research approaches evolve over the years. In such cases, some data collection is only done at certain points in time. In this study, for example, only two rounds of assessing the Big5 personality profiles (NEO-FFI) were conducted over the years. As every assessment tends to require works council and senior management approval, efforts are bundled and limited. The fact that some team members had already left the company or had not yet joined, led to a few cases where certain data pieces, like the Big5, are missing. In this study two cases were identified with such a condition.

The last two cases require different handling than the first case since list-wise deletion would result in distortion of the analysis based on the arguments listed for the first case. Therefore, some more detailed research was done around missing values in the cases at hand. The problem is usually called wave nonresponse, as discussed in Enders (2010) and Schafer and Graham (2002). While the underlying statistical theory around modern handling of missing data is quite complex and theoretical, the recommendations for the case here are relatively straight forward. As a missing response is considered its own statistical process, it is assumed that it underlies either random processes or not. If it is random data, it is classified as Missing Completely at Random (MCAR). In this case, a statistical model can be used to fill random values. The data is analyzed and randomly distributed, and therefore MCAR. Based on this, averaging the missing values is the recommended method. In the data itself, the missing values are coded as a -99 signed integer to indicate missing data to the software processing the data to force appropriate handling.

10.1.2.6OPERATIONALIZING DIVERSITY

The data described up to this point is comprised of absolute measures of surface-level and deep-level diversity and performance measures. The data itself does not provide readily available insights into diversity as a measure. In this chapter some special aspects of how to operationalize diversity are discussed as this is the core of this study.

To operationalize diversity, Harisson and Kleins's (2007 p.1203) descriptions of minimum and maximum definitions of the categories in their typology is used. Describing the variation for this study means that separation minimum diversity is reached when members agree on a topic. If members separate into two oppositional groups in opinions around a topic, diversity maximizes. In disparity, a minimum is defined as e.g. all members having exactly even pay. Maximum is defined by the fact that e.g. one team member has a distinctly higher pay than all the rest. Variety is minimal when team members, for example, all share the same knowledge about a subject. It maximizes if knowledge around that subject is dispersed and different. Following these definitions and recommendations by Harrison and Klein, it can be derived that variety, in contrast to the other two diversity types, tends to have positive influence on predicted outcomes, while the other two have negative influence. The data in the study describes variety rather than separation or disparity. This conclusion is vital to the decision on how to further treat diversity in the data. Therefore, it was decided to use evenness to operationalize diversity as variety in this study. Compare also chapter 9.5.

Evenness measures are used on all dimensions describing surface-level diversity in the data. These dimensions, or categories in Blaus terminology, are ordinal (categorical) values. The Blaus Index is ideal to describe the variety within ordinal values. It was decided to convert one more attribute into categorical values: age. While age is usually a ratio variable, it can also be treated as a categorical value without losing much fidelity in the study. The age variable was converted into an ordinal value to allow for calculating evenness for every project. The full range conveniently reached from 25 to 70 years of age for the included persons. Thus, a 5 years range cluster was formed and coded into the data, e.g., Category 1 being 25-30 years of age. Such a method is also supported by Solanas et al. (2012).

For all measures in deep-level diversity, it was decided to stay in the commonly used schema of mean levels of dimension (see chapter 9.2.4). While it does simplify the aspects of

diversity measures drastically, using the evenness method on personality and culture dimensions would not allow an analysis of the derived hypotheses and would not allow for situating the results in the existing research body. Following the debate from chapter 9.2.4.1 and the arguments from Moynihan and Peterson (2001) that we reflected earlier a combination of level and variance (SD) is used and calculated for all deep-level data parameters. This will allow for one to ensure a compatibility with earlier research being focused more on levels and second to connect to more modern level and variance driven concepts.

10.2 ANALYSIS AND TECHNIQUE

The following chapter describes the four steps of analysis performed in this study. The first step was the data conversion and formatting; the second the descriptive parametric statistics; the third the correlation analysis; and last the analysis for regression including some control analysis and moderation analysis.

When it comes to the methods of assessing and analyzing diversity and its differences in the team context, two visible trends exist.

One trend is the approach to find data that has enough volume, and as such enough statistical power, N, to allow for a statistical analysis along individual factors that contribute to the compositional features of teams. This approach is widely used. A meta study summarizing many of them is done by Bell (2007). In this paper, many of them listed in context of more detailed contributions to team performance. One problem with this approach is that it often forces a compromise in the context of the data for making sufficiently large data sets which fit the need of a large N for statistical power. Some studies use multiple data sources of different companies and combine them to allow for a large statistical base, e.g. Peters (2015), Beal et al. (2003), Ely and Thomas (2009) and Brickson (2000). As argued earlier, this bares some potential weakness in the fidelity of results, as some measures used in the studies and data might not be comparable or of enough direct quality to draw precise conclusions.

The second approach is a wide variety of statistical methods to describe and analyze diversity as mathematical constructs and aggregations. There are some research pieces contributing temporal summaries of the efforts, such as Mathieu et al. (2014). Mathieu et al. group the

methods of assessing individuals and their relations to teams into four groups along two dimensions: individual focus and team focus, which is a model based on theoretical work from Kozlowski and Klein (2000).

1. Traditional Personnel- Position Fit Model (Individual Focus- Individual Models)
2. Personnel Model with Teamwork considerations (Team Focus- Individual Models)
3. Team profile Model (Individual Focus- Team Models)
4. Relative Contribution Model (Team Focus- Team Models)

Following that classification, the study design of the work here falls into the second category: Personnel Model with Teamwork considerations. The data has a strong focus on individual characteristics, and at the same time provides indices into the team processes (peer review aspects). It has a robust performance model to test against. The individual characteristics are described by Mathieu et al., as a team member's knowledge, skills, abilities and composition (KSAO). KASO conceptually represents the data that is available to this study.

When reviewing some frameworks listed by Mathieu et al. (2014 p.134 ff) for all four categories, some trends arose in methods used to analyze within these categories. Categories three and four (team models focus) require complex methods, such as the multilevel analysis in Zyphur (2009), Mackenzie (2015), and Maynard et al. (2012), factor analysis and Faultline methods by Zanutto et al. (2010) and Homan et al. (2007), cluster-theory by Tonidandle and Overall (2004) and round-robin by Nestler et al. (2015) and Schönbrodt et al. (2012). Categories one and two tend to use more traditional methods, such as regression analysis and moderation. The modeling of the constructs used to analyze are simpler and more linear in categories one and two. Given the category two design of this study, correlation and regression analysis were used while diving into some aspects of mediation effects and controls for personality and culture parameters as suggested by more modern concepts. Some round-robin respectively within group agreements analysis has been done to consolidate data into one level as the original data consisted from two levels.

All diversity attributes were used as the independent variable in the following correlation and regression analysis. The three available performance measures were the dependent variables.

10.2.1 DATA CONVERSION

Data collection, as that collected in this study, poses additional challenges to researchers and analysts when trying to decode and gather insights from them. Data is usually not readily structured for analysis and answering questions directly. This study required extensive work to organize, structure and analyze the data. Assuming a realistic scenario of management posed with the task to compose teams in a work environment, some limitations arose. In this scenario, the pool of resources that a manager could choose from is limited and not infinite, as suggested by many models. Exchanging within the pool or adapting the pool is penalized with high effort and cost, such as organizational change, hire and fire decisions and performance management over long periods of time. A manager is less interested in the ideal team member as an individual, but rather interested in the most ideal combination of members in the pool at hand. Taking these considerations and coming back to the data structure, fundamental decisions were taken on how to organize the data for analysis and how to deal with diversity to support that analysis. Two aspects are key.

First, the goal of this work does not require an analysis of how an individual member fits a team. The individual contribution, or fit, is rather irrelevant to the manager facing the practical task to influence team performance. They will have to focus on factors like: will a team with higher diversity in certain factors perform better or worse in comparison to a more homogeneous team? For example, is there any indication that teams speaking the same native language are performing better than such with a high variety in native languages? Common sense and managerial street smartness would suggest that. But is that really the case? What if such variety fosters creativity and leads to better results?

Secondly, the way that data is organized and analyzed should allow for comparison with other studies around team efficiency.

For the reasoning above, data was organized and aggregated into project-based data sets. This means that each individual project carries a set of diversity factors and performance results to feed further analyses. If not otherwise stated, data aggregation and conversion was performed with R Studio and small software converters written by the author in the Python programing language. Data conversion here is also necessary as the data is not managed in a relational database. Such a database can be used with the R software package as well. As this study is a one-time effort, a more conventional way has been used for cost/benefit reasons.

The first necessary data conversion step was to assemble all the data collected into a master file of data. Files containing peer review results, NEO-FFI results and cultural dimensions from the Globe Smart tool were consolidated and converted. Simple keying of data has a data quality of 10^{-3} error rate, according to the Radio Technical Commission for Aeronautics (RTCA) Special Committee 181 (2015). Therefore, automated digital data processing was used in this study to retain data quality.

The second step was to convert all survey responses into a numerical, machine-readable format. Categorical data formats are needed when textual coding is present. A full coding schema was developed to do so and can be found in Appendix B.4 Table 18.

The third data conversion step was to aggregate the data per person within a project. In peer review data, multiple records of feedback from multiple team members can exist, since multiple people answer for each member. Therefore, these records needed to be consolidated into one record per project per person. The method for merging the records to limit it to one per project was to build the average over the individual peer review responses. A partial view can be found in Appendix B.4 Table 19.

Lastly, personality and culture dimensions data required calculation of SD values on the full scale data frames for inclusion into project level data.

All other diversity attributes described in chapter 9 needed to be brought to a state that one can perform numerical, statistical analysis on them. As previously discussed, some, such as GlobeSmart and Big5, are represented by continuous measures and are therefore directly usable for correlation and regression analysis. For the categorical values, mainly in the surface-level category, the decided concept of evenness needed some pre-work to allow for analysis, in contrast to the continuous measurements. For this reason, several small steps were required to convert and calculate the appropriate values for these data types. A software sub-package in R named “Diverse” by Guevara et al. (2017) was used to compute the evenness for the data. Unfortunately, that package requires data branching and a diagonal mirroring/vectoring of data fields as Diverse requires a frequency count list as input. For that reason, the data needed to be broken out and converted to the above format. Most of it could be automated in R and Python. After the split of data for calculations, it was brought back together and merged into a similar format as before. The difference was that the new data

format was described per project in terms of evenness per data field. Examples can be found in Appendix B.4 Table 20 and Table 21.

Lastly, the peer performance data, the expectations result from the POs and the team results needed to be brought into a usable format. The expectation results and peer results comprise the dependent variables in the study. All values were already in continuous measures 0-100.

Through the process of preparing and initially checking data quality, one issue with the data types arose. The variable “age” is systematically related with tenure in the company. As tenure in the data is defined as tenure with the company and not with the team itself, it was decided to focus on age only, as tenure does not relate to any relevant team aspects seen in the research here. One possible explanation is that recruiting and hiring into the organization at hand is mainly done through a strong university involvement program and as such mostly graduates directly from university are hired.

10.3 STRUCTURE OF ANALYSIS

The overall statistical analysis was done in the software package R and R Studio by the R Foundation (2020). R is an open-source software package designed to perform statistical computing. TBC also provides enterprise hosting for the full suite of R. Two aspects of structure of analysis are contextually important.

One is the structure of dependent and independent variables. The figure below shows what data was used as dependent variables and independent variables. It also shows grouping of data on the independent side. This structure is used to further process the data.

Dependent Variables		
Input		
	PO_Exp	Performance rated by PO based on pre-negotiated Matrix
	PO_Team	Performance rated by PO based on Team set goals
	Peer	Peer review team agreement
Independent Variables		
Surface-Level		
	Gender	
	Age	
	Origin	Country of origin
	Residence	Country of Residence
	Language	Native Language
	Education	Education Level
Deep-Level		
	Neuroticism Level	Personality parameters Level and SD
	Neuroticism SD	
	Extraversion Level	
	Extraversion SD	
	Openness Level	
	Openness SD	
	Agreeableness Level	
	Agreeableness SD	
	Conscientousness Level	
	Conscientousness SD	
	Independence Level	Culture parameters Level and SD
	Independence SD	
	Egalitarianism Level	
	Egalitarianism SD	
	Risk Level	
	Risk SD	
	Direct Level	
	Direct SD	
	Task Level	
	Task SD	

Table 2 Dependent and independent Variables

The second is to address the missing data. While missing data cases on the raw data level were addressed earlier, aggregating and converting data led to missing data cases as well. In some cases, some pieces of data are only available for one team member even though two, three or four participated, e.g., only one person took the culture traits assessment. As the processing software does not care about those aspects and simply assumes one result, or averages, to process, it would lead to wrong assumptions and results for this study. So, a data clean-up was performed. The data is best grouped into three data blocks with different N to preserve statistical validity and power, but still be able to analyze to full depth where possible. The first is surface-level diversity and education level. All these sets share an N=111. The next block, with an N= 109, contains the first block plus the personality traits. The final and third block include all variables with an N=87. In summary, three rounds of list-wise deletions were performed based on incomplete data factor lists to preserve statistical power.

10.4 PARAMETRIC TESTING AND DESCRIPTIVE STATISTICS

This section describes the parametric analysis and descriptive statistics analysis. Both steps are important to understand implications towards correlation testing and regressions analysis. Two steps were performed. One was the analysis of the raw data and the second was the analysis of the consolidated, aggregated data set.

10.4.1 RAW DATA RESULTS

Descriptive statistics and parametric testing on data provides knowledge on base data which can be used to drive decisions on what and how to use other statistical analyses and validate the null hypotheses. Many correlation tests require normally distributed, or closely resembling normally distributed, data. As it is represented in Appendix C, all data groups were subject to density/histogram plots, Q-Q plots and boxplots. These show that most data is normally distributed and shows no artifacts like outliers or a spread between average and mean. The exceptions are PO performance and team goal performance. Measurements are subject to some skew in the data, as the results are negotiated by teams and POs. Hence, it is not a fully random effect. Other research shows that grades in education are generally not normally distributed. A recent study at Stanford from 2019 by Arthur et al. (2019) analyses 4000 assignment grades and comes to the conclusion that, despite a wide-spread assumption, grades are not normally distributed. Comparing the grand average distribution from that study with descriptive statistic from the study here, it resembles the same behavior. To get further clarity on the data at question, two more tests were done to check for normal distribution more thoroughly. A Shapiro-Wilk Test and, in case the data resembles only parts of a distribution curve, a Wilcox Signed Rank test were performed on the data groups (see chapter C.3 and C.4). None of the tests suggested that the data follows a normal distribution. The tests performed are all significant. Analogue was true for the Wilcox test for partial normality.

Aggregating and converting data can therefore be done without distribution related limitations on the raw data. PO performance and team goals performance must be treated or excluded, accordingly, when normally distributed data is expected. As no conversion is necessary on these two measures, no action was needed.

10.4.2 AGGREGATED DATA RESULTS

The aggregated data sets form the basis for regression and correlation analysis. Here a Shapiro-Wilk test is done on all applicable aggregated data. The results in Appendix C.5 show that only two variables are normally distributed, but the test is not significant. One is the dependent variable, peer performance. This is unchanged from the raw data. The second is the independent variable, agreeableness, in the Big5. The calculation of the Shapiro-Wilk was repeated for all three sample sizes (N=11, N=109, N=87) and showed consistent results.

This analysis suggests not using the Pearson's correlation (Pearson's R), as it requires the base data to be normally distributed. The Spearman correlation coefficient/Spearman Rank Correlation (ρ) does allow for non-parametric input, if it shows linearity (monotonic relations), which is the case for all the data at hand. Another reason for using Spearman's correlation is the presence of continuous and ordinal values, which is also the case here. Therefore, further testing was parameterized using a "spearman" *method* in R Studio (2020). For the two normally distributed data groups, one check run was done using Pearson's R. Results will be discussed in the next chapter.

10.5 CORRELATION ANALYSIS

Following derived knowledge from the previous chapter, three correlation calculations were performed, one each for N = 111, 109, 87.

	PO_Exp	*PO_Team*	*Peer*	*Language*	*Orgin*	*Residence*	*Age*	*Gender*	*Education*
PO_Exp									
PO_Team	0.817 *(.011)*								
Peer	0.033 *(.948)*	-0.100 *(.810)*							
Language	-0.367 *(.336)*	-0.417 *(.270)*	-0.683 *(.050)*						
Orgin	-0.567 *(.121)*	-0.517 *(.162)*	-0.650 *(.067)*	0.967 *(<.001)*					
Residence	-0.367 *(.336)*	-0.367 *(.336)*	-0.733 *(.031)*	0.950 *(<.001)*	0.917 *(.001)*				
Age	-0.583 *(.108)*	-0.717 *(.037)*	0.367 *(.336)*	-0.100 *(.810)*	0.050 *(.912)*	-0.167 *(.678)*			
Gender	0.200 *(.613)*	0.550 *(.133)*	-0.650 *(.067)*	0.233 *(.552)*	0.217 *(.581)*	0.317 *(.410)*	-0.617 *(.086)*		
Education	-0.600 *(.097)*	-0.733 *(.031)*	-0.083 *(.843)*	0.667 *(.059)*	0.717 *(.037)*	0.617 *(.086)*	0.500 *(.178)*	-0.200 *(.613)*	

Computed correlation used spearman-method with listwise-deletion.

Table 3 Correlation matrix (n=111) (Surface-Level Diversity, Education, Performance)

	PO_Exp	PO_Team	Peer	Language	Origin	Residence	Age	Gender	Education	N	NS	E	ES	O	OS	A	AS	C	CS
PO_Exp																			
PO_Team	0.528 *(.022)*																		
Peer	0.312 *(.193)*	0.025 *(.922)*																	
Language	-0.181 *(.457)*	-0.100 *(.683)*	-0.475 *(.041)*																
Origin	-0.219 *(.365)*	-0.109 *(.657)*	-0.456 *(.051)*	0.988 *(<.001)*															
Residence	-0.182 *(.453)*	-0.046 *(.854)*	-0.568 *(.012)*	0.961 *(<.001)*	0.933 *(<.001)*														
Age	-0.175 *(.471)*	-0.482 *(.038)*	0.070 *(.776)*	-0.068 *(.781)*	0.021 *(.934)*	-0.214 *(.377)*													
Gender	0.070 *(.776)*	0.432 *(.066)*	-0.409 *(.083)*	0.661 *(.003)*	0.649 *(.003)*	0.749 *(<.001)*	-0.418 *(.077)*												
Education	-0.160 *(.512)*	-0.212 *(.381)*	-0.163 *(.503)*	0.775 *(<.001)*	0.835 *(<.001)*	0.688 *(.002)*	0.326 *(.173)*	0.477 *(.041)*											
N	0.430 *(.068)*	-0.188 *(.440)*	0.328 *(.170)*	0.147 *(.546)*	0.154 *(.527)*	0.042 *(.865)*	0.281 *(.243)*	-0.104 *(.673)*	0.411 *(.082)*										
NS	0.282 *(.240)*	-0.109 *(.657)*	0.379 *(.110)*	-0.826 *(<.001)*	-0.833 *(<.001)*	-0.818 *(<.001)*	0.175 *(.471)*	-0.765 *(<.001)*	-0.728 *(.001)*	0.005 *(.986)*									
E	-0.439 *(.062)*	0.075 *(.759)*	-0.626 *(.005)*	0.719 *(.001)*	0.714 *(.001)*	0.781 *(<.001)*	-0.182 *(.453)*	0.718 *(.001)*	0.442 *(.060)*	-0.472 *(.043)*	-0.756 *(<.001)*								
ES	0.346 *(.147)*	0.160 *(.512)*	0.351 *(.141)*	-0.875 *(<.001)*	-0.896 *(<.001)*	-0.863 *(<.001)*	0.005 *(.986)*	-0.658 *(.003)*	-0.856 *(<.001)*	-0.054 *(.826)*	0.867 *(<.001)*	-0.733 *(.001)*							
O	-0.279 *(.247)*	0.207 *(.393)*	-0.654 *(.003)*	0.707 *(.001)*	0.698 *(.001)*	0.779 *(<.001)*	-0.330 *(.168)*	0.784 *(<.001)*	0.418 *(.077)*	-0.423 *(.073)*	-0.702 *(.001)*	0.937 *(<.001)*	-0.707 *(.001)*						
OS	-0.140 *(.565)*	0.112 *(.646)*	-0.184 *(.449)*	-0.649 *(.003)*	-0.654 *(.003)*	-0.567 *(.013)*	0.042 *(.865)*	-0.347 *(.145)*	-0.698 *(.001)*	-0.584 *(.010)*	0.577 *(.011)*	-0.105 *(.667)*	0.609 *(.007)*	-0.077 *(.754)*					
A	-0.188 *(.440)*	0.268 *(.265)*	-0.612 *(.006)*	0.486 *(.037)*	0.488 *(.036)*	0.575 *(.011)*	-0.305 *(.203)*	0.784 *(<.001)*	0.286 *(.234)*	-0.432 *(.066)*	-0.560 *(.014)*	0.818 *(<.001)*	-0.511 *(.027)*	0.900 *(<.001)*	0.130 *(.595)*				
AS	0.249 *(.302)*	0.102 *(.678)*	0.398 *(.092)*	-0.874 *(<.001)*	-0.877 *(<.001)*	-0.863 *(<.001)*	0.154 *(.527)*	-0.714 *(.001)*	-0.774 *(<.001)*	-0.077 *(.754)*	0.833 *(<.001)*	-0.677 *(.002)*	0.905 *(<.001)*	-0.682 *(.002)*	0.654 *(.003)*	-0.521 *(.024)*			
C	-0.347 *(.145)*	0.135 *(.580)*	-0.518 *(.025)*	0.718 *(.001)*	0.746 *(<.001)*	0.749 *(<.001)*	-0.079 *(.748)*	0.770 *(<.001)*	0.586 *(.010)*	-0.360 *(.131)*	-0.789 *(<.001)*	0.949 *(<.001)*	-0.796 *(<.001)*	0.912 *(<.001)*	-0.195 *(.423)*	0.828 *(<.001)*	-0.742 *(<.001)*		
CS	0.242 *(.317)*	0.149 *(.541)*	0.339 *(.156)*	-0.882 *(<.001)*	-0.904 *(<.001)*	-0.854 *(<.001)*	-0.011 *(.968)*	-0.605 *(.007)*	-0.854 *(<.001)*	-0.240 *(.320)*	0.826 *(<.001)*	-0.605 *(.007)*	0.916 *(<.001)*	-0.582 *(.010)*	0.754 *(<.001)*	-0.337 *(.159)*	0.921 *(<.001)*	-0.674 *(.002)*	

Computed correlation used spearman-method with listwise-deletion.

Table 4 Correlation matrix (n=109)

	PO_Exp	PO_Team	Peer	Language	Origin	Residence	Age	Gender	Education	I	IS	EG	EGS	R	RS	D	DS	T	TS	N	NS	E	ES	O	OS	A	AS	C	CS
PO_Exp																													
PO_Team	0.295 (.120)																												
Peer	0.096 (.621)	-0.113 (.559)																											
Language	-0.242 (.205)	0.017 (.930)	-0.436 (.019)																										
Origin	-0.407 (.029)	-0.066 (.733)	-0.410 (.028)	0.963 (<.001)																									
Residence	-0.257 (.178)	0.193 (.315)	-0.458 (.013)	0.920 (<.001)	0.921 (<.001)																								
Age	-0.273 (.151)	-0.724 (<.001)	-0.015 (.938)	0.201 (.294)	0.270 (.156)	-0.009 (.962)																							
Gender	-0.160 (.405)	0.538 (.003)	-0.251 (.188)	0.271 (.154)	0.289 (.128)	0.529 (.004)	-0.508 (.005)																						
Education	-0.449 (.015)	-0.469 (.011)	-0.227 (.235)	0.613 (.001)	0.696 (<.001)	0.482 (.009)	0.678 (<.001)	-0.041 (.833)																					
I	-0.083 (.667)	0.472 (.010)	0.056 (.774)	-0.288 (.130)	-0.271 (.154)	-0.098 (.612)	-0.572 (.001)	0.589 (.001)	-0.464 (.012)																				
IS	0.550 (.002)	0.475 (.010)	0.089 (.644)	-0.276 (.146)	-0.389 (.038)	-0.139 (.469)	-0.618 (<.001)	0.128 (.508)	-0.531 (.004)	0.098 (.613)																			
EG	0.425 (.022)	-0.110 (.569)	0.156 (.417)	-0.284 (.135)	-0.389 (.038)	-0.458 (.013)	0.041 (.831)	-0.748 (<.001)	-0.175 (.363)	-0.472 (.010)	0.410 (.028)																		
EGS	0.493 (.007)	0.542 (.003)	-0.023 (.905)	-0.365 (.053)	-0.493 (.007)	-0.242 (.204)	-0.623 (<.001)	0.072 (.708)	-0.604 (.001)	0.194 (.313)	0.809 (<.001)	0.418 (.025)																	
R	0.387 (.039)	-0.314 (.098)	0.429 (.021)	-0.392 (.036)	-0.462 (.012)	-0.558 (.002)	0.194 (.312)	-0.772 (<.001)	-0.170 (.375)	-0.383 (.041)	0.267 (.162)	0.830 (<.001)	0.204 (.286)																
RS	0.360 (.056)	0.377 (.045)	-0.085 (.661)	-0.284 (.135)	-0.409 (.028)	-0.220 (.251)	-0.476 (.010)	-0.083 (.667)	-0.411 (.028)	-0.028 (.885)	0.773 (<.001)	0.593 (.001)	0.914 (<.001)	0.338 (.073)															
D	0.553 (.002)	-0.097 (.617)	0.045 (.817)	-0.285 (.134)	-0.405 (.030)	-0.400 (.032)	-0.070 (.718)	-0.630 (<.001)	-0.268 (.159)	-0.393 (.036)	0.468 (.011)	0.930 (<.001)	0.469 (.011)	0.760 (<.001)	0.624 (<.001)														
DS	0.367 (.051)	0.500 (.006)	0.124 (.520)	-0.514 (.005)	-0.584 (.001)	-0.361 (.055)	-0.569 (.002)	0.079 (.683)	-0.595 (.001)	0.275 (.148)	0.784 (<.001)	0.354 (.060)	0.916 (<.001)	0.214 (.263)	0.834 (<.001)	0.372 (.047)													
T	0.218 (.254)	0.332 (.079)	0.299 (.116)	-0.349 (.064)	-0.400 (.032)	-0.350 (.063)	-0.413 (.027)	0.149 (.438)	-0.522 (.004)	0.687 (<.001)	-0.040 (.837)	-0.090 (.643)	0.033 (.863)	0.036 (.851)	-0.129 (.503)	-0.040 (.835)	0.034 (.859)												
TS	0.383 (.041)	0.538 (.003)	0.256 (.179)	-0.328 (.083)	-0.455 (.014)	-0.236 (.216)	-0.592 (.001)	0.143 (.457)	-0.563 (.002)	0.321 (.090)	0.716 (<.001)	0.266 (.163)	0.843 (<.001)	0.123 (.525)	0.689 (<.001)	0.239 (.211)	0.771 (<.001)	0.189 (.324)											
N	0.176 (.359)	-0.301 (.112)	0.085 (.659)	-0.351 (.062)	-0.396 (.034)	-0.394 (.035)	0.138 (.474)	-0.559 (.002)	-0.113 (.559)	-0.329 (.082)	0.457 (.013)	0.709 (<.001)	0.392 (.036)	0.726 (<.001)	0.578 (.001)	0.698 (<.001)	0.454 (.014)	-0.274 (.150)	0.227 (.235)										
NS	0.485 (.008)	0.255 (.181)	-0.172 (.371)	-0.220 (.251)	-0.321 (.090)	-0.274 (.150)	-0.114 (.555)	-0.451 (.015)	-0.320 (.091)	-0.167 (.384)	0.237 (.214)	0.667 (<.001)	0.516 (.005)	0.412 (.027)	0.550 (.002)	0.698 (<.001)	0.429 (.021)	0.052 (.789)	0.272 (.153)	0.353 (.061)									
E	-0.314 (.097)	0.214 (.263)	-0.250 (.190)	0.461 (.013)	0.510 (.005)	0.591 (.001)	-0.091 (.637)	0.670 (<.001)	0.168 (.381)	0.353 (.061)	-0.501 (.006)	-0.836 (<.001)	-0.393 (.036)	-0.790 (<.001)	-0.509 (.005)	-0.724 (<.001)	-0.419 (.025)	0.165 (.392)	-0.342 (.070)	-0.800 (<.001)	-0.337 (.074)								
ES	0.207 (.280)	0.353 (.061)	-0.192 (.317)	-0.270 (.156)	-0.333 (.078)	-0.263 (.168)	-0.191 (.319)	-0.219 (.253)	-0.304 (.109)	0.203 (.290)	0.174 (.364)	0.406 (.030)	0.601 (.001)	0.190 (.322)	0.583 (.001)	0.386 (.039)	0.536 (.003)	0.161 (.402)	0.458 (.013)	0.251 (.188)	0.795 (<.001)	-0.217 (.258)							
O	-0.237 (.214)	0.326 (.085)	-0.360 (.056)	0.556 (.002)	0.568 (.002)	0.704 (<.001)	-0.181 (.346)	0.743 (<.001)	0.154 (.423)	0.293 (.123)	-0.327 (.084)	-0.769 (<.001)	-0.281 (.139)	-0.772 (<.001)	-0.371 (.048)	-0.650 (<.001)	-0.341 (.071)	0.111 (.564)	-0.266 (.163)	-0.695 (<.001)	-0.312 (.099)	0.946 (<.001)	-0.237 (.215)						
OS	0.266 (.163)	-0.305 (.107)	-0.056 (.772)	-0.030 (.877)	-0.079 (.682)	-0.238 (.212)	0.495 (.007)	-0.673 (<.001)	0.071 (.714)	-0.412 (.027)	-0.399 (.033)	0.374 (.046)	-0.122 (.526)	0.359 (.056)	-0.072 (.710)	0.373 (.047)	-0.183 (.340)	0.040 (.835)	-0.272 (.152)	0.163 (.396)	0.568 (.002)	-0.100 (.604)	0.357 (.058)	-0.176 (.359)					
A	-0.278 (.144)	0.247 (.195)	-0.172 (.370)	0.318 (.093)	0.390 (.037)	0.542 (.003)	-0.271 (.155)	0.848 (<.001)	0.024 (.901)	0.500 (.006)	-0.097 (.617)	-0.853 (<.001)	-0.275 (.149)	-0.703 (<.001)	-0.437 (.019)	-0.758 (<.001)	-0.238 (.213)	0.111 (.566)	-0.181 (.346)	-0.606 (.001)	-0.694 (<.001)	0.701 (<.001)	-0.498 (.007)	0.733 (<.001)	-0.615 (<.001)				
AS	0.313 (.099)	0.228 (.233)	-0.251 (.189)	-0.008 (.966)	-0.117 (.545)	-0.089 (.644)	0.070 (.718)	-0.401 (.032)	-0.075 (.697)	-0.214 (.264)	0.087 (.654)	0.519 (.004)	0.391 (.037)	0.314 (.097)	0.451 (.015)	0.499 (.006)	0.317 (.094)	-0.074 (.700)	0.160 (.407)	0.299 (.115)	0.883 (<.001)	-0.209 (.276)	0.736 (<.001)	-0.176 (.360)	0.601 (.001)	-0.637 (<.001)			
C	-0.161 (.404)	-0.065 (.737)	-0.300 (.113)	0.021 (.916)	0.005 (.981)	-0.074 (.700)	0.086 (.655)	-0.247 (.196)	0.025 (.897)	-0.053 (.786)	-0.145 (.450)	0.297 (.118)	0.176 (.360)	0.099 (.610)	0.289 (.128)	0.332 (.079)	0.140 (.468)	0.030 (.879)	0.015 (.940)	0.201 (.294)	0.488 (.008)	-0.054 (.780)	0.545 (.003)	-0.103 (.594)	0.413 (.027)	-0.436 (.019)	0.485 (.008)		
CS	0.128 (.508)	-0.217 (.256)	-0.380 (.043)	0.274 (.149)	0.238 (.213)	0.136 (.479)	0.456 (.014)	-0.492 (.007)	0.317 (.094)	-0.654 (<.001)	-0.180 (.349)	0.407 (.029)	0.074 (.702)	0.195 (.310)	0.217 (.256)	0.397 (.034)	-0.039 (.839)	-0.471 (.011)	-0.182 (.342)	0.265 (.165)	0.626 (<.001)	-0.116 (.547)	0.451 (.015)	-0.092 (.635)	0.719 (<.001)	-0.579 (.001)	0.706 (<.001)	0.524 (.004)	

Computed correlation used spearman-method with listwise-deletion.

Table 5 Correlation matrix (n=87)

Results from the three sets of calculations show that all values are fairly stable across the different N cases. That means the model and data preparation is stable and robust. Bootstrapping was also used as a validation method, and showed similar results. An R of up to 2000 was used in bootstrapping.

When looking at the raw results with N=111 (the most reduced variable set), country of residence negatively correlates with Peer review results. This reveals that when diversity in country of residence is higher (i.e. members of the project are living in different countries), performance perceived by the team (Peer) is lower. The correlation factor is not high, however, and as such is considered a weak effect. In addition, country of origin, country of residence and native language show very high correlation amongst themselves. This can be explained by the fact that not many team members joined the company outside of their native country. They still live in the country of origin, speaking their native language. These effects would distort the findings in regression, so the multicollinearity was calculated through the Variance Inflation Factor (VIF) before moving on to regressions analysis. The diversity in education shows negative correlation to the team set goal performance (PO_Team). At the same time there is a correlation between education diversity and country of origin visible. This indicates that education diversity shows differences to where the team members come from. At the same time with slightly higher p-values the same effect is visible with language and residence. As such it can be assumed it is an unequal distribution of education diversity amongst the location included in the data. Another negative correlation is present for age diversity and PO_Team. This means that more age diverse groups perform worse on their team set goals than more homogenous groups. One effect is visible related to the output variables themself. PO_Team positively correlates to PO_Exp. As such, a high rating based on PO_Exp goals correlates strongly with a high rating for team set goals, PO_Team.

The second calculation with N = 109 reveals a negative correlation between the PO_Team performance and age diversity, meaning a homogeneous team age-wise tends to achieve higher PO_Team team set goals performance, which is the same effect as with the N=87 above. Effects around country language, and residence are also in the bigger set of data. The N=109 set here also shows some weak links similar to the education correlation from above for gender. There seems to be a slight uneven distribution of gender per location. Also the set shows some small effect between education diversity and gender diversity suggesting also a not evenly distributed level here.

The additional personality data, added in this frame, only shows significant correlation with the Peer data. On the level of medium effect sizes high extraversion levels negatively correlates with Peer results. Same is true for openness, agreeableness, and conscientiousness

levels. All SD variants of all five personality parameters do not show any correlations that are significant even though p-levels are relatively low and come close to 90% and 95% ranges.

Lastly, gender diversity shows a positive correlation with the team achieving their team goals PO_Teams. A team with a high gender diversity performs better regarding their self-motivated and self-defined goals. For this frame, all effect sizes are small.

The third calculation on the N=87 frame shows similar findings, and effect sizes are still small and significance values are decreasing as expected with a larger data set. The data set shows significant negative correlation between PO_Team and the diversity in age in a team. PO_Exp positively relates to three culture traits: egalitarism, risk, and directness. Only risk correlates positively with Peer results. This means that when a team is more relationship driven and/or shows a higher task average orientation, the results tend to be better. When examining the variances on the culture traits, independence SD shows positive mild correlation with PO_Exp and PO_Team. Same is true for egalitarism SD, risk SD, and directness SD. The positive effect of gender diversity on PO_Team persists in this data frame.

The personality traits show less impact in this larger frame as they did in the smaller one. Three significant effects on correlation can be seen. Higher levels of openness negatively relate to Peer results. The effect is small though. The other two are related to neuroticism SD where it shows that higher variation in neuroticism positively relate to PO_Exp. The second is that lower levels of variance in conscientiousness relate positively to Peer results.

Pair-wise analyses on effects over the three data frames resulted in additional findings. When comparing the results from the first and second frames, the effect size goes up from frame one (N=111) to frame two (N=109). One possible, mathematically driven explanation is that while some variables are removed from the calculation, the reduction of N from 111 to 109 is not as big and has small effects on the overall calculations. As such, results driven by fewer variables weigh in stronger and show higher effects. When comparing frame two and three, despite shrinking N, the effects are confirmed and are stable.

In conclusion, the three calculations show stability in the model and show relative consistent results, but also a few differences. The differences in correlations might be explainable by

different N and influences by variables included and excluded in the different frames. For better readability, a cross-section of the relations is presented in tabular form.

Factor (higher)	relates to	Output factor
Residence diversity	Negatively	Peer
Education level diversity	Negatively	PO_Team
Age diversity	Negatively	PO_Team
Gender diversity	Positively	PO_Team
Neuroticism SD	Positively	PO_Exp
Extraversion level	Negatively	Peer
Conscientiousness level	Negatively	Peer
Conscientiousness SD	Negatively	Peer
Egalitarism level	Positively	PO_Exp
Egalitarism SD	Positively	PO_Exp
Egalitarism SD	Positively	PO_Team
Risk level	Positively	PO_Exp
Risk level	Positively	Peer
Risk SD	Positively	PO_Exp
Risk SD	Positively	PO_Team
Directness level	Positively	PO_Exp
Independence SD	Positively	PO_Exp
Independence SD	Positively	PO_Team
Directness SD	Positively	PO_Exp
Directness SD	Positively	PO_Team
PO_Team	Positively	PO_Exp

Table 6 Factor correlations cross section (all N). By Input

Output factor (higher)	relates to	Factor
Peer	Negatively	Residence diversity
Peer	Negatively	Extraversion level
Peer	Negatively	Conscientiousness level
Peer	Negatively	Conscientiousness SD
Peer	Positively	Risk level
PO_Team	Negatively	Education level diversity
PO_Team	Negatively	Age diversity
PO_Team	Positively	Gender diversity
PO_Team	Positively	Egalitarism SD

PO_Team	Positively	Directness SD
PO_Team	Positively	Risk SD
PO_Exp	Positively	Neuroticism SD
PO_Exp	Positively	Egalitarism level
PO_Exp	Positively	Egalitarism SD
PO_Exp	Positively	Risk level
PO_Exp	Positively	Risk SD
PO_Exp	Positively	Directness level
PO_Exp	Positively	Independence SD
PO_Exp	Positively	Directness SD
PO_Exp	Positively	PO_Team

Table 7 Factor correlations cross section (all N). By Output

10.6 REGRESSION ANALYIS

In regression analysis, one must avoid effects of multicollinearity. Multicollinearity describes the state when one predictor can be predicted linearly from any/or multiple other ones (Fields at et al. (2012). One, or more, such linearly coupled predictions can distort the results, since predictions in regression models are usually based on proportional relations to each other.

To avoid such an effect in this study, the Variance Inflation Factor (VIF) was calculated to identify and avoid multicollinearity. The VIF is calculated by:

$$VIF = \frac{1}{1-R_i^2}.$$

The VIF calculation resulted in a few findings that will now be addressed before moving onto regression analysis. Generally, a VIF value higher than 10 is considered too high, whereby a researcher should remove the affected variables from the final calculation, compare Cohen et al. (2003). Some experts even suggest going as low as 2.5 as a threshold value for VIF. Therefore the author decided to address some values that stand out.

The VIFs were calculated for all frames of data. For the first frame, N=111, values for language and origin were high, from ca. eight to just under ten. Both show very high correlation within, as spoken native language is very likely correlated to the country of origin. Therefore, it was assumed the diversity of language has the same effects as diversity of

country of origin. Subsequently, these values were excluded from further calculations to eliminate the effect of the multicollinearity.

For the second frame, N=109, the VIFs were calculated as well. Conscientiousness had a VIF of 5.25, which was notably higher than other personality traits. As it is still within the bounds of the suggested VIF < 10, and does not show large differences over other values, it was kept in the regression. Similar considerations were taken in the third data frame N=87 for egalitarianism, conscientiousness and extraversion, and, for the same reasoning as in the second data frame, were kept in the regression analysis.

Table 8 VIF results

		PO_Exp	PO_Team	Peer	Age	Education	Gender	Residence
N=111	Expectation Performance	NA	1,05	1,19	1,17	1,38	1,14	1,33
	Team Performance	1,04	NA	1,20	1,18	1,38	1,11	1,34
	Peer Performance	1,05	1,07	NA	1,15	1,31	1,12	1,20

		PO_Exp	PO_Team	Peer	Age	Education	Gender	Residence	N	NS	E	ES	O	OS	A	AS	C	CS
N=109	Expectation Performance	NA	1,12	1,26	1,24	1,48	1,25	1,53	1,37	2,46	3,91	2,31	2,42	2,43	1,41	3,03	4,32	3,26
	Team Performance	1,15	NA	1,27	1,25	1,48	1,23	1,56	1,37	2,47	3,90	2,32	2,42	2,43	1,39	3,02	4,41	3,37
	Peer Performance	1,16	1,13	NA	1,24	1,41	1,24	1,38	1,39	2,44	3,84	2,32	2,42	2,44	1,39	3,04	4,35	3,37

		PO_Exp	PO_Team	Peer	Age	Education	Gender	Residence	I	IS	EG	EGS	R	RS	D	DS	T	TS	N	NS	E	ES	O	OS	A	AS	C	CS
N=87	Expectation Performance	NA	1,58	1,51	1,52	2,11	1,81	1,79	3,45	3,07	6,16	3,48	1,84	3,78	3,52	2,78	1,90	1,37	2,11	3,62	3,79	3,20	2,08	3,83	2,23	3,40	1,29	4,63
	Team Performance	1,39	NA	1,51	1,53	2,03	1,79	1,96	3,47	3,02	6,33	3,01	1,85	3,38	3,55	2,75	1,92	1,39	2,10	3,16	3,77	2,14	2,08	3,69	2,23	3,39	1,29	4,52
	Peer Performance	1,44	1,63	NA	1,53	2,10	1,82	1,93	3,35	3,06	6,27	3,37	1,78	3,77	3,47	2,78	1,92	1,29	2,04	3,60	3,61	3,13	2,08	4,03	2,11	3,38	1,28	4,53

In Table 8, the results, after removing the strongly correlating locality values, for the final round of VIF calculation are shown. The new calculation shows much lower values, mostly below two. The conscientiousness, extraversion and culture dimensions of risk, egalitarianism and directness are between three and five. As this is likely not posing a problem, the dataset was kept at that state to further progress.

Finally, a multi-regression analytic calculation was performed. A linear stepwise model was applied. For the problem space at hand, a backward stepping model is suggested by literature like by Fields et al. (2012) and Cohen et al. (2003). The calculation was done for all three data frames, each relating to all three performance measures of PO_team, PO_Exp and Peer review. That leads to nine data result frames (3x3). The reformatted, simplified output is shown in Table 9, Table 10, and Table 11. To test for possible better fit, bootstrapping on regression was performed, as Fields et al. (2012) suggest. The methods used followed mathematics laid out by Fox (2002) with an R of up to 2000 iterations. The results are very similar, with small residuals when comparing with the stepwise linear model.

	Peer		**PO_Exp**		**PO_Team**	
Predictors	*std. Beta*	*p*	*std. Beta*	*p*	*std. Beta*	*p*
Peer			0.12	0.246	0.08	0.419
PO_Exp	0.11	0.246			0.15	0.136
PO_Team	0.07	0.419	0.15	0.136		
Age	0.15	0.118	-0.12	0.253	-0.04	0.708
Education	0.25	**0.015**	0.05	0.682	-0.07	0.506
Gender	0.11	0.250	-0.01	0.931	0.17	0.086
Residence	-0.34	**0.001**	0.11	0.320	-0.01	0.940
Observations	111		111		111	
R^2 / R^2 adjusted	0.171 / 0.123		0.058 / 0.003		0.068 / 0.014	

Table 9 Regression analysis (n = 111)

	Peer		PO_Exp		PO_Team	
Predictors	*std. Beta*	*p*	*std. Beta*	*p*	*std. Beta*	*p*
Peer			0.10	0.336	0.06	0.576
PO_Exp	0.10	0.336			0.12	0.240
PO_Team	0.06	0.576	0.12	0.240		
Age	0.13	0.204	-0.12	0.270	-0.10	0.375
Education	0.25	**0.026**	0.08	0.492	-0.06	0.595
Gender	0.09	0.401	0.02	0.862	0.15	0.172
Residence	-0.38	**0.001**	0.16	0.196	0.03	0.811
N	-0.03	0.751	-0.15	0.174	-0.14	0.238
NS	0.17	0.239	0.10	0.510	-0.04	0.807
E	-0.22	0.223	0.04	0.841	-0.07	0.708
ES	0.01	0.949	-0.11	0.465	0.04	0.774
O	-0.01	0.948	0.01	0.923	0.08	0.592
OS	-0.16	0.283	-0.19	0.212	-0.19	0.214
A	0.12	0.270	-0.01	0.959	-0.12	0.287
AS	-0.02	0.923	-0.09	0.593	0.11	0.529
C	0.22	0.248	-0.29	0.150	-0.06	0.753
CS	0.05	0.776	0.31	0.078	-0.02	0.930
Observations	109		109		109	
R^2 / R^2 adjusted	0.217 / 0.081		0.146 / -0.002		0.122 / -0.031	

Table 10 Regression analysis (n = 109)

	Peer		PO_Exp		PO_Team	
Predictors	*std. Beta*	*p*	*std. Beta*	*p*	*std. Beta*	*p*
Peer			0.07	0.612	0.06	0.646
PO_Exp	0.06	0.612			0.18	0.134
PO_Team	0.06	0.646	0.20	0.134		
Age	0.02	0.848	-0.06	0.633	0.01	0.959
Education	0.10	0.515	-0.06	0.695	0.22	0.124
Gender	-0.02	0.868	0.07	0.613	0.13	0.342
Residence	-0.15	0.315	0.34	**0.020**	-0.03	0.813
I	0.28	0.145	0.10	0.603	0.04	0.828
IS	0.13	0.478	0.12	0.540	0.21	0.244
EG	0.21	0.430	0.36	0.186	-0.07	0.797
EGS	-0.27	0.167	-0.05	0.797	0.54	**0.003**
R	0.21	0.136	0.09	0.558	-0.02	0.892
RS	0.07	0.750	0.05	0.806	-0.49	**0.010**
D	-0.23	0.241	-0.14	0.500	-0.01	0.965
DS	-0.04	0.818	-0.06	0.756	-0.14	0.416
T	-0.05	0.732	0.14	0.331	0.03	0.854
TS	-0.26	**0.030**	0.11	0.376	-0.01	0.939
N	-0.25	0.095	-0.13	0.392	-0.16	0.268
NS	0.13	0.515	-0.02	0.924	0.53	**0.005**
E	-0.34	0.088	-0.03	0.884	-0.09	0.651
ES	0.22	0.248	-0.01	0.945	-0.19	0.302
O	-0.02	0.906	-0.02	0.917	0.05	0.746
OS	0.06	0.773	0.38	0.076	-0.46	**0.020**
A	0.29	0.063	-0.07	0.671	-0.07	0.641
AS	0.14	0.463	-0.09	0.660	-0.11	0.558
C	-0.08	0.505	0.01	0.925	0.00	0.989
CS	-0.31	0.164	-0.19	0.423	0.31	0.151
Observations	87		87		87	
R^2 / R^2 adjusted	0.342 / 0.057		0.310 / 0.011		0.390 / 0.126	

Table 11 Regression analysis (n = 87)

	PO_Exp			PO_Team			Peer		
Predictors	*std. Beta*	*p*	*std. p*	*std. Beta*	*p*	*std. p*	*std. Beta*	*p*	*std. p*
N * NS	0.04	0.777	0.777	0.01	0.923	0.923	0.21	0.145	0.145
E * ES	-0.09	0.473	0.473	-0.08	0.549	0.549	-0.02	0.849	0.849
O * OS	-0.11	0.385	0.385	-0.16	0.190	0.190	-0.14	0.255	0.255
A * AS	-0.18	0.143	0.143	-0.06	0.581	0.581	-0.11	0.375	0.375
C * CS	0.22	0.177	0.177	0.43	**0.007**	**0.007**	0.10	0.545	0.545
Observations	109			109			109		
R^2 / R^2 adjusted	0.119 / -0.023			0.142 / 0.003			0.121 / -0.021		

Figure 11 Regression interaction for personality levels and SD

While results show consistent and stable parameters, the overall effects are comparatively small in nature. The small size of adjusted R^2 may suggest rejecting the overall regression analysis and revert to the correlation analysis. But small R values have been subject to a lot of discussions lately in the wake of large data sets in data science, as in R Foundations regression documentation (2011), Frost (2020), MiniTab statistics forum discussions MiniTab (2014) and Fields et al. (2012). A small R means two things. First, the effect is small and data is scattered, but a trend is still visible. Residuals are higher than in other models. Second, it means that when checking for fit of the model, further analysis is recommended to ensure significance and correct interpretation. Therefore, additional F statistics were calculated, which suggest rejecting some of the null hypotheses. Detailed output can be found in Appendix C. In cases with high probability confirmed through F testing, the R Value still suggest small effect sizes to the overall model. It is recommended to inspect regressions lines and scatterplots visually to see effects that might contribute, such as curved linear effects or effects significantly increasing residuals on one or two sides. The corresponding plots are also included in Appendix C. In addition, numerical testing on the models is suggested. The Durbin-Watson test is a suggested method. The test above confirmed reasonable robustness for all nine data frames and their regression. Hence regression analysis is used to further investigate the results.

In all the data frames, Peer review is negatively influenced by diversity in residence. Teams composed of members from different locations negatively influences team performance. In contrast, frame three shows that PO_Exp is positively influenced by a higher degree of diversity in residence. There seems to be a difference in how a team perceives results and the

way these results are rated from an outside perspective. One other effect is that diversity in education significantly increases Peer review in frame one and two. Diversity in the way team members approach a problem seems to positively influence the perception on how teams work together. Confirming findings in the correlation analysis, gender diversity has a positive influence on team performance. Higher diversity in gender increases the ratings in PO_Team. In no data frame the levels of culture traits or personality traits show any significance for the effects. A small exception is lower levels of neuroticism have positive influence on Peer review with a at least 0.09 p-Value. The variances in culture and personality show similar results than the correlation analysis. A lower variance in task SD orientation boosts Peer review data. Lower SD for risk increases PO_Team and higher egalitarism SD increases PO_Team. For the personality traits, higher SD in neuroticism in teams positively influences PO_Team results. In contrast, lower levels of SD in openness increase PO_Teams. When controlling for the influence of personality factor levels, as suggested by Homan et al. (2008), and the respective SD values only the relation conscientiousness shows a mild positive relation that is of significance. Almost half of the effect of consciousness on PO_Team is explained by the SD of the same trait. Having said this, neither of the models shows relations of significance for consciousness or its SD with any of the outcome variables.

A table of the effects found in regression is summarized below.

Factor	predicts	Factor
Residence diversity	Negatively	Peer
Residence diversity	Positively	PO_Exp
Education level diversity	Positively	Peer
Gender diversity	Positively	PO_Team
Neuroticism level	Negatively	Peer
Neuroticism SD	Negatively	PO_Team
Openness SD	Positively	PO_Team
Task SD	Positively	Peer
Egalitarism SD	Positively	PO_Team

Table 12 Result summary from regression analysis (by input factor)

Factor	predicts	Factor
Peer	Positively	Education level diversity

Peer	Negatively	Residence diversity
Peer	Negatively	Neuroticism level
Peer	Positively	Task SD
PO_Exp	Positively	Residence diversity
PO_Team	Positively	Gender diversity
PO_Team	Negatively	Neuroticism SD
PO_Team	Positively	Openness SD
PO_Team	Negatively	Risk SD
PO_Team	Positively	Egalitarism SD

Table 13 Result summary from regression analysis (by output factor)

Following up on the conceptualization decisions in chapter 9.1 one more set of statistical analysis was performed. Indicated by Information/Decision process models, following conceptually Bell (2007) and Moynihan & Peterson (2001), we earlier saw that the Peer review data could either be viewed as an outcome, or in the definition of Hackman as an indication of team viability, or as an indication on how these internal team processes performed. When assuming the Peer review data represent process measures around information/decision processes the Peer data would act as a moderator in regression analysis. (compare Figure 4). Therefore, an additional multi regression analysis has been performed. Peer data was set as a moderator between the multiple input parameters and the two remaining output variables PO_Exp and PO_Team. These calculations were done on all three sets of data (N=87, N=109, N=111). Due to the lack of significance in analysis otherwise, only data frame N=87 will be reviewed. Full summaries of the analysis can be found in appendices C.9.4 and C.9.5. Two sets of summaries are presented, one for PO_Team and PO_Exp for better readability. Summarizing the findings, the analysis showed only a few relations with statistical significance. All of these cases do not show any noteworthy effect size of Peer as a moderator on the overall effects. The maximum range is about 3% of the effect explained by the moderation of Peer review results. Openness, neuroticism SD, egalitarism SD, risk SD show direct influence without relevant moderation from Peer for the outcome of PO_Team. Task, independence SD, residence diversity and openness SD show direct influence without relevant moderation from Peer for the outcome of PO_Exp.

11 DISCUSSIONS OF RESULTS, HYPOTHESES TESTING, AND LIMITATIONS

The following chapter discusses the findings from the analyses in regards to the postulated hypotheses. It will also discuss some limitations on these findings.

11.1 RESULTS

The research question to be answered is: what surface-level and deep-level diversity factors contribute to the effectiveness of work teams? Also, what can a manager do to positively influence teams based on these findings? To answer these questions, a set of hypotheses developed from literature research was presented in chapter 9. The first step is to see if results from the regression and correlation analyses support or contradict the hypotheses. As a general reference a consolidated table of the findings from analysis is provided here.

Output		***Input***
PO_Exp	Positively influenced by higher	Neuroticism SD
		Egalitarism Level
		Egalitarism SD
		Risk Level
		Risk SD
		Directness Level
		Independence SD
		Directness SD
		Residence Diversity
	Negatively influenced by higher	N/A
PO_Team	Positively influenced by higher	Gender Diversity
		Egalitarism SD
		Risk SD
		Openness SD
	Negatively influenced by higher	Age Diversity
		Education level Diversity
		Directness SD
		Neuroticism SD

Peer	Positively influenced by higher	Risk Level
		Education level Diversity
		Task SD
	Negatively influenced by higher	Residence Diversity
		Extraversion Level
		Conscientiousness Level
		Conscientiousness SD
		Neuroticism Level

Table 14 Combined results reference

Hypothesis 1A suggests that higher levels of agreeableness will predict higher team performance (PO_Team). This is not directly confirmed by the findings of this study. Nevertheless, some of the studies from which the hypothesis is derived, like Mathieu et al. (2015) approach the influence of personality traits from a group cohesion perspective. As such a higher team performance would require a higher level of group cohesion. That would suggest that other than the PO_Exp or PO_Team performance value, the Peer values might be influenced positively. That is not the case. While the claim still stands in many other studies that we have looked at previously that higher agreeableness leads mainly through group cohesion to a higher performance of a team this claim could not be confirmed here. This effect not being present is also backed up by the moderation analysis not showing any connection on agreeableness.

Higher average extraversion combined with low variance in extraversion (SD) in teams will predict lower team performance. Hypothesis 1B makes out this combination of trait distribution along the personality factor of extraversion. In simpler words this could be framed as: Too many extravert people are hindering group cohesion or task related processes by either people being stuck in discussions or not being able to focus on work. Following that logic the SMWT-Research teams in this study show negative influence by higher levels of extraversion on the Peer results. The effect does not show, or carry, through to PO_Teams and PO_Exp which is confirmed also by mediation analysis. Therefore, 1B is partially confirmed. This also somewhat is supported by the ideas and prerequisites of emergent leadership in team studies, like Carte et al. (2006), that see a limited amount of extraverted team members as subsequently positive for teams and such processes. The finding is much in

line with an empirical study from Erez et al (2002). While the study takes a different viewpoint from a research design perspective, it shows that rotated leadership and workload sharing are both connected to team satisfaction and team output performance. Inner team processes also using peer reviews by Erez et al., show results very similar to the study here. It indicates that the factors need to be balanced to produce good outcomes and maintain team satisfaction. Combining their findings with the finding from this study, it can be concluded that a higher level of extraversion might lead to a perceived imbalance in workload sharing, thus negatively affecting team peer performance while not significantly affecting outcomes. The effect may manifest either through conflict or slowed emergent leadership, following Erez et al.'s logic. The finding that team outcome is not affected could be due to the short nature of the projects in the study here. As Erez et al. indicate, some effects develop over time. So, three to four months might not be enough to see deterioration effects on outcomes from lower team satisfaction. Also, the team members know that timeframes are limited, and setups will change in short time frames. Behavior may be tolerated more easily, as it is not a permanent situation going forward.

Hypothesis 1C says that higher average conscientiousness while minimizing their SD in teams will predict higher team performance (PO_Team) except for creative work. Which leads to the hypothesis for this study that the effect will not be present. The teams at question in this study define as creative work teams. While structure is still present as they do not execute fundamental exploratory research but product focused entrepreneurial work in an enterprise controlled environment, it still qualifies as creative work. The effect connecting the two could not be show in the study here for PO_Team. Main reason most likely being the creative work. Having said this, both aspects conscientiousness and its SD show effects on Peer reviews. Higher levels and higher disparity (higher SD) show negative influence on Peer data. While the higher levels of conscientiousness leading to lower Peer reviews is difficult to explain in the context of this study as it contradicts most research, like Barrick & Mount (1991), Humphry et al. (2007), it might be related to findings from Humphry et al. (2011) predicting that minimizing SD, or in other words minimizing disparity, for conscientiousness increases the team's ability to produce output. This would be in line with the other researchers mentioned before. It is assumed that a higher disparity in conscientiousness is detrimental for group task related processes. While this aspect is not confirmed by this study

for the two output related measures PO_Team and PO_Exp it is showing in relation to Peer results. As such the hypothesis is partially confirmed.

Hypothesis 1D suggests higher average openness in teams will predict higher team performance (PO_Team). This is not directly confirmed by this study. Openness SD at a higher level is showing higher PO_Team results though. This effect could be related to the group sampling. The sample from which the study's data is collected from shows a slightly low openness and a comparatively low openness SD compared to reference groups. Following that logic a person showing more openness could be perceived as more helpful to group effects or creative processes.

Hypothesis 1E claims lower average neuroticism and/or lower neuroticism SD in teams will predict higher team performance (PO_Team) and Peer reviews. This hypothesis is partially confirmed. Peer review data shows to be negatively influenced by higher levels of neuroticism levels. For PO_Team the data shows no relation to neuroticism levels but negative relations to higher SD for neuroticism. This finding is in line with research suggesting that in addition to low levels of neuroticism it is also apparent that even one or few team members with low levels can have detrimental effects on team performance. Very much in line with many findings along of variance/disparity in conscientiousness. The widely assumed effect is that this one member disrupts interpersonal and task related processes within the team. Controversially the data here shows that regarding PO_Exp higher levels of neuroticism SD predict better results. Again, more in the field of inside knowledge and assumptions, the author suggests that this might be due to the effect that PO_Exp is rated from the outside (by the PO) as such they might be seeing it positive if results are transmitted by lower levels of neuroticism team members. So they feel not stuck with the higher levels. The mediation of the Peer results do not show any relation to PO_Team though.

Hypothesis 1F claims lower variance in conscientiousness will predict higher team performance. Analogue to finding discussed up to here around neuroticism, conscientiousness SD in higher levels are prone to lowered team performance according to existing studies. The same logic seems to apply in which even one team member showing larger distance to the others in terms of levels of conscientiousness can significantly influence a team negatively. In this study no relation to PO_Team and PO_Exp for conscientiousness was found. For Peer review the inverse effect was found. Higher levels of conscientiousness SD predict lower

Peer reviews. Explanation could be very similar to the one in 1D previously. While the average level of conscientiousness of the sample is in this case slightly higher than reference it also shows comparatively low SD for the sample group. This could overemphasize the visibility of the effect to the teams in case there are differences perceived. Another contributing factor might be that due to the nature of the rather short project durations of three to four months, a conscientious team is not as fast as the external POs expect when creating expectations and rating results. As such thorough conscientious behavior might lead to the team to feel slowed down in a task related manner.

The first hypothesis around culture traits 2A claims higher average collectivism in teams will predict higher team performance in terms of outcome and team internal aspects. This is not confirmed by this study. No effect with the collectivism trait is found. As we have debated earlier the aspects of culture and team performance is not a very deeply researched field. Also the dimensions are showing a rather large variance when studies operationalize culture. One possible reason for not seeing effects on collectivism as defined by GlobeSmart in this study is that the boundaries in definition for what this dimension means are somewhat wide and fuzzy. There is wide variety of research showing the importance of group cohesion processes, be it socially or task related and how these processes interact, like CEM, Continuous Configuration (CCA) and others. One could see quite a big overlap with collectivism in a broad definition with cohesion in groups. This could influence the interpretation of results in many ways.

Hypothesis 2B suggest higher levels of egalitarism will predict higher team performance even with lower conscientiousness present. The effect for egalitarism is fully confirmed while the influence and relation with lower levels of conscientiousness could not be confirmed. Higher levels of egalitarism show positive influence on PO_Exp.

Hypothesis 3A shows higher educational diversity in teams will predict higher team performance. The findings here are controversial at first glance. While higher levels of educational diversity show positive influence to Peer reviews they show detrimental effects PO_Team results (and to PO_Exp with lower p-value though). It seems that the diversity is helping the cohesion, communication parts represented partly by the peer data. It seems to have negative influence on achieving the self-set goals. The data unfortunately does not allow for any insights why that might be. Teams see value and use education variance, but it does

not influence the output of the team from an external perspective. The effect is different from most literature findings in that it is based on educational level diversity, not on highest education level or subject matter expertise diversity. The hypothesis is partially confirmed.

Hypothesis 3B claims higher gender related diversity in teams will predict higher team performance in terms of outcome. This hypothesis is fully confirmed. Gender diversity shows positive relation to higher PO_Team results. This finding is in line with most other modern research in this field.

Hypothesis 3C shows higher gender related diversity in teams will predict higher team performance in terms of outcome (PO_Exp) through the moderation effects of information/decision-making in teams. This effect could not be found in the data analysis. Moderation analysis did not show an equivalent effect for gender diversity and Peer review towards PO_Teams or PO_Exp.

Hypothesis 3D claims higher age-based diversity in teams will predict lower team performance in terms of outcome. This effect is confirmed by the analysis. Higher age diversity in the teams is linked to lower PO_Team results. One conclusion from that finding is that a common notion of managers to let a very experienced person, and therefore usually older, explain and lead the team to success is most likely not a successful model in SMWTs. One way of thinking would be to change the situation into a coaching role or more formal leadership perceived role to avoid such problems. Conceptually that is supported by Zhang et al. (2012).

Hypothesis 3E shows gender and age diversity will have higher direct effects on team performance than education diversity. This hypothesis is confirmed. The correlations found for age diversity and gender diversity are higher in size than the one for education diversity. In fact age diversity is with a distance higher than gender and education in its effect sizes.

Hypothesis 3F indicates higher diversity in country of residence of team members will predict lower team performance (Peer, PO Exp and PO_Team). Finding is that higher diversity in residence, or in other words office location, is predicting lower Peer review results. Language diversity and residence diversity factors show a negative correlation with the Peer review as well, meaning that teams with a higher diversity in native language spoken

or country of residence relate to lower self-reflected team view. This intrinsic viewpoint of teams is confirmed by regression analysis, which predicted lower Peer review results with higher diversity in residence. The second part of the hypothesis, which predicts lower external performance, was contradicted by regression analysis. More diversity in residence for teams do not show any relation to PO_Exp and PO_Team. This finding is the most contradicting to the literature review. One possible explanation is that research predicting both effects are based on team internal factors and are using team output as a secondary measure or proxy. As such, the confirmation of the internal vs. external effects would be explainable. Another prevailing topic with this hypothesis is that the effect vanishes over time. Therefore, it could be argued, in the case of this study, the team pool size, around 30 people, might reach a point when the effect can be seen in assessing the team's internal processes but no longer in terms of outcome. The team member pool might encompass effects of convergence already. This would suggest that managers can pick the best skillset needed for the task and not be concerned with the internal team perception. One finding is also that the factor of residence diversity is mainly influenced by diversity in language. In addition, not being located at the same office and an existing or perceived language barrier leads to lower Peer performance. Language diversity shows stronger effects than diversity in location only. It seems to hinder and slow down the effective collaboration and inner processes of coherence within teams. It is suspected that the effect is related to team composition. When POs and management compose teams, one focus is the skillset that is available and the needs for the task at hand. For that reason, team composition might cross over into two or three locations to bring the needed experts to the team. This action is also seen as a positive influence to teams and their outcomes by Kristof-Brown & Stevens (2001). It ensures highly valued output in the expectations since the right experts are on the team, but also leaves the team with a challenging and diverse location situation to deal with. As such, the situation is assessed as detrimental to peer performance, but potentially positive for results produced.

Hypothesis 4A predicts higher rated Peer performance leads to higher levels of team outcome performance (PO_Teams and/or PO_Exp). This would also follow the IMOI model by Mathieu et al. (2008) which shows a relationship between the inner processes quality in teams and the outcomes of such teams. This is not confirmed. None of the analysis shows any signs of relation between the Peer review results and the outcome related parameters.

Hypothesis 4B predicts higher rated team goals performance (PO_Team) leading to higher levels of team outcome performance (PO_Exp). A strong positive significant relation is confirmed between the two parameters. Two effects could justify this correlation. The first is that a PO, positively influenced by having their external expectations met, overshadows their judgment on the team goals, like a halo effect. Positive perception of one set of goals might generate a halo effect for the additional sets of goals. Such an effect is not uncommon, according to Bechger et al. (2010). The second may be that teams which work together efficiently, have effective task setting and focus and have effective inner processes of teamwork lead to better outcomes on the set goals. The data shows two contradicting findings here, however. If teams are successful in task focus, it might be indicative that the culture dimension of a task should be high for high achieving teams. Higher task orientation predicts higher external ratings. On the other hand, when teams have higher internal coherence, according to hypothesis 4A, Peer performance should be related to team goals performance (PO_Team) or external rated performance (PO_Exp). Both are not the case. The contradicting findings in regression points to the halo effect explanation as the cause. A review of meta studies addressing this question by Beal et al. (2003), Shippers et al. (2013) and Markova and Perry (2014) is not conclusive either. Most studies combine and aggregate results in ways which do not allow interpretation for either effect. Hence a relation as indicated in the IMOI model could not be shown in pure data analytics here.

In addition to the hypotheses derived earlier some more findings showed in the data. Maybe not surprising, they show mainly in the areas of culture dimensions. As culture dimensions are not deeply researched in the field of team performance it is quite obvious that if new insights show they would be in this field. Not very many previously existing insights would lead to less hypotheses in earlier chapters. Also the fact that the systematic GlobeSmart dimensions are used could lead to further findings as a portion of the culture based research is not too focused on systematic dimensions. The more recent ones are more focused on models like Hofstede. We will systematically go through the findings by output parameter that they influence. Starting with PO_Exp there are only positive effects from higher levels of risk and directness. At the same time higher variance (SD) in the dimensions risk, directness and independence are adding positively to PO_Exp. For PO_Team higher variance (SD) in risk dimension leads to higher results. Higher directness SD on the other hand leads to lower performance for PO_Team. For Peer review results a higher risk level and a higher variance

(SD) for task helps to achieve better reviews. Making interpretations on what the factors are that drive these results is quite challenging given the fact that the data does not include and qualitative insights. One possible explanation is yet again in the nature the team sampled. The overall population sample shows a higher certainty (lower risk) and higher indirectness (lower direct values) than the population of the POs requesting, setting, and grading the projects. In fact when looking at the risk and directness dimension the POs show significant different values than the teams in average. The POs, even though mostly German, resemble very closely the US culture, being very risk taking and direct according to the GlobeSmart dimensions. It could be assumed that this also resembles a certain expectation on how teams should performed or generate output and respectively preset and generate content. It might be more appreciated by the POs to see action that resembles their own preferred cause of action. This would also explain the results for the higher SDs for the same dimensions influencing positively the outcome. When teams contain a few team members that fit these expectations it helps the views of the POs when generating the PO_Exp and PO_Team ratings. Such a positive "closeness in dimension" effect would also be supported in general by culture dimension theory like Hofstede (2001) and Doherty (2016).

Hypothesis 1A	not confirmed
Hypothesis 1B	partially confirmed
Hypothesis 1C	partially confirmed
Hypothesis 1D	not confirmed
Hypothesis 1E	partially confirmed
Hypothesis 1F	not confirmed
Hypothesis 2A	not confirmed
Hypothesis 2B	confirmed
Hypothesis 3A	partially confirmed
Hypothesis 3B	confirmed
Hypothesis 3C	not confirmed
Hypothesis 3D	confirmed
Hypothesis 3E	confirmed
Hypothesis 3F	confirmed
Hypothesis 4A	not confirmed
Hypothesis 4B	confirmed

Table 15 Hypotheses confirmation table

Overall the results show mixed results regarding the hypotheses set up. About half are confirmed and a few more are partially confirmed. But there are also a number of not

confirmed ones. The results allow insights into a real world case modern SWMT in an Agile software environment. Even though the study uses a more simplistic IPO model as a basic framework it provides valuable new insights into how parameters on a comprehensive set of surface-level and deep-level diversity influence certain levels of performance of teams in such an environment. It contributes reflection on team diversity IPO research onto modern SMWT teams. Due to the unique setup of three different performance parameters and the knowledge and review of more modern and recent models of team performance and diversity, like CEM, there are learnings that will allow to setup, design and eventually collect data to use with such models in the given environment to do even deeper analysis. A few new systematic findings along culture traits are also shown by the study.

11.2 LIMITATIONS AND FUTURE RESEARCH

The research conducted has some limitations which will be described and discussed in this section. This chapter will also show future research opportunities based on the results of this study.

One possible limitation to the research and its interpretation is the type of teams used in the study. Namely SMWTs which are specialized in the field of software and services with Agile processes. SMWTs are a special breed of teams, as they do not rely on formally assigned leadership or processes imposed from the outside. Thus, the wide variety of leadership in teams research is only partially applicable, and limited to the informal leadership aspects of research. Few researchers focus solely on the aspects of SMWTs or break down their findings or constructs to reflect that. Meta studies deal with varying types of team on a conceptual level, but are ambiguous on what findings can be reflected on what types of team. While this is a clear limitation of this study in terms of generalization of findings, it is also a contribution to the field of SMWT and software-based Agile teams.

Another important limitation is the sample population studied, which is a comparatively homogeneous group of hand-picked, mostly science, engineering and technology-based educated team members. While there is a significant range in subject matter expertise, from user experience design, to mechanical engineering, to data scientists, the population can be described as a low-context sample in the original definition by Hall (1959) and many more scholars after him. Recently, despite efforts to unify the cultural and personality models with

Hall's viewpoints, no conclusive positive or negative results have yet been found, as Cardon (2008) concludes. The sample population does also not reflect the full spectrum of personality traits. It is a fairly homogeneous group in terms of personality traits. It is closer to a representative sample for software related service and product development teams than it is to many other forms of teams. In Table 1, the personality profile summary shows that team members are pre-selected for certain traits, such as low neuroticism, which is different to the general population. The sample population also does not include any person without any type of formal education. The environment is a technical, highly-educated R&D organization. Team members are pre-selected by hiring processes for their high levels of expertise and qualification. The median education level of over 65 persons is an above master level education. This describes a strong characteristic of the team that should be considered when comparing results from this research to others. It could be argued that the sample is not representative of many team situations in general.

The analysis presented leaves a few questions open as the data does not allow for deeper analysis of some subjects. One example being the educational diversity being good for Peer review but not for PO_EXP. The other example being the rationales and causation behind the findings regarding cultural traits that were discussed earlier. Both cases show that the influence on the team internally contrasts with the outcome performance of the team. As the data does not contain any subjective insights in why the team members or the POs see things in a certain light, this poses a limitation to the research. At the same time, it is a future opportunity to be researched. A newly structured sub-study and data collection could also help to investigate the relationship that residence shows with language. The same is true for the finding that high levels of extraversion negatively influence Peer review data. With further analysis in the data, a design could be set up to interview the team members and POs to investigate what the motivations and perceptions are around these findings. Qualitative research should investigate which findings are specific to the given situation and which ones will have broader validity. The same is true for effects changing over time. Also formal measures like TCI could be assessed at the same time. Due to the short-lived nature of the teams in the given organizational context, subjective insights might be helpful to understand how the super-structure team of around 30-40 people sees their relationships and inner team dynamics. The reality is that project teams are not disconnected from overall organizational effects in the enterprise as well as subgroups within an enterprise, like departments or

operational units. Some scholars started to pay attention to such time effects like Harrison and Mathieu et al. (2014) and Harrison et al. (2000). In another study Harrison et al. (1998) stress the fact that the surface-level diversity effects lessen over time, while deep-level effects strengthen.

Research on inner dynamic and social models within teams has developed significantly over the last years. It has started using constructs and designs that cannot be reflected by the data in this study, as data design and collection for this study began eight years ago. New, inner team constructs, like social loafing (Prince et al. (2006)), social integration, showing time effects between surface-level and deep-level diversity (Guillaume et al. (2012)), attempts to model the complex systems like CEM (van Knippenberg et al. (2007)), diversity in beliefs (Van Dick et al. (2008)), perception of diversity and actual diversity (Harrison et al. (2000)), identity in teams concepts (Brickson (2000)), conflict in teams (LePine et al. (2008)) and stereotypes (Liebermann et al. (2013)), are emerging. Some efforts to model the dynamics in teams have been classified as "complex conceptualizations" and are considered "theoretical research rather than empirical research," as stated by van Knippenberg et al. (2007). But substantial progress is being made to allow the models to be used in empirical studies. It opens the future for detailed insight studies in sub-mechanics within teams using within-teams-models. Explicit design of such a study should match the model, intended insights and dynamic nature of the subject, as suggested by Tuma and Hannan (1984). In particular the conceptual models that we looked at earlier like the Categorization Elaboration Model by van Knippenberg et al. (2004; 2007) or building block to such models like social categorization and information/decision making, tasking are coming into focus. The frame of teams like the ones in this study might allow based on work done here for deeper analysis on projects and team members going forward. The current work is limited by the data collected and how the collection structures has been setup many years ago. One more imminent step for the near future is to reconsider the data and its collection and start collecting structures that are more focused on a model like CEM. This type of work is high up on the authors list of future management related research within TBC.

12 PRACTICAL RECOMMENDATIONS

This chapter will discuss practical implications and recommendations to managers and personas responsible for SMWT Teams. The chapter provides several viewpoints into the problem space and tries to connect the academic/scientific perspective that we saw over the last chapters with the managerial practice often applied and encouraged behavior that is present in large scale companies.

Before we go into implications and recommendations it will be beneficial to put some light on the schemas and models that managers widely use to get a better understanding on why certain recommendations are important. The managerial practice present in most larger scale enterprises today is a very asset focused one. This means, that team members and teams themself, and many other things, are viewed as assets and resources being used to produce desired output, results and gaining return on invest. These results must be produced in the most efficient and fastest possible way. Another common perspective regarding this thinking model is the notion of control and repeatability, or scale when it is related to growth. Specifically, that results and matters around teams are measures by key performance indicators (KPIs) and the scale and repeatability is attempted to be ensured by describing how teams operate through the use of process, rules and regulations. There is even its own industry and even academic research around such efforts stemming from enterprise controlling theory. Compare e.g. Kieniger (2017) or Gleich (2017). While this is certainly a valid perspective for an enterprise to have, it only reflects a somewhat one-sided picture to a multifaceted problem. The controller/manager viewpoint that is described here is leading to a black-box thinking and model when it comes to team performance. It overemphasizes the aspects of how the composition of assets, in the case here team members, leads to higher or lowered output. While this is by no question one important aspect of teams as we have seen over the course of this thesis, there are three more learnings from the efforts here that should feed the actions and models of a modern manager being responsible for such teams. Apart from the compositional aspects the theme of inner dynamics and process should be a concern and the aspects of how to actively help teams and members to improve such processes and dynamics should be as well. Also what can be done proactively by managers to ensure such positive outcomes. We will have a look at all three aspects here. None of the three aspects should be used or implemented without paying attention to the other two and without gaining

understanding how they influence each other. The recommendations have all to be seen in the context of this study. That means in the context of SMWTs and R&D/creative work. Several contingencies and differences for other teams are discussed.

Team compositional aspects

We will lead the discussion by the aspect that is most visible and easiest to comprehend. In simpler words answering the question: "Who should be on a team and why?". The input to the black box before we move into the black box and see how input relates to the inner processes of the box. The software world is under immense pressure to recruit and retain top talent, even during the COVID-19 pandemic, and perhaps even more with the resultant commercial impact and market shift. Being global, even for small companies, is a necessity to find talent, manage cost and be flexible in reacting to market needs, as Zborowska and Kolding (2018) conclude. The struggle for top talent also pushes into recruiting a young work force, which brings new challenges around work motivation. Millennials entering the work force en masse now have different needs and expectations of the work environment and their teams, as Thompson and Gregory (2012) as well as van Knippenberg et al. (2004) describe. Assuming a manager has access to a limited number of team members within an enterprise with which to staff projects or tasks, an approach is needed to build effective teams. To structure the approach, a simple concept, previously developed by the author (2011), of viewing employee development and selection will be used. The structure is loosely based on the knowledge classification of Nonaka an Konno (1998).

Figure 12 Employee assessment

The structure assumes that three factors contribute to the success of an employee in an organization. The first is the technical and/or business skills they bring to the table, the second their personality and culture/background, and the third being their g factor as defined by Carroll (1993) or intelligence quotient (IQ) as well as their interpersonal and communication skills sometimes referenced as emotional quotient (EQ). While the g factor

selection is subject to hiring processes to maximize its average across a team, the other two allow for combinatory selection when staffing teams. Following that logic, the first step for a manager when staffing projects or tasks is to evaluate the technical and/or business skills needed to perform a given task or project. That process is outside of the work of this study. The second block of aspects are those reflected in the surface-level and deep-level diversity factor analyses in this study. Any combination of team members along one of these factors can follow three basic principles: maximize (or minimize), diversify or create disparity. A first aspects for a team design is that managers should minimize diversity in residence. While findings in this study suggest that higher diversity does not affect team output negatively, other studies find negative effects. The main reason to do so is rather in the inner dynamic of teams. Diversity in residence is influencing Peer results negatively. More detailed the results of the study suggest that in particular diversity in language has a strong relation to location diversity. Therefore, it can be said that a manager should very carefully decide in how many locations her team resides and how these collaborate. Practically talent availability, business footprint and how teams are designed to work together must be carefully balanced. If a multi-location team setup is not avoidable, guidance derived from the finding here is to minimize the native languages spoken. In other words, multi-location teams in one country, like multi cities in the US, is less impactful to Peer reviews than across countries with different languages. In managerial language the recommendation is to implement a location strategy considering these aspects. Classic location strategies usually do not include such aspects and are more focused on cost and talent.

The next aspects in surface-level diversity that requires attention is the age diversity in teams. Age is higher than gender and education in effect size. As such should be on a higher priority for managers. As we have seen from analysis results lower age diversity has positive impact on the Peer reviews. Practically, team compositions should focus on similar age levels within the designed groups rather than diverse age distribution. Gender diversity is another relatively simple composition parameter. Not saying it is in effect and why, but as for a manager to pay attention to. The study's results suggest to strive for maximum gender diversity for teams to improve team outcome. Maximizing in this case means in most cases achieving parity between female and male members. In the context of the SMWT R&D teams this might be easier said than done. The domain of software and engineering based research is unfortunately still a male dominated world. As such there is a more strategic recommendation

similar to the aspect of location strategy. Adjusting an existing staff basis for the ability to achieve these goals can be mainly done through the hiring and recruiting processes. Therefore, the overall composition of the wider team should keep such a factor in mind. That is not only true for gender but also for age. The last surface-level recommendation is around education diversity. It must be mentioned again that diversity here is not in terms of subject matter expertise but on the dimension of education level. Education variety will negatively influence Peer reviews. Therefore, it is advisable to have group team members with similar educational levels rather than not.

Getting on to the deep level factors in diversity a first focus goes to character traits. One outcome of the study is the recommendation to try to get a noticeable extraversion disparity in teams. A small amount of highly extraverted members should be matched with a remaining mix of less extraverted members to balance between team processes and outcome. While the mechanisms are not simple and straight forward as we see in a bit, the reasoning here is simple when just looking at it top down. Such an effort helps to address two issues and findings in the analysis. One is that some level of extraversion is positive for getting results and outcome communicated and disseminated while too many similarly extraverted team members seem to hinder the team to achieve higher performance and getting things done. Looking for a certain level of disparity in teams along the dimension of extraversion helps those issues no matter what the overall level of extraversion of the team might be. Very much in line with most research to this point and also with the study here in general, minimization of the level of neuroticism in the team should be a goal for a manager. Be it through the composition of teams or through managing the wider staff available to staff teams. For cultural dimensions the effects that the study found are following a common schema. Higher levels and higher SD show positive levels of influence towards external outcome of teams. That effect is visible for risk and directness. One interpretation and recommendation here is similar to extraversion before. Such traits seem to be positive for teams if they are present in general but it is preferable to even have some “outliers” in the team to boost the positive effect. This would go in line with a common notion for character traits and culture traits that once groups are becoming more similar/homogenous and extreme in their dimensions other effects can start prevailing. A common sense example would be that a group that maximized to very high levels on agreeability could very much become victim to high levels of group thinking and cognitive bubbles. An added variance would counteract such unwanted effects.

Recommendation to managers here is to maximize for risk and directness while still maintaining a significant variance in the teams. Higher SD (variance) for independence, openness, egalitarism, and task-orientation also have positive influence on PO_Task and PO_Exp of teams. Even though the general levels for these traits do not show significance in influence a similar argument than before can be made. That said, the direction of the level cannot be determined though. General higher or lower levels could be beneficial. Recommendation here is to maintain a significant variance of these dimensions within the teams.

Aspects of inner dynamics of a team

Moving on to the inner section of the managers black box. The importance of the inner dynamics of teams already became visible during the last section highlighting compositional recommendations. As we have seen in the theoretical conception (chapter 9.1) and literature review (chapter 9.2) recent years provided more focus on the internal dynamics and processes as moderation and mediation on team performance. We reviewed concepts like CEM and Contingent Configuration (CCA). In this chapter we will not repeat the concepts but rather highlight how the results from this study in light of such concepts influence recommendations to managers in practical terms.

Taking a step back to the black-box model in management as an approach to teams it becomes apparent that one very high level recommendation to managers is to not ignore the inner processes and dynamic of teams but rather balance the two viewpoints, composition and inner, with each other. In context of this study this is a particularly important point as SMWT are very different in the way they relate to managers and leaders than traditional project leaders who typically have more insights and direct feedback on what is going on within the team and how the dynamics work. Based on that observation it is recommended for managers of SMWT teams to keep monitoring and retaining insights around a few aspects of the teams working together. First aspects are based on insights and the model that is developed by van Knippenberg et al. (2004), the CEM model. Fundamentally, it integrates the social categorization and the information/decision making perspective of teams in relation to performance. As we started to discuss the compositional parameters have influence into these processes and also have time effects attached to them that we discussed earlier in this study. As such a manager/leader should stay in touch with the groups inner dynamics of SMWT to

gain insights how such processes develop. Without repeating full aspects here as we have reviewed them earlier, van Knippenberg et al. argue that diversity in the type of information and social categorization are not diversities themselves but rather aspects of diversity. That suggests to leaders to keep assessing the quality and flow of theses aspects on a continuous basis with SMWTs. They also argue that eventually these effects lead to identity, cohesion, and commitment in the team which then positively influences performance. It is difficult to formally assess these aspects in a real world enterprise context for a manager continuously. But it is suggested that proxies like identity/cohesion and how conflicts gets handled by the teams indicate how the more complex mechanisms are working.

One aspect supporting this approach empirically derived from this study is that the diversity in residence/language might not only be a communication problem. Another possible explanation is that there are identity/categorization processes contributing to the effect. As there are more people and departments in the respective locations where the teams of this study reside, there are overarching and contrasting location identities influencing the teams identify. Similarly Moynihan & Peterson (2001) suggest that the team process effects team performance with the influence of composition dimensions in personality when proposing their Continuous Configuration Approach. In line with van Knippenberg et al. they see effects that are task related and the behavior that the teams bring towards them. They also see interpersonal and resource behavior as factors influencing the outcome of teams. While they see task related behavior mainly influencing outcome performance, interpersonal is influencing member satisfaction with the group, and resource behavior is influencing team viability. While it is more natural and widely known to managers to monitor and critique task related processes and to ensure viability it is as well important to monitor the interpersonal aspects. This is also supporting the conclusions from earlier. In summary, the recommendation to a manager of SMTW teams is to establish methods and routines to monitor social processes related to conflict, cohesion, commitment and other interpersonal behavior in addition to compositional aspects that we have discussed before and reflect their relation to each other. This is important in regards to the interactions and corresponding effects that diversity dimensions themselves and across team members are showing. The les investigated cultural parameters included in this study also suggest to monitor their influence. These interactions are very complex and not fully understood. Moynihan and Peterson (p.355) for example list a wide variety of interaction of different mediators that are possible.

The multi-dimensional nature, such as the agreeability example mentioned earlier, while being still somewhat trivial, are illustrating this. This interaction is also the reason why the recommendation of this study is to use and pay attention to compositional parameters but at the same time not ignoring inner processes and dynamic of teams which are far more complex and strongly influenced by them. One empirical bias for such a recommendation is that about half of the significant relations between independent and dependent variables are effects related to Peer review data. Peer review data in this study represents the teams internal view of the situation. As such a significant impact is present in the compositional parameters influencing internal affairs and over time ultimately influencing team performance. Another indication that is important to monitor relating team internals is related to the Peer rating performance effects. A few compositional parameters show their effects only for the Peer reviews, like conscientiousness or neuroticism. For a manager it is important to understand that a black-box approach is hiding such effects from her as this is a particular threat with SMWTs and/or Agile approach. A possible future research or management project building upon this would be to expand the Peer review method to be more formally and more theory focused on models like CEM and CCA. This would help to manage the situation in the long run to be more efficiently and more relevant. However, for the case here, it is a strong support to maintain Peer reviews and also train managers to understand the connection between composition and inner processes better.

Aspects of managerial actions

Now having identified the need to understand inner matters of teams, the question remains what to do with such insights and how to action on them. One path to action is a classical management approach which is to work with individuals and their approach to the situation through methods such as coaching or similar methods. A discussion on what options are available is left outside of this study. What is part of this study for review and recommendations are two aspects of more strategic nature to address such insights into team processes in relation to compositional parameters as this is the focus of this study.

First aspect is around one of the major strategic tasks of a manager to manage talent for an organization. This aspect represents the biggest lever in the process of retaining talent through hiring or managing out staff. The process most significant to team composition in this context is the hiring process. Two considerations can be influenced by a manager when

hiring. One is the hiring location and the other is who to hire. The factor of location could have significant influence in relationship to talent availability and corporate policies. But in general, it should be attempted to minimize locations and with that also language diversity when recruiting for talent that is supposed to work in SMWTs. As we have seen before. Selecting talent is a more difficult task. First and foremost, the mix between subject matter expertise and surface-level and deep-level factors needs to be balanced. Guidance to her comes from composition parameters and team internal processes as discussed earlier. The second recommendation for hiring talent is the creation and maintenance of the right mix in the pool of people hired for teamwork. This is highly dependent on employees already on payroll, pool size and team size, but some guidance can still be given e.g. gender related is the easiest to recommend. A pool of team members should strive for parity in gender to increase the statistical likelihood of being able to staff a project in maximized gender diversity. General advice derived from the study is to adopt the recruiting process to include the individual considerations and factors and to manage a pool mix. Most processes today in enterprise hiring processes are not designed to do so. They often only allow for maximizing certain parameters in general for one candidate or position.

When diversity is rated important for an organization, another effect arises: “people attract similar people,” as Schneider (1987) postulates. Desirable people will attract more of the same desirable people, and it will make team members stay with the organization long term, leading to lower turnover as Halfhill et al. (2005) conclude. Assessing factors such as culture traits, personality traits or even base selections on age factors might be subject to local regulations and laws and therefore difficult to implement as tools in a direct fashion in some cases.

The second set of aspects is the more strategic employee/team members talent management and training. Most team member factors assessed in this study do not change much over time. Nevertheless, it is possible to improve the way that organizations treat diversity and function with diversity present through internal dynamic and processes. In recent years, some scholars have started to introduce the idea of diversity management within an organization. The motivation to do so, as stated by Guillaume et al. (2010), is: “When mismanaged such diversity can undermine employee social integration and effectiveness and lead to lower work group performance; when managed effectively, however, as well as facilitating social

integration and effectiveness, diversity can also promote creativity and innovation." The same paper suggests addressing work group level factors, organizational and leadership factors as well as societal factors when attempting to manage diversity. Kearney and Gebert (2009) conclude in their study with 62 R&D teams that a transformational leadership style with teams can successfully mitigate some negative effects from diversity found in this study. Adverse team performance, through age diversity for example, could be mitigated through the presence of transformational leadership. For SMWTs, however, it is only partially possible to provide a high level of transformational leadership as a manager. Therefore, it seems advisable to provide training and awareness to teams and their members to such style differences so as to ensure appropriate style within the team. Such an approach is also backed by Ely and Thomas (2009), as they find positive effects when doing so. They found that integration and learning for an organization had a positive influence on how tensions related to diversity is managed. They suggest addressing the possible issues and training team members and managers to more efficiently deal with the effects of diversity. Mathieu et al. (2017 p.458) also suggest that not always can a team be ideally configured. But there is some research and evidence around that one can compensate for that as we have seen. Training might help to address individuals' approach to team can help. Overall, a positive summary is drawn by Guillaume et al. (2013) in a meta study on the current state of diversity research for the existence and implementation of diversity management efforts. Ohunakin et al. (2019) have similar conclusions in assessing parts of the Shell corporation and the positive effects of diversity management.

Lastly, as this might has impact on leadership and management practice/considerations, a few contingencies to the recommendations will be discussed. The first that is most broadly in its impact is the question if such recommendations could be used, or are applicable, outside the R&D, Software/Service SMWT context. In summary most of the general recommendations are applicable in the mind of the author. Most recommendations are fundamentally based on models that have broader applicability than the case of this study. Having said this, one needs to be careful when applying detailed findings that involve relations of dimensions or priorities of dimensions. As we have seen earlier, teams in this study do represent a somewhat special case in two dimensions. One is the special way of operating and organizational context. The other one being the sample that differs from general population in character and education for example. These two factors also pose potential problems as

fundamental defining parameters might change when stepping outside of the context of this study. One example here is that once we leave the field of SMWTs we usually enter the field of teams that include a leadership figure in its configuration. While this still might be the case in SMWTs that a de facto leader gets chosen by the team explicitly or implicitly, the situation of an externally assigned leader for the tactical work of a team represents a different situation than in this study. Following scholars in the field of SMWTs (e.g. Hackmann (1987) or Renn (1998)) effects here can be positive and negative depending on what type of leader acts with the team. Renn (p.817) argues in line with more recent work like the models we have reviewed that teams can suffer for example from conflict on team decisions. Those examples, also similar to e.g. Yeats & Hyten (1994), suggest that effects relating a leader in a team are generally following the same principles of cohesion, social categories analogue to the ones reviewed earlier. The argument is that if the assigned leader is acting in a way that the team needs to excel in these dimensions then performance is increasing. In other words the leader must act complimentary to the needs or deficiencies of a team. In that respect the situation is similar to the second recommendation of this study which suggests that managers of SMWT stay abreast with dynamics in their team. Yet it might be a very different situation when factors like high status leadership or non-cooperative leadership styles are used by that team leader. Then teams might fall apart into more individual task driven individuals supporting and integrating work with such a leader as Renn analyses. Then insights and recommendation from this study are most likely irrelevant and not applicable.

A second topic in terms of contingencies and applicability is a very current one. The topic of teams working virtual. The teams in context of the work here are mostly working virtual across locations and face to face at each location. Travels are usually in schedule for teams to at least partially meet over the course of a project. With the entrance of travel restrictions due to the COVID19 pandemic these travels completely seized for over a year now. At the same time the teams at questions are to the greatest extent working in home office environments for the same time frame. The interesting question now is, if that virtual aspect of working together has any deeper impact to team performance. Kirkman et al. (2004) analyze empirically that a lower number of face to face meetings, “being more virtual” has negative influence on team empowerment and process improvement and hence on team performanc. In regards to the work presented here such an effect might explain or contribute to the diversity in residence findings. Teams in this study work mostly virtual across locations. Having said

this, the effect is also present with native language. As such this might be true or not. Overall it is agreed knowledge that surface-level diversity is less salient when working remotely (e.g. Humphrey (2007), Homan (2008)) and deep-level effects, respectively team dynamics, start playing a bigger role. The study could not confirm this through data. The data and the context of the teams does not allow to test for such effect directly. Yet again the location effect is very dominant. It might very well be driven effects like identity as a team dynamic effect. With the move to full virtual work for all team members it might be very interesting to use that same form of data that is still being collected for the teams to the current day and start a comparative analysis of partially virtual location based data collected so far and full virtual data to see how such effects change and might have time effects.

13 SUMMARY AND CONCLUSIONS

This study modeled and analyzed data from a multi-national R&D SMWT team regarding surface-level and deep-level diversity factors and their contribution to three different aspects of team performance. It modeled performance measures based on recent studies that describe team performance models, like IMOI by Ilgen et al. (2005), and derived three performance measures from it. One is the inner process performance of teams through establishing a Peer review process and data collection. The second is the performance on self-set goals (PO_Team) defined intrinsically by teams. The third is an objective output performance measuring process that includes pre-negotiating goals and expectations (PO_Exp). Surface-level and deep-level diversity measures were operationalized and 111 projects over seven years converted into analytical data. Through correlation, regression and moderation analysis, the individual influence of the diversity factors on each of the three performance measures was investigated. The influence of several diversity factors from previously conducted research was confirmed.

Additionally, the work presented here contributed new insights in two areas. One area is with diversity factors found in earlier studies, and confirmed by this study, it could be shown how each of these factors relates in different ways to the three levels of performance. Previous studies overwhelmingly use only one construct of performance for teams, missing new findings on processes influencing team performance. An example here is that previously it was agreed that gender diversity has positive influence on team performance in general. The work here was able to show that its positive influence is mainly on the team inner processes and self-set goals. No direct positive influence was found on externally set content goals performance. Secondly, the work presented also contributed new diversity factors which have influence on team performance. Examples include the positive influence of educational level diversity and the culture dimensions of risk and task orientation being uplifting to inner processes. Effect sizes are small but vary to some degree. That variation also plays a role when recommending action from the findings. Finally, practical recommendations were developed for the practitioner dealing with multi-national SMWTs in a software and service-related R&D context. Guidance for team composition, hiring and talent management and diversity management as an organizational practice were given based on the findings of the

study. The study also incorporated the findings with recently developed models of diversity and performance focusing more in the internal dynamics and effects of teams.

Looking ahead, some potential research questions which look deeper into the findings arose beyond statistical analysis. Unfortunately, most of these questions are outside of what the data collected here can provide in this study. An example here is that moving into assessing some of the "why" questions with structured qualitative research within teams and POs might prove a valuable exercise to further improve the modeling of processes and dynamics around teams and their performance. These models are prerequisite to dive deeper into the field using modern data science techniques such as machine learning and simulations of what does and doesn't work for teams regarding diversity. It would also be beneficial to use similar data collection structures on a wider variety of teams to ensure more generalizable results, as the sample of team members in this study posed potential sample restrictions in terms of the nature of work, education levels being comparatively high and the hiring selection being done by mainly two managers. The study provided new insights into the field of team diversity and team performance and could fulfill its original motivation: "the need for more empirical attention to the climates or cultures that facilitate the positive effects of diversity on work and organizational outcomes.", as stated by Guillaume et al. (2013). The approach taken and contribution to knowledge leads to further research in the same form and fashion to push forward the field of team diversity research as it will stay an important field with vast practical impact for many years to come.

14 BIBLIOGRAPHY

Agile Alliance (2016) *Agile Alliance*. 2016.

Ancona, D.G. and Caldwell, D.F. (1992) Demography and Design: Predictors of New Product Team Performance. *Organization Science*. 3 (3), 321–341.

Anderson, N.R. and West, M.A. (1998) Measuring climate for work group innovation: Development and validation of the team climate inventory. *Journal of Organizational Behavior*. 19 (3), 235–258.

Anderson, N.R. and West, M.A. (1994) *Team climate inventory: Manual and user's guide*. Windsor, Berkshire, NFER Nelson.

Aperian Global (2020) *Discover the GlobeSmart learning Plattform by Aperian Global*. 2020.

Aqeel, Iqbal, M., Aldaihani, A.R. and Shah, A. (2019) Big-five personality traits mapped with software development tasks to find most productive software development teams. *International Journal of Innovative Technology and Exploring Engineering*. 8 (12), 965–971.

Armstrong, S.J. and Priola, V. (2001) Individual differences in cognitive style and their effects on task and social orientations of self-managed work teams. *Small Group Research*. 32 (3), 283–312.

Aronoff, J., Messe, L.A. and Wilson, J.P. (1983) Personality factors in small group functioning. In: H.H. Blumberg, P. Hare, V. Kent, and M. Davies (eds.). *Small Groups and Social Interaction*. Chichester, England, Wiley. pp. 79–88.

Arthurs, N., Stenhaug, B., Karayev, S. and Piech, C. (2019) Grades are not normal: Improving exam score models using the logit-normal distribution. *EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining*. 252–257.

Baker, D.P. and Salas, E. (1992) Principles for measuring teamwork skills. *Human Factors*. 34 (4), 469–475.

Bantel, K.A. and Jackson, S.E. (1989) Top Management and Innovations in Banking : Does the Composition of the Top Team Make a Difference? *Strategic Management Journal*. 10 (S1), 107–124.

Barrick, M.R. and Mount, M.K. (1991) The Big Five Personality Dimensions and Job Performance: a Meta-Analysis. *Personnel Psychology*. 44 (1), 1–26.

Barrick, M.R., Stewart, G.L., Neubert, M.J., Mount, M.K., et al. (1998) Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*. 83 (3), 377–391.

Barry, B. and Stewart, G.L. (1997) Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology*. 82 (1), 62–78.

Bartram, D. (2013) Scalar Equivalence of OPQ32: Big Five Profiles of 31 Countries. *Journal of Cross-Cultural Psychology*. 44 (1), 61–83.

Beal, D.J., Cohen, R.R., Burke, M.J. and McLendon, C.L. (2003) Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations. *Journal of Applied Psychology*. 88 (6), 989–1004.

Bechger, T.M., Maris, G. and Hsiao, Y.P. (2010) Detecting halo effects in performance-based examinations. *Applied Psychological Measurement*. 34 (8), 607–619.

Bell, S.T. (2007) Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*. 92 (3), 595–615.

Bell, S.T. and Marentette, B.J. (2011) Team viability for long-term and ongoing organizational teams. *Organizational Psychology Review*. 1 (4), 275–292.

Bettenhausen, K.L. (1991) Five years of Group Research: What we have learned

and what needs to be addressed. *Journal of Management*. 17 (2), 345–381.

Blau, P.M. (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure.* New York, The Free Press.

Bliese, P.D. (2000) Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. pp.349–381.

Bouchard, T.J., Barsaloux, J. and Drauden, G. (1974) Brainstorming procedure, group size, and sex as determinants of the problem-solving effectiveness of groups and individuals. *Journal of Applied Psychology*. 59 (2), 135–138.

Bowlin, G. (2011) *Corporate Goals*

Bradley, B.H., Postlethwaite, B.E., Klotz, A.C., Hamdani, M.R., et al. (2012) Reaping the benefits of task conflict in teams: The critical role of team psychological safety climate. *Journal of Applied Psychology*. 97 (1), 151–158.

Brickson, S. (2000) The Impact of Identity Orientation on Individual and Organizational Outcomes in Demographically. *Academy of Management Review*. 25 (1), 82–101.

Brodbeck, F.C., Anderson, N. and West, M.A. (2000) Das Teamklima-Inventar: Handanweisung und Validierung der deutschsprachigen Version. *Hogrefe*.

Brodbeck, F.C., Guillaume, Y.R.F. and Lee, N.J. (2010) Ethnic diversity as a multilevel construct: the combined effects of dissimilarity, group diversity, and societal status on learning performance in work groups. *Journal of Cross-Cultural Psychology*. 42 (7), 1198–1218.

Bunderson, S.J. and Sutcliffe, K.M. (2002) Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. *Academy of Management Journal*. 45 (5), 875–893.

Burns, L.R. (1989) Matrix management in hospitals: testing theories of matrix

structure and development. *Administrative science quarterly*. 34 (3), 349–368.

Burton, R.M., DeSanctis, G. and Obel, B. (2006) *Organizational Design: A Step-by-Step Approach.* Cambridge, Cambridge University Press.

Campion, M.A. and Medsker, G.J. (1993) Relations between work group characteristics and effectivness: Implications for designing effective work groups. *Personnel Psychology*. 46 (1), 823–851.

Cardon, P.W. (2008) A critique of Hall's contexting model: A meta-analysis of literature on intercultural business and technical communication. *Journal of Business and Technical Communication*. 22 (4), 399–428.

Carroll, J.B. (1993) *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* Cambridge, Cambridge University Press.

Carte, T.A., Chidambaram, L. and Becker, A. (2006) Emergent leadership in self-managed virtual teams: A longitudinal study of concentrated and shared leadership behaviors. *Group Decision and Negotiation*. 15 (4), 323–343.

Chatman, J. a. and Flynn, F.J. (2001) The influence of demographic heterogeneity on the emergence and consequences of cooperative norms in work teams. *Academy of Management Journal*. 44 (5), 956–974.

Chen, G., Kirkman, B.L., Kanfer, R., Allen, D., et al. (2007) A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology*. 92 (2), 331–346.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2003) *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd edition. New York, Routledge Taylor & Francis Group.

Cohen, S.G. and Bailey, D.E. (1997) What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*. 23 (3), 239–290.

Costa, P.T. and McCrae, R.R. (1992) *Revised NEO Personality Inventory (NEO PIR) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL, Psychological Assessment Resources.

Coutu, D. (2009) Why teams don't work: An Interview With J. Richard Hackman. *Havard Business Review*. 87 (5), 98–105.

Curral, L.A., Forrester, R.H., Dawson, J.F. and West, M.A. (2001) It's what you do and the way you do it:team task, team size, and innovation-related group processes. *European Journal of Work and Organizational Psychology*. 10187–204.

Dahlin, K.B., Weingart, L.R. and Hinds, P.J. (2005) Team diversity and information use. *Academy of Management Journal*. 48 (6), 1107–1123.

Devine, D.J. and Philips, J.L. (2001) Do smarter teams do better: A meta-analysis of cognitive ability and team performance. *Small Group Research*. 32 (5), 507–532.

Van Dick, R., Van Knippenberg, D., Hägele, S., Guillaume, Y.R.F., et al. (2008) Group diversity and group identification: The moderating role of diversity beliefs. *Human Relations*. 61 (10), 1463–1492.

Diekmann, J. and König, C.J. (2015) Personality testing in personnel selection: Love it Leave it? Understand it! In: *Employee recriutment, selection, and assessment: Contemporary issues for theory and prectice*. Hove, UK, Phychology Press. p.

Dionne, S.D., Randel, A.E., Jaussi, K.S. and Chun, J.U. (2004) Diversity and Demography in Organizations: A Levels of Analysis Review of the Literature. *Research in Multi-Level Issues*. 3 (1), 181–229.

Doherty, W.J. (2016) *Psychometric Properties of the GlobeSmart Profile*

Driskell, J.E., Hogan, R. and Salas, E. (1987) Personality and group performance. In: *Personality and Social Psychology Review*. Beverly Hills, Sage Publications Ltd. pp. 91–112.

Durham Business School (2015) *DBA Program*

Durham University (2020) *Ethical considerations*. [Online]. 2020. Available at: https://www.dur.ac.uk/research.innovation/governance/ethics/considerations/ (Accessed: 7 September 2020).

Earley, C. and Mosakowski, E. (2000) Creating Hybrid Team Cultures: An Empirical Test of Transnational Team Functioning. *Academy of Management Journal*. 43 (1), 26–49.

Ellwart, T., Bündgens, S. and Rack, O. (2013) Managing knowledge exchange and identification in age diverse teams. *Journal of Managerial Psychology*. 28 (7), 950–972.

Elron, E. (1997) Top management teams within multinational corporations: Effects of cultural heterogeneity. *Leadership Quarterly*. 8 (4), 393–412.

Ely, R.J., Padavic, I. and Thomas, D.A. (2012) Racial Diversity, Racial Asymmetries, and Team Learning Environment: Effects on Performance. *Organization Studies*. 33 (3), 341–362.

Ely, R.J. and Thomas, D.A. (2009) Cultural Diversity at Work : The Effects of Diversity Perspectives on Work Group Processes and Outcomes. *Administrative Science Quarterly*. 46 (2), 229–273.

Enders, C.K. (2010) *Applied Missing Data Analysis*. New York, Guildford Press.

Erez, A., LePine, J. a. and Elms, H. (2002) Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*. 55 (4), 929–948.

Eseryel, U.Y., Crowston, K. and Heckman, R. (2020) Functional and Visionary Leadership in Self-Managing Virtual Teams. *Group and Organization Management*.

Fields, A., Miles, J. and Field, Z. (2012) *Discovering Statistics using R*. Los Angeles, Sage Publications Ltd.

Fine, A. (2009) *InsideOut Coaching. A program for leaders*

Fox, J. (2002) *Bootstrapping Regression Models. Appendix to An R and S-PLUS Companion to Applied Regression* pp.1–14.

Fritzsche, L., Wach, D., Jungmann, F. and Wegge, J. (2017) Zur Bedeutung von Faultlines als Grenzen bei der Teamarbeit. *Gruppe. Interaktion. Organisation. Zeitschrift fur Angewandte Organisationspsychologie*. 48 (2), 127–135.

Frost, J. (2020) *How To Interpret R-squared in Regression Analysis* [Online]. Available at: https://statisticsbyjim.com/regression/interpret-r-squared-regression/.

Funder, D.C. (2010) *The personality puzzle*. 5th edition. New York, London, W. W. Norton.

Gist, M.E., Locke, E.A. and Taylor, M.S. (1987) Organizational Behaviour: Group Structure, Process, and Effectivness. *Journal of Ma*. 13 (2), 237–257.

Gleich, R., Losbichler, H. and Möller, K. (2017) *Standards in Controlling*. Martin Tschandl (ed.). Freiburg, Haufe Gruppe.

Global Aperian (2016) *GlobeSmart ® Profile User Guide*

Guevara, M. r., Hartmann, D. and Mendoza, M. (2017) *Package" 'diverse'*

Guillaume, Y.R.F., Brodbeck, F.C. and Riketta, M. (2012) Surface- and deep-level dissimilarity effects on social integration and individual effectiveness related outcomes in work groups: A meta-analytic integration. *Journal of Occupational and Organizational Psychology*. 85 (1), 80–115.

Guillaume, Y.R.F., Dawson, J.F., Priola, V., Sacramento, C.A., et al. (2010) *Managing Diversity in Organizations: An Integrative Model and Agenda for Future Research*

Guillaume, Y.R.F., Dawson, J.F., Woods, S.A., Sacramento, C. a., et al. (2013) Getting diversity at work to work: What we know and what we still don't know. *Journal of Occupational and Organizational Psychology*. 86 (2), 123–141.

Guzzo, R.A. and Dickson, M.W. (1996) Teams in Organizations: Recent Research on Performance and Effectiveness. *Annual Review of Psychology*. 47 (1), 307–338.

Hackman, R.J. (1982) *A set of methods for research on work teams*

Hackman, R.J. (1976) *The design of self-managing work groups*

Hackman, R.J. (1987) The design of work teams. *Handbook of organizational behavior*. 315–342.

Hackman, R.J. (1986) The psychology of self-managed teams in organizations. In: M.S. Pallak and R.O. Perloff (eds.). *Phychology and Work: Productivity, change, and employment*. Washington, DC, Phychology Association. pp. 89–136.

Hackman, R.J. (1998) Why teams don't work *Leader to Leader*. (Winter) pp.24–32.

Halfhill, T., Sundstrom, E., Lahner, J., Calderone, W., et al. (2005) Group Personality Composition and Group Effectiveness: An Integrative Review of Empirical Research. *Small Group Research*. 36 (1), 83–105.

Hall, E.T. (1959) *The silent language*. Greenwood, Fawcett.

Hannan, M.T. and Freeman, J. (1984) Structural Inertia and Organizational Change. *American Sociological Review*. 49 (2), 149.

Hannan, M.T., Polos, L. and Carroll, G.R. (2007) *Logics of Organization Theory*. Princeton, NJ USA, Princeton University Press.

Harrison, D. a., Price, K.H., Gavin, J.H. and Florey, A.T. (2000) Time, Teams, and Task Performance: a Longitudinal Study of the Changing Effects of Diversity on Group Functioning. *Academy of Management Proceedings*. 2000 (5), C1–C6.

Harrison, D.A. (2018) *Variation into Elevation* p.37.

Harrison, D.A. and Klein, K.J. (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*.

32 (4), 1199–1228.

Harrison, D.A., Price, K.H. and Bell, M.P. (1998) *Beyond Relational Demography : Time and the Effects of Surface- and Deep-Level Diversity on Work Group*. 41 (1), 96–107.

Hiller, N.J., Day, D. V. and Vance, R.J. (2006) Collective enactment of leadership roles and team effectiveness: A field study. *Leadership Quarterly*. 17 (4), 387–397.

Hirst, M. (2008) *The air transport system*. 1st edition. Cambridge, Woodhead Publishing Limited.

Hofstede, G. (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Second Edi. Thousand Oaks, CA, Sage Publications Ltd.

Hofstede, G. (2011) Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*. 2 (1), 1–26.

Hofstede, G.J. and Bond, M.H. (1984) Hofstede's culture dimensions. An idependent validation using Rokeach's value survey. *Journal of Cross-Cultural Psychology*. 15 (4), 417–433.

Hollenbeck, J.R., Beersma, B. and Schouten, M.E. (2012) Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review*. 37 (1), 82–106.

Homan, A.C., Hollenbeck, J.R., Humphrey, S.E., Van Knippenberg, D., et al. (2008) Facing Differences with an open Mind: Openenss to Experience, Salience of Intragroup Differences and Performance of Diverse Work Groups. *Academy of Management Journal*. 51 (6), 1204–1222.

Homan, A.C., van Knippenberg, D., Van Kleef, G.A. and De Dreu, C.K.W. (2007) Bridging Faultlines by Valuing Diversity: Diversity Beliefs, Information Elaboration, and Performance in Diverse Work Groups. *Journal of Applied Psychology*. 92 (5), 1189–1199.

Humphrey, S.E., Hollenbeck, J.R., Meyer, C.J. and Ilgen, D.R. (2011) Personality Configurations in Self-Managed Teams: A Natural Experiment on the Effects of Maximizing and Minimizing Variance in Traits. *Journal of Applied Social Psychology*. 41 (7), 1701–1732.

Humphrey, S.E., Hollenbeck, J.R., Meyer, C.J. and Ilgen, D.R. (2007) Trait configurations in self-managed teams: A conceptual examination of the use of seeding for maximizing and minimizing trait variance in teams. *Journal of Applied Psychology*. 92 (3), 885–892.

Ilgen, D.R. (1999) Teams embedded in organizations: Some implications. *American Psychologist*. 54 (2), 129–139.

Ilgen, D.R., Hollenbeck, J.R., Johnson, M. and Jundt, D. (2005) Teams in organizations: from input-process-output models to IMOI models. *Annual review of psychology*. 56 (1), 517–543.

ISO (2004) *5218:2004*

ISO (2019) *8601-1:2019*

ISO (1999) *ISO 3166 standard – Codes for the representation of names of countries and their subdivisions*

Jackson, S.E. (1992) Team composition in organizational settings: Issus in managing an increasingly diverse workforce. In: *Group processes and productivity*. Newbury Park, CA, SAGE Publications, Inc. pp. 136–180.

Jackson, S.E., Brett, J.F., Sessa, V.I., Cooper, D.M., et al. (1991) Some Differences Make a Difference: Individual Dissimilarity and Group Heterogeneity as Correlates of Recruitment, Promotions, and Turnover. *Journal of Applied Psychology*. 76 (5), 675–689.

Jackson, S.E., Joshi, A. and Erhardt, N.L. (2003) Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management*. 29 (6), 801–830.

Joshi, A. and Roh, H. (2009) The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*. 52 (3), 599–627.

Karasek, R.A. (1979) Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. *Administrative Science Quarterly*. 24 (2), 285.

Katzenbach, J.R. and Smith, D.K. (2011) The discipline of teams *Harvard Business Review, Building better teams*. pp.19–44.

Kearney, E. and Gebert, D. (2009) Managing Diversity and Enhancing Team Outcomes: The Promise of Transformational Leadership. *Journal of Applied Psychology*. 94 (1), 77–89.

Keller, R.T. (2001) Cross-Functional Project Groups in Research and New Product Development : Diversity , Communications , Job Stress , and Outcomes. *Academy of Management Journal*. 44 (3), 547–555.

Kieniger, M. and Schimank, C. (2017) *Digitalisierung der Unternehmenssteuerung*. Michael Kieninger (ed.). Stuttgart, Schäffer-Poeschel Verlag.

Kirkman, B.L., Rosen, B., Tesluk, P.E. and Gibson, C.B. (2004) The Impact of Team Empowerment on Virtual Team Performance: The Moderating Role of Face-to-Face Interaction. *Academy of Management Journal*. 47 (2), 175–192.

Klaus, D.J. and Glaser, R. (1968) Increasing team proficiency through training. *Defense Technical Information Center*. (8), 65.

van Kleef, G.A., Homan, A.C., Beersma, B. and van Knippenberg, D. (2010) On angry leaders and agreeable followers: How leaders' emotions and followers' personalities shape motivation and team performance. *Psychological Science*. 21 (12), 1827–1834.

Klein, K.J., Bliese, P.D., Kozlowski, S.W.J., Dansereau, F., et al. (2000) Multilevel Analytical Techniques. Commonalities, differences, and continouing questions. In: K. J. Klein & S. W. J. Kozlowski (ed.). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. Jossey-Bass. pp. 512–

553.

van Knippenberg, D., Dawson, J.F., West, M.A. and Homan, A.C. (2011) Diversity faultlines, shared objectives, and top management team performance. *Human Relations*. 64 (3), 307–336.

Van Knippenberg, D., De Dreu, C.K.W. and Homan, A.C. (2004) Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*. 89 (6), 1008–1022.

Van Knippenberg, D. and Schippers, M.C. (2007) Work group diversity. *Annual Review of Psychology*. 58 (1), 515–541.

Kogut, B. and Singh, H. (1988) The Effect of National Culture on the Choice of Entry Mode. *Journal of International Business Studies*. 19 (3), 411–432.

Koo, T.K. and Li, M.Y. (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 15 (2), 155–163.

Kozlowski, S.W.J. and Klein, K.J. (2000) A multi-level approach to theory and research in organizatuons: Contextual, temporal, and emergent procesesses. In: Katherine J. Klein and S. W. J. Kozlowski (eds.). *Multilevel theory, research, and methods in organizations*. 3rd–90th edition. San Francisco, Jossey-Bass. p.

Kristof-Brown, A.L. and Stevens, C.K. (2001) Goal congruence in project teams: Does the fit between members' personal mastery and performance goals matter? *Journal of Applied Psychology*. 86 (6), 1083–1095.

Launer, M. (2011) *Effective organization of a research department for a global transportation information provider*.

Lawler, E.E. (1978) The new plant revolution. *Organizational Dynamics*. 6 (3), 3–12.

Lawrence, J.R., Demaree, R.G. and Wolf, G. (1993) rwg: An assessment of within-group interrater agreement. *American Psychologist*. 78 (2), 306–309.

LeBreton, J.M. and Senter, J.L. (2008) Answers to 20 Questions about interrater reliability and interpreter agreement. *Organizational Research Methods*. 11 (2), 815–852.

LePine, J.A., Buckman, B.R., Crawford, E.R. and Methot, J.R. (2011) A review of research on personality in teams: Accounting for pathways spanning levels of theory and analysis. *Human Resource Management Review*. 21 (4), 311–330.

LePine, J.A., Hanson, M.A., Borman, W.C. and Motowildo, S.J. (2000) Contextual performance and teamwork: Implicationsfor staffing. *Resarch in Personnel and Human Resources Management*. 1953–90.

LePine, J.A., Piccolo, R.F., Jackson, C.L., Mathieu, J.E., et al. (2008) A Meta-Analysis of Teamwork Processes : Tests of a Multidimensional Model and Relationships With Team Effectiveness Criteria. *Personnel Psychology*. 61 (1), 273–307.

Levine, J.M. and Moreland, R.L. (2006) *Small groups*. John M. Levine and Richard L. Moreland (eds.). Hove, UK, Taylor & Francis.

Liebermann, S.C., Wegge, J., Jungmann, F. and Schmidt, K.H. (2013) Age diversity and individual team member health: The moderating role of age and age stereotypes. *Journal of Occupational and Organizational Psychology*. 86 (2), 184–202.

Mackenzie, K.D. (2015) *The Process Approach To Multi-Level Organizational Behavior*.

Markova, G. and Perry, J.T. (2014) Cohesion and individual well-being of members in self-managed teams. *Leadership & Organization Development Journal*. 35 (5), 429–441.

Mathieu, J.E., Goodwin, G.F., Heffner, T.S., Salas, E., et al. (2000) The influence of shared mental models on team process and performance. *Journal of Applied Psychology*. 85 (2), 273–283.

Mathieu, J.E., Hollenbeck, J.R., Knippenberg, D. Van and Ilgen, D.R. (2017) A century of work teams in the journal of applied psychology. *Journal of Applied Psychology*. 102 (3), 452–467.

Mathieu, J.E., Kukenberger, M.R., D'Innocenzo, L. and Reilly, G. (2015) Modeling reciprocal team cohesion – Performance relationships as impacted by shared leadership and members ' competence. *Journal of Applied Psychology*. 100 (3), 713–734.

Mathieu, J.E., Maynard, M.T., Rapp, T. and Gilson, L.L. (2008) Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future. *Journal of Management*. 34 (3), 410–476.

Mathieu, J.E., Tannenbaum, S.I., Donsbach, J.S. and Alliger, G.M. (2014) A Review and Integration of Team Composition Models: Moving Toward a Dynamic and Temporal Framework. *Journal of Management*. 40 (1).

Mathisen, G.E., Martinsen, Ø. and Einarsen, S. (2008) The relationship between creative personality composition, innovative team climate, and team innovativeness: An input - process - output perspective. *Journal of Creative Behavior*. 42 (1), 13–31.

Matthews, B. (2019) *Statics and Analytical Geometry*. 2020th edition. ED-Tech Press (ed.). Waltham Abbey, Essex, ED-Tech Press.

Maynard, M.T., Gilson, L.L. and Mathieu, J.E. (2012) Empowerment-Fad or Fab? A Multilevel Review of the Past Two Decades of Research. *Journal of Management*. 38 (4), 1231–1281.

McCrae, R.R. and Costa, P.T. (2007) Brief Versions of the NEO-PI-3. *Journal of Individual Differences*. 28 (3), 116–128.

McCrae, R.R. and Costa, P.T. (1990) *Personality in adulthood*. New York, London, Guildford Press.

McCrae, R.R., Martin, T.A. and Costa, P.T. (2005) Age trends and age norms for the NEO personality inventory-3 in adolescents and adults. *Assessment*. 12 (4), 363–

373.

McGarth, J.E. (1964) *Social psychology:A brief Introduction*. New York, Holt, Rinehart and Winston.

Milliken, F.J. and Martins, L.L. (1996) Searching for common threads: Understnading the multiple effects of diversity in organizational groups. *Academy of Management Review*. 21402–433.

MiniTab (2014) *No TitleHow to Interpret a Regression Model with Low R-squared and Low P values*. 2014.

Mohammed, S. and Angell, L.C. (2003) Personality Heterogeneity in Teams: Which Differences Make a Difference for Team Performance? *Small Group Research*. 34 (6), 651–677.

Mohammed, S. and Angell, L.C. (2004) Surface- and deep-level diversity in workgroups: Examining the moderating effects of team orientation and team process on relationship conflict. *Journal of Organizational Behavior*. 25 (8), 1015–1039.

Mohammed, S. and Harrison, D. a. (2013) The clocks that time us are not the same: A theory of temporal diversity, task characteristics, and performance in teams. *Organizational Behavior and Human Decision Processes*. 122 (2), 244–256.

Moynihan, L.M. and Peterson, R.S. (2001) *A Contingent configuration approach to understanding the role of personality in organizational groups*.

Nestler, S., Grimm, K.J. and Schönbrodt, F.D. (2015) The Social Consequences and Mechanisms of Personality. *European Journal of Personality*. 29 (2), 272–295.

Neuman, G.A., Wagner, S.H. and Christiansen, N.D. (1999) The Relationship between Work-Team Personality Composition and the Job Performance of Teams. *Group and Organization Management*. 24 (1), 28–45.

Neuman, G.A. and Wright, J. (1999) Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*. 84 (3), 376–389.

Nishii, L.H. and Mayer, D.M. (2009) Do Inclusive Leaders Help to Reduce Turnover in Diverse Groups? The Moderating Role of Leader-Member Exchange in the Diversity to Turnover Relationship. *Journal of Applied Psychology*. 94 (6), 1412–1426.

Nonaka, I. and Konno, N. (1998) The concept of 'BA'. Building a foundation for knowledge management. *California Management Review*. 40 (3), 40–54.

Ohunakin, F., Adeniji, A., Ogunnaike, O.O., Igbadume, F., et al. (2019) The effects of diversity management and inclusion on organisational outcomes: A case of multinational corporation. *Business: Theory and Practice*. 20 (1), 93–102.

Peters, T. (2015) *Personality Traits Modeling of Operations Control Center (OCC) Personnel and Vocational Implications*. (December 2015), 54.

Phillips, K.W., Northcraft, G.B. and Neale, M.A. (2006) Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group Processes and Intergroup Relations*. 9 (4), 467–482.

Price, K.H., Harrison, D.A. and Gavin, J.H. (2006) Withholding inputs in team contexts: Member composition, interaction processes, evaluation structure, and social loafing. *Journal of Applied Psychology*. 91 (6), 1375–1384.

R Foundation (2011) *Regression Analysis Manual* pp.1–6.

R Foundation (2020) *the R Project for Statistical Computing*. [Online]. 2020. Available at: https://www.r-project.org/foundation/.

Rapp, T.L., Gilson, L.L., Mathieu, J.E. and Ruddy, T. (2016) Leading empowered teams: An examination of the role of external team leaders and team coaches. *Leadership Quarterly*. 27 (1), 109–123.

Renn, R.W. (1998) A Review: High-Performing Self-Managed Work Teams : A Comparison of Theory and. *Academy of Management Review*. (October), 816–819.

Roberts, B.W., Walton, K.E. and Viechtbauer, W. (2006a) Patterns of mean-level

change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*. 132 (1), 1–25.

Roberts, B.W., Walton, K.E. and Viechtbauer, W. (2006b) Personality traits change in adulthood: reply to Costa and McCrae (2006). *Psychological bulletin*. 132 (1), 29–32.

RTCA SC 181 (2015) *DO-200B, Standards for Processing Aeronautical Data*

Rumsey, D. (2011) *Statistics for Dummies*. 2nd edition. Hobocken, Whiley Publishing Inc.

Schafer, J.L. and Graham, J.W. (2002) Missing data: Our view of the state of the art. *Psychological Methods*. 7 (2), 147–177.

Schiefele, J. (2011) A comparison, application, and critical review of Organizational Culture Analysis Models *The Boeing Company*.

Schippers, M.C., Den Hartog, D.N., Koopman, P.L. and Wienk, J.A. (2003) Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. *Journal of Organizational Behavior*. 24 (6), 779–802.

Schippers, M.C., Homan, A.C. and Knippenberg, D. Van (2013) To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance†. *Journal of Organizational Behavior*. 34 (1), 6–23.

Schneider, B. (1987) The People Make the Place. *Personnel Psychology*. 40 (3), 437–453.

Schönbrodt, F.D., Back, M.D. and Schmukle, S.C. (2012) TripleR: An R package for social relations analyses based on round-robin designs. *Behavior Research Methods*. 44 (2), 455–470.

Seeber, I., Maier, R. and Weber, B. (2014) Opening the black box of team processes

and emergent states: A literature review and agenda for research on team facilitation. *Proceedings of the Annual Hawaii International Conference on System Sciences*. 473–482.

Seo, E., Kang, H. and Song, J. (2020) Blending talents for innovation: Team composition for cross-border R&D collaboration within multinational corporations. *Journal of International Business Studies*. 51 (5), 851–885.

Shin, S.J. and Zhou, J. (2007) When Is Educational Specialization Heterogeneity Related to Creativity in Research and Development Teams? Transformational Leadership as a Moderator. *Journal of Applied Psychology*. 92 (6), 1709–1721.

Simpson (1949) April 30, 1949 vol. 163. *Nature*. 163 (1946), 688.

Smith-Crowe, K., Burke, M.J., Cohen, A. and Doveh, E. (2013) Statistical Significance Criteria for the rwg and Average Deviation Interrater Agreement Indices. *Journal of Applied Psychology*. 99239–261.

Smith, P.B., Dugan, S. and Trompenaars, F. (1996) National Culture and the Values of Organizational Employees. A dimensional Analysis across 43 Nations. *Journal of Cross-Cultural Psychology*. 27 (2), 231–264.

Solanas, A., Selvam, R.M. and Leiva, D. (2012) Common indexes of group diversity: upper boundaries. *Psychological Reports*. 111 (3), 777–796.

Stewart, G.L. (2006) A meta-analytic review of relationships between team design features and team performance. *Journal of Management*. 32 (1), 29–55.

Studio, R. (2020) *R Pubs*. [Online]. 2020. Available at: https://rpubs.com/ (Accessed: 11 September 2020).

Sy, T. (2005) Challenges and Strategies of Matrix Organizations. *Human Resource Planning*. 2839.

The European Parliament and the Council of the European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on

the protection of natural persons with regard to the processing of personal data and on the free movement of such data *Official Journal of the European Union*. pp.1–88.

Thompson, C. and Gregory, J.B. (2012) Managing Millennials: A Framework for Improving Attraction, Motivation, and Retention. *Psychologist-Manager Journal*. 15 (4), 237–246.

Tonidandle, S. and Overall, J.E. (2004) Determining the number of clusters by sampling with replacement. *Psychological Methods*. 9 (2), 238–249.

Triandis, H.C., Kurowski, L.L. and Gelfand, M.J. (1994) Workplace diversity. In: Harry C. Triandis, M.P. Dunnette, and L.M. Hough (eds.). *Handbook of industrial and organizational psychology*. 2nd edition. Palo Alto, CA, Consulting Psychologists Press. pp. 769–827.

Triandis, H.C., McCusker, C. and Hui, C.H. (1990) Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology*. 59 (5), 1006–1020.

Tuckman, B. (1965) Development sequence in small groups. *Phychological Bulletin*. 63 (3), 384–399.

Tuckman, B. and Jensen, C.M. (1977) Stages of Small-Group Development Revisited. *Group & Organizational Studies*. 419–427.

Tuma, N.B. and Hannan, M.T. (1984) *Social Dynamics*. Orlando, Academic Press.

UNESCO (1997) *International Standard Classification of Education 97*

Van der Vegt, G.S. and Janssen, O. (2003) Joint impact of interdependence and group diversity on innovation. *Journal of Management*. 29 (5), 729–751.

Wageman, R. (1997) Critical success factors for creating superb self-managing teams. *Organizational Dynamics*. 26 (1), 49.

Watson, W.E., Johnson, L. and Merritt, D. (1998) Team orientation, self-orientation

and diversity in task groups. *Groups & Organization Management*. 23 (2), 161–188.

Watson, W.E., Kumar, K. and Michaelsen, L.K. (1993) Cultural Diversity's Impact On Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups. *Academy of Management Journal*. 36 (3), 590–602.

West, M.A. and Hirst, G. (2003) Cooperation and teamwork for Innovation. In: Michael A. West and D. Tjosvold (eds.). *International Handbook of Organizational Teamwork and Cooperative Working*. Chichester, Whiley Publishing Inc. pp. 297–319.

William, K.Y. and O'Reilly, C.A. (1998) Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behaviour*. 2077–140.

Wu, C. (2016) Personality change via work : A job demand – control model of Big-five personality changes. *Journal of Vocational Behavior*. 92157–166.

Yeatts, D.E., Hipskind, M. and Barnes, D. (1994) Lessons learned from Self-managed work teams. *Business Horizons*. July-Aug (1), 11–18.

Yeatts, D.E. and Hyten, C. (1998) *High-performing Self-managed work teams : a comparison of theory to practice*. London, UK, SAGE Publications, Inc.

Zanutto, E.L., Bezrukova, K. and Jehn, K.A. (2010) Revisiting faultline conceptualization: measuring faultlin strenght and distance. *Online, Melbourne Business School, Springer Science*.

Zarraga, C. and Bonache, J. (2005) The Impact of Team Atmosphere on Knowledge Outcomes in Self-managed Teams. *Organization Studies*. 26 (5), 661–681.

Zborowska, E. and Kolding, M. (2018) *Outsourcing : Da gibt es nichts zu überlegen*

Zhang, Z., Waldmann, D.A., Wang, Z., Waldman, D.A., et al. (2012) A multilevel invesitgation of leader-member exchange, informal leader emergence, and individual and team performance. *Personnel Psychology*. 65 (1), 49–78.

Zyphur, M.J. (2009) When mindsets collide: Switching analytical mindsets to advance organization science. *Academy of Management Review*. 34 (4), 677–688.

15 ABBREVIATIONS

aka	Also known as
AOC	Airline Operations Center
CO	Colorado
DBA	Doctorate in Business Administration
DPO	Data Privacy Officer
DU	Durham University
e.g.	„exempli gratia“ for example
EAR	European Arms Regulation
EEA	European Economic Area
EQ	Emotional quotient
ESE	Employee Survey Evaluation Tool
Etc.	Etcetera
EU	European Union
GDPR	General Data Protection Regulation
GMA	General Mental Ability
GROW	Goals-Reality-Options-Way forward
HR	Human Resources
IMOI	Input-Mediator-Output-Input
IPO	Input-Process-Output
IQ	Intelligence quotient
ITAR	International Trade and Arms Regulation
JEPP	Jeppesen Sanderson Inc.
KASO	Knowledge, Ability, Skills, Composition
PII	Personal Identifiable Information
PMO	Program Management Office
PO	Project Owner
R&D	Research & Development
SD	Standard Deviation
SMWT	Self-managed work team

TBC	The Boeing Company
TCI	Team Climate Index
TMT	Top Management Teams
UML	Unified Markup Language
USA	United States of America
WOC	Workers Council

APPENDIX A DATA COLLECTION AND FEEDBACK MATERIAL

This appendix lists forms and questionnaires related to data collection and feedback to team members.

A.1 PERFORMANCE FEEDBACK FORM

EY Review 20xx (Values based on 1. Jan 20xx- 30. Dec 20xx; All projects consolidated)

Fred Normal

Overall Result (Project Results + Peer Reviews)

	Result (in %)	Overall Average
Total	80.5	79.7

Project Results

	Result (in %)	Overall Average
PO Expectations	86.0	80.2
Team Goals	83.1	82.9

Peer Review Results

Questions	Result (1-5)	Overall Average
1. Did this team member actively participate in team discussions and activities? (1-5)	3.6	3.7
2. Did this team member contribute ideas that moved the project forward? (1-5)	3.7	3.8
3. Did this team member carry out his/her specified tasks for the group? (1-5)	3.2	3.5
4. Did this team member work in collaboration (seek input, get feedback) with the team on his/her assigned tasks? (1-5)	3.9	3.7
5. Did this team member use a communication style that had a positive influence on the team? (1-5)	2.3	3.6
6. Did this person use/model a communication style that moved the project/effort forward? (1-5)	4.7	3.6
7. The proportion of project work assigned to this team member was fair relative to the rest of the team members? (0 (No) or 1 (Yes))	1.0	1.0
8. How would you rate this team member's overall contribution to the team? (1-5)	4.2	3.6

Comments

Fred is an excellent leader!
Fred has a very conciliative attitude in controversary discussions.
Sometimes, it was not clear, what was going on in a certain task/deliverable.
During this project Fred sometimes seemed to be distracted. I know his work from previous projects and therefore this time he was not always reliable in delivering his work. Apart from that he actively contributed to the project and also raised concerns when for example the team moved towards being too location centric. I always appreciate his feedback, which is always constructive.

Figure 13 End-year evaluation hand out

A.2 PEER REVIEW QUESTIONNAIRES

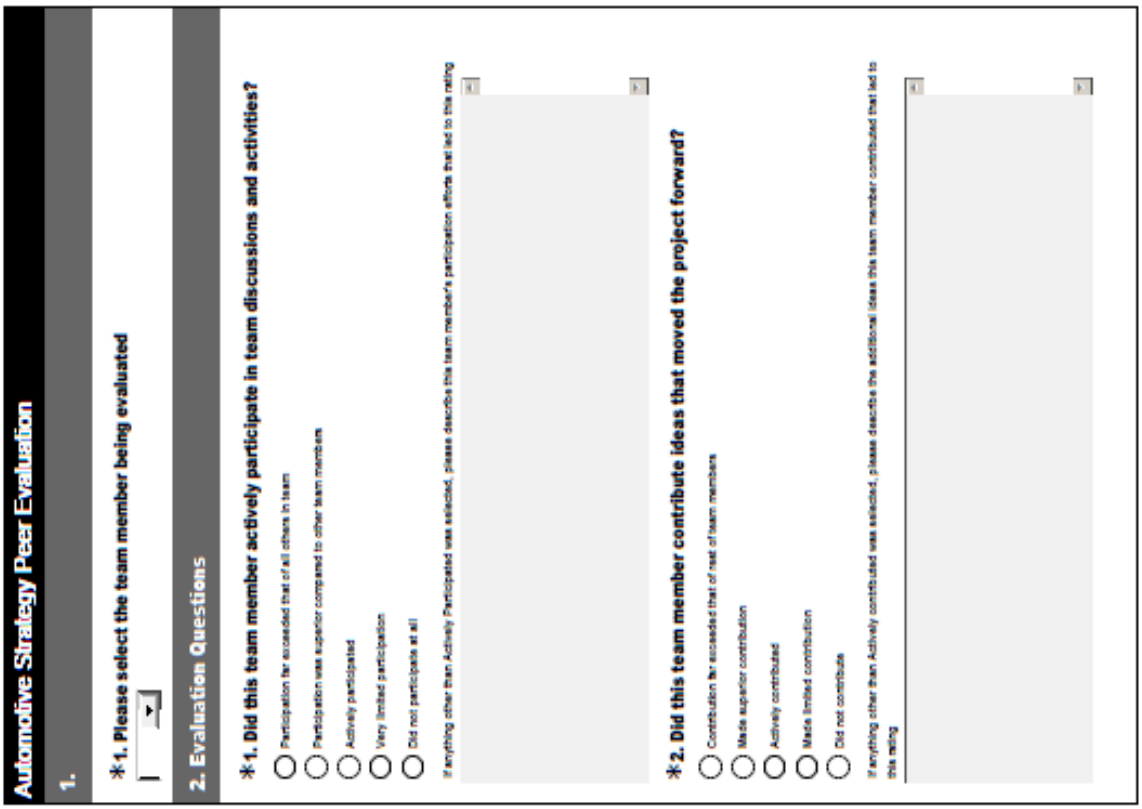


Figure 14 Peer evaluation survey questions. Page 1

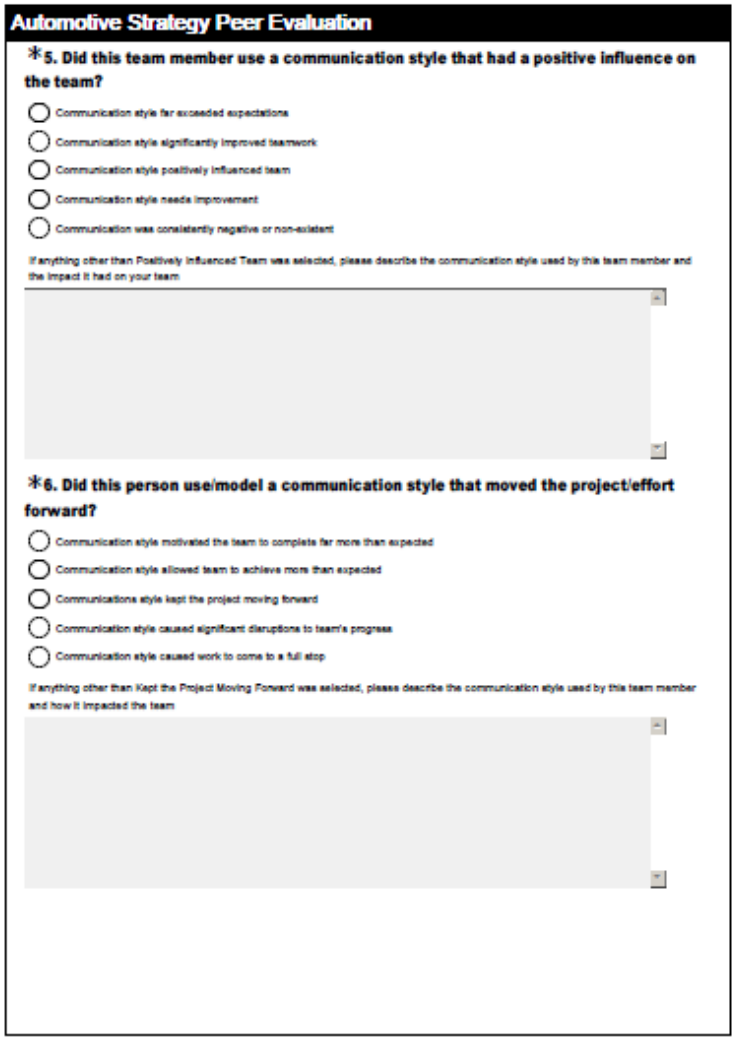
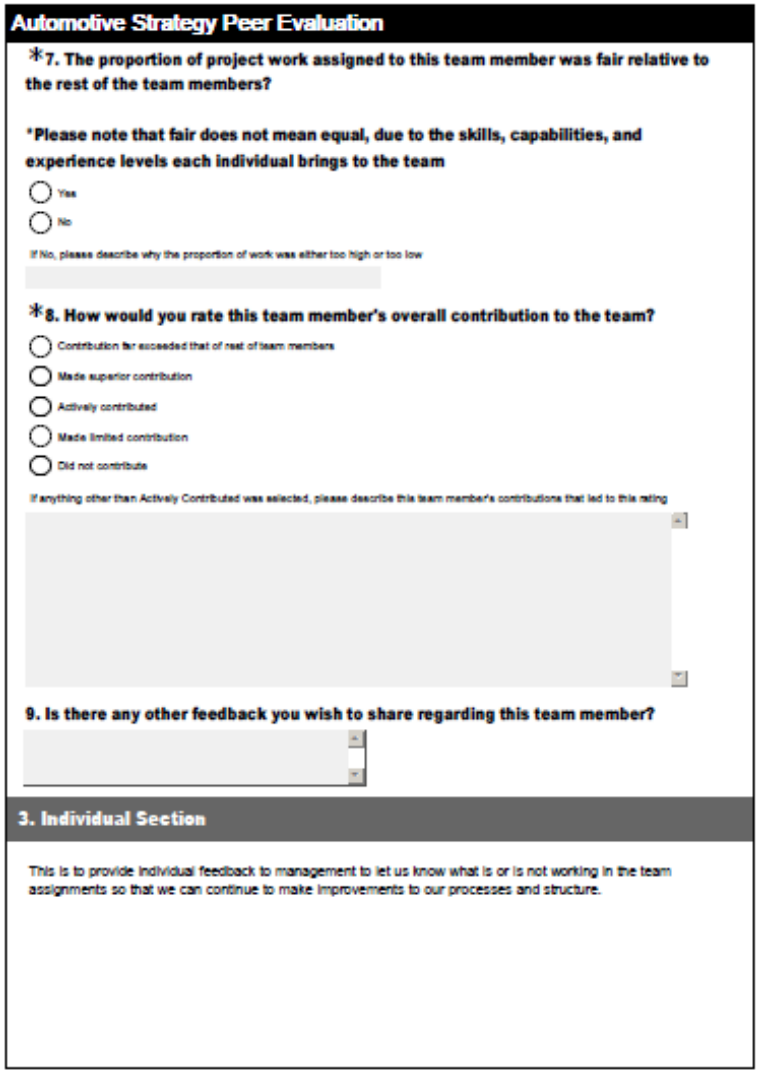


Figure 15 Peer evaluation survey questions. Page 2

Figure 16 Peer evaluation survey questions. Page 3



Automotive Strategy Peer Evaluation

*7. The proportion of project work assigned to this team member was fair relative to the rest of the team members?

*Please note that fair does not mean equal, due to the skills, capabilities, and experience levels each individual brings to the team

- ◯ Yes
- ◯ No

If No, please describe why the proportion of work was either too high or too low

*8. How would you rate this team member's overall contribution to the team?

- ◯ Contribution far exceeded that of rest of team members
- ◯ Made superior contribution
- ◯ Actively contributed
- ◯ Made limited contribution
- ◯ Did not contribute

If anything other than Actively Contributed was selected, please describe this team member's contributions that led to this rating

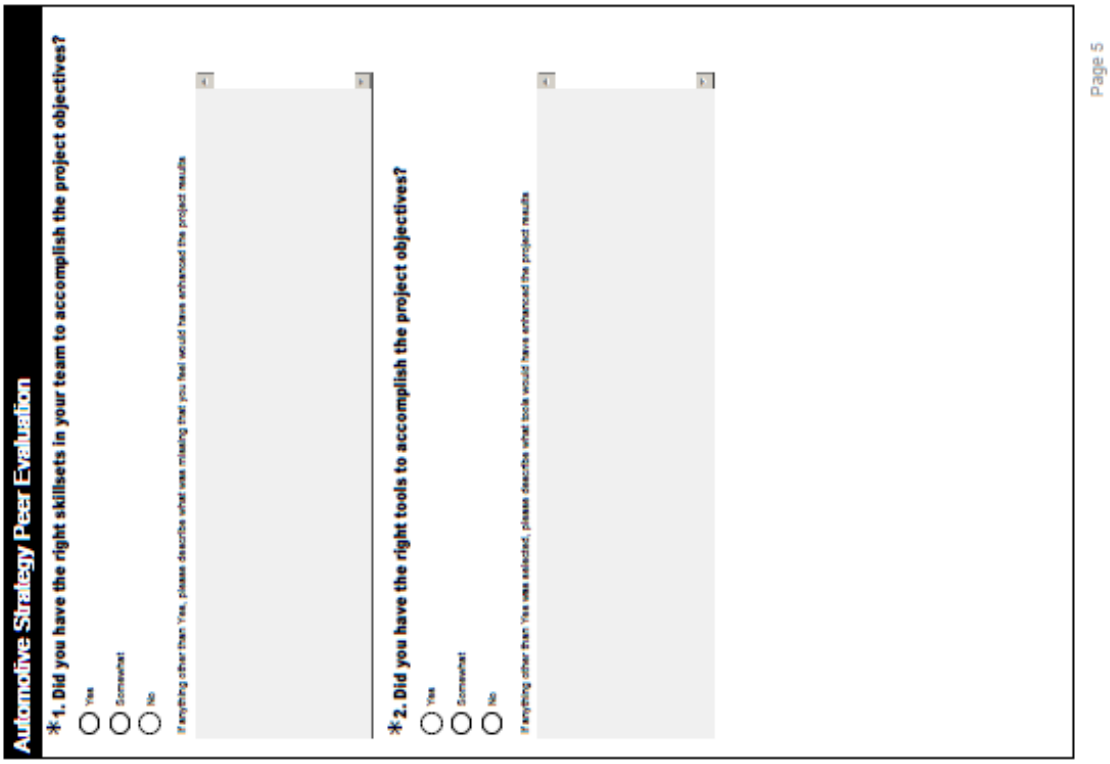
9. Is there any other feedback you wish to share regarding this team member?

3. Individual Section

This is to provide individual feedback to management to let us know what is or is not working in the team assignments so that we can continue to make improvements to our processes and structure.

Page 4

Figure 17 Peer evaluation survey questions. Page 4



Automotive Strategy Peer Evaluation

*1. Did you have the right skillsets in your team to accomplish the project objectives?

- Yes
- Somewhat
- No

If anything other than Yes, please describe what was missing that you feel would have enhanced the project results

*2. Did you have the right tools to accomplish the project objectives?

- Yes
- Somewhat
- No

If anything other than Yes was selected, please describe what tools would have enhanced the project results

Page 5

Figure 18 Peer evaluation survey questions. Page 5

A.3 EXPECTATION MATRIX

Future Interaction -- Expectations Matrix: 25% of team evaluation

	Features:	Weight	% Complete	
1	**Analyse plans**			
1.1.	Analyze potential Interaction Models for Aviation (Cockpit/AOC) in 5 - 10 years	15%	90%	13.5%
1.2.	Analyze what is visible as Interaction Technologies/Concepts in the mass market/gaming industry (xBox, Wii, Military, …)	10%	90%	9.0%
1.3.	Look at 3D, voice, gestures, tracking, technology, training, personal interaction, brain interfaces	5%	90%	4.5%
1.4.	Buy selected technology and try	5%	100%	5.0%
2	**Aviation**			
2.1.	How are findings "viable" to the Aviation domain and Jeppesen	10%	90%	9.0%
3	**Results**			
3.1.	Build a presentation (other than ppt ?) to show results	5%	100%	5.0%
3.2.	Build Technology and Industry roadmap on Future Interaction topic	5%	80%	4.0%
3.3.	Build video on results showing the Aviation impact	5%	90%	4.5%
3.3.	Knowledge Wiki about Future interaction technology and the trends	5%	90%	4.5%
	Subtotal	**65%**		**59.0%**
	Functionality:	**Weight**	**% Complete**	
1	Build out one demonstrator showing one future Interaction technology (In Aviation context, Cockpit, AOC)	14%	100%	14%
	Subtotal	**14%**		**14%**
	Innovation:	**Weight**	**% Complete**	
1	Broadness of proposal	5%	80%	4%
2	Innovation potential of proposals	5%	90%	5%
3	Industry know-how demonstrated	5%	80%	4%
5	Degree of innovation (No. of IP submissions (4) , papers (1) , stakeholder presentations (6))	6%	70%	4%
	Subtotal	**21%**		**17%**
		100%		**90%**
	Team Goals Matrix: 25% of team evaluation	**Weight**	**% Complete**	
1	This project will generate 2 IP's	30%	100%	30%
2	The features above will be completed in 2 iterations	5%	100%	5%
3	The expenses of this project will not exceed $4900	5%	100%	5%
4	The internal man hours estimated to complete the above features is 482	5%	100%	5%
5	The number of stakeholder meetings planned is 2	10%	100%	10%
6	conduct 2 external interviews	15%	100%	15%
7	build 2 demonstrators	30%	100%	30%
		100%		**100%**

Figure 19 Example for an Expectation Matrix. Compare Launer (2011)

APPENDIX B DATA AND DATA CONVERSION DETAILS

This section contains data frames of data used in this study

B.1 DATA CONSOLIDTED ON PROJECT LEVEL

Project	CUL_Interdependent	CUL_egalitarism	CUL_risk	CUL_direct	CUL_task	Peer Review	PRO_Exp	PRO_Team	Neuroticism	Extroverision	Openness	Agreeablness	Conscientiousness	Language	Origin	Residence	Age	Gender	Education
1	2,00	3,50	3,20	3,20	2,20	2,97	80,33	72,67	17,00	29,00	28,50	28,50	32,50	0,00	0,00	0,00	1,00	0,00	1,00
2	3,20	2,35	2,15	2,95	2,70	2,83	93,00	98,00	16,67	32,67	29,33	30,67	37,00	0,81	0,81	0,81	1,00	0,81	0,95
3	3,80	2,95	2,50	2,65	3,45	2,93	85,00	92,00	20,25	25,75	27,50	32,75	29,50	0,92	0,92	0,92	0,92	0,00	0,96
4	4,00	1,20	1,10	2,70	3,20	2,85	79,00	79,00	18,13	32,88	29,25	33,88	37,75	1,00	1,00	1,00	1,00	1,00	0,00
5	3,15	2,80	3,35	3,15	1,95	2,83	91,00	85,00	21,00	31,50	37,67	31,00	34,00	0,97	0,97	0,97	0,87	0,97	0,95
6	3,15	2,80	3,35	3,15	1,95	2,80	91,00	84,00	21,00	31,50	37,67	31,00	34,00	0,97	0,97	0,97	0,87	0,97	0,96
8	2,65	2,65	2,95	2,35	1,30	3,06	78,00	78,00	19,50	21,33	24,00	23,33	25,67	0,81	0,81	0,00	0,95	0,00	1,00
11	3,00	2,60	2,50	2,80	3,10	3,03	75,00	99,00	23,50	27,33	22,00	29,67	27,33	0,00	0,00	0,00	0,92	0,00	0,92
12	2,60	2,85	3,10	2,90	2,45	2,84	75,00	99,00	18,50	27,00	25,75	27,75	28,50	0,00	0,00	0,00	0,95	0,00	0,95
14	3,60	2,50	3,00	3,40	2,70	3,25	79,00	96,00	19,00	31,00	37,00	29,50	34,00	0,00	0,00	0,00	0,00	1,00	1,00
16	4,00	1,20	1,10	2,70	3,20	2,50	79,00	86,00	16,00	40,00	31,00	35,00	46,00	1,00	1,00	1,00	1,00	1,00	1,00
18	2,90	3,13	2,77	3,47	1,80	2,89	90,00	85,00	21,25	27,67	33,25	29,75	31,50	0,97	0,97	0,97	0,87	0,72	0,96
19	2,40	3,50	3,20	3,20	2,20	3,06	93,00	95,00	19,00	23,50	21,00	22,00	27,00	0,00	0,00	0,00	1,00	0,00	1,00
21	3,13	1,87	2,10	2,37	2,43	2,88	80,00	90,00	17,25	32,00	28,75	30,00	34,00	0,00	0,00	0,00	0,98	0,59	0,73
22	2,40	3,50	3,20	3,20	2,20	3,50	94,00	96,00	19,00	23,50	21,00	22,00	27,00	0,00	0,00	0,00	1,00	0,00	1,00
23	3,00	4,00	3,20	4,10	3,60	2,80	91,00	98,00	21,00	23,00	23,00	27,00	30,00	0,00	0,00	0,00	0,00	0,00	0,00
24	2,45	3,00	2,90	2,30	1,90	2,84	87,00	85,00	19,00	23,33	24,67	31,00	28,67	0,81	0,81	1,00	0,95	0,00	0,81
29	3,00	3,15	2,30	3,75	2,10	3,07	84,00	68,00	20,00	25,50	31,33	28,33	30,67	0,00	0,00	0,00	0,81	0,81	0,81
30	3,00	4,00	3,20	4,10	3,60	2,92	94,00	85,00	22,67	23,00	22,00	30,25	26,00	0,00	0,00	0,00	0,89	0,00	0,97
31	3,13	1,87	2,10	2,37	2,43	2,82	87,00	73,00	17,25	32,00	28,75	30,00	34,00	0,00	0,00	0,00	0,98	0,59	0,73
32	2,00	3,50	3,20	3,20	2,20	3,54	82,00	81,00	21,67	25,67	27,67	29,00	28,67	0,00	0,00	0,00	1,00	0,00	1,00
33	3,20	2,60	2,50	2,80	3,10	2,50	84,00	75,00	20,00	27,00	28,50	23,50	25,50	0,00	0,00	0,00	1,00	0,00	0,00
35	2,00	3,50	3,20	3,20	2,20	3,03	89,00	84,00	21,67	24,75	27,67	26,50	28,25	0,72	0,72	0,72	0,87	0,00	0,96
36	2,90	3,13	2,77	3,47	1,80	3,12	78,00	92,00	20,00	25,50	31,33	28,33	30,67	0,81	0,81	0,81	0,95	0,81	0,95
38	3,00	4,00	3,20	4,10	3,60	3,15	85,00	75,00	22,67	23,00	22,00	30,25	26,00	0,00	0,00	0,00	0,87	0,00	0,86
39	3,13	1,87	2,10	2,37	2,43	3,30	87,00	95,00	17,25	32,00	28,75	30,00	34,00	0,00	0,00	0,00	0,96	0,72	0,72
43	2,40	3,50	3,20	3,20	2,20	3,27	90,00	75,00	19,50	21,33	22,00	22,67	25,33	0,95	0,95	0,81	1,00	0,00	0,95
44	2,00	3,50	3,20	3,20	2,20	2,90	95,00	92,00	19,50	25,00	30,00	31,00	31,50	1,00	1,00	1,00	0,92	0,92	1,00
45	2,70	2,20	2,60	2,20	2,05	2,67	86,00	99,00	19,00	27,50	24,00	26,50	26,00	0,00	0,00	0,00	0,92	0,92	0,92
47	4,00	1,20	1,10	2,70	3,20	3,81	87,00	73,00	15,50	36,50	33,50	33,50	42,00	0,00	0,00	0,00	1,00	0,00	0,00
48	2,00	3,00	3,70	3,40	2,70	2,96	74,00	72,00	21,00	23,50	24,00	26,50	26,00	0,92	0,92	0,92	1,00	0,00	0,92
50	2,60	2,63	2,80	2,53	2,10	3,65	92,00	97,00	19,25	26,67	27,00	28,75	28,75	0,81	0,97	0,81	0,95	0,81	1,00
54	3,00	4,00	3,20	4,10	3,60	3,09	89,00	74,00	19,50	23,33	25,00	27,75	24,25	0,00	0,00	0,00	0,95	0,00	0,00
55	3,15	2,80	3,35	3,15	1,95	2,75	92,33	81,33	18,00	26,50	35,00	20,50	29,50	0,92	0,92	0,92	0,92	0,00	0,92

57	2,90	3,13	2,77	3,47	1,80	3,44	86,00	84,00	18,67	28,00	27,67	29,67	31,67	0,81	0,81	0,81	1,00	0,00	0,95
58	2,00	3,00	3,70	3,40	2,70	3,03	92,00	81,00	21,00	23,50	24,00	26,50	26,00	0,92	0,92	0,92	1,00	0,00	0,92
59	2,40	3,80	1,60	4,10	1,50	2,88	86,00	91,00	22,00	20,00	20,00	26,00	24,00	1,00	1,00	1,00	1,00	0,00	0,92
60	3,10	4,00	3,20	4,10	3,60	3,38	91,00	83,00	21,00	21,50	25,50	25,50	26,00	0,00	0,00	0,00	1,00	0,00	0,00
61	2,00	3,80	1,60	4,10	1,50	2,77	84,00	88,00	18,50	26,50	28,00	29,00	31,00	0,95	0,95	0,95	1,00	0,00	0,81
62	3,80	2,10	3,20	2,10	2,70	2,89	90,00	76,00	18,50	21,67	24,25	29,75	21,75	0,00	0,00	0,00	0,96	0,72	0,00
63	3,03	2,60	3,23	2,97	2,20	2,65	91,00	87,00	18,50	30,50	29,00	24,00	28,00	1,00	1,00	1,00	0,81	0,00	1,00
64	4,00	1,20	1,10	2,70	3,20	3,13	90,00	68,00	18,00	40,00	31,00	32,50	37,00	1,00	1,00	1,00	0,92	0,00	1,00
66	2,60	2,20	2,20	1,80	1,40	3,81	85,00	87,00	19,50	25,00	32,00	32,00	32,00	1,00	1,00	1,00	1,00	0,00	1,00
67	2,40	3,80	1,60	4,10	1,50	2,30	85,00	73,00	19,33	23,33	23,67	30,33	28,33	0,00	0,00	0,92	1,00	0,00	0,92
68	3,00	3,30	3,70	3,50	1,50	3,36	93,00	89,00	21,67	23,67	17,50	31,67	25,00	0,00	0,00	0,00	0,95	0,00	0,81
69	2,60	2,20	2,20	1,80	1,40	3,94	80,00	97,00	19,50	25,00	32,00	32,00	32,00	0,00	1,00	1,00	1,00	1,00	1,00
70	2,80	2,20	3,00	2,60	2,70	3,38	86,00	83,00	19,50	30,00	31,00	31,50	30,00	1,00	1,00	1,00	0,00	1,00	1,00
71	4,00	2,10	3,20	2,10	2,70	3,63	91,00	91,00	20,00	21,50	20,33	28,67	22,00	0,00	0,00	0,00	1,00	1,00	0,81
72	2,80	2,20	3,00	2,60	2,70	3,38	89,00	85,00	19,50	30,00	31,00	31,50	30,00	1,00	1,00	1,00	0,00	1,00	1,00
73	2,93	3,47	3,37	3,40	2,73	3,73	90,00	100,00	14,00	35,00	28,50	20,00	37,50	0,00	0,00	0,00	0,92	0,00	0,00
74	2,93	3,47	3,37	3,40	2,73	3,60	91,00	85,00	18,33	29,50	23,50	26,50	32,00	0,00	0,00	0,00	0,96	0,00	0,97
75	2,80	3,20	3,55	3,30	1,65	3,53	78,00	82,00	19,33	23,00	20,60	28,80	24,50	0,00	0,00	0,00	0,98	0,00	0,59
76	2,50	2,85	2,70	2,50	1,80	2,80	81,00	76,00	18,50	24,40	24,25	27,20	28,60	0,97	0,97	0,72	0,96	0,00	0,96
77	4,00	2,10	3,20	2,10	2,70	3,58	75,00	68,00	20,00	21,50	18,50	29,00	22,00	0,00	0,00	0,00	1,00	0,92	0,92
78	4,00	2,10	3,20	2,10	2,70	3,56	90,00	84,00	21,33	22,00	23,00	31,00	23,00	0,00	0,00	0,00	0,90	0,65	0,65
80	2,20	2,10	3,20	2,10	2,70	2,75	91,00	87,00	21,00	23,50	24,00	26,50	26,00	0,92	0,92	0,92	1,00	0,00	0,92
81	2,70	2,20	2,60	2,20	2,05	2,89	88,00	90,00	19,00	27,50	24,00	26,50	26,00	0,92	0,92	0,92	1,00	0,00	0,92
82	2,95	3,55	3,20	3,35	3,35	3,40	94,00	85,00	14,00	29,25	23,25	24,75	31,00	0,00	0,00	0,00	0,95	0,00	0,81
83	2,40	3,50	3,20	3,20	2,20	3,44	90,00	70,00	19,50	21,00	22,00	24,50	24,50	1,00	1,00	0,00	1,00	0,00	1,00
85	2,55	3,45	2,65	3,50	1,35	2,88	83,33	90,83	19,33	25,00	29,00	30,75	31,50	0,97	0,97	0,97	0,96	0,72	0,96
86	3,00	3,30	3,70	3,50	1,50	3,54	87,00	98,00	16,50	22,00	20,00	30,00	21,50	0,00	0,00	0,00	0,92	0,00	0,92
88	2,40	3,30	3,70	3,50	1,50	3,31	82,00	87,00	19,50	21,00	22,00	24,50	24,50	1,00	1,00	0,00	1,00	0,00	1,00
90	2,60	2,85	3,10	2,90	2,45	3,13	82,00	80,00	19,33	25,33	23,33	25,50	25,33	0,00	0,92	0,92	1,00	0,00	1,00
91	2,00	3,80	1,60	4,10	1,50	3,30	93,00	100,00	19,75	20,33	26,75	31,25	26,25	0,92	0,81	0,95	0,81	0,81	0,81
92	2,20	3,00	3,70	3,40	2,70	2,75	87,67	78,00	21,00	23,50	24,00	26,50	26,00	1,00	1,00	1,00	1,00	0,00	1,00
94	2,60	2,20	2,20	1,80	1,40	2,60	87,74	82,22	17,50	29,00	26,50	34,00	33,50	1,00	1,00	1,00	1,00	0,00	1,00
95	2,70	3,03	2,77	2,80	2,97	3,05	96,00	88,00	20,00	25,25	21,75	27,25	25,00	1,00	1,00	1,00	1,00	0,00	0,95
96	2,00	3,50	3,20	3,20	2,20	3,65	84,00	74,00	18,33	24,25	22,25	29,75	28,00	0,95	0,95	0,95	1,00	0,00	0,95
97	2,90	2,30	2,55	1,95	1,85	3,19	89,33	91,33	19,00	24,00	25,50	27,00	27,00	0,00	0,00	0,00	0,00	1,00	1,00
98	2,73	2,93	2,73	3,27	3,00	3,40	91,33	92,58	15,50	26,33	25,33	29,33	30,67	0,92	0,92	0,92	1,00	0,92	1,00
99	2,80	2,20	3,00	2,60	2,70	3,19	93,00	94,00	18,33	31,50	30,00	34,67	33,00	0,92	0,92	0,92	0,92	0,92	0,92
100	2,87	3,37	2,67	3,37	1,97	3,09	87,00	82,00	21,00	22,00	20,00	28,50	25,50	0,92	0,92	0,92	0,92	0,92	0,92
101	2,37	3,10	2,67	3,50	1,60	3,46	90,00	67,00	20,00	23,33	24,33	30,67	31,00	1,00	1,00	0,95	1,00	0,00	1,00
102	2,90	2,65	2,98	3,20	2,90	3,23	88,00	90,00	18,60	23,33	26,40	28,00	29,33	0,92	0,92	1,00	0,90	1,00	0,92
103	2,90	2,68	2,74	2,46	2,60	3,16	91,00	96,00	19,17	25,14	23,86	29,33	26,71	0,91	0,91	0,91	0,92	0,59	0,87
104	3,50	3,20	2,70	3,10	3,20	3,56	90,33	90,67	20,00	24,00	39,00	34,00	31,50	0,00	0,00	0,00	1,00	0,00	1,00
105	3,15	2,85	3,45	3,45	2,65	3,44	84,11	82,22	21,00	23,50	24,00	27,00	26,00	1,00	1,00	1,00	1,00	0,00	1,00

106	3,10	2,78	2,95	2,95	2,48	3,42	93,00	100,00	20,00	24,00	24,33	28,00	26,75	0,95	0,95	1,00	0,81	0,81	0,81
107	2,58	2,86	2,74	3,72	2,60	3,01	88,83	100,00	19,00	20,75	25,20	31,80	27,80	0,87	0,87	0,87	0,96	0,00	0,87
108	3,10	2,30	2,55	3,35	2,80	3,02	91,47	100,00	17,75	25,67	30,75	33,67	33,25	0,96	0,96	0,96	0,96	0,97	0,87
109	2,30	3,20	2,65	2,45	2,40	3,27	92,75	96,67	20,00	25,00	24,00	30,00	27,40	0,96	0,96	0,96	0,96	0,00	0,96
110	3,00	2,35	2,90	3,75	3,55	3,24	94,25	92,22	17,00	25,50	33,00	32,00	35,50	1,00	1,00	1,00	0,00	0,00	0,00
111	2,95	3,50	2,15	3,60	2,35	3,06	88,08	80,74	21,00	22,00	20,00	28,50	25,50	0,00	0,00	0,00	0,00	1,00	0,00
112	2,97	3,07	2,80	3,30	3,03	3,15	91,00	90,00	20,00	20,67	25,00	28,00	25,33	1,00	0,00	1,00	0,92	0,92	0,92
113	3,80	2,10	3,20	2,10	2,70	3,44	88,33	80,67	20,00	21,00	20,00	29,00	20,00	0,00	0,00	0,00	1,00	0,00	1,00
114	2,72	2,44	3,08	2,74	2,54	3,26	92,00	87,00	19,00	24,83	26,20	31,50	30,00	0,96	0,96	0,92	0,90	1,00	0,96
115	2,93	3,37	2,67	2,97	3,13	3,55	89,00	78,00	20,33	23,33	28,00	30,25	28,50	0,95	0,95	0,95	1,00	1,00	0,95

Table 16 Consolidated/Aggregated data

B.2 DATA STRUCTURE

Project	Member	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	PQ1	PQ2	PQ3	...
N	NS	E	ES	O	OS	A	AS	C	CS	Birthdate	Age	Startdate	...
Gender	Education	Origin	Residence	Language	Job Grade	I	IS	EG	EGS	R	RS	D	...
DS	T	TS	PO_Exp	PO_Team	Peer								

Figure 20 Data Structure

B.3 RAW DATA EXAMPLE

	ear	Trac k	Proje ctinTr ack	TMe mber	Q1	Q2	Q3	Q4	Q5	Q6	Q7	...	P...	PQ2	P...	N	E	O	A	C	Birthd ate	age	Startc ate	Ge nd er	Ed uc ...	Count ryofor gin	Count ryofre si...	native langu age	Gra...	Indepe ndent	egalitar ism	risk	direct	task	Exp	Team Goals	PO
682	015	2	2	39	5	5	5	5	4	4	2	4	1	1	1	-99	22	17	29	24	1975	41	2010	2	1	3	3	3	3	-99.0	-99.0	-99.0	-99.0	-99.0	84	74	2
683	015	2	2	37	5	1	4	5	4	5	2	5	2	2	2	20	17	23	24	22	1978	38	2007	2	1	2	1	1	4	-99.0	-99.0	-99.0	-99.0	-99.0	84	74	2
684	015	3	1	21	4	3	4	4	3	3	2	4	2	2	2	19	23	26	27	26	1983	33	2012	1	4	1	1	1	3	3.2	2.4	2.9	2.1	2.3	92	100	2
685	015	3	1	48	3	3	3	3	4	3	2	3	2	1	2	19	25	25	27	28	1983	33	2012	2	1	1	1	1	3	2.6	2.2	2.2	1.8	1.4	92	100	2
686	015	3	2	11	5	1	4	5	5	4	2	4	1	2	2	12	31	33	37	44	1985	31	2013	1	1	2	3	1	3	3.6	2.4	2.9	4.9	4.2	92	93	1
687	015	3	2	11	3	1	4	3	3	4	2	3	0	0	2	12	31	33	37	44	1985	31	2013	1	1	2	3	1	3	3.6	2.4	2.9	4.9	4.2	92	93	1
688	015	3	2	52	3	3	3	3	3	3	2	3	0	0	2	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	92	93	1
689	015	3	2	52	3	3	3	2	3	3	2	3	0	0	2	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	92	93	1
690	015	3	2	40	3	5	4	4	4	3	5	0	5	0	0	2	-99	23	22	26	21	1978	38	2010	2	2	1	1	1	3	2.2	2.9	2.1	1.7	2.6	92	93	1
691	015	3	2	40	5	5	5	5	4	5	2	5	1	2	2	-99	23	22	26	21	1978	38	2010	2	2	1	1	1	3	2.2	2.9	2.1	1.7	2.6	92	93	1
692	015	3	3	17	3	2	2	4	3	2	2	2	2	2	2	16	33	28	41	39	1966	50	2010	2	3	2	2	1	3	-99.0	-99.0	-99.0	-99.0	-99.0	93	94	1
693	015	3	3	17	3	3	3	3	3	3	2	3	2	2	2	16	33	28	41	39	1966	50	2010	2	3	2	2	1	3	-99.0	-99.0	-99.0	-99.0	-99.0	93	94	1
694	015	3	3	24	4	5	4	4	4	4	2	4	2	2	2	19	30	23	26	24	1975	41	2007	2	2	1	1	1	5	2.8	2.2	3.0	2.6	2.7	93	94	1
695	015	3	3	15	3	3	3	3	3	3	2	4	2	2	2	20	-99	39	37	36	1976	40	2010	1	3	2	2	1	4	-99.0	-99.0	-99.0	-99.0	-99.0	93	94	1
696	015	3	4	35	4	3	3	3	3	3	2	3	2	2	2	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	87	82	2
697	015	3	4	35	3	3	3	3	4	3	2	4	1	2	2	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	87	82	2
698	015	3	4	26	3	3	3	2	3	3	2	3	-99	-99	-99	20	24	-99	31	27	1984	32	2014	1	1	2	2	1	3	3.5	3.2	2.7	3.1	3.2	87	82	2
699	015	3	4	57	4	1	4	3	4	3	2	4	1	2	2	-99	-99	-99	-99	-99	1974	42	2008	2	2	1	1	1	5	2.7	3.1	3.7	2.9	1.2	87	82	2
700	016	1	1	35	3	3	3	4	4	4	4	4	1	1	2	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	90	67	2
701	016	1	1	35	3	3	-99	-99	-99	-99	-99	...	-99	-99	-99	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	90	67	2
702	016	1	1	35	3	2	-99	-99	-99	-99	-99	...	-99	-99	-99	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	90	67	2
703	016	1	1	35	3	1	2	2	2	2	5	2	2	2	2	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	90	67	2
704	016	1	1	35	3	3	4	4	4	4	1	4	-99	-99	-99	22	20	20	26	24	1981	35	2012	2	1	2	2	1	3	2.4	3.8	1.6	4.1	1.5	90	67	2
705	016	1	1	1	4	3	4	4	3	3	3	4	-99	-99	-99	18	25	32	40	41	1988	28	2016	2	2	6	1	6	3	2.0	2.4	2.7	3.5	2.1	90	67	2
706	016	1	1	22	4	4	4	4	4	4	1	4	1	2	1	-99	25	21	26	28	1979	37	2010	2	1	3	3	3	1	-99.0	-99.0	-99.0	-99.0	-99.0	90	67	2
707	016	1	1	22	4	2	3	3	3	3	4	3	-99	-99	-99	-99	25	21	26	28	1979	37	2010	2	1	3	3	3	1	-99.0	-99.0	-99.0	-99.0	-99.0	90	67	2
708	016	1	1	57	4	4	4	4	4	4	2	5	2	1	2	-99	-99	-99	-99	-99	1974	42	2008	2	2	1	1	1	5	2.7	3.1	3.7	2.9	1.2	90	67	2
709	016	1	2	52	2	2	3	3	4	4	4	3	1	2	1	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
710	016	1	2	52	3	2	3	3	3	4	3	3	2	2	2	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
711	016	1	2	52	4	2	3	3	3	3	3	3	2	2	2	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
712	016	1	2	52	2	2	3	3	3	3	4	2	2	0	2	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
713	016	1	2	52	2	3	2	4	2	3	4	2	1	0	0	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
714	016	1	2	52	2	3	-99	-99	-99	-99	-99	...	-99	-99	-99	19	25	21	25	27	1968	48	2012	2	4	1	1	1	3	2.4	3.5	3.2	3.2	2.2	88	90	2
715	016	1	2	11	4	3	4	4	3	4	2	4	2	2	2	12	31	33	37	44	1985	31	2013	1	1	2	3	1	3	3.6	2.4	2.9	4.9	4.2	88	90	2
716	016	1	2	11	4	3	4	4	4	4	1	4	-99	-99	-99	12	31	33	37	44	1985	31	2013	1	1	2	3	1	3	3.6	2.4	2.9	4.9	4.2	88	90	2
717	016	1	2	11	4	4	3	3	4	4	1	5	2	2	2	12	31	33	37	44	1985	31	2013	1	1	2	3	1	3	3.6	2.4	2.9	4.9	4.2	88	90	2

Table 17 Raw data example listing

B.4 DATA CONVERSION

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Project #	Year	Track	Project in Track	Team	icipate	ideas	Did th	Did th	Did th	Did th	The p	How	Did you have th	Did you have thi	Were the hour	N	E	O	A	C	Birthdate	age	Startdate	Gender	Education	Country of origir	Country c	native lang
2	Project	Year	Track	CountIT	TM	PQ1	PQ2	PQ3	PQ4	PQ5	PQ6	PQ7	PQ8	TQ1	TQ2	TQ3	N	E	O	A	C	Birth	Age	Start	Gender	Edu	Corg	Cres	Lang
3																													
4	Project #	year	track #	InTrack #	#	1-5	1-5	1-5	1-5	1-5	1-5	1-5	1-5	0-2	0-2	0-2	0-50	0-50	0-50	0-50	0-50	year	number	year	female=1	d=doctoral=4	1=USA	1=USA	1=USA
5														0 No	0 No	0 No									male=2	m=master=3	2=GER	2=GER	2=GER
6				a=1										1 Somewhat	1 Somewhat	1 Somewhat									unknown=3	b=bachelor=2	3=POL	3=POL	3=POL
7				b=2										2 Yes	2 Yes	2 Yes										a=college=1	4=DEN	4=DEN	4=DEN
8				c=3																							5=COL	5=COL	5=COL
9				d=4																							6=ESP	6=ESP	6=ESP
10				e=5																							7=POR	7=POR	7=POR
11				f=6																									
12				g=7																									
13				h=8																									
14				i=9																									
15				j=10																									
16				k=11																									
17				l=12																									
18				m=13																									
19				n=14																									
20																													

Table 18 Data coding table

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Project #	year	Track	Sub	Team Member	member actively participate in team discussions a	e ideas	Did this	put, ge	en style	Did this p	The prop	How we	Did you	Did you	Were the	N	E
2	3	2012	1	1	6	2,5	2,5	3,0	3,0	2,5	3,0	2,0	2,5	2,0	2,0	2,0	25	3
3	3	2012	1	1	42	3,0	3,0	3,7	3,0	3,3	3,0	2,0	3,3	2,0	2,0	2,0	-99	-9
4	3	2012	1	1	36	3,0	3,0	3,7	3,7	3,7	3,7	2,0	3,3	2,0	2,0	2,0	21	2
5	3	2012	1	1	25	3,0	2,3	3,0	3,3	2,7	3,0	2,0	3,0	2,0	2,0	2,0	-99	-9
6	3	2012	1	1	17	3,7	1,7	3,0	3,7	3,0	3,3	2,0	3,7	1,7	1,7	2,0	16	3
7	3	2012	1	1	63	3,0	1,0	3,5	3,5	3,5	3,0	2,0	3,5	2,0	2,0	2,0	19	1
8	3	2012	1	1	6	3,0	2,3	3,3	3,4	3,1	3,2	2,0	3,2	1,9	1,9	2,0	20,3	25
9	4	2012	1	2	9	2,9	2,6	2,9	2,7	3,0	2,7	3,1	3,3	2,0	1,8	2,0	-99	-9
10	4	2012	1	2	4	3,0	2,4	2,9	2,8	2,9	2,8	2,6	3,3	2,0	2,0	2,0	16	4
11	4	2012	1	2	2	2,9	2,5	2,9	2,7	2,9	2,7	2,9	3,3	2,0	1,9	2,0	16,0	40
12	5	2012	1	3	19	3,0	3,0	3,0	3,0	3,0	3,0	2,0	3,7	1,7	1,7	2,0	18	3

Table 19 Project level aggregation

Project Number	Language	Value
1	1	4
1	2	2
1	3	0
1	4	0
1	5	0
1	6	0
1	7	0
2	1	0
…	…	…

Table 20 Data input coding for Diverse/R

Project	EvennessConscientiousness	EvennessLanguage	EvennessOrigin	EvennessResidence	EvennessAge	Evenness Gender	Evenness Education Level
3	0	0,918	0,918	0,918	0,921	0	0,959
4	0,918	1	1	1	1	1	0
5	1	0,971	0,971	0,971	0,865	0,971	0,946
6	1	0,971	0,971	0,971	0,865	0,971	0,96
8	1	0,811	0,811	0	0,946	0	1
11	0,946	0	0	0	0,918	0	0,918
12	0	0	0	0	0,946	0	0,946
14	0,918	0	0	0	0	1	1
16	0,918	1	1	1	1	1	1
18	0,918	0,971	0,971	0,971	0,865	0,722	0,96
19	1	0	0	0	1	0	1
21	0,921	0	0	0	0,975	0,592	0,725
22	0,914	0	0	0	1	0	1
23	0	0	0	0	0	0	0
24	1	0,811	0,811	1	0,946	0	0,811
29	0,946	0	0	0	0,811	0,811	0,811
30	0,865	0	0	0	0,887	0	0,971
31	0,961	0	0	0	0,975	0,592	0,725
32	0,86	0	0	0	1	0	1

Table 21 Data structure Evenness

B.5 SAMPLE POPULATION STATISTICS AND REFERENCES

	I/I	E/S	R/C	D/I	T/R
	M	M	M	M	M
Germany	2,3	3,1	3,2	1,0	2,9
United States	1,0	1,5	1,2	1,5	1,0
Poland	2,9	2,8	2,2	3,5	3,4
Sample	3,0	2,8	2,8	3,0	2,7

Figure 21 Descriptive Statistics comparison for culture dimensions

APPENDIX C STATISTICS

C.1 DESCRIPTIVE STATISTCS FOR RAW PEER REVIEW DATA. N=901

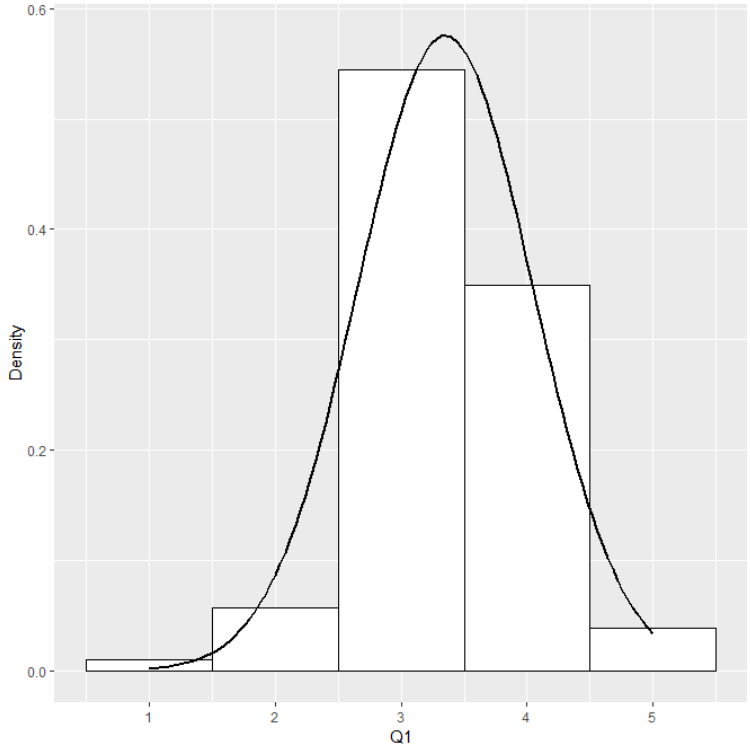


Figure 22 Histogram Q1

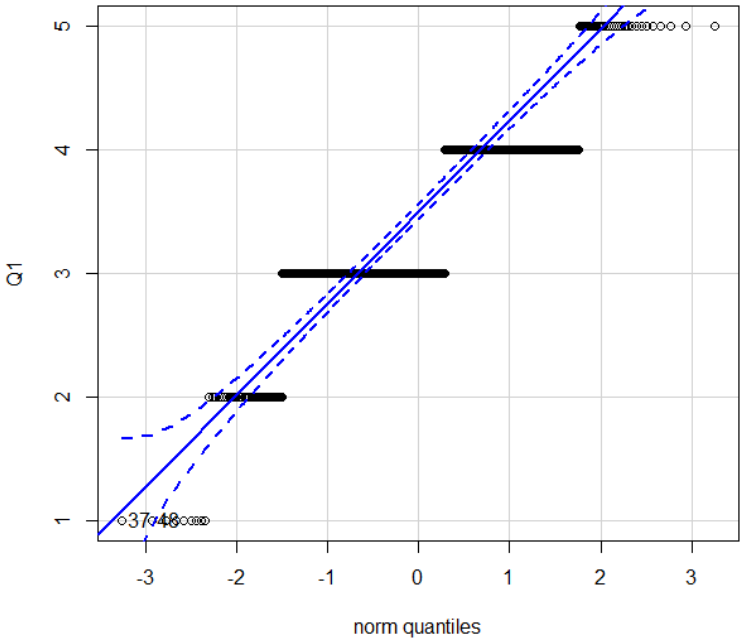


Figure 23 Q-Q Plot Q1

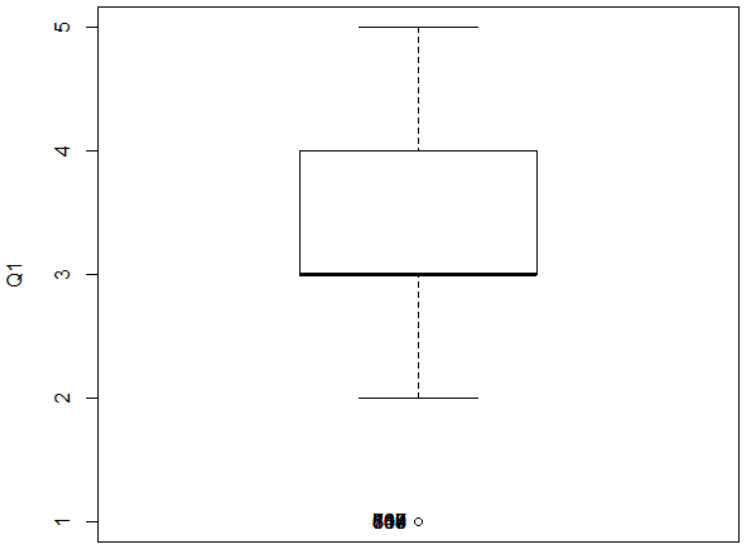


Figure 24 Box plot Q1

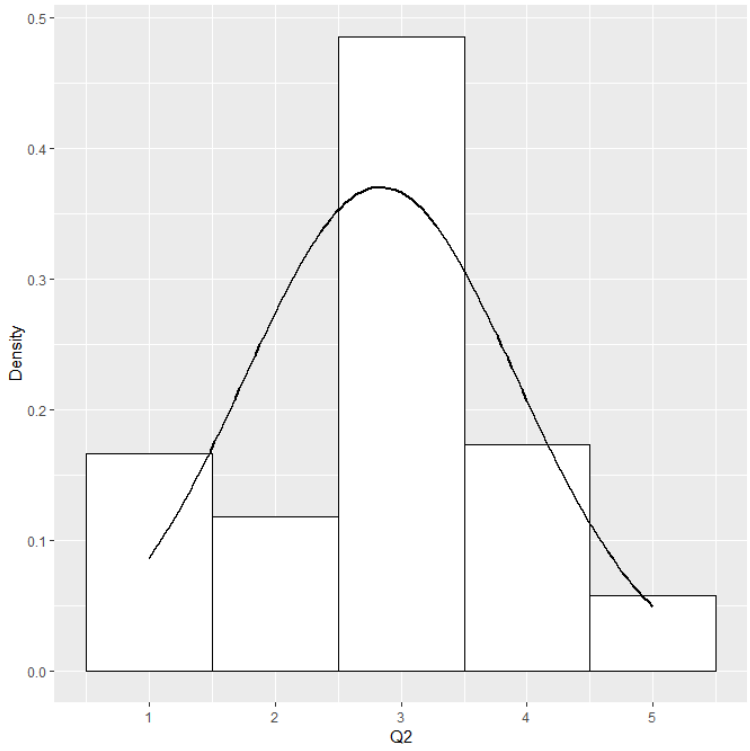


Figure 25 Histogram Q2

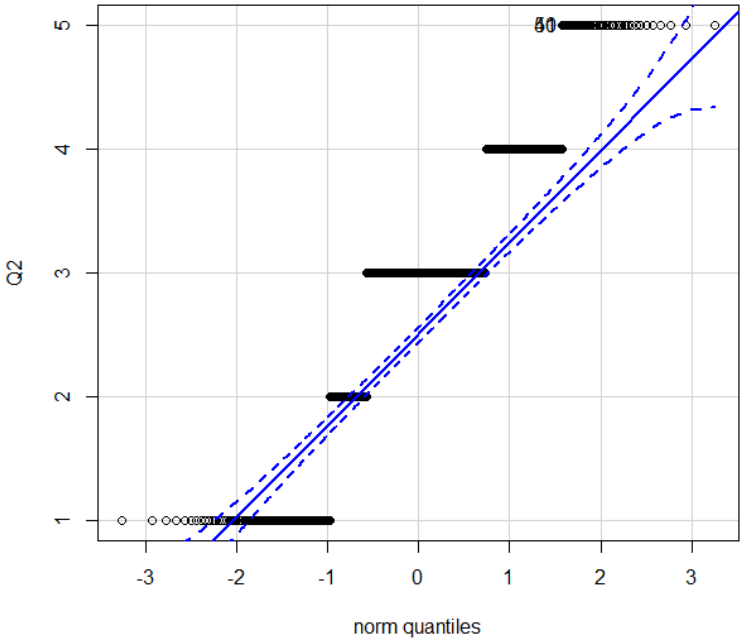


Figure 26 Q-Q Plot Q2

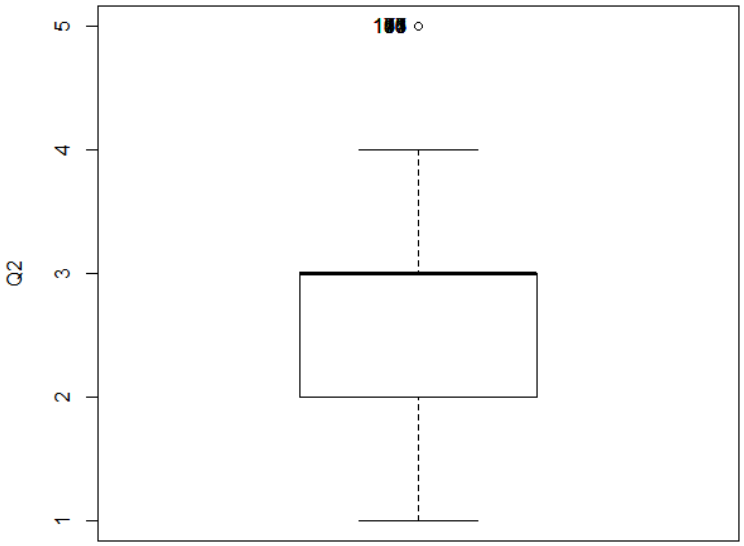


Figure 27 Box plot Q2

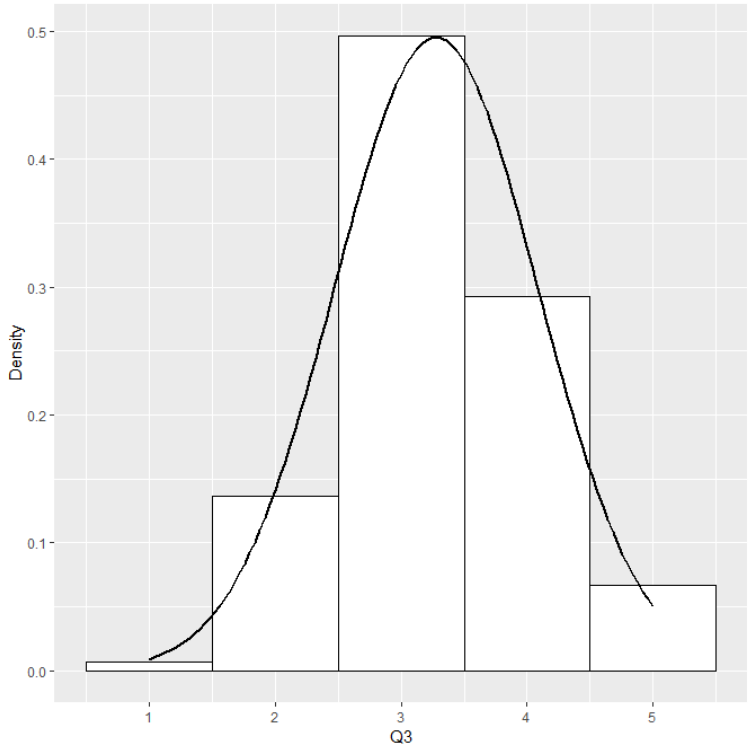


Figure 28 Histogram Q3

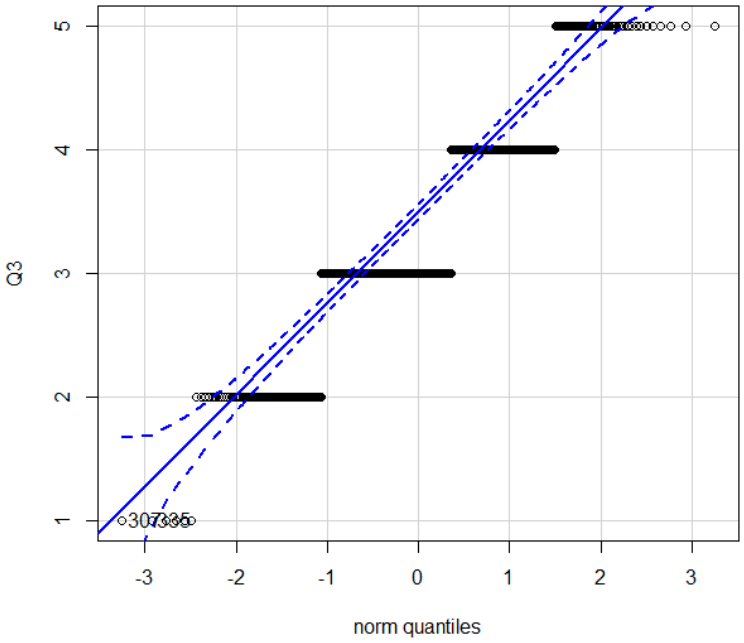


Figure 29 Q-Q Plot Q3

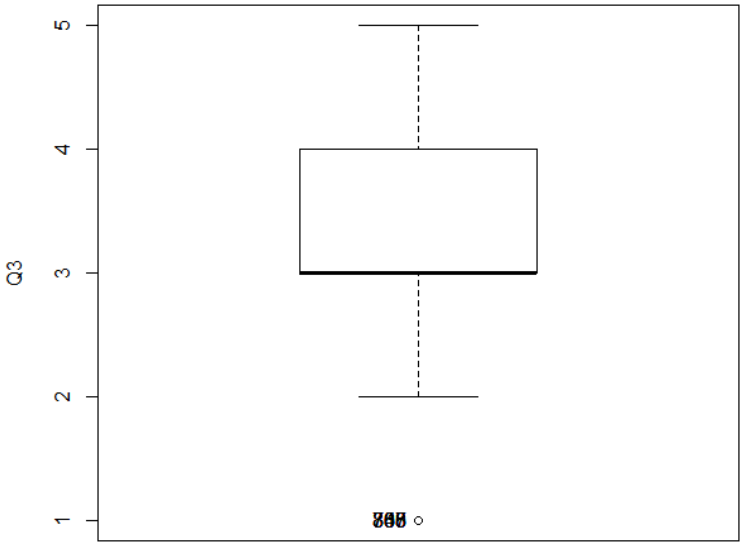


Figure 30 Box plot Q3

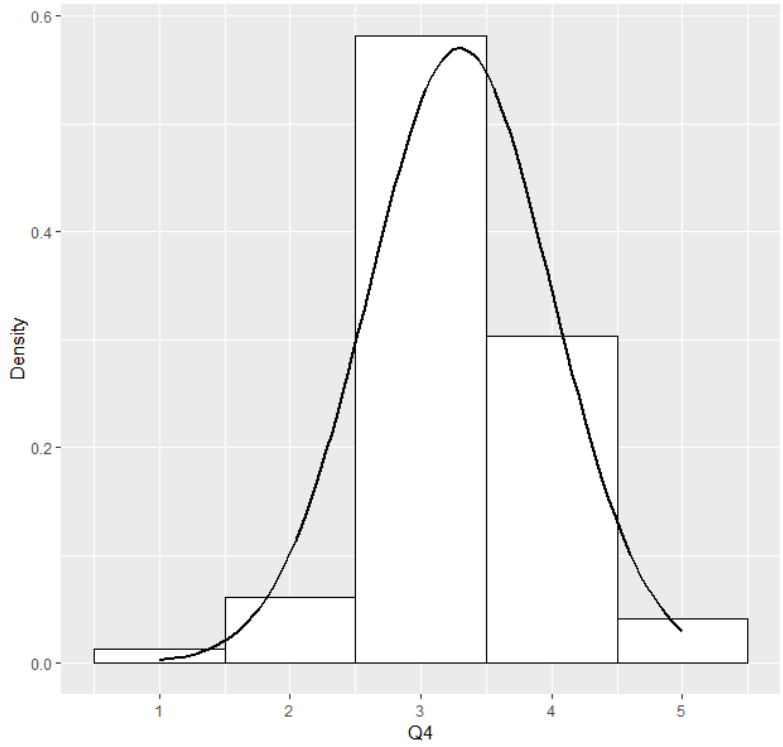


Figure 31 Histogram Q4

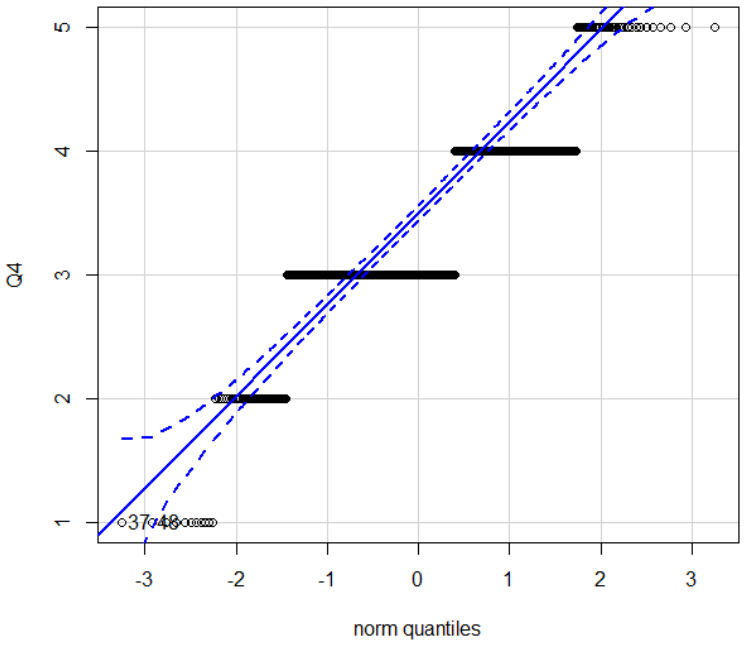


Figure 32 Q-Q Plot Q4

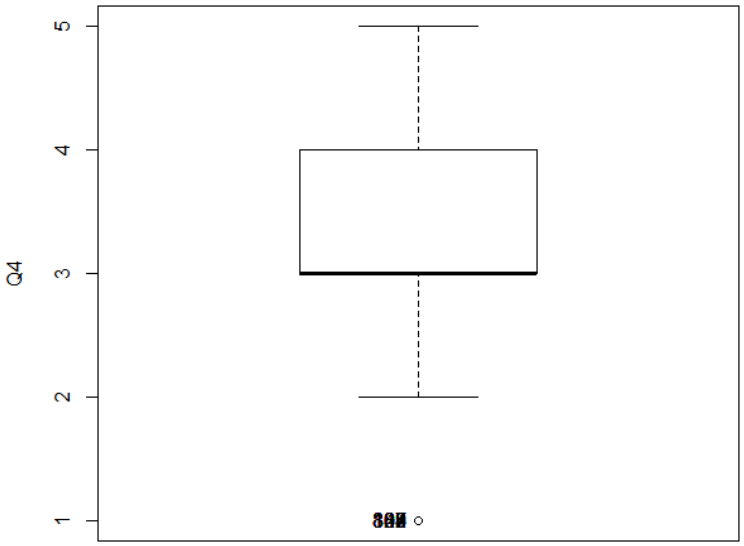


Figure 33 Box plot Q4

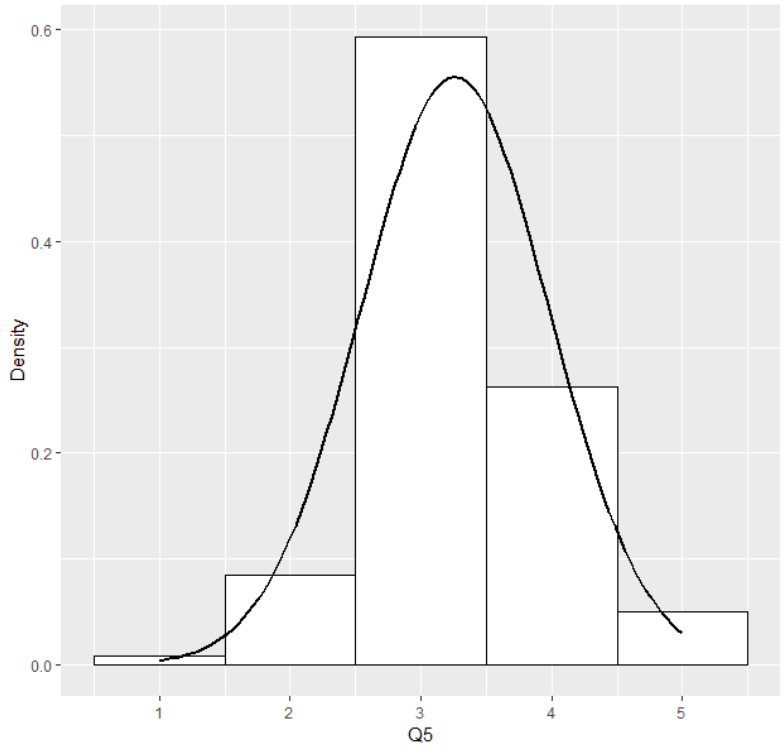


Figure 34 Histogram Q5

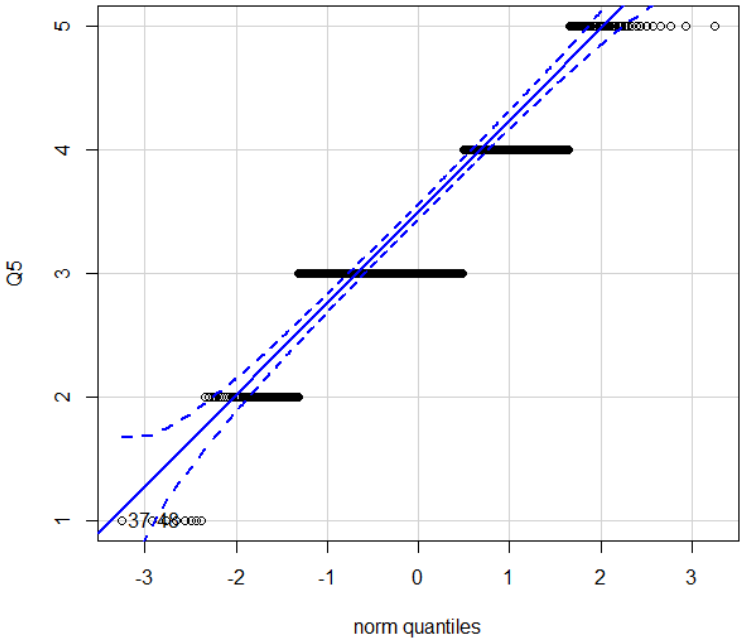


Figure 35 Q-Q Plot Q5

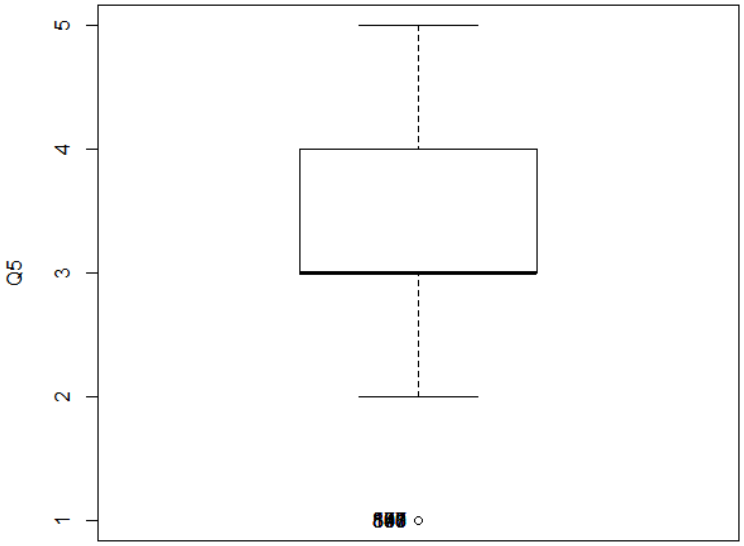


Figure 36 Box plot Q5

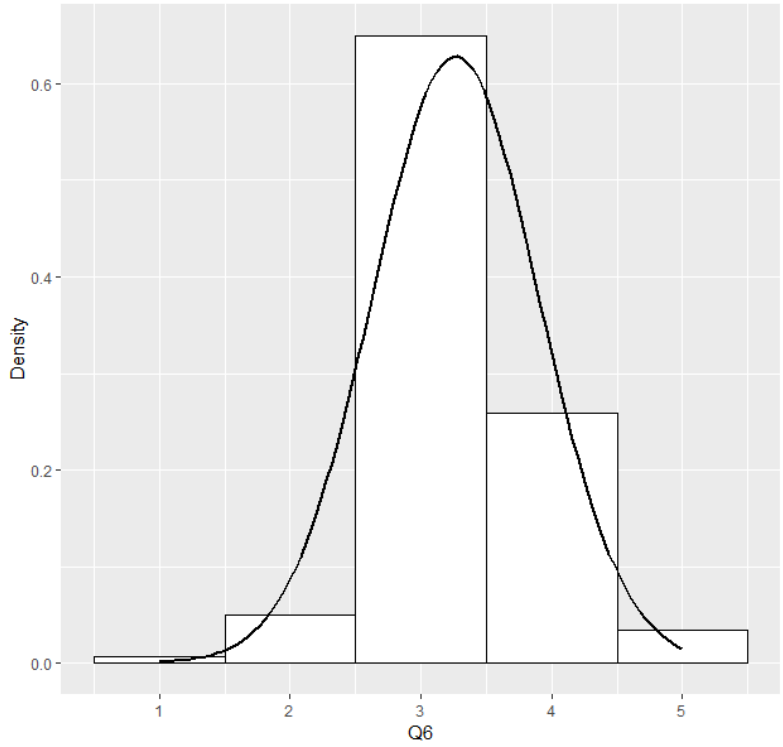


Figure 37 Histogram Q6

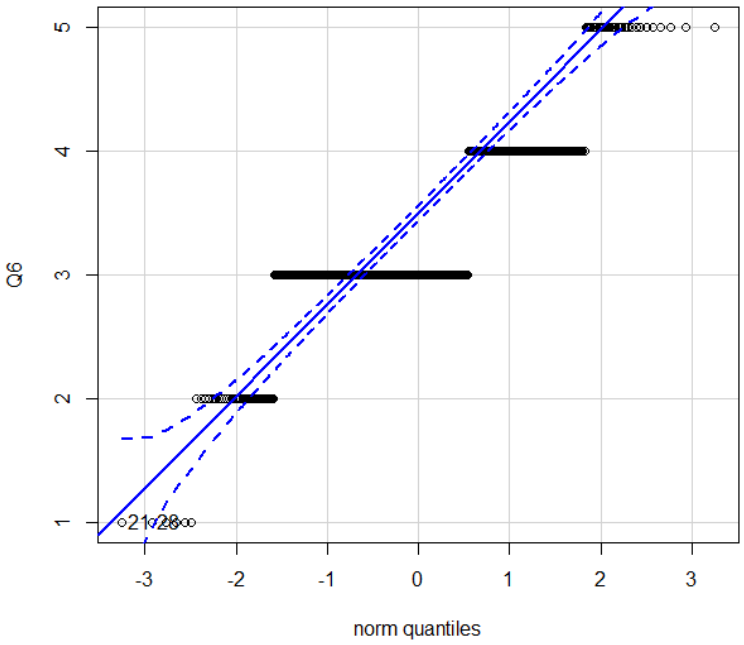


Figure 38 Q-Q Plot Q6

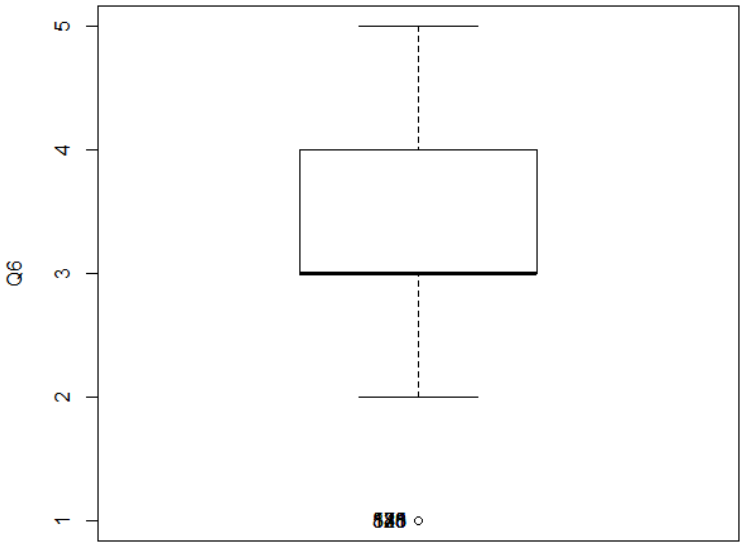


Figure 39 Box plot Q6

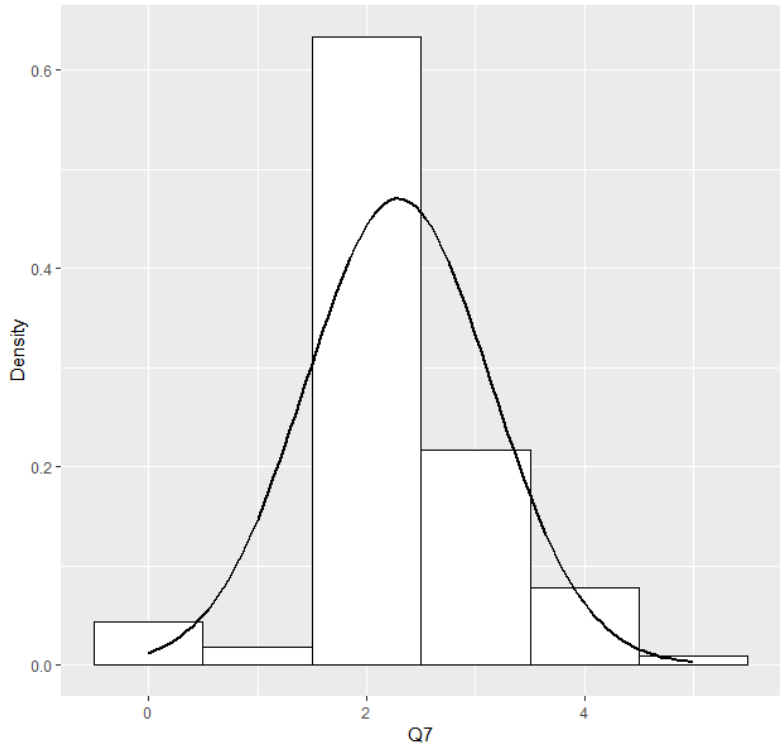


Figure 40 Histogram Q7

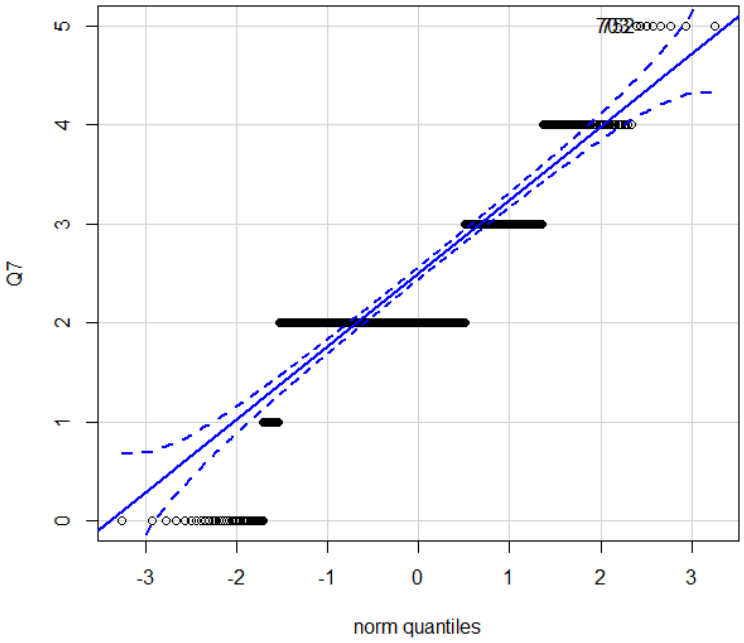


Figure 41 Q-Q Plot Q7

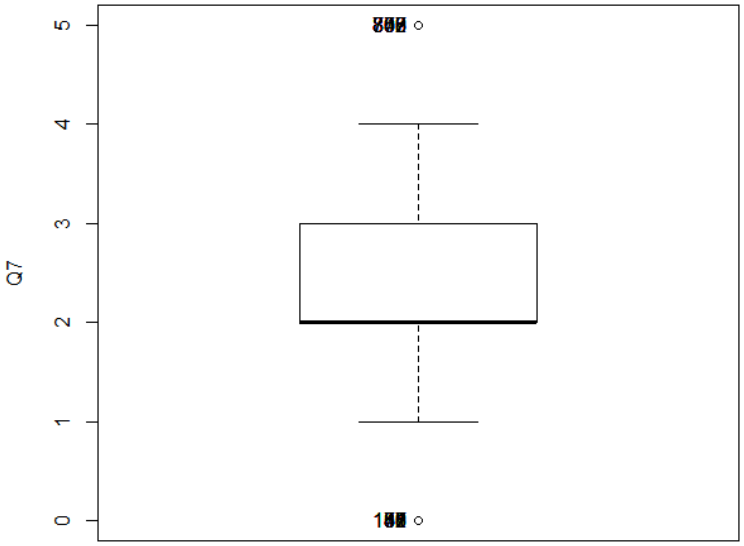


Figure 42 Box plot Q7

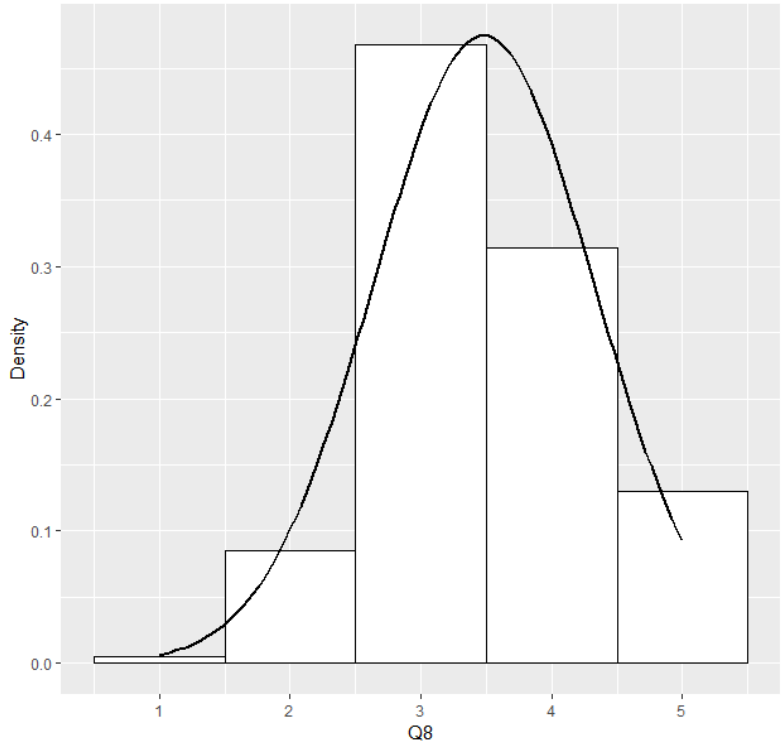


Figure 43 Histogram Q8

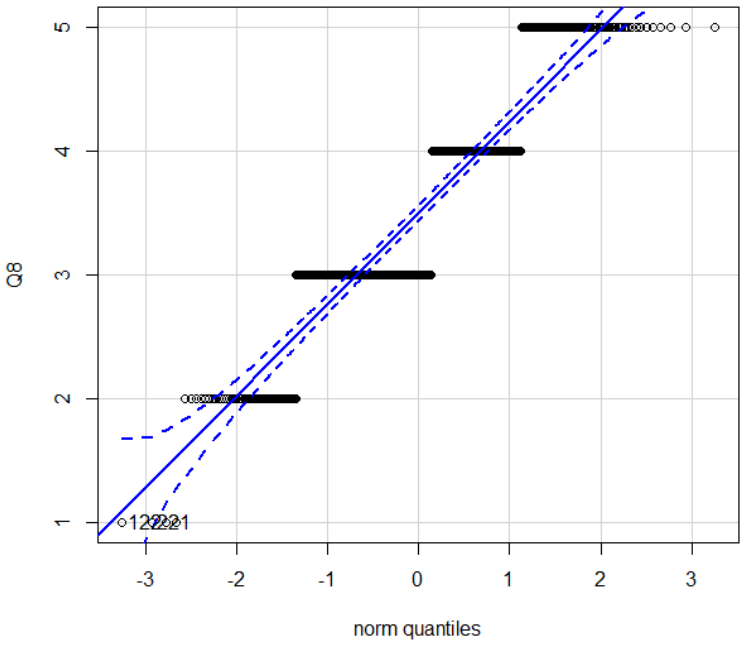


Figure 44 Q-Q Plot Q8

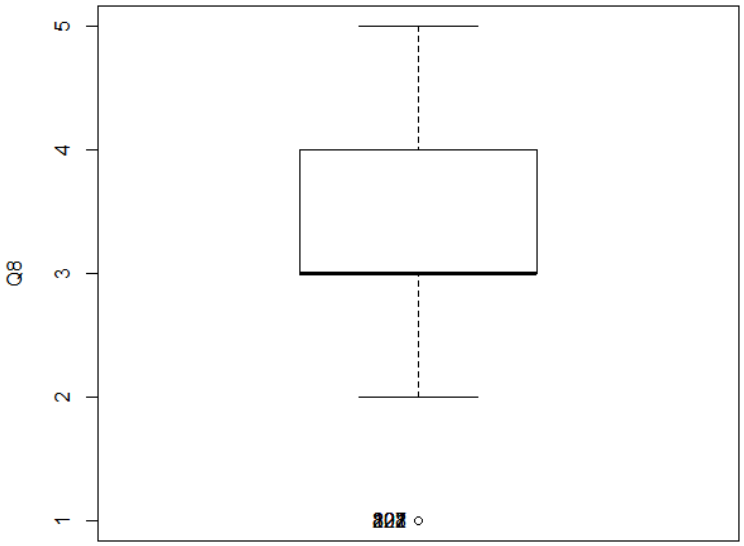


Figure 45 Box plot Q8

Descriptive Statistics over Persons involved n=65

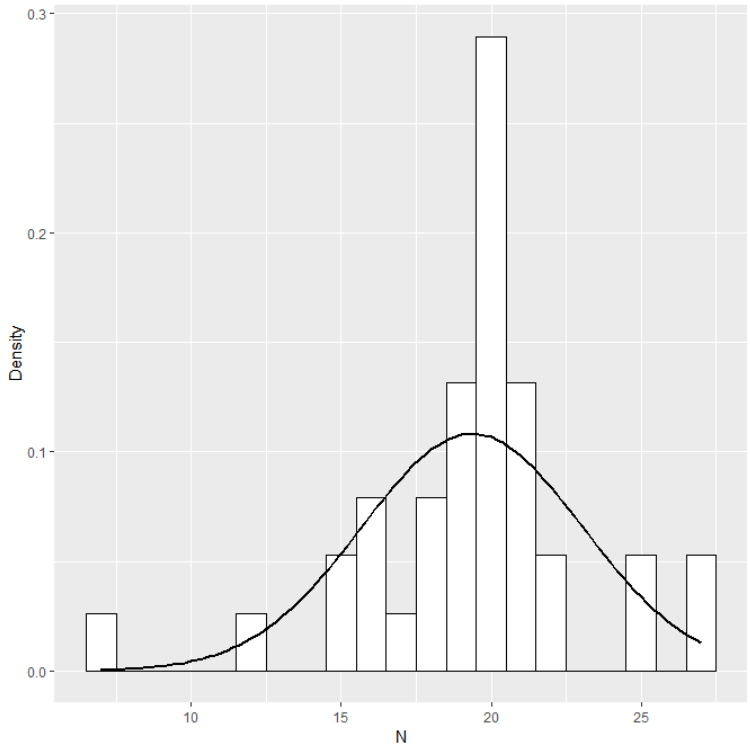


Figure 46 Histogram N

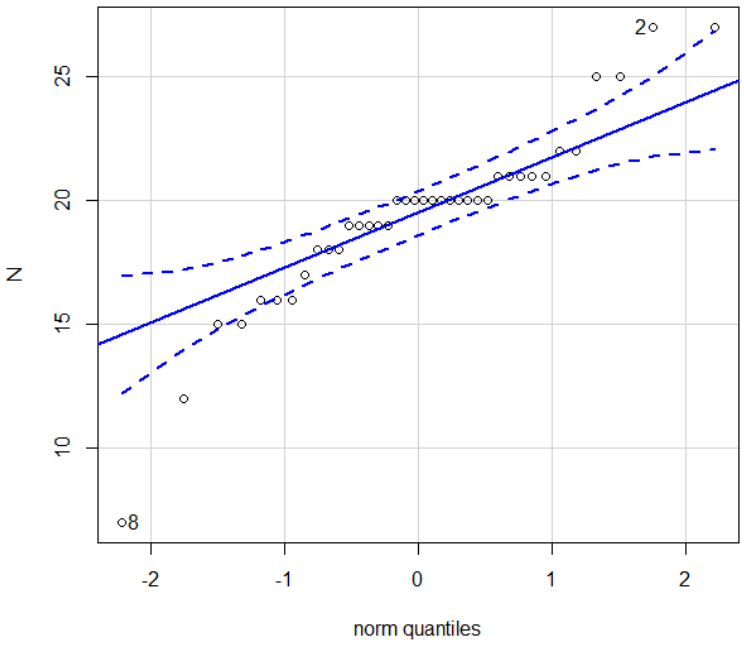


Figure 47 Q-Q plot N

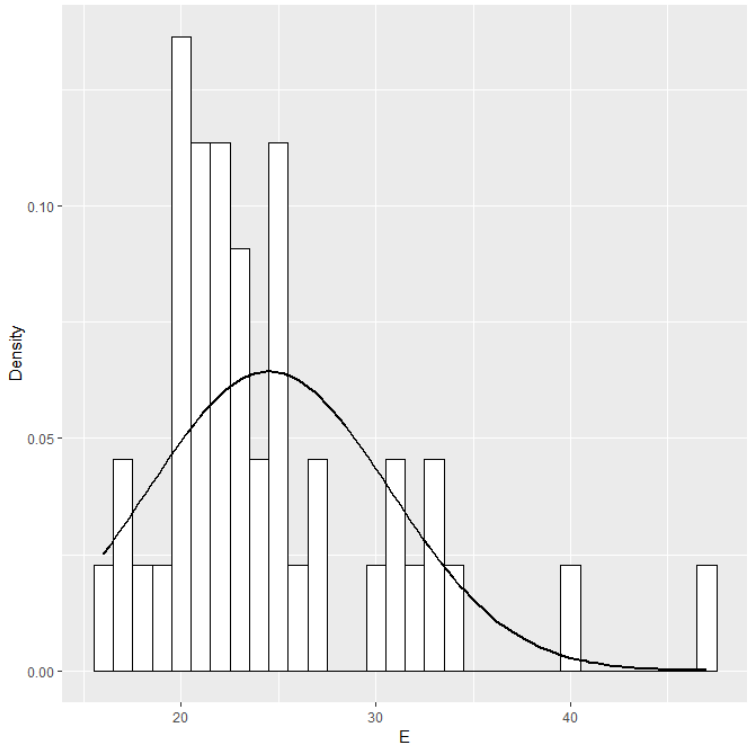


Figure 48 Histogram E

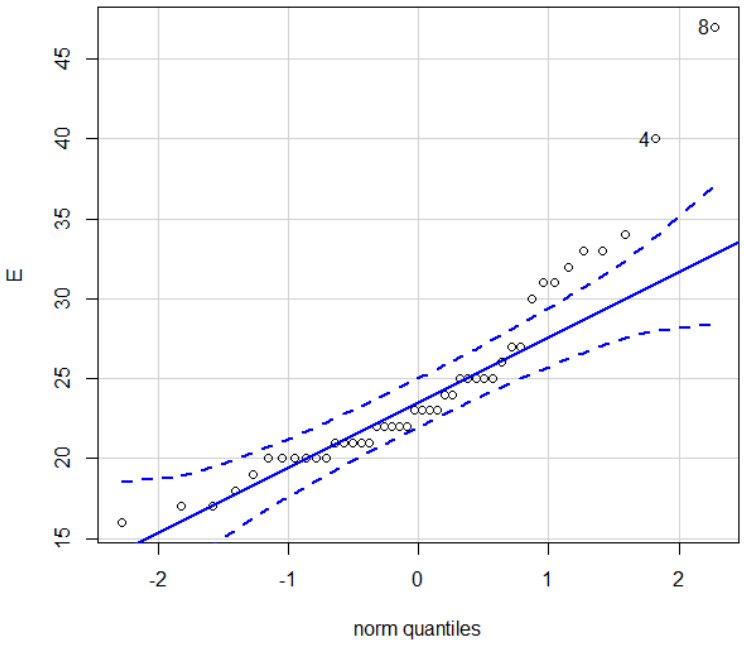


Figure 49 Q-Q plot E

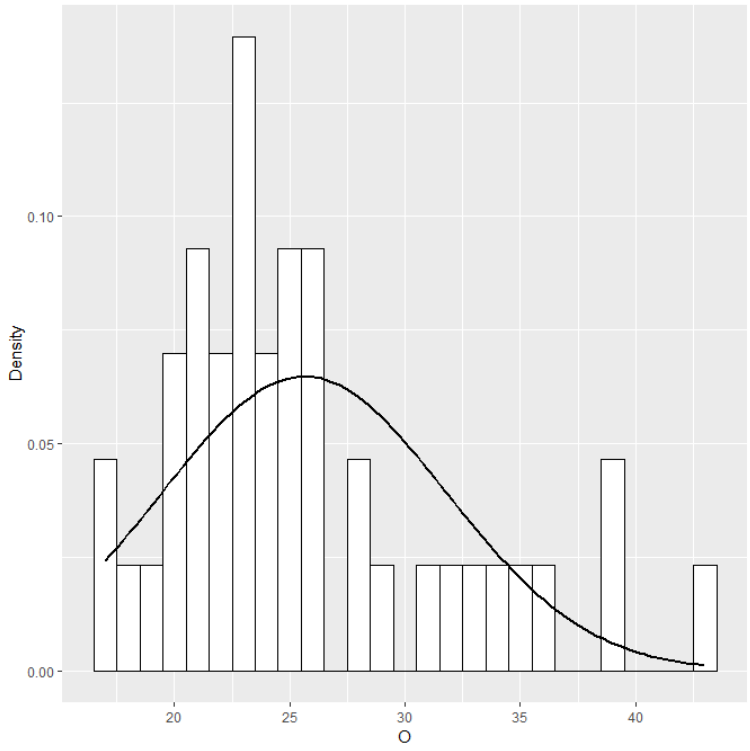


Figure 50 Histogram O

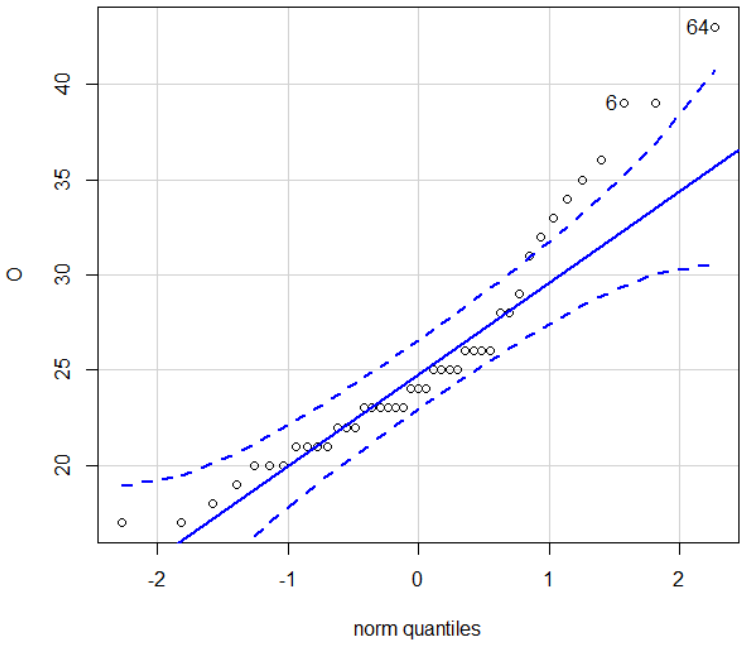


Figure 51 Q-Q plot O

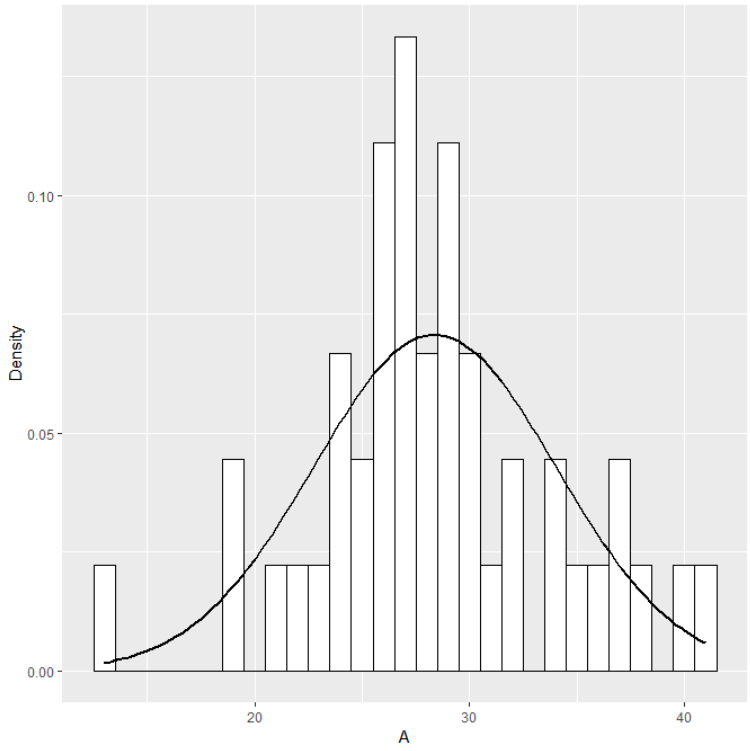


Figure 52 Histogram A

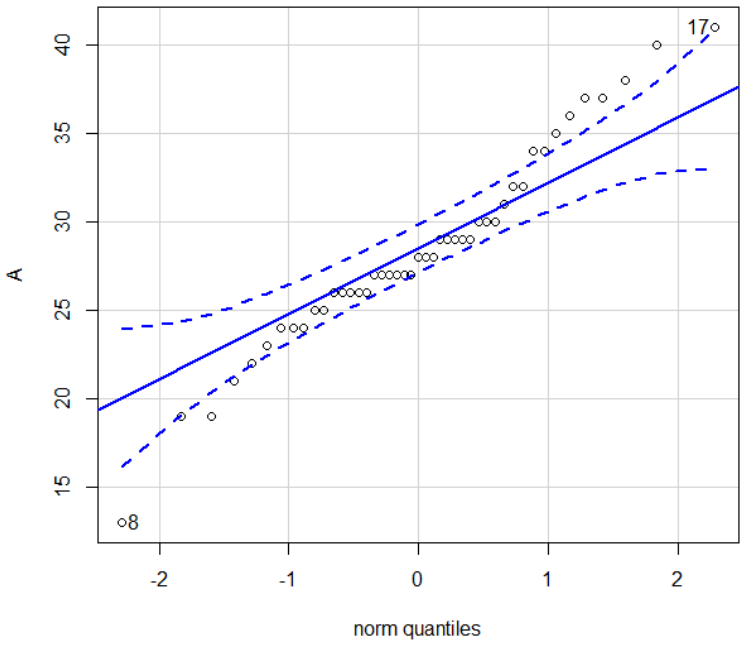


Figure 53 Q-Q plot A

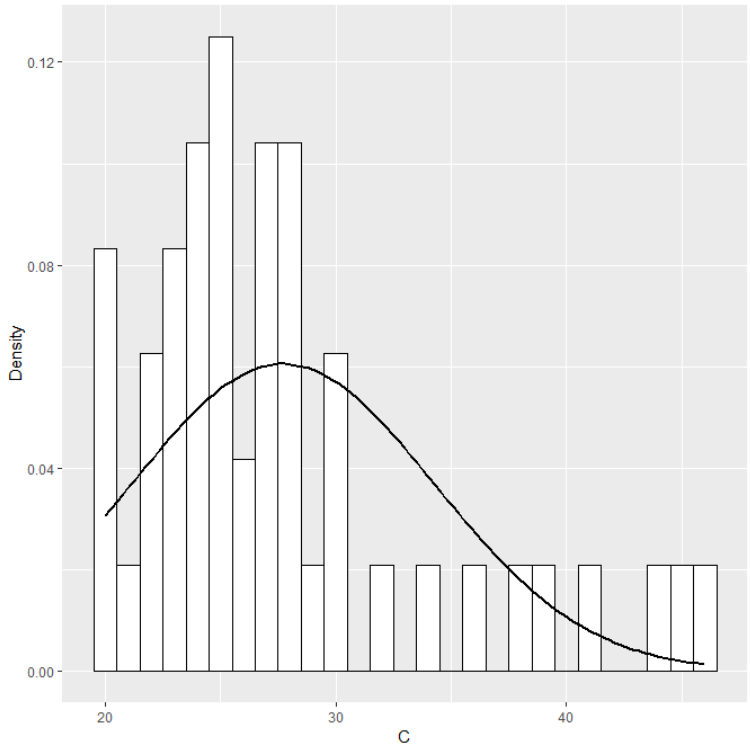


Figure 54 Histogram C

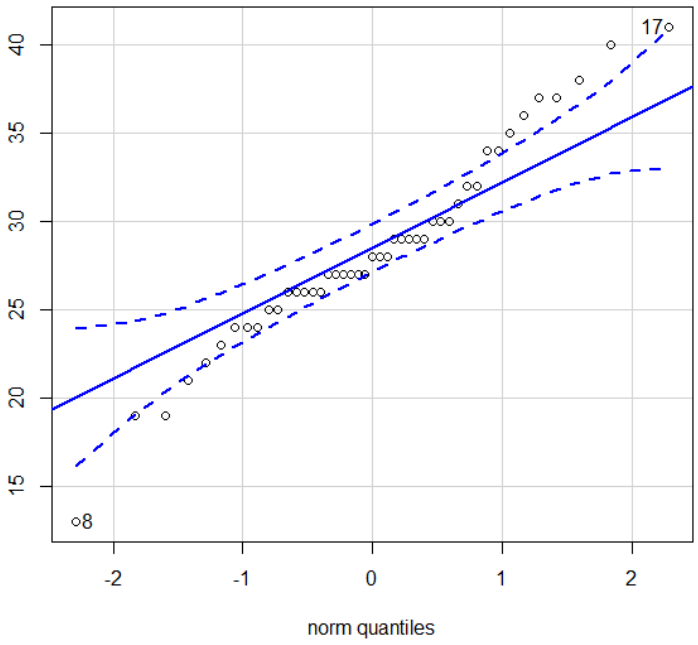


Figure 55 Q-Q plot C

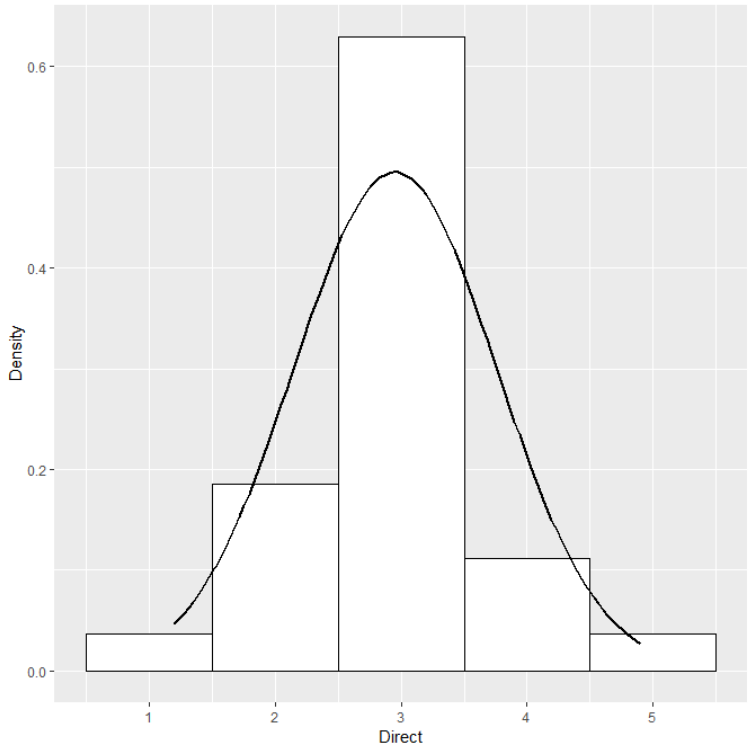


Figure 56 Histogram Culture directness

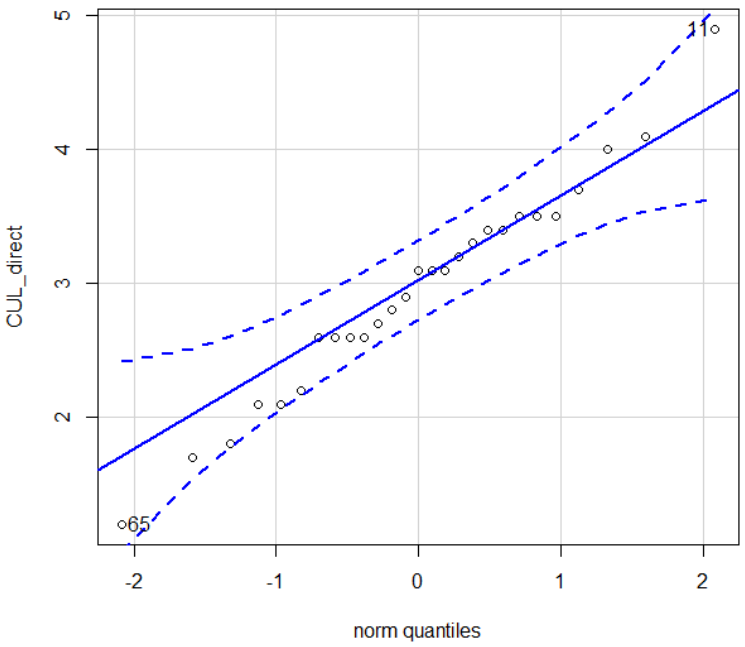


Figure 57 Q-Q plot CUL_direct

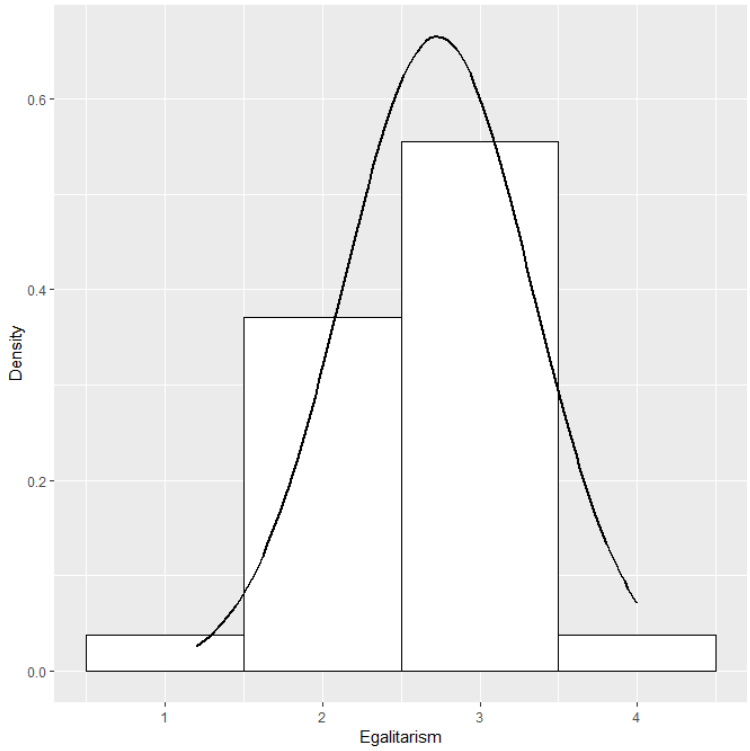


Figure 58 Histogram Egalitarism

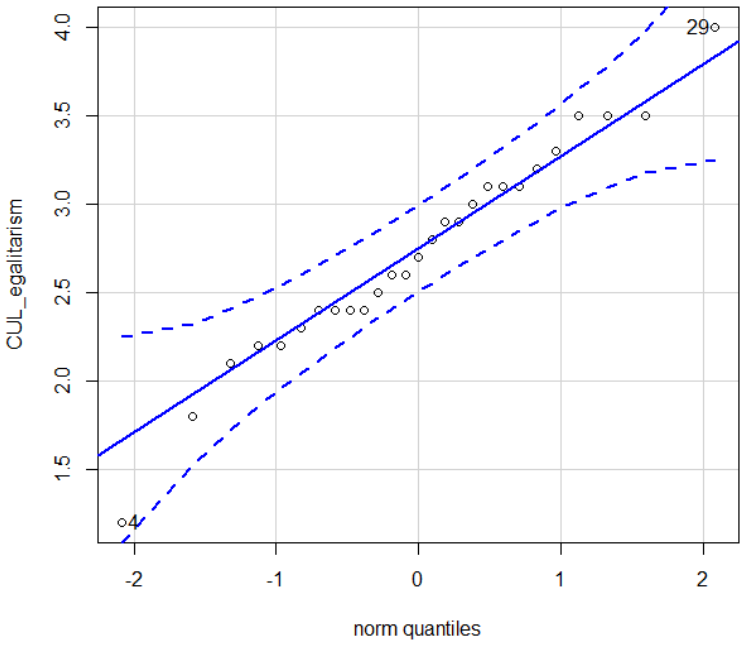


Figure 59 Q-Q plot CUL_egalitarism

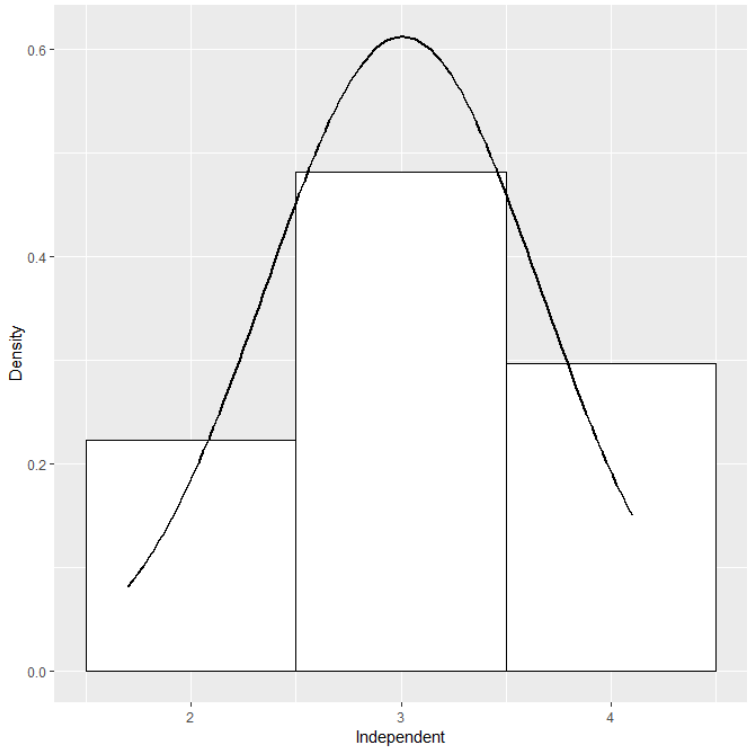


Figure 60 Histogram Independence

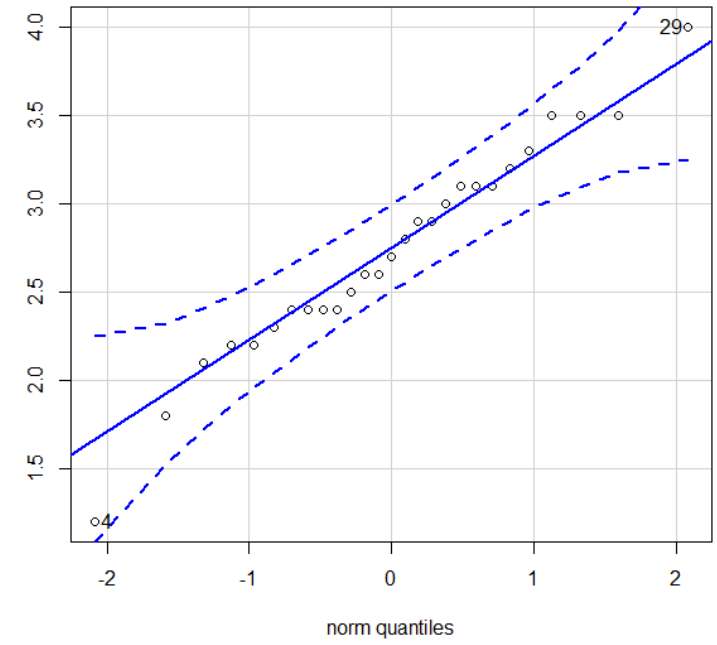


Figure 61 Q-Q plot CUL_independence

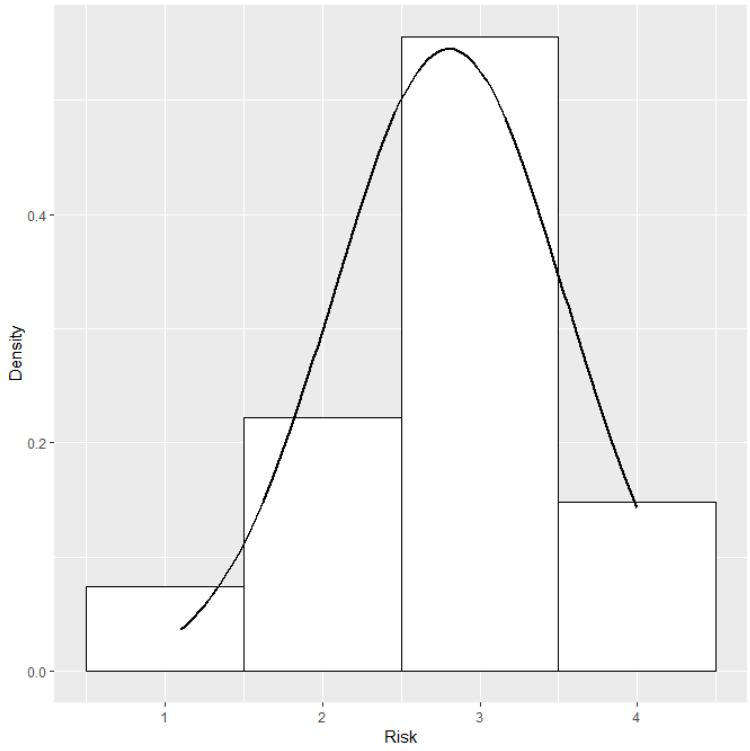


Figure 62 Histogram Risk

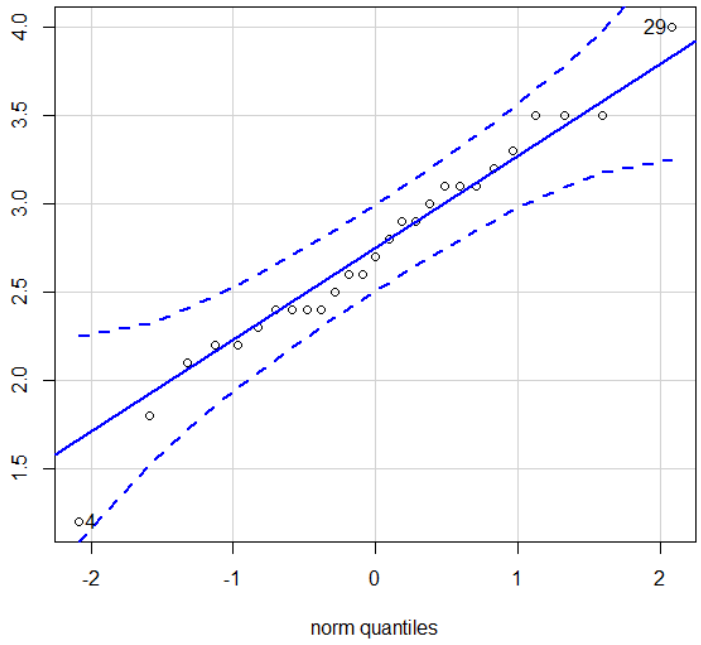


Figure 63 Q-Q plot CUL_risk

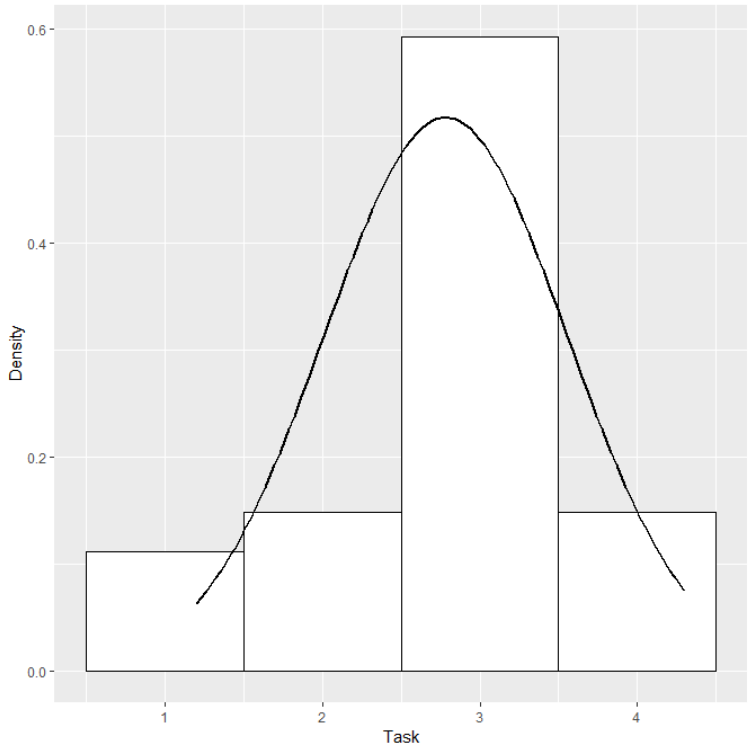


Figure 64 Histogram Task

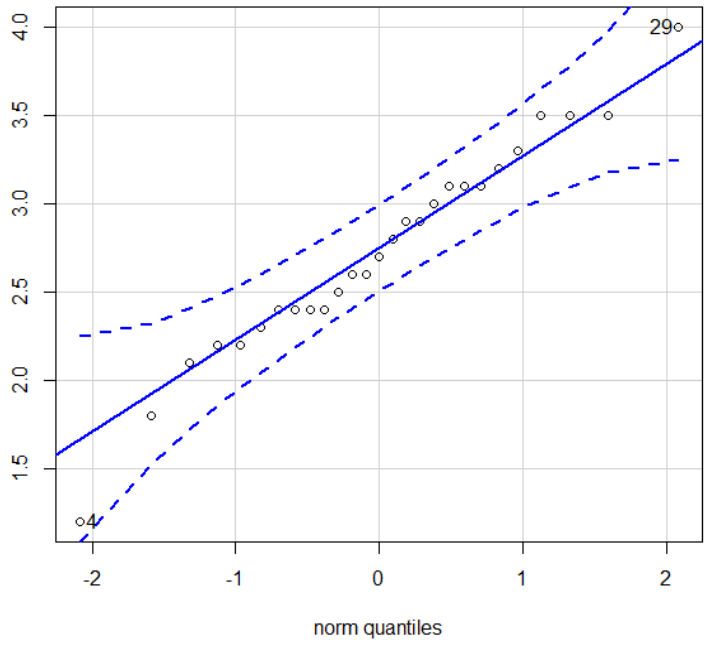


Figure 65 Q-Q plot CUL_egalitarism

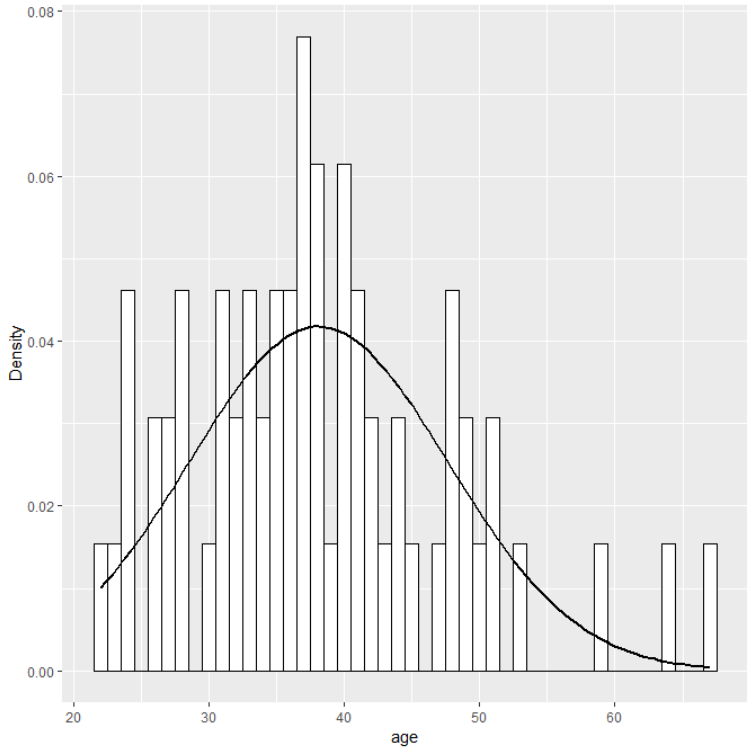


Figure 66 Histogram Age

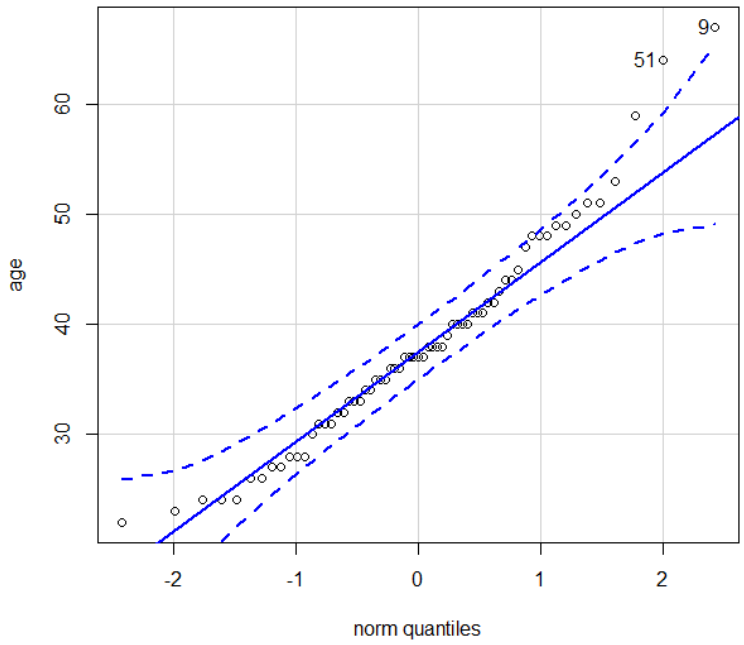


Figure 67 Q-Q plot Age

C.2 DESCRIPTIVE STATISTICS FOR PROJECT BASED DATA N=111

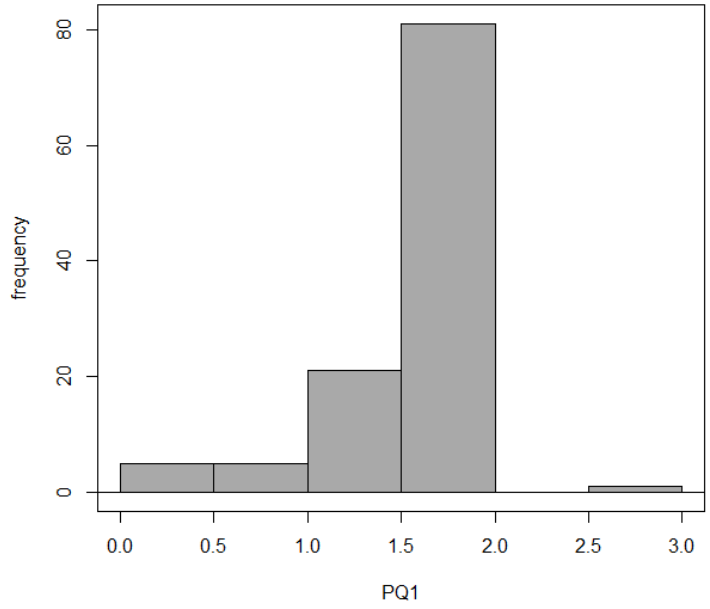


Figure 68 Histogram PQ1

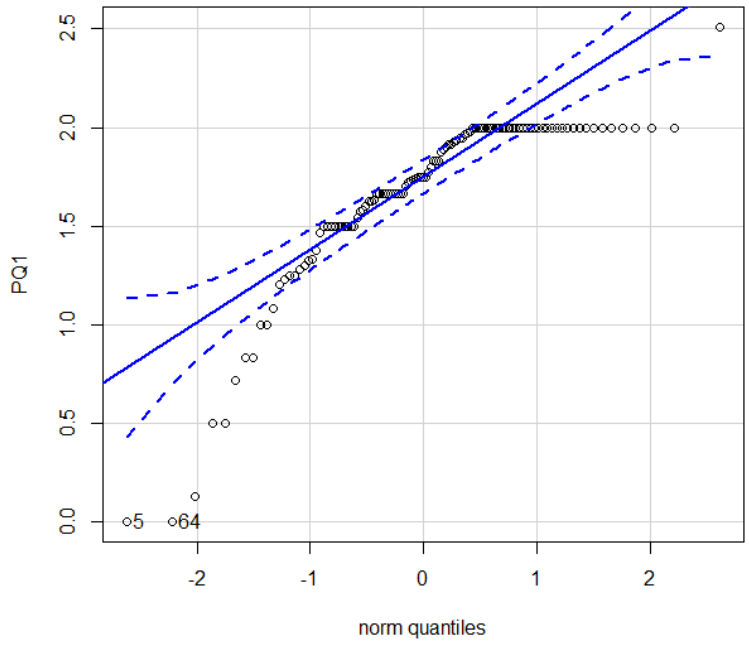


Figure 69 Q-Q plot PQ1

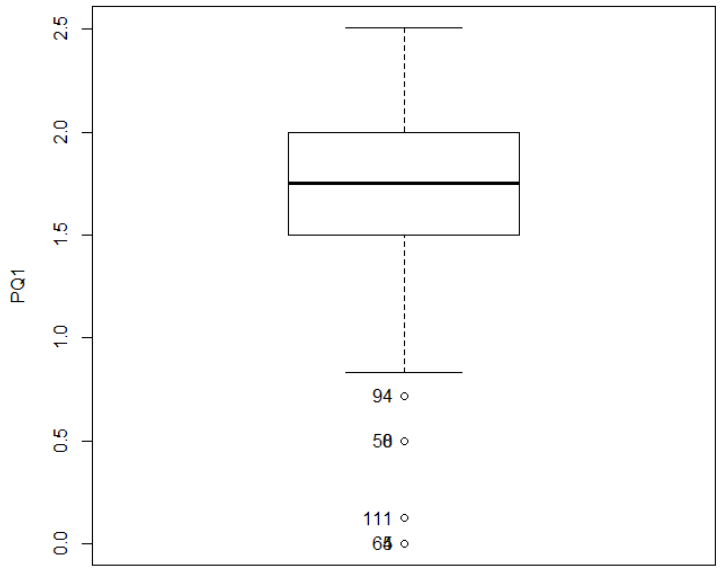


Figure 70 Box plot PQ1

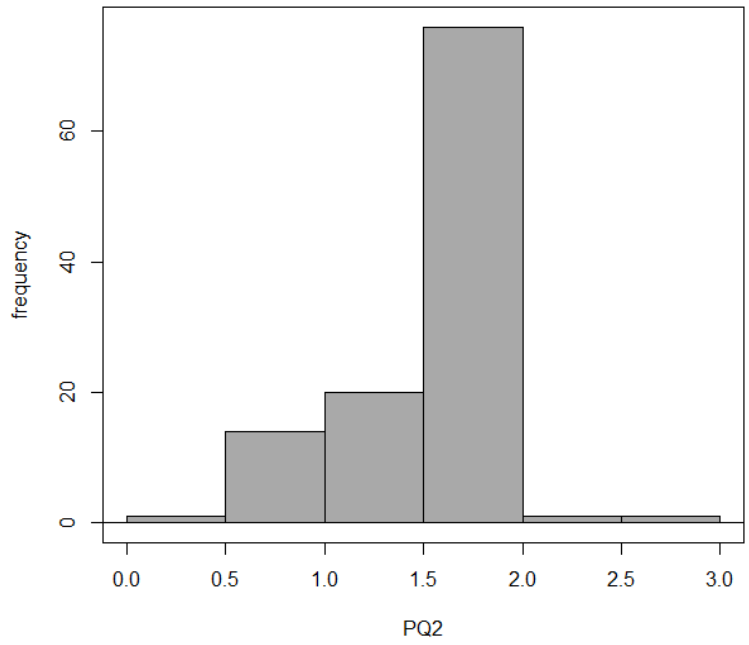


Figure 71 Histogram PQ2

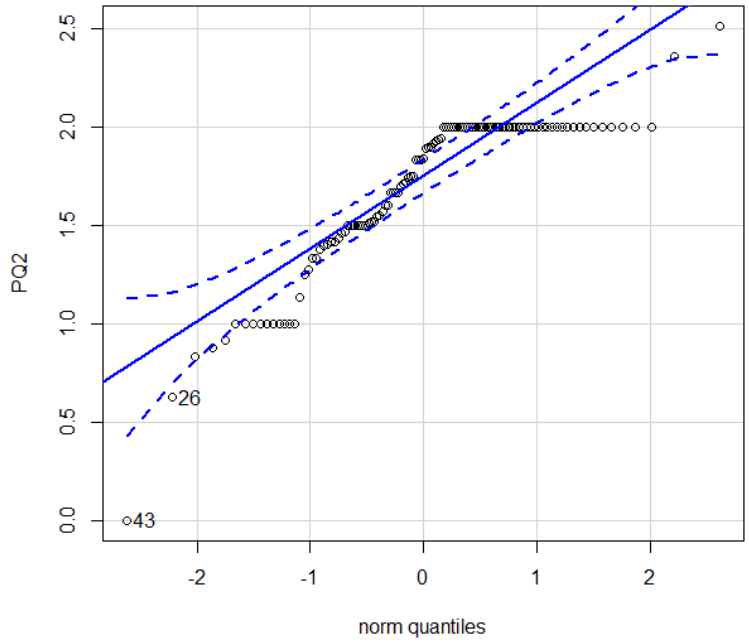


Figure 72 Q-Q plot PQ2

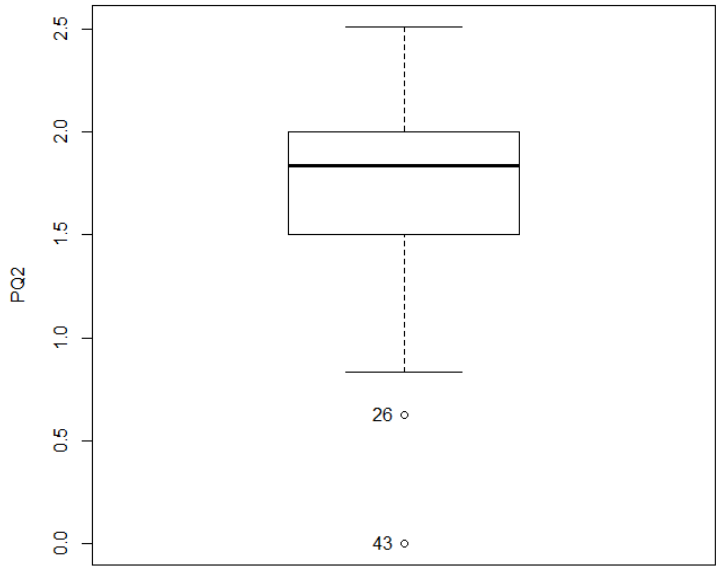


Figure 73 Box plot PQ_2

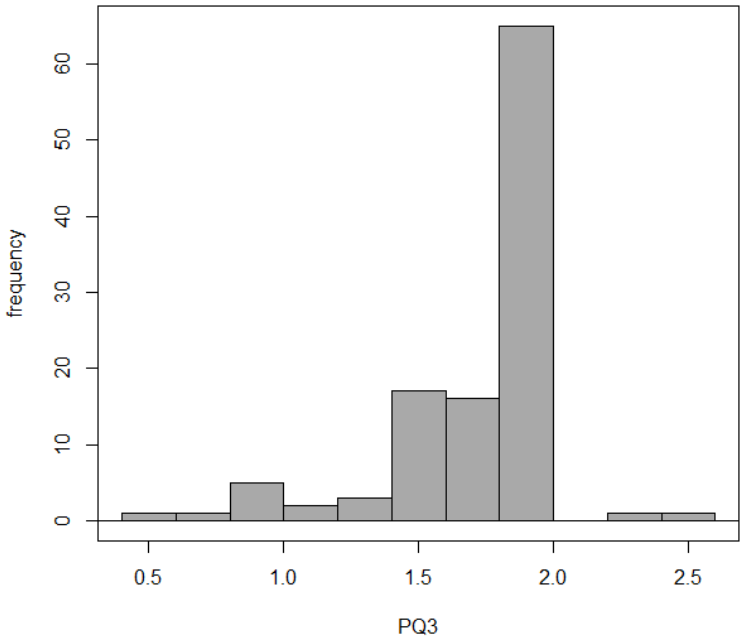


Figure 74 Histogram PQ3

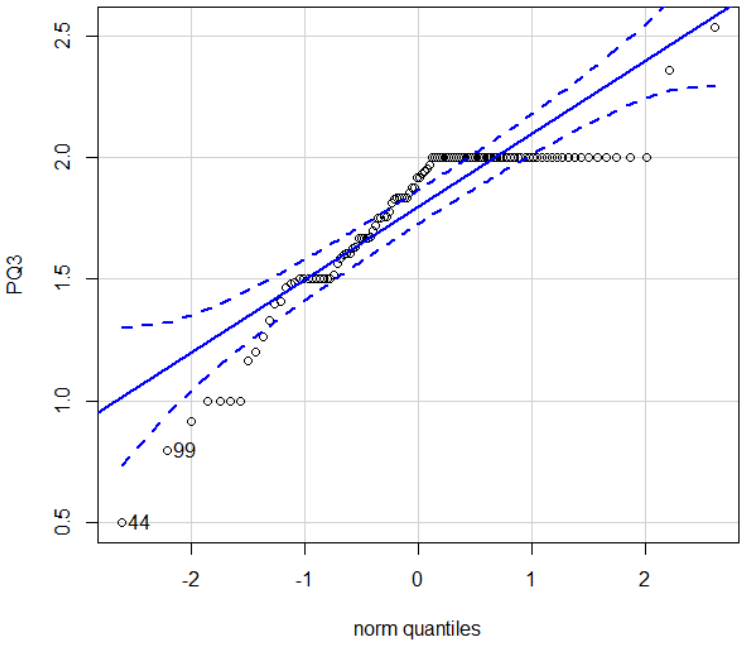


Figure 75 Q-Q plot PQ3

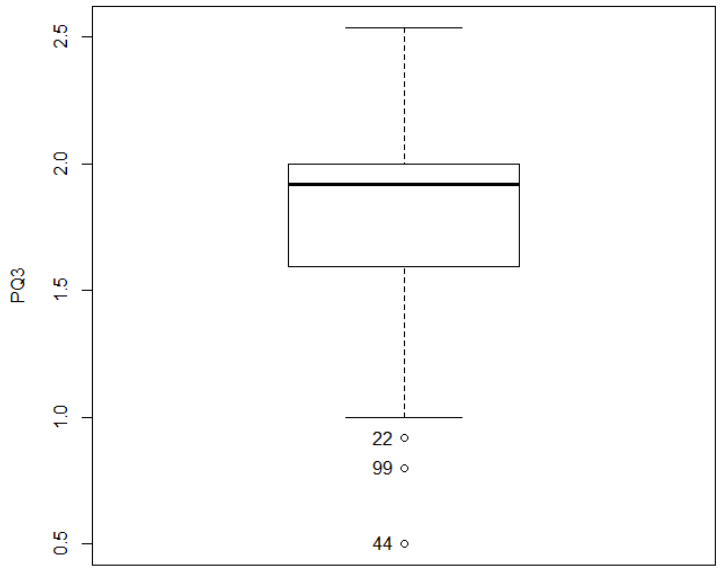


Figure 76 Box plot PQ3

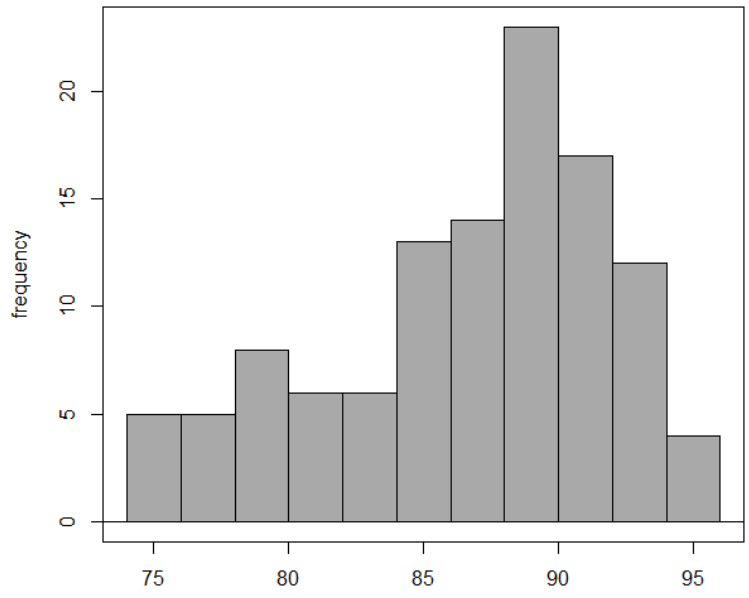


Figure 77 Histogram PO_Exp

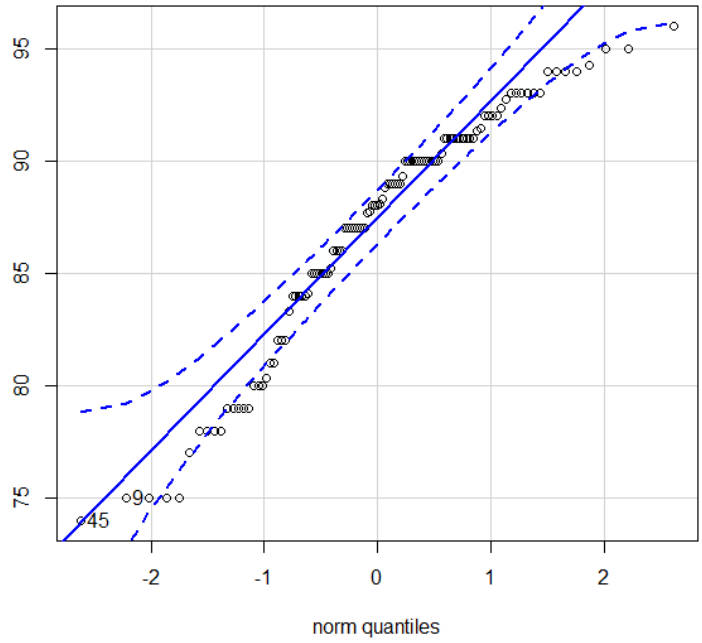


Figure 78 Q-Q plot PO_Exp

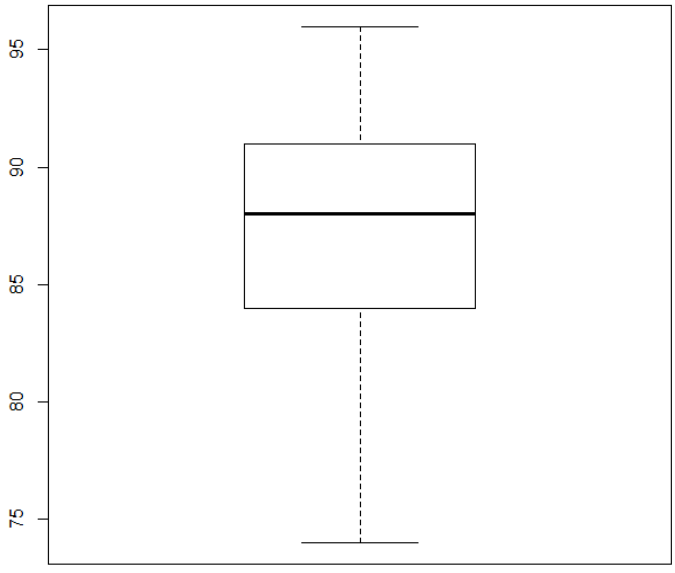


Figure 79 Box plot PO_Exp

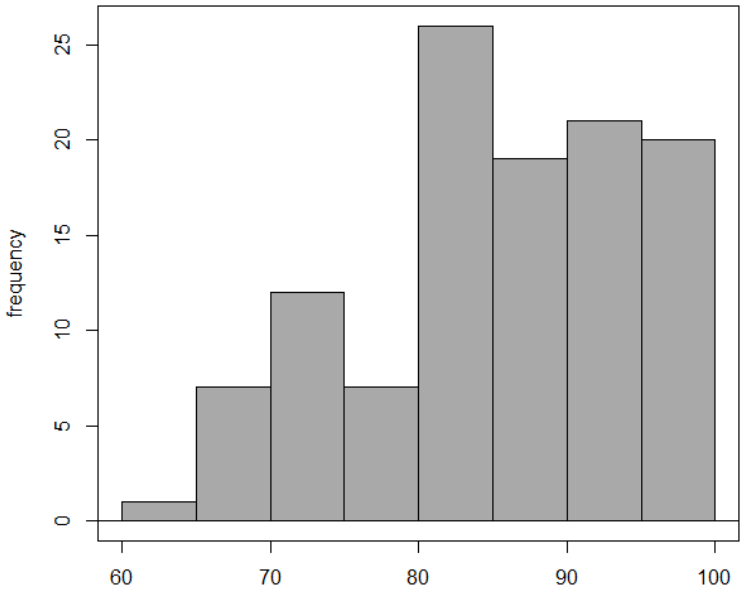


Figure 80 Histogram PO_Team

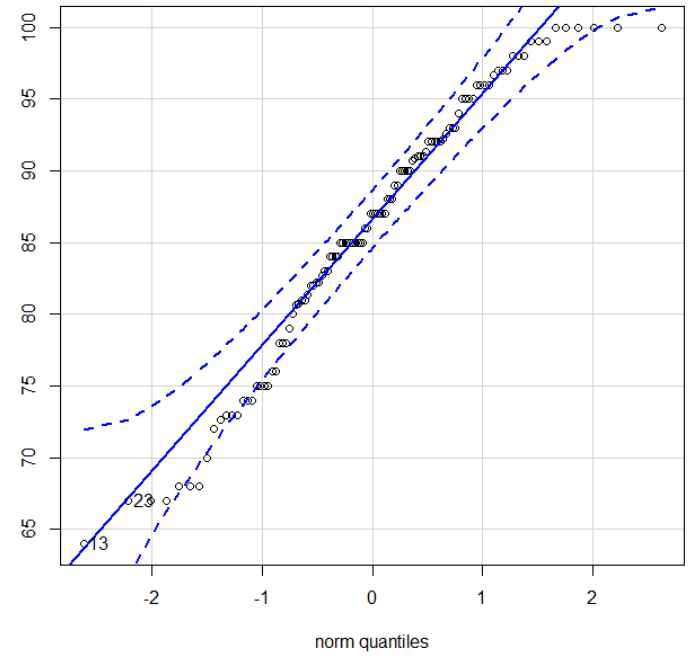


Figure 81 Q-Q plot PO_team

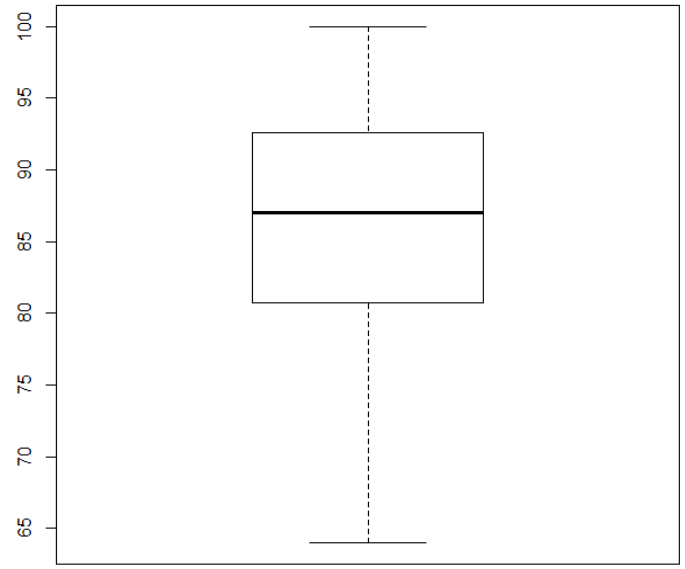


Figure 82 Box plot PO_Team

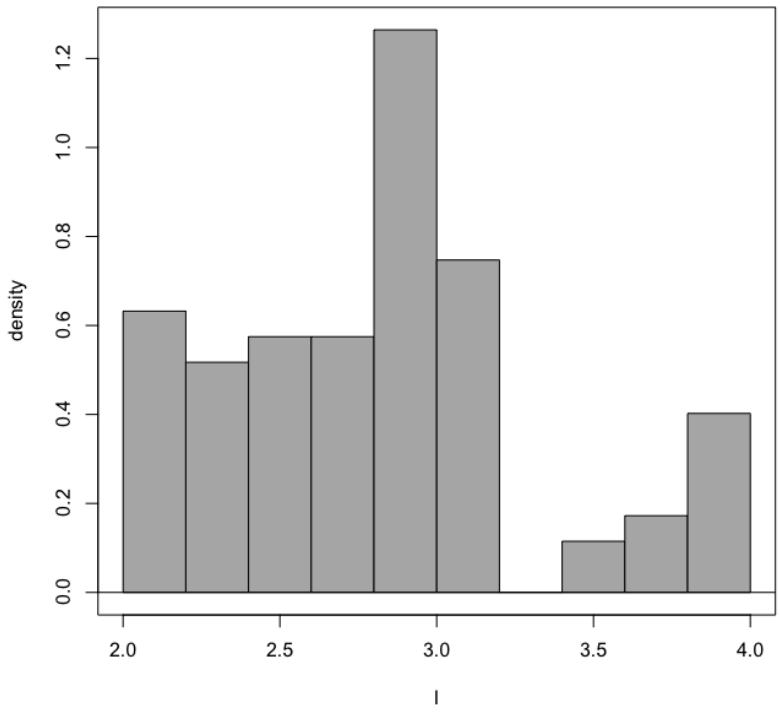


Figure 83 Histogram Independence

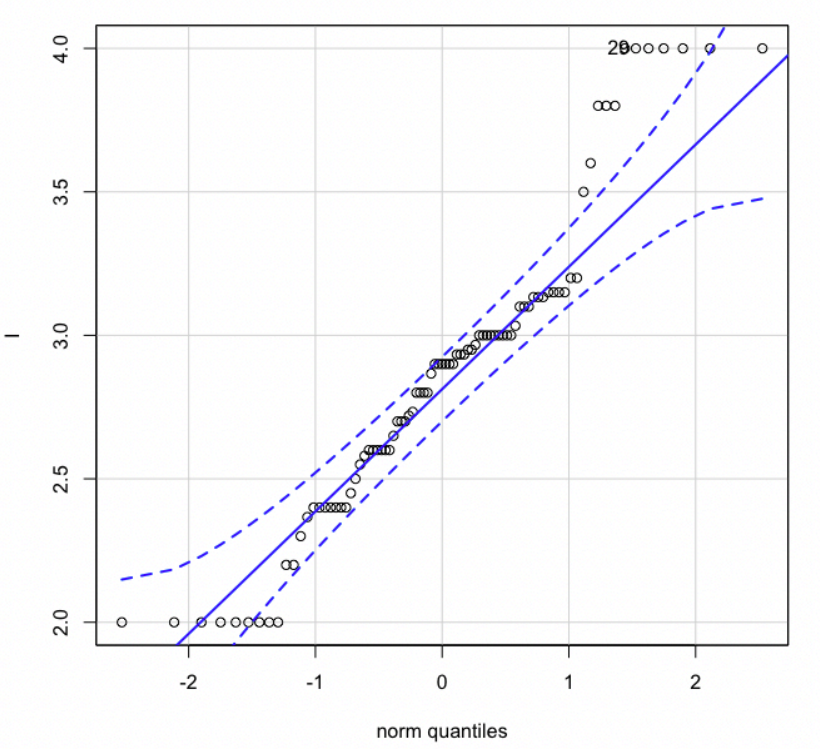


Figure 84 Q-Q plot Independence

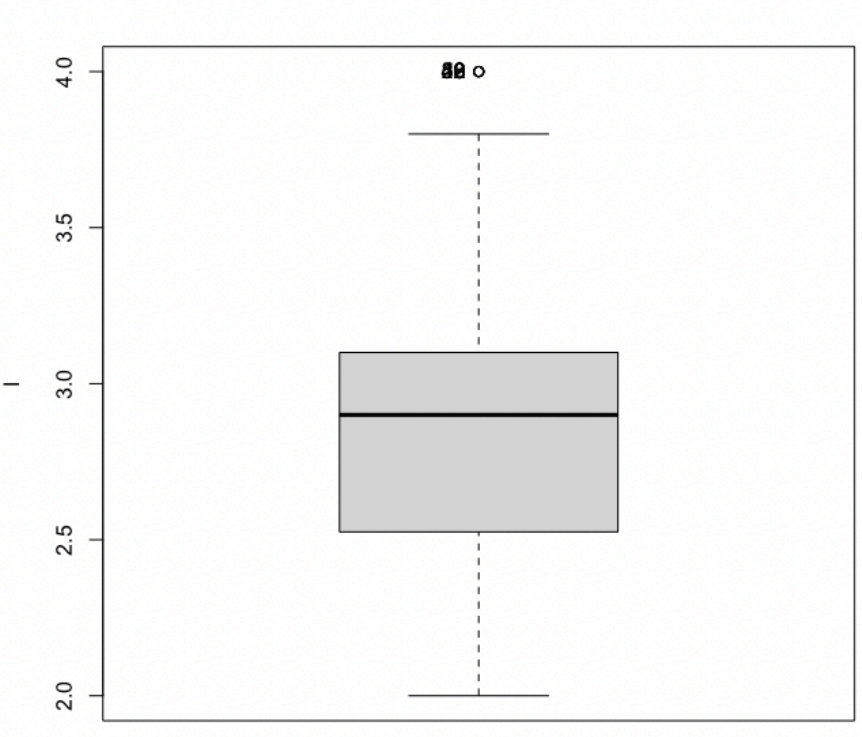


Figure 85 Box plot Independence

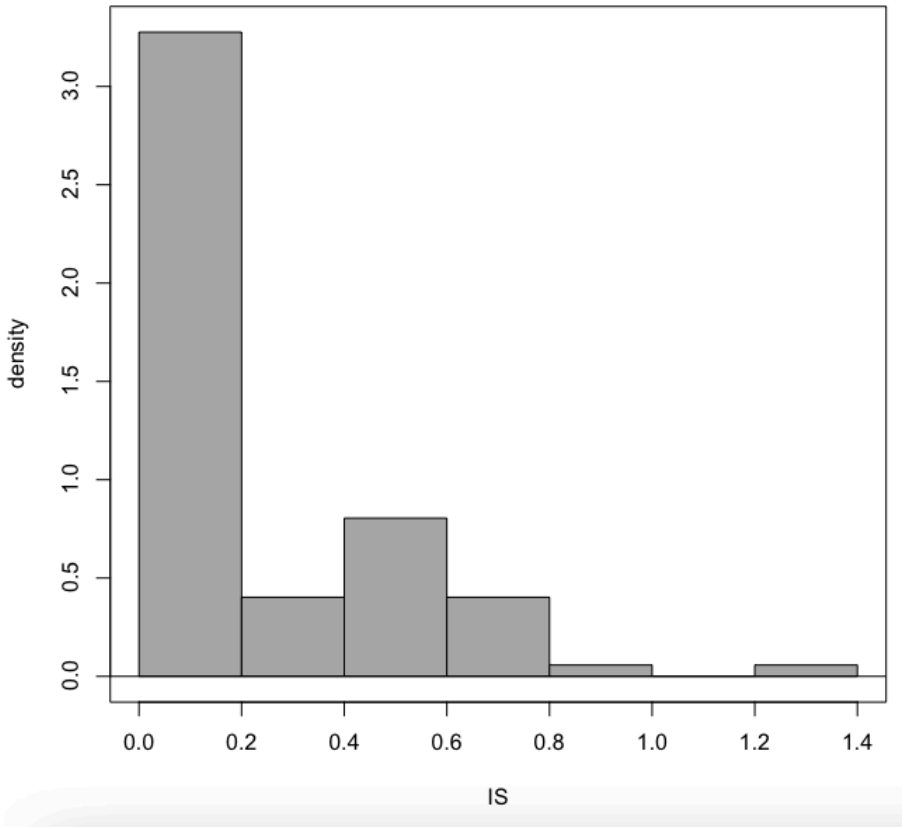


Figure 86 Histogram Independence Std Deviation

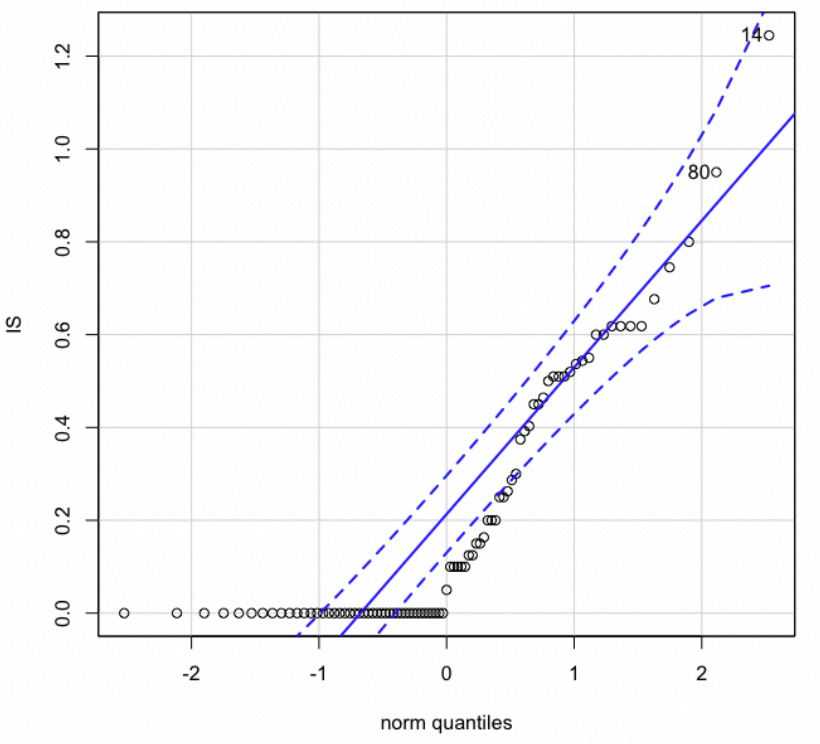


Figure 87 Q-Q plot Independence Std Deviation

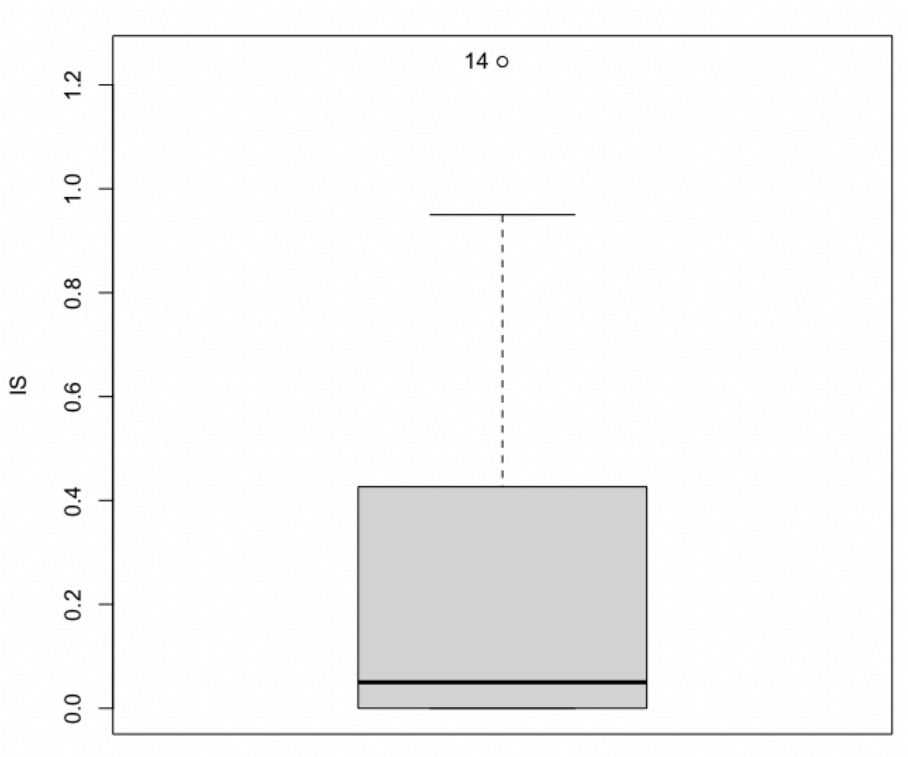


Figure 88 Box plot Independence Std Deviation

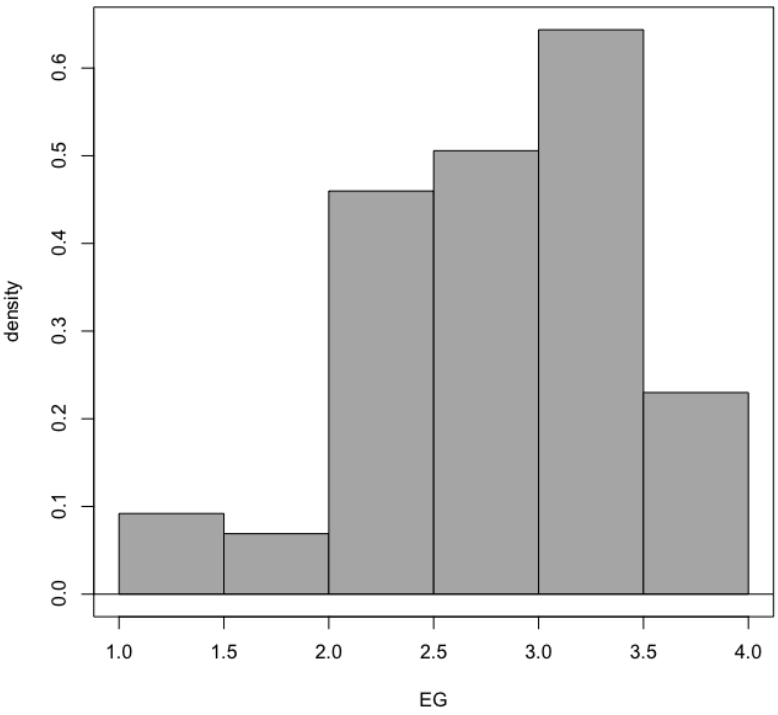


Figure 89 Histogram Egalitarism

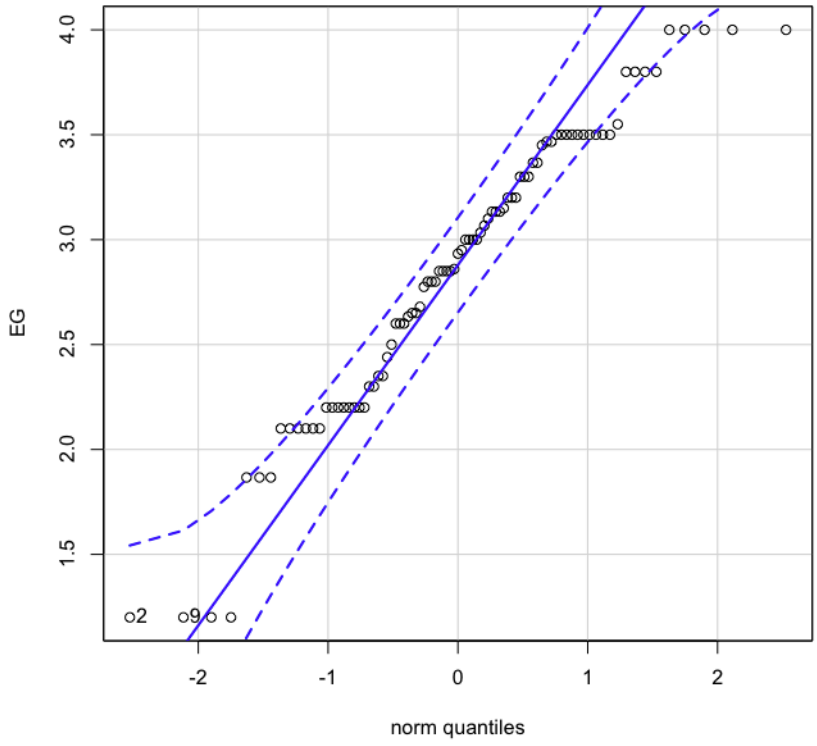


Figure 90 Q-Q plot Egalitarism

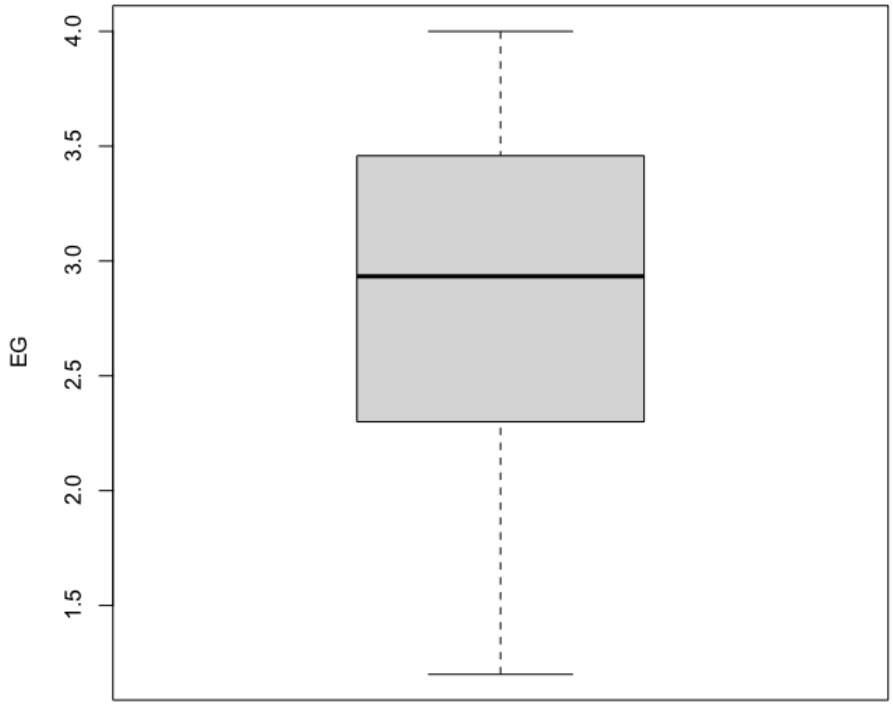


Figure 91 Box plot Egalitarism

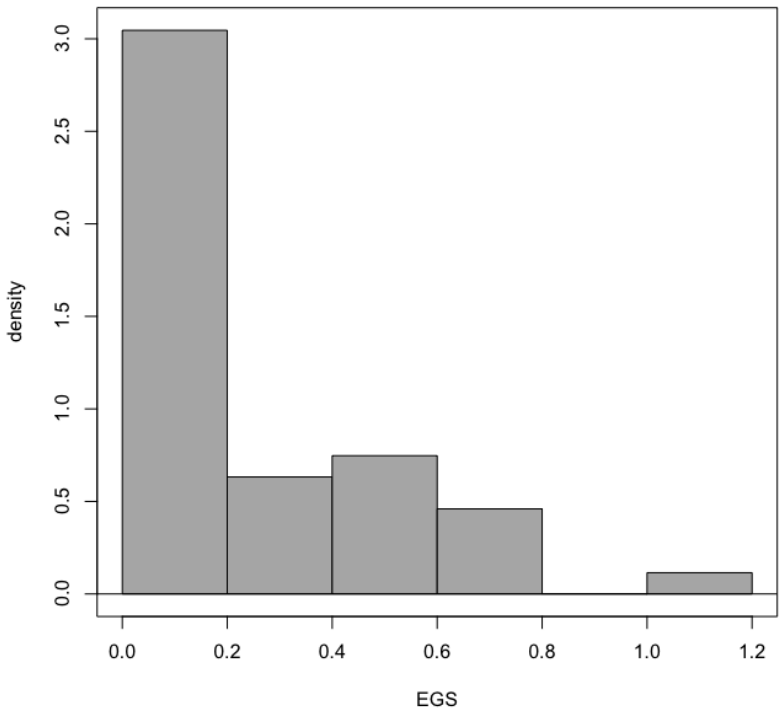


Figure 92 Histogram Egalitarism Std Deviation

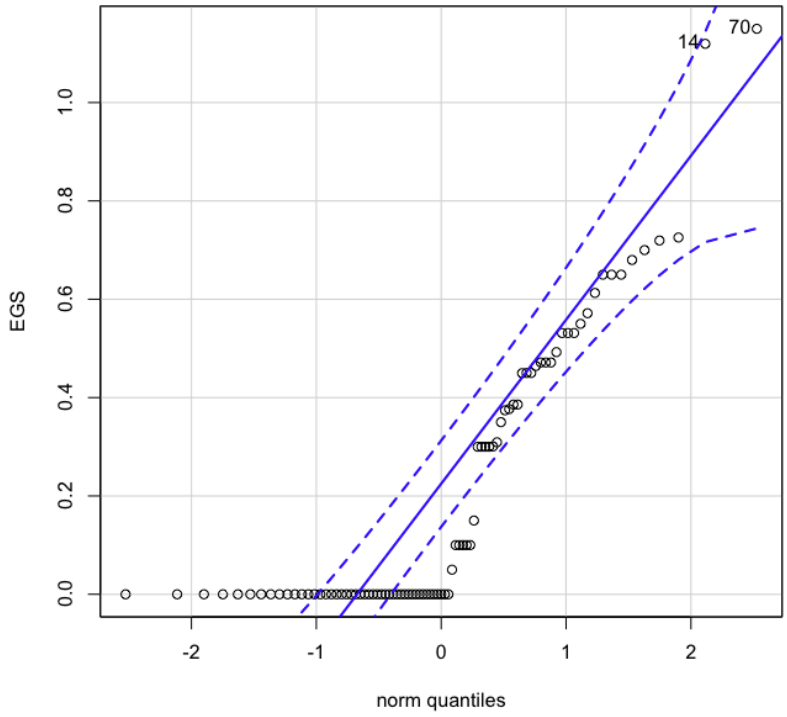


Figure 93 Q-Q plot Egalitarism Std Deviation

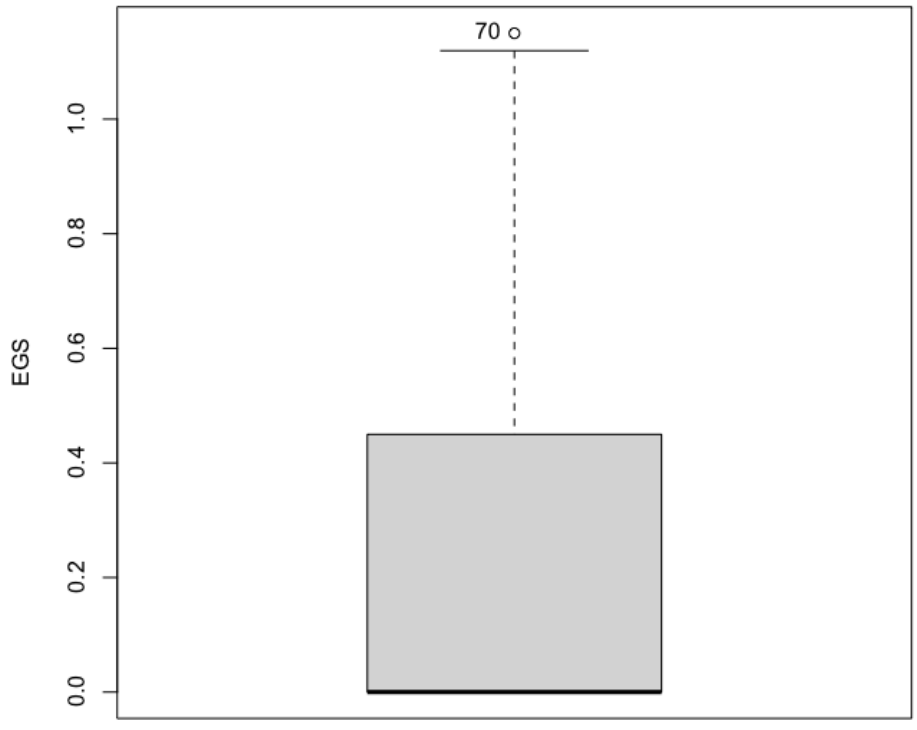


Figure 94 Box plot Egalitarism Std Deviation

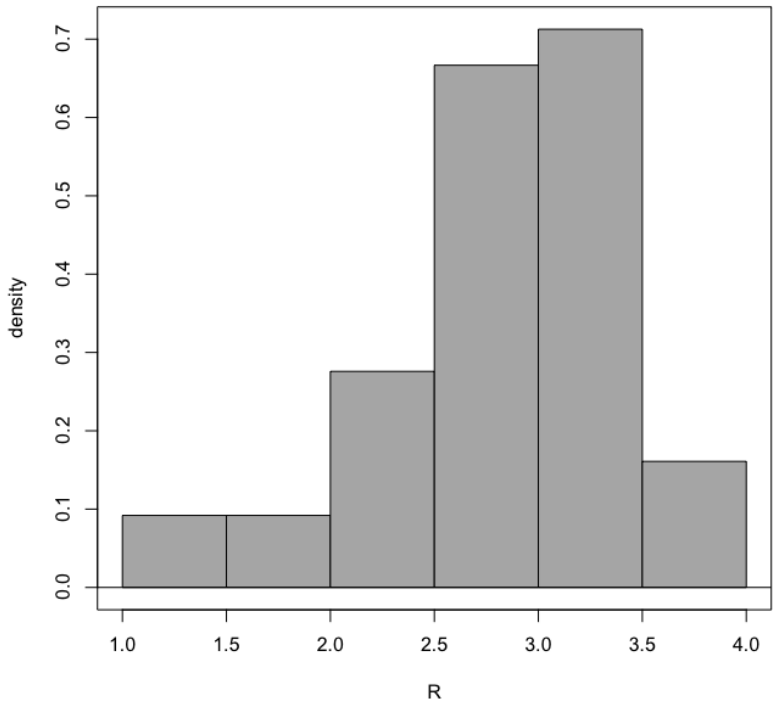


Figure 95 Histogram Risk

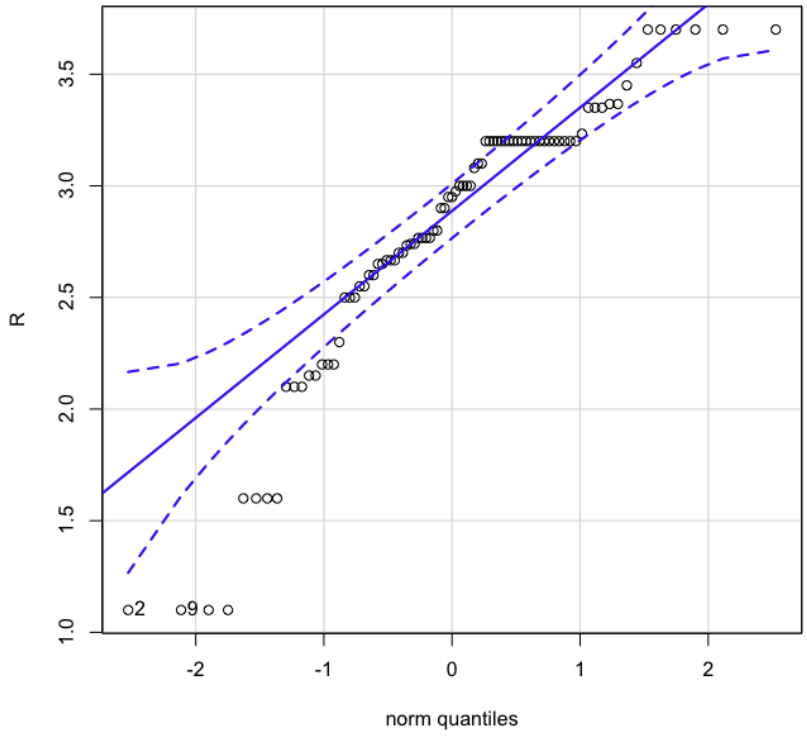


Figure 96 Q-Q plot Risk

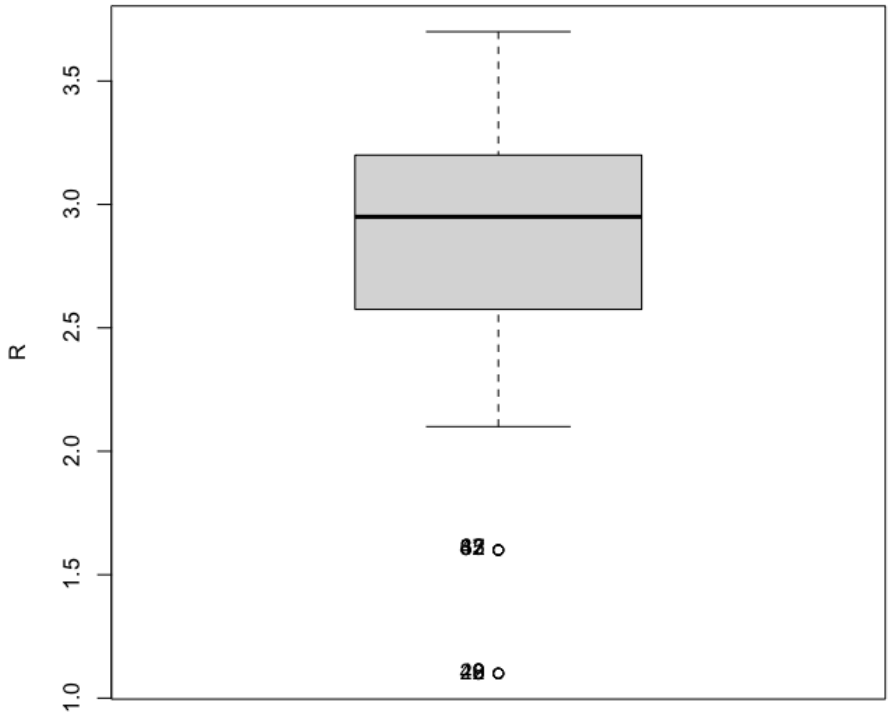


Figure 97 Box plot Risk

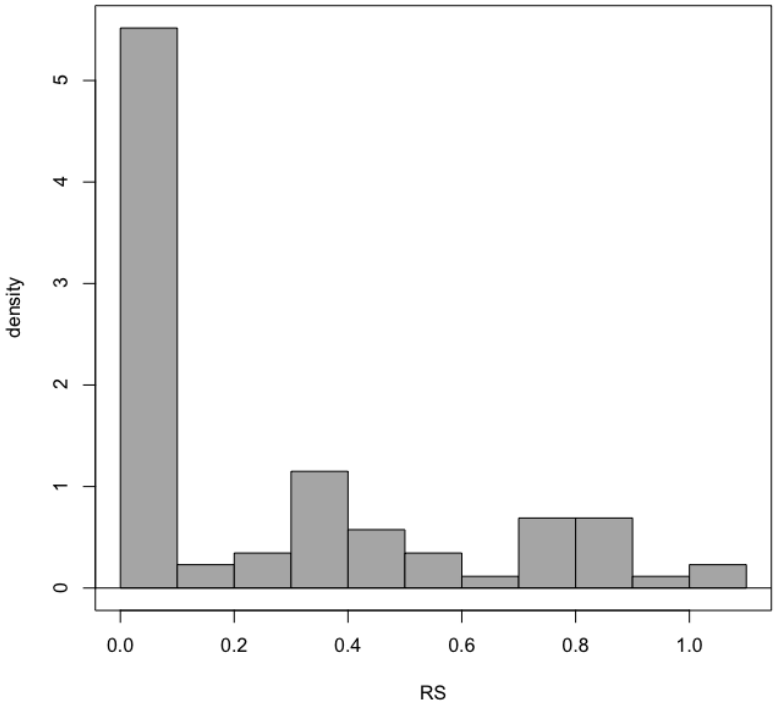


Figure 98 Histogram Risk Std Deviation

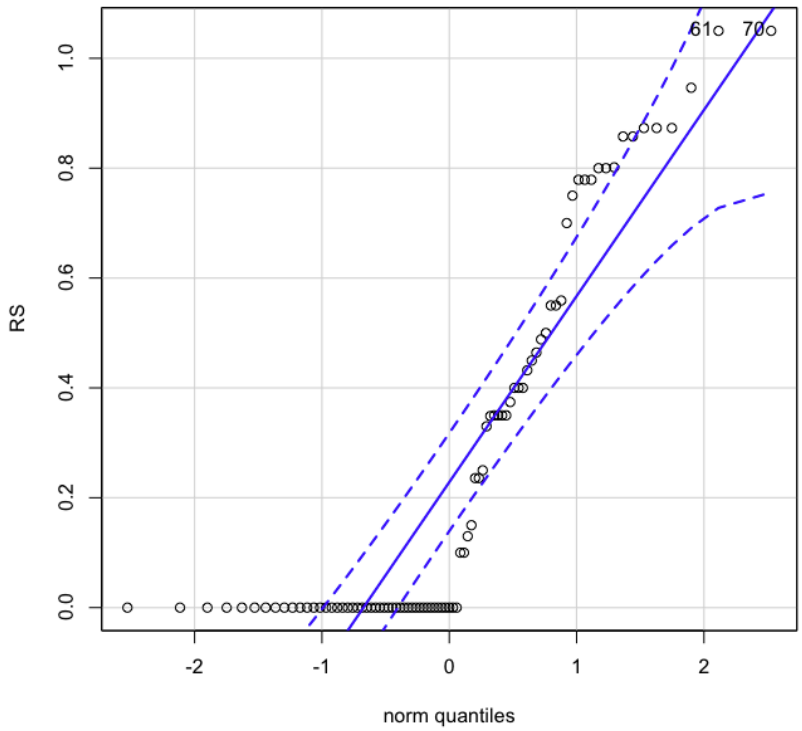


Figure 99 Q-Q plot Risk Std Deviation

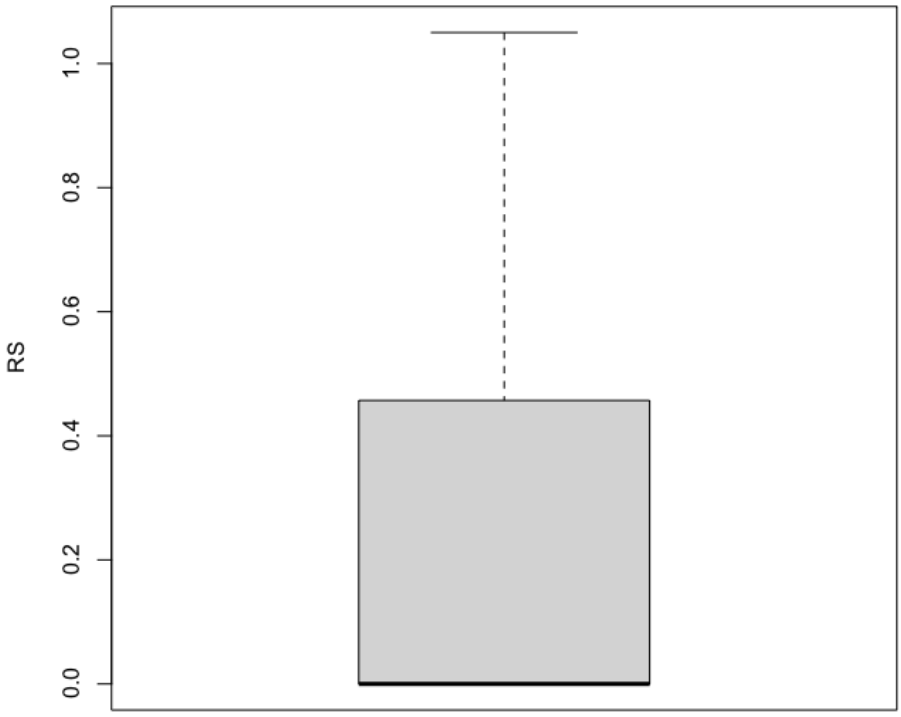


Figure 100 Box plot Risk Std Deviation

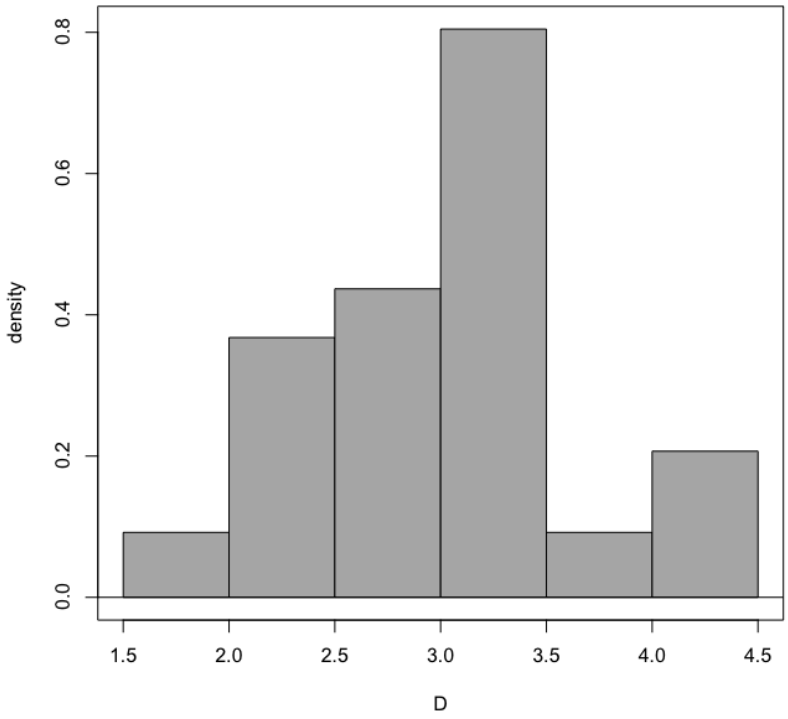


Figure 101 Histogram Directness

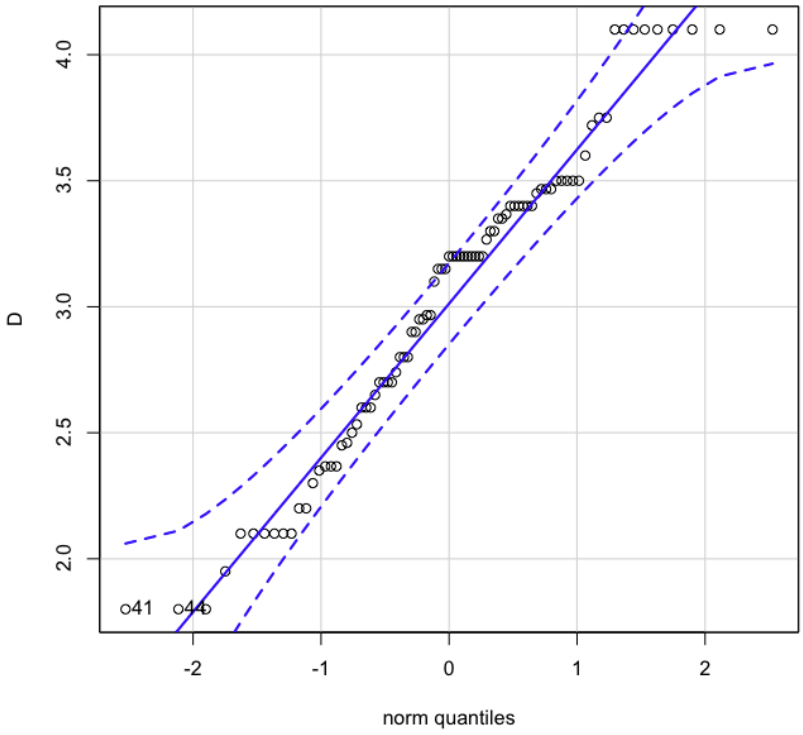


Figure 102 Q-Q plot Directness

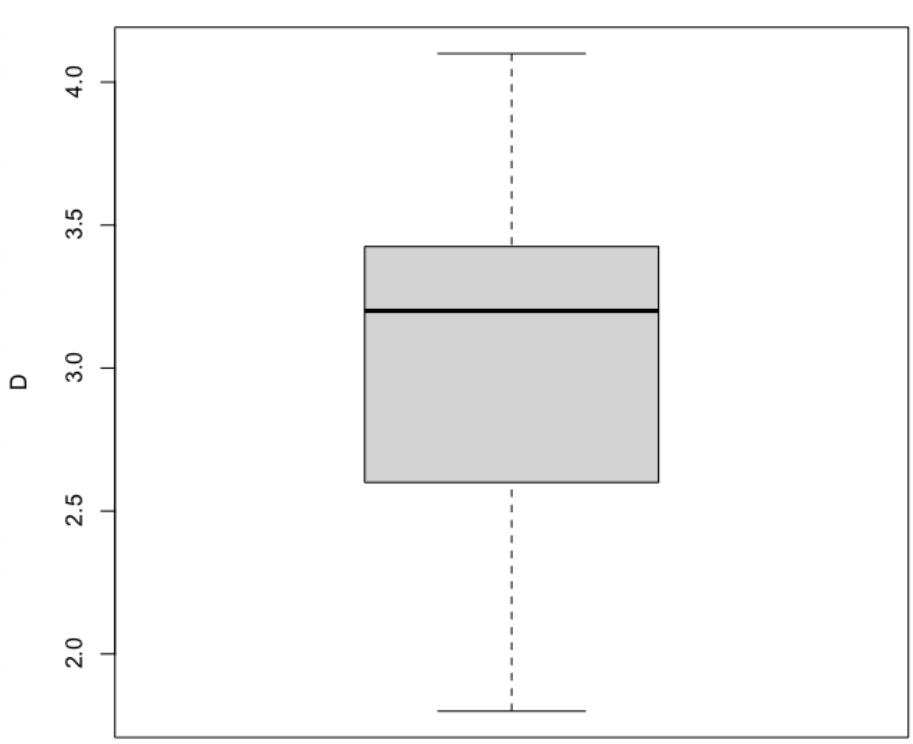


Figure 103 Box plot Directness

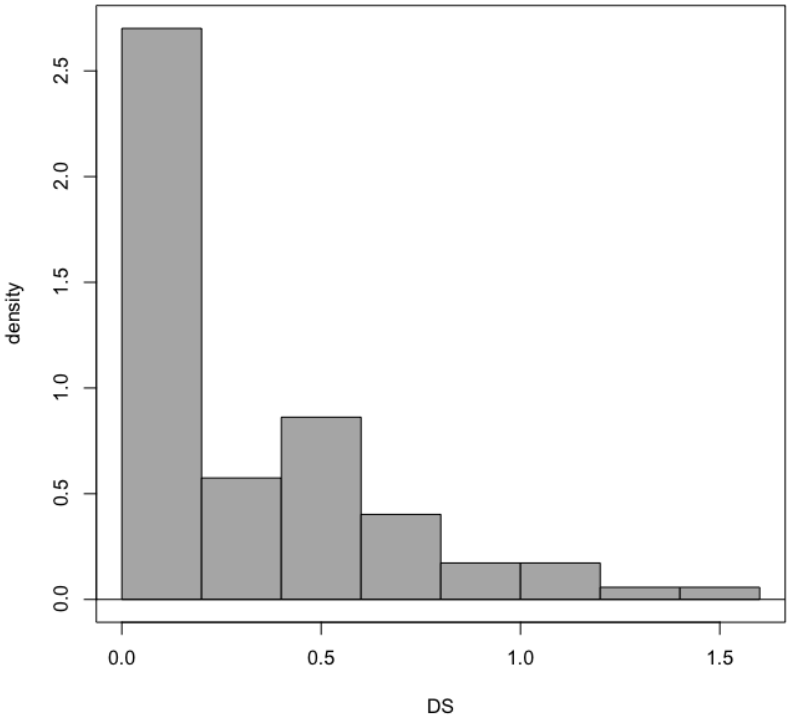


Figure 104 Histogram Directness Std Deviation

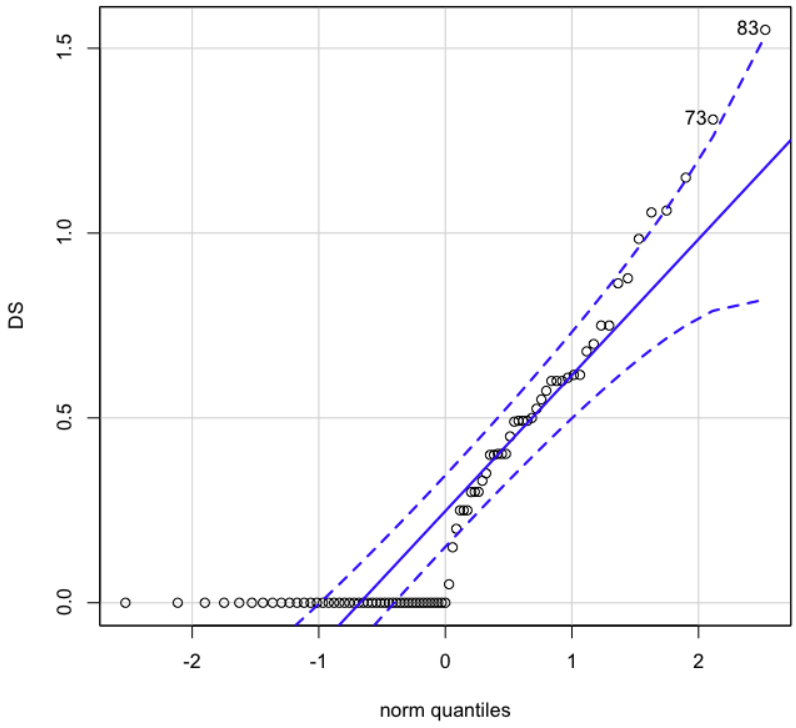


Figure 105 Q-Q plot Directness Std Deviation

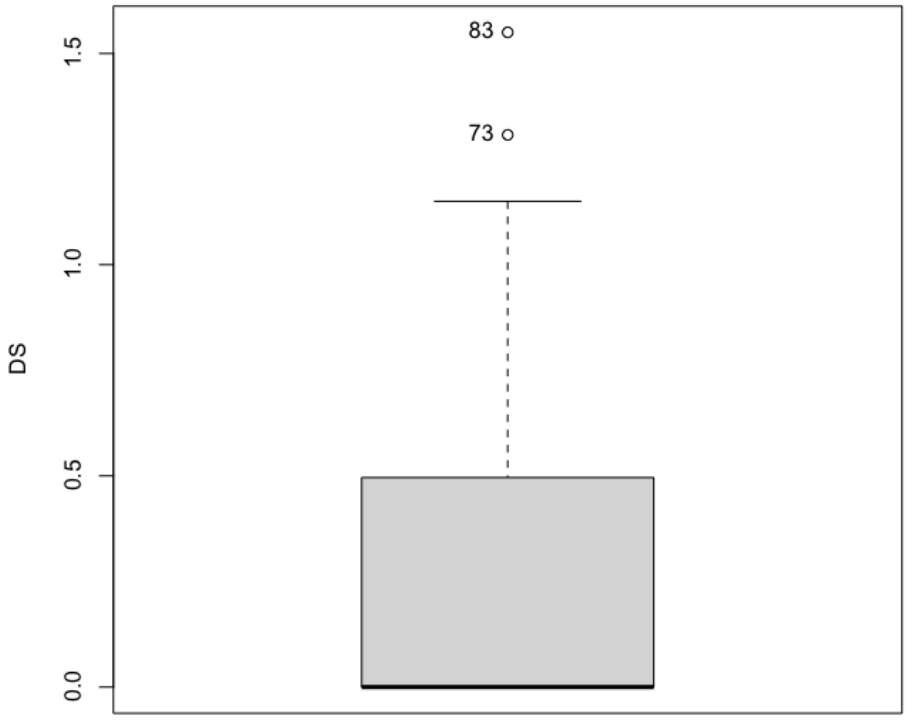


Figure 106 Box plot Directness Std Deviation

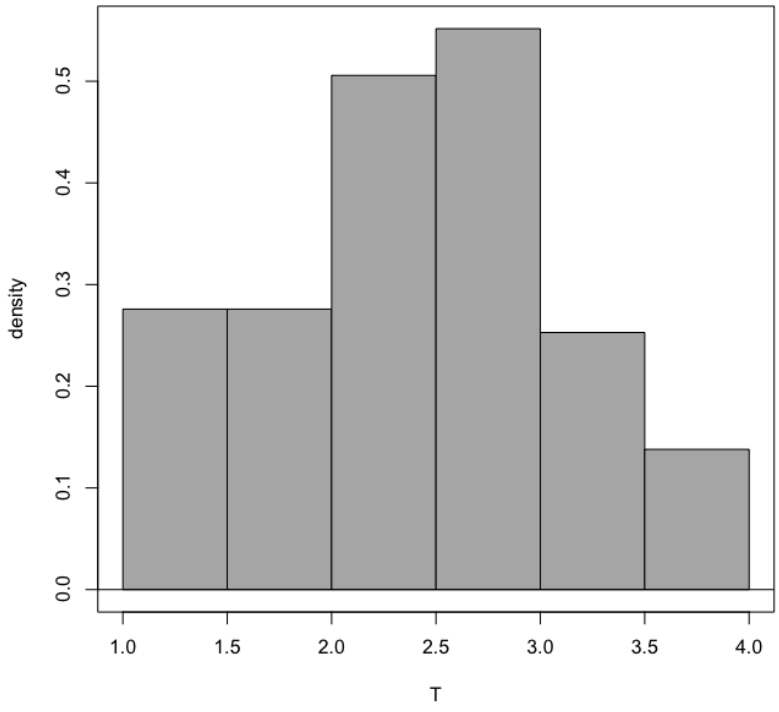


Figure 107 Histogram Task

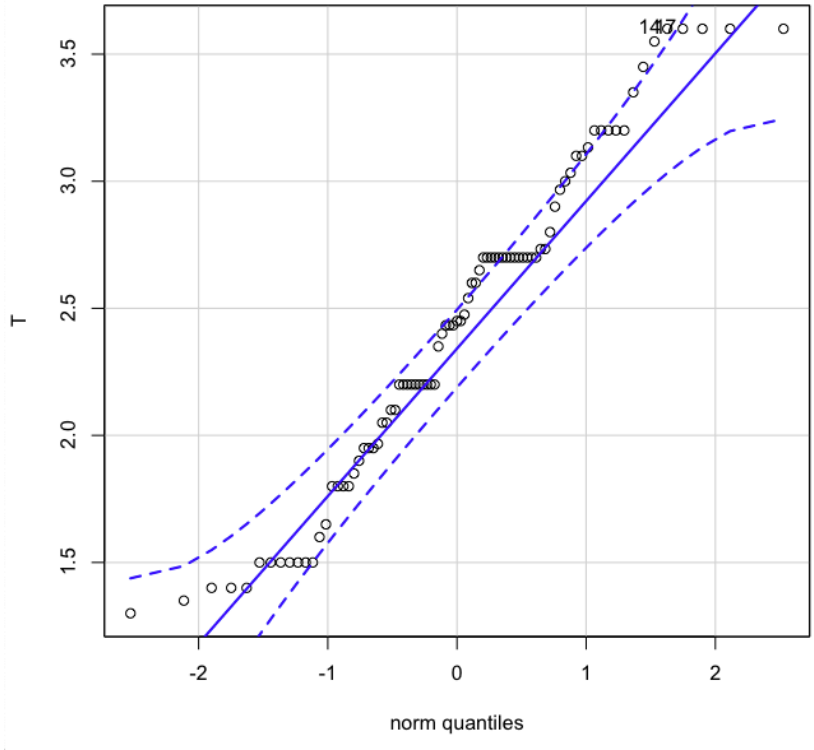


Figure 108 Q-Q Task

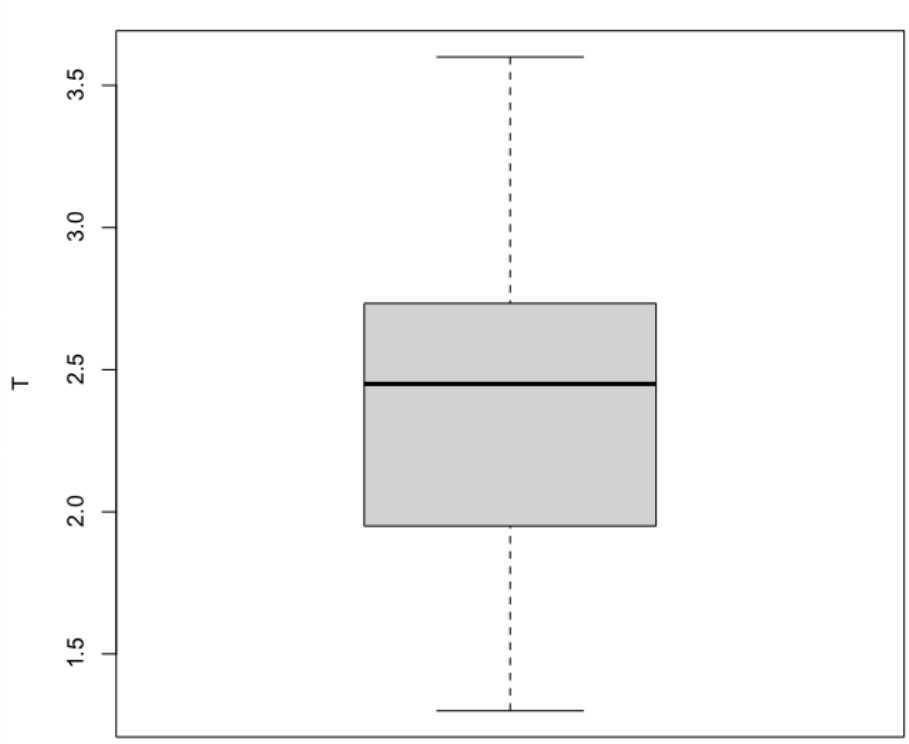


Figure 109 Box plot Task

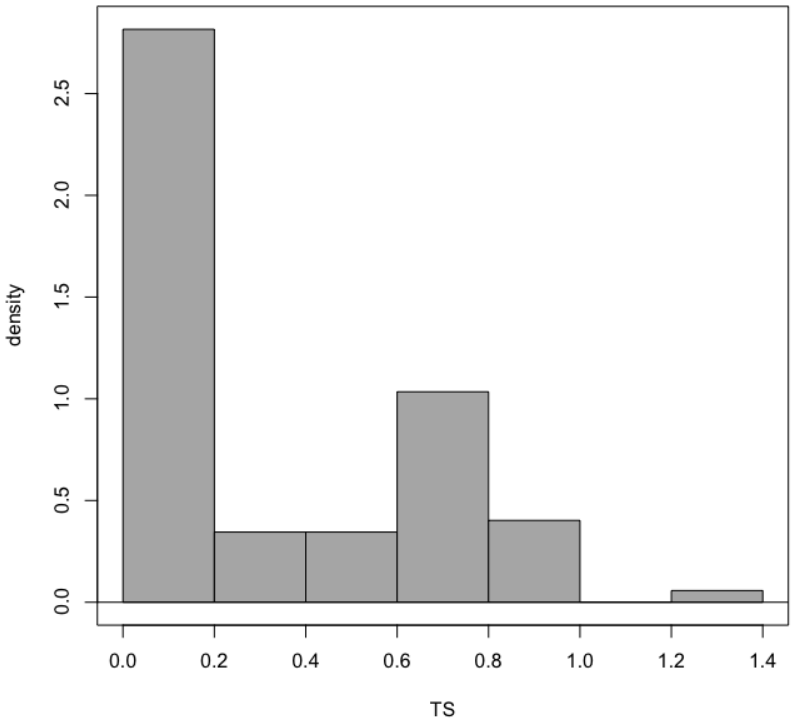


Figure 110 Histogram Task Std Deviation

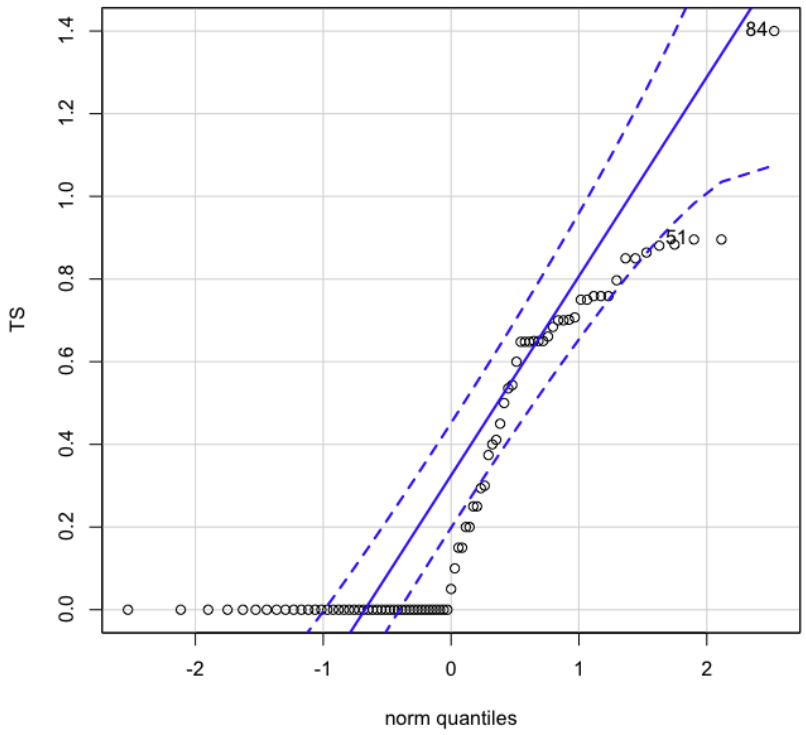


Figure 111 Q-Q plot Task Std Deviation

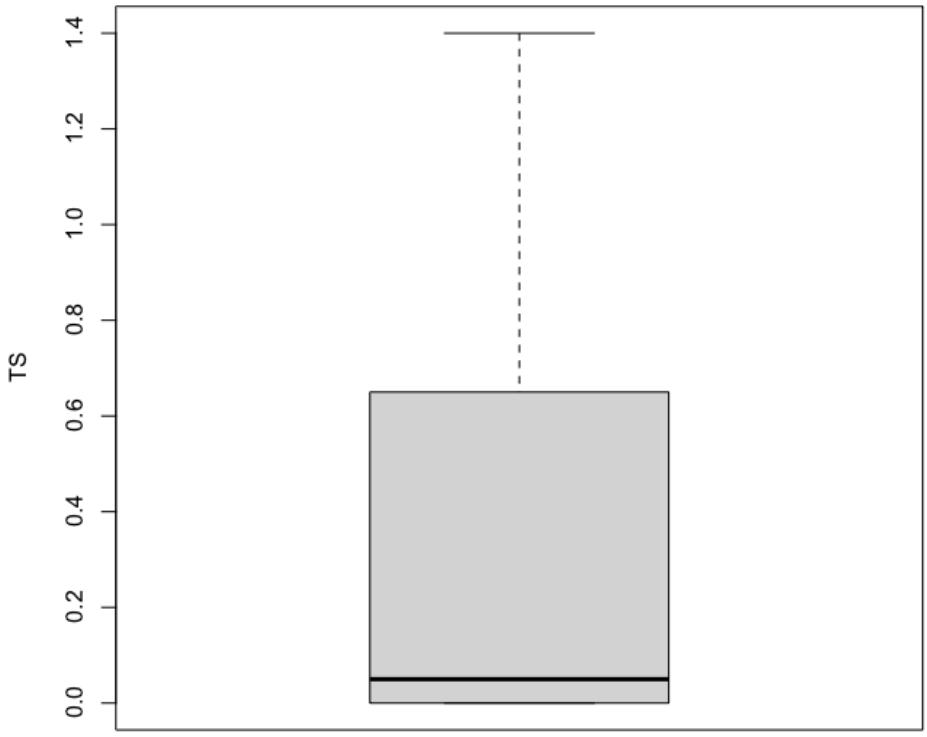


Figure 112 Box plot Task Std Deviation

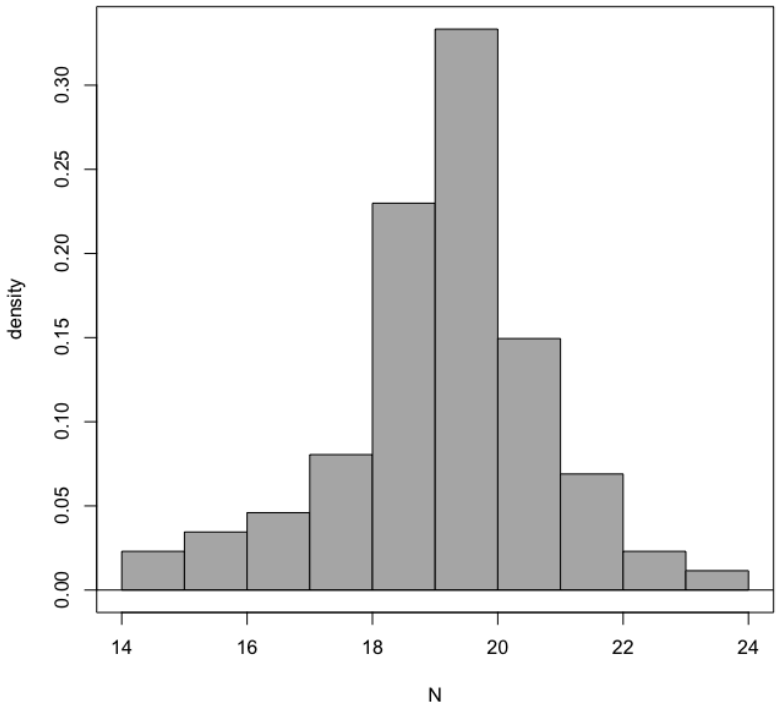


Figure 113 Histogram Neuroticism

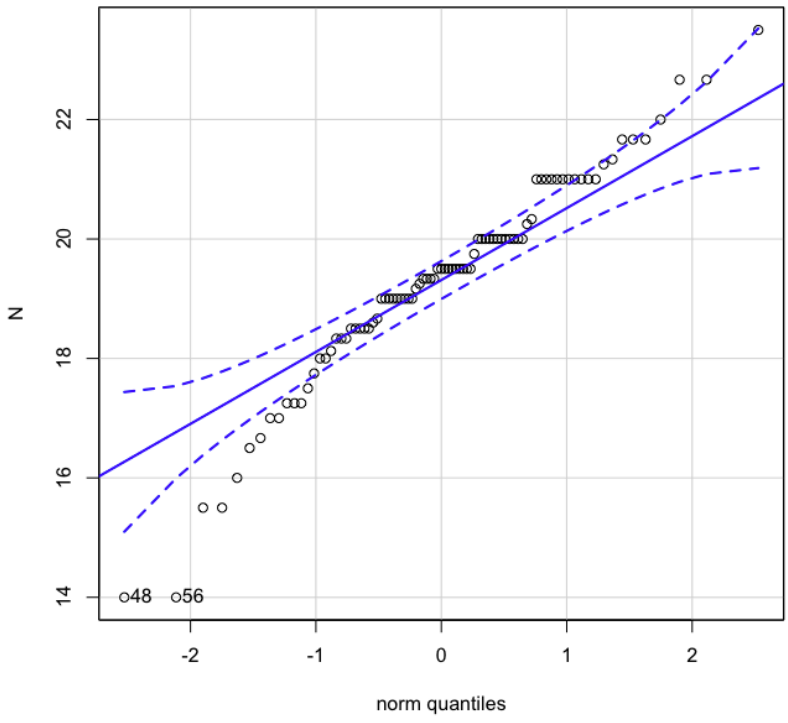


Figure 114 Q-Q plot Neuroticism

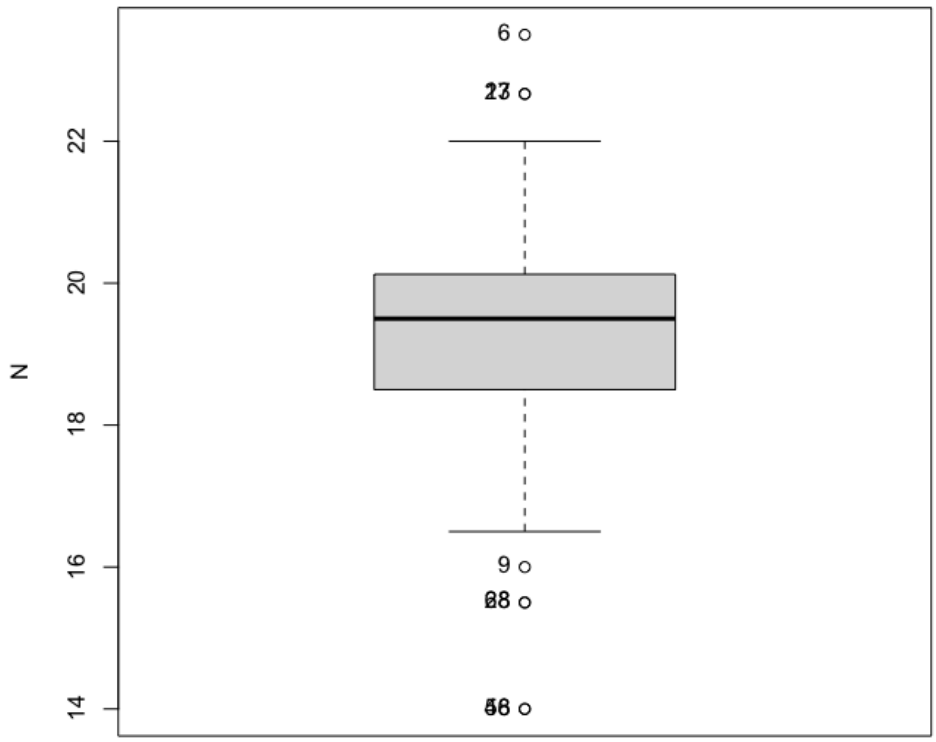


Figure 115 Box plot Neuroticism

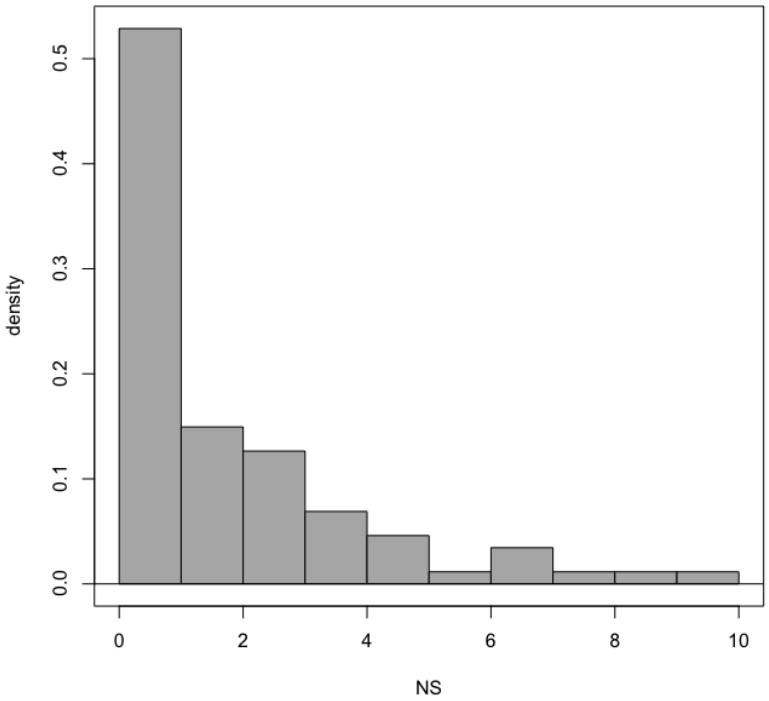


Figure 116 Histogram Neuroticism Std Deviation

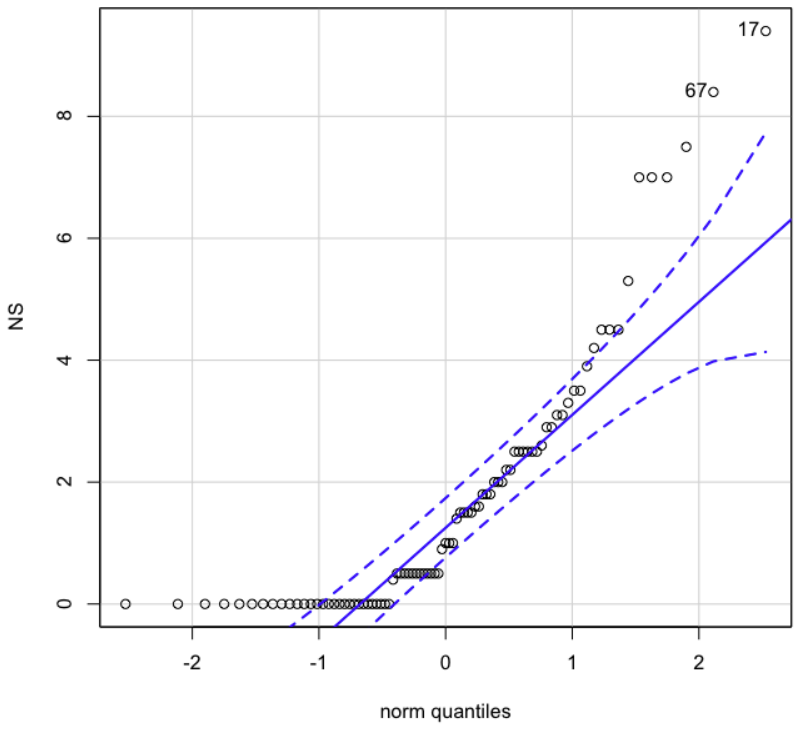


Figure 117 Q-Q plot Neuroticism Std Deviation

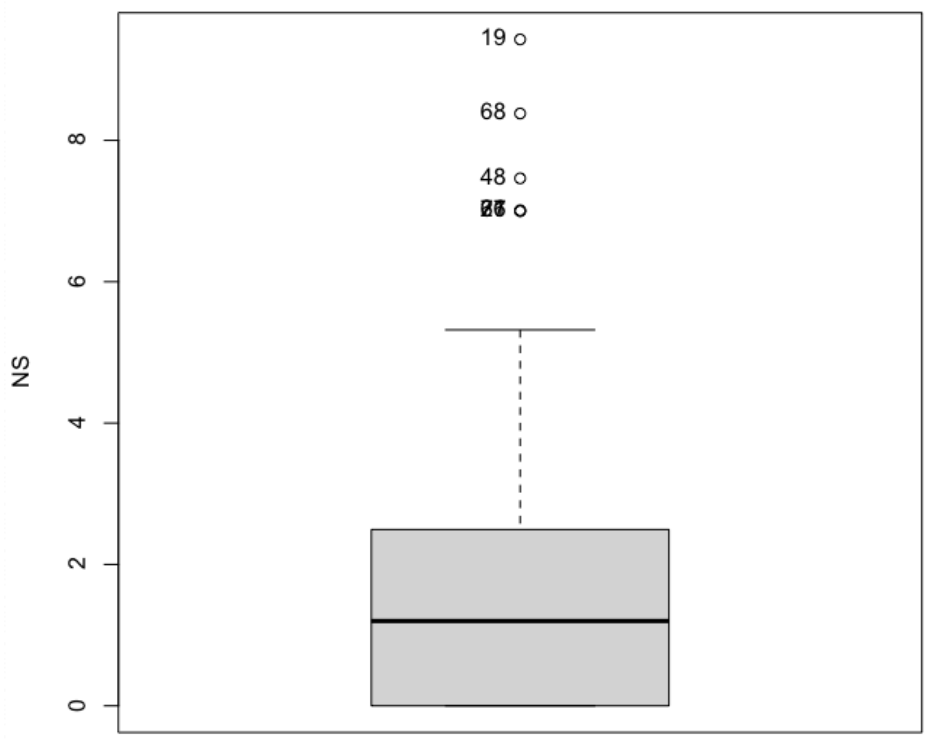


Figure 118 Box plot Neuroticism Std Deviation

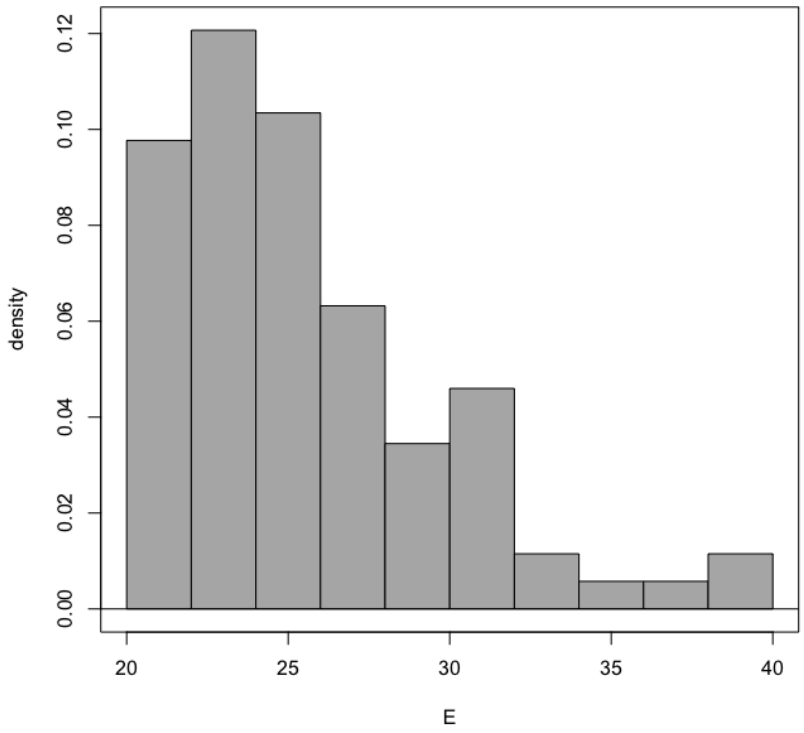


Figure 119 Histogram Extraversion

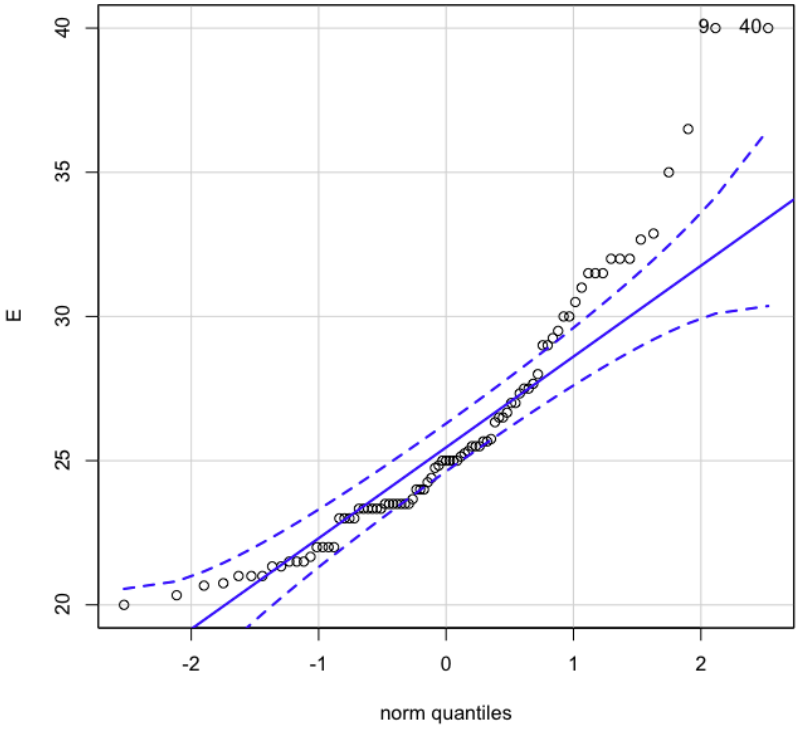


Figure 120 Q-Q Plot Extraversion

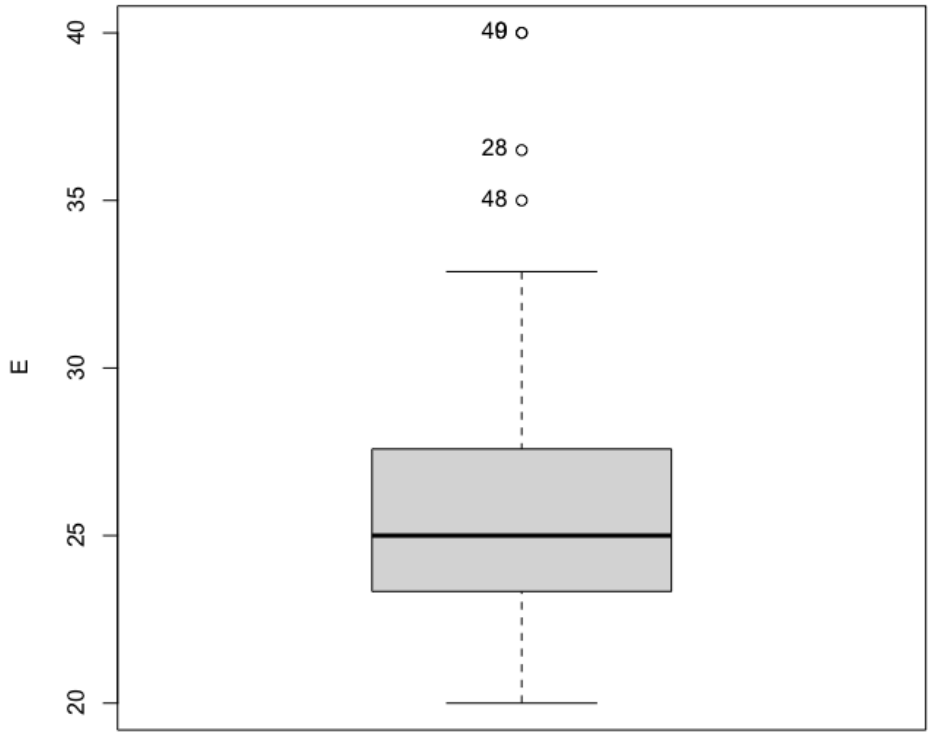


Figure 121 Box Plot Extraversion

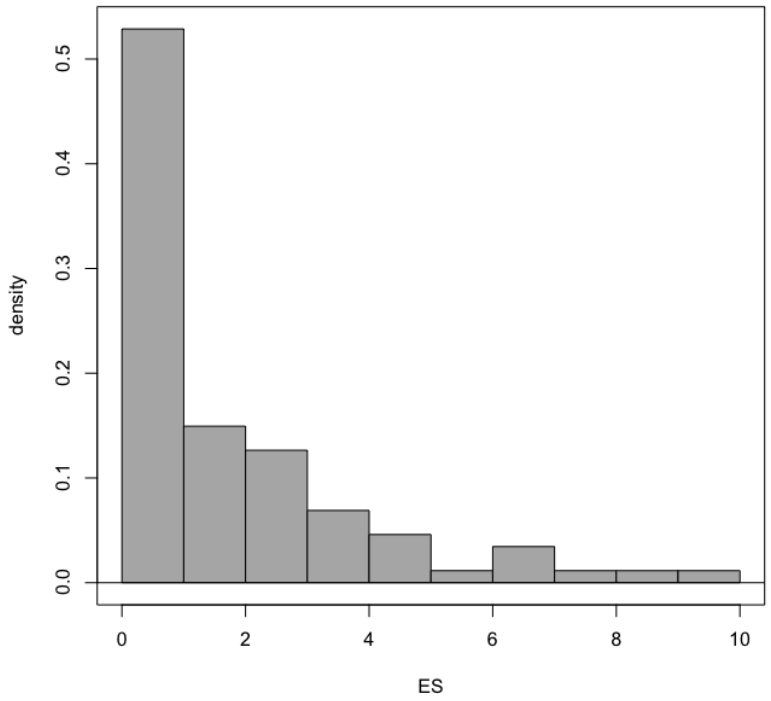


Figure 122 Histogram Extraversion Std Deviation

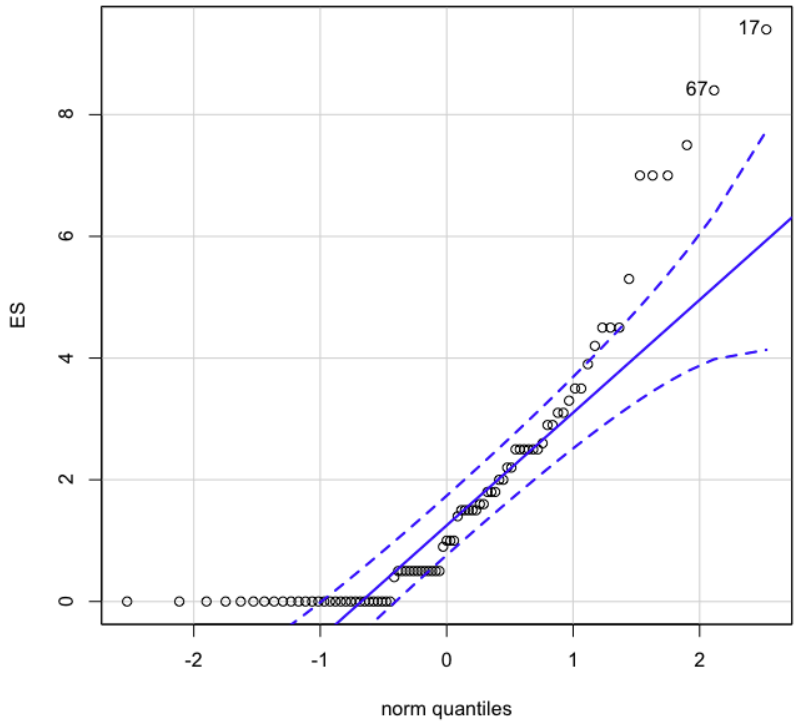


Figure 123 Q-Q Plot Extraversion Std Deviation

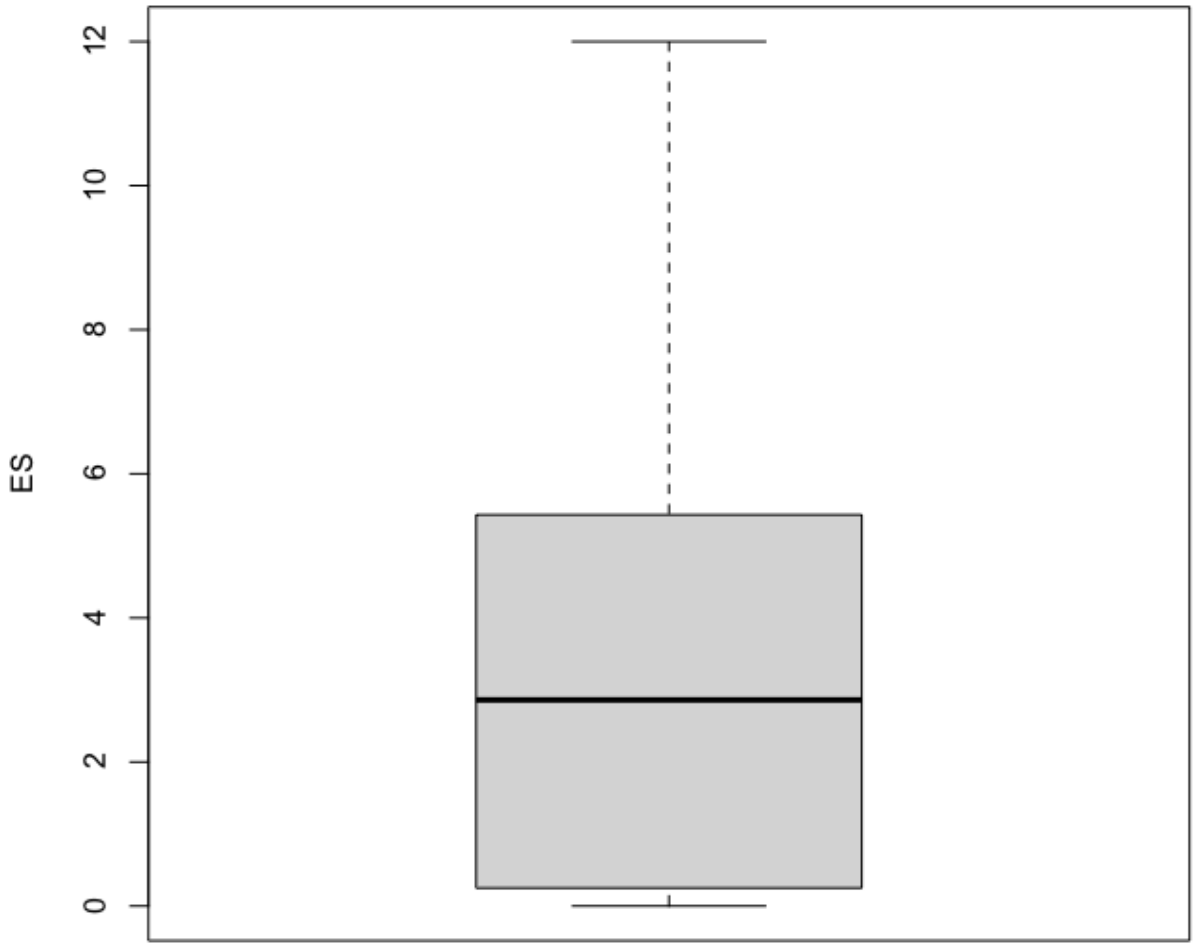


Figure 124 Box plot Extraversion Std Deviation

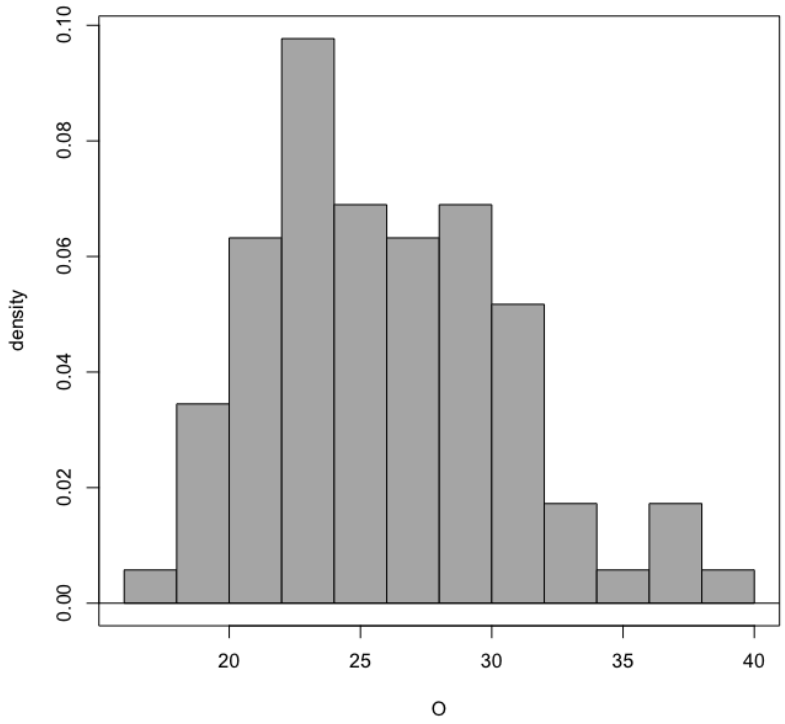


Figure 125 Histogram Openness

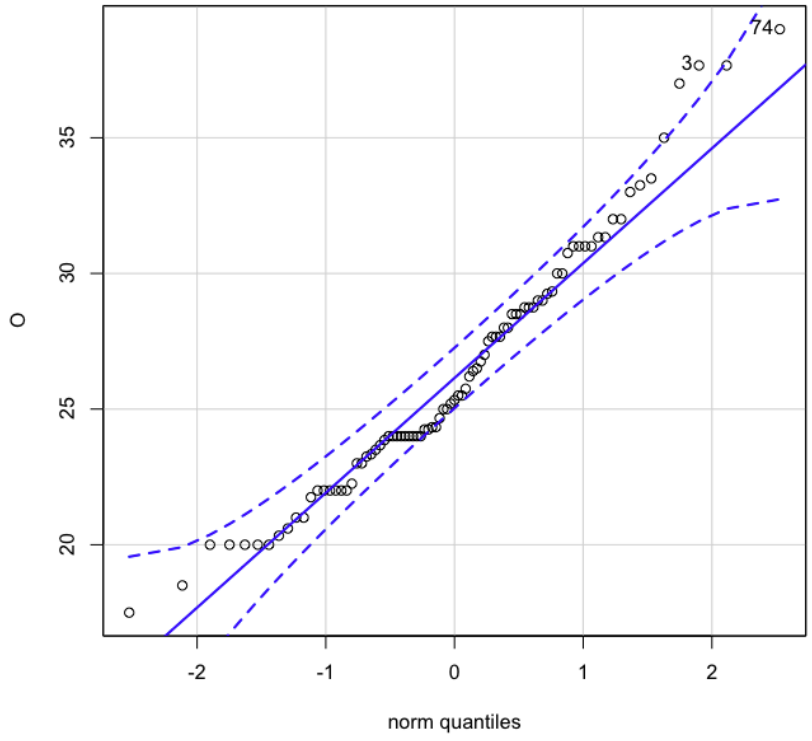


Figure 126 Q-Q Plot Openness

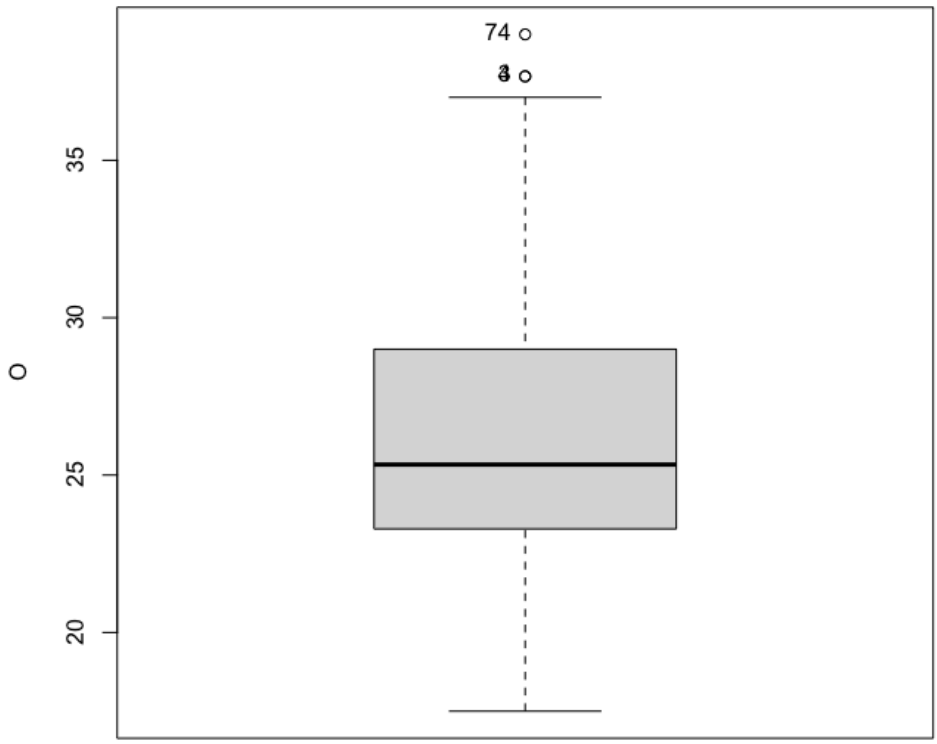


Figure 127 Box Plot Openness

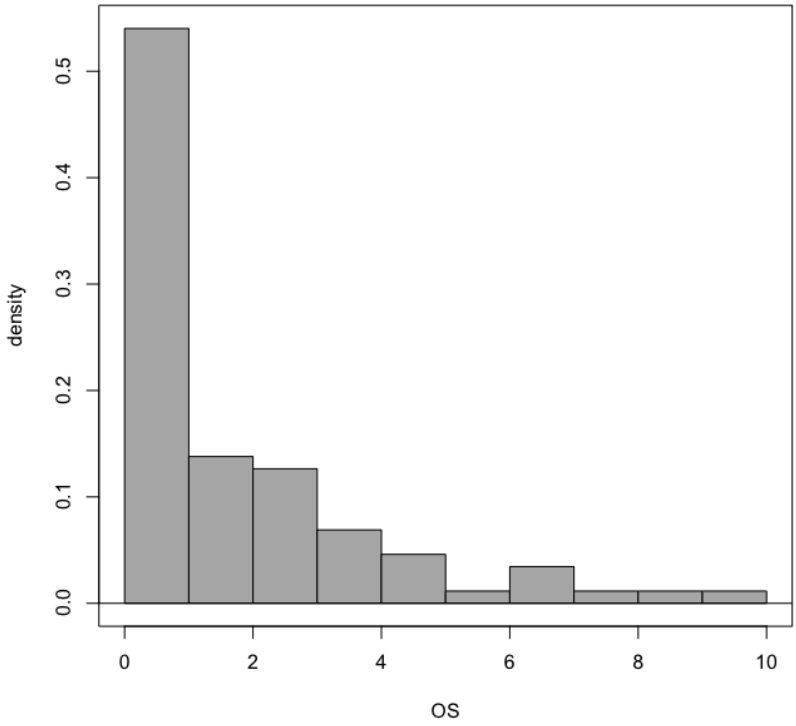


Figure 128 Histogram Openness Std Deviation

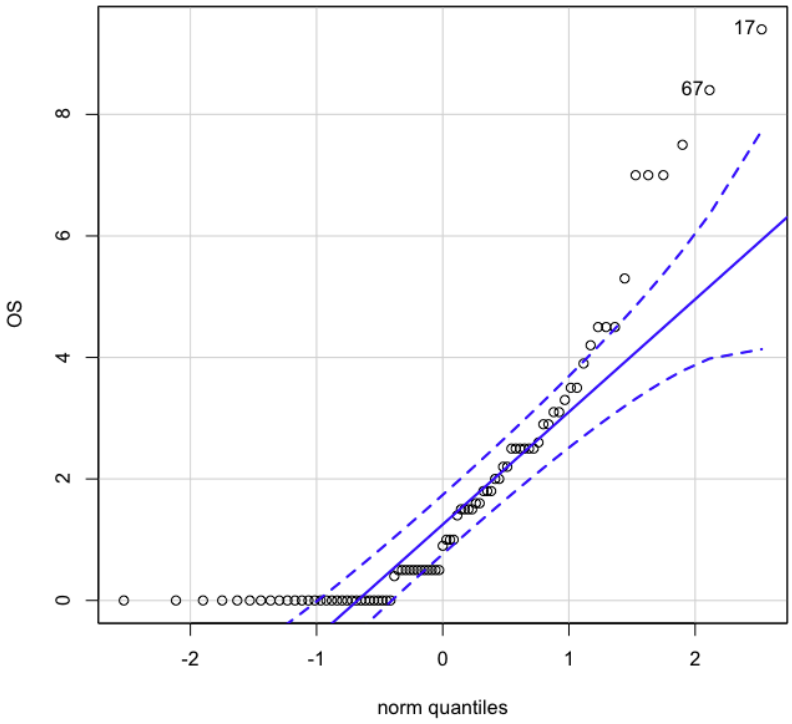


Figure 129 Q-Q Plot Openness Std Deviation

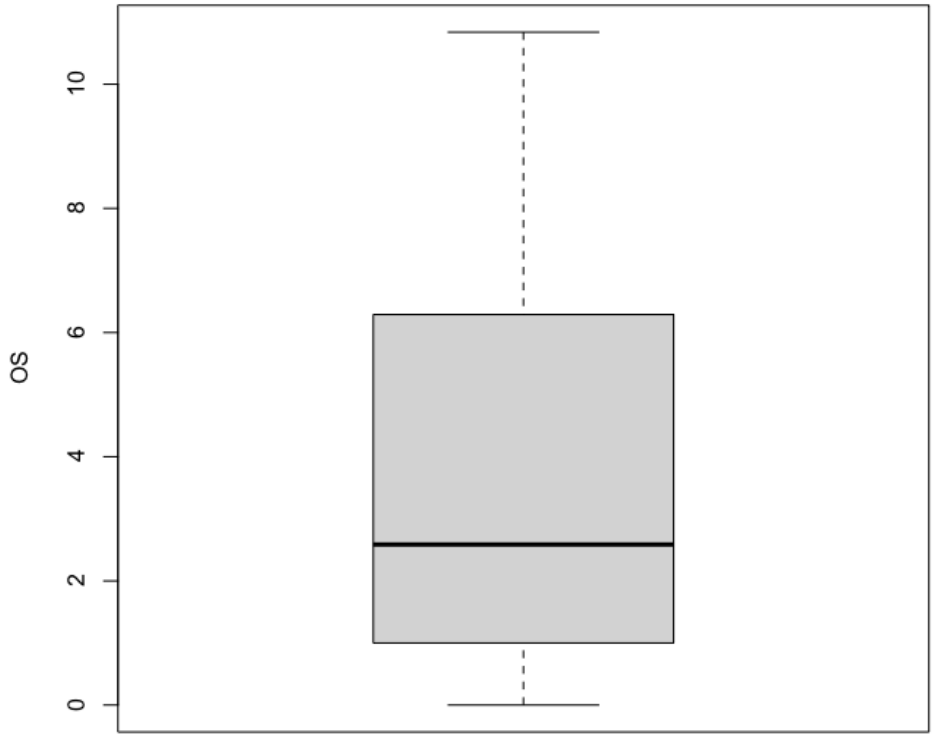


Figure 130 Box plot Openness Std Deviation

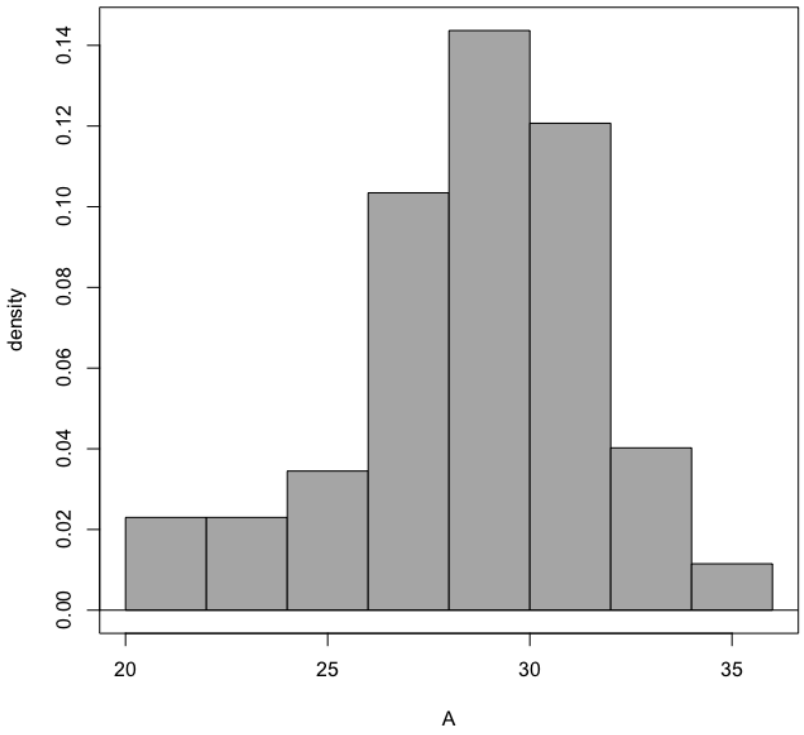


Figure 131 Histogram Agreeability

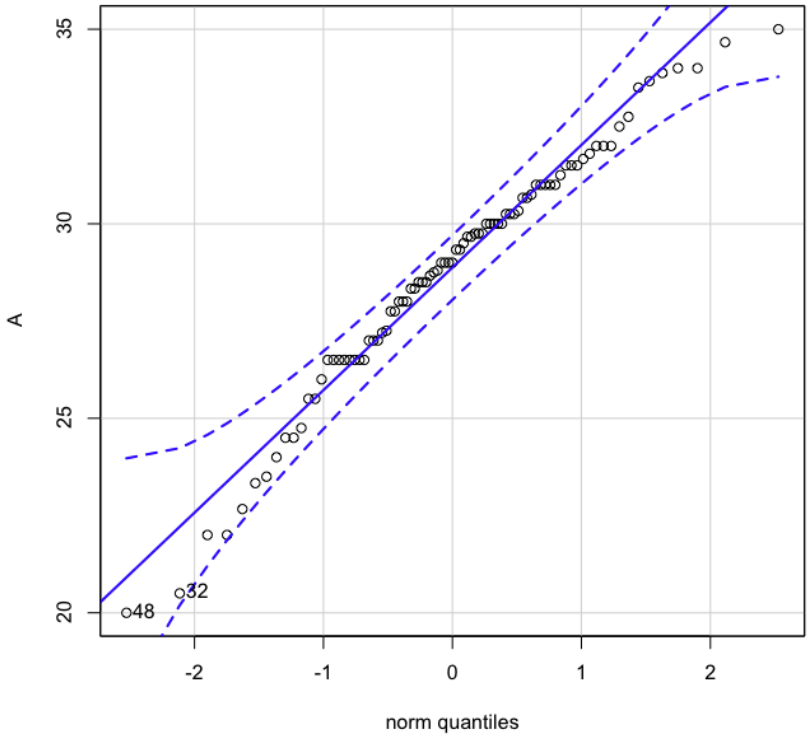


Figure 132 Q-Q Plot Agreeabiliy

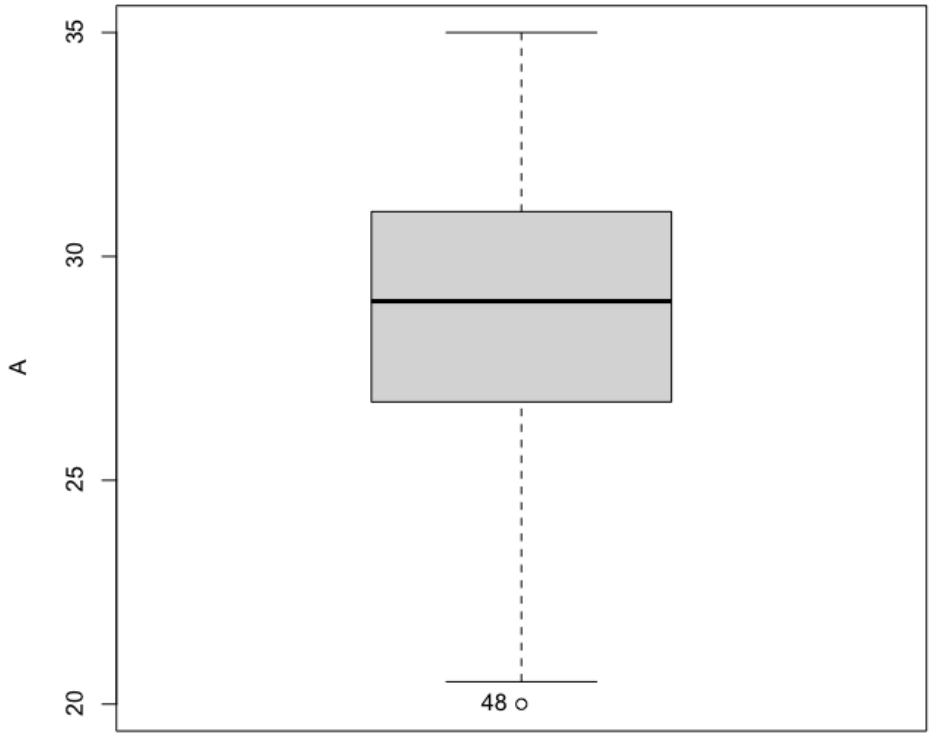


Figure 133 Box Plot Agreeabiliy

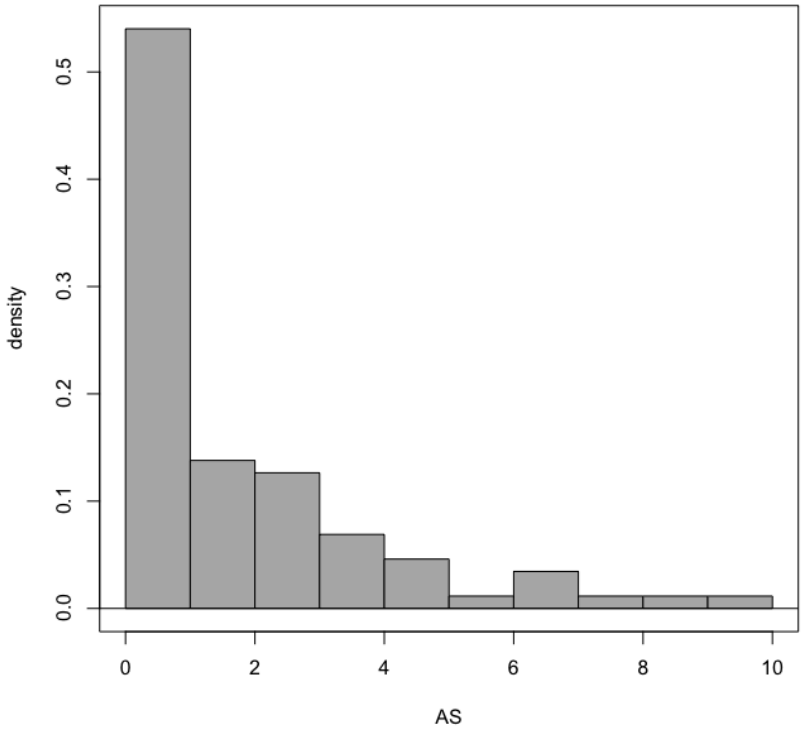


Figure 134 Histogram Agreeabiliy Std Deviation

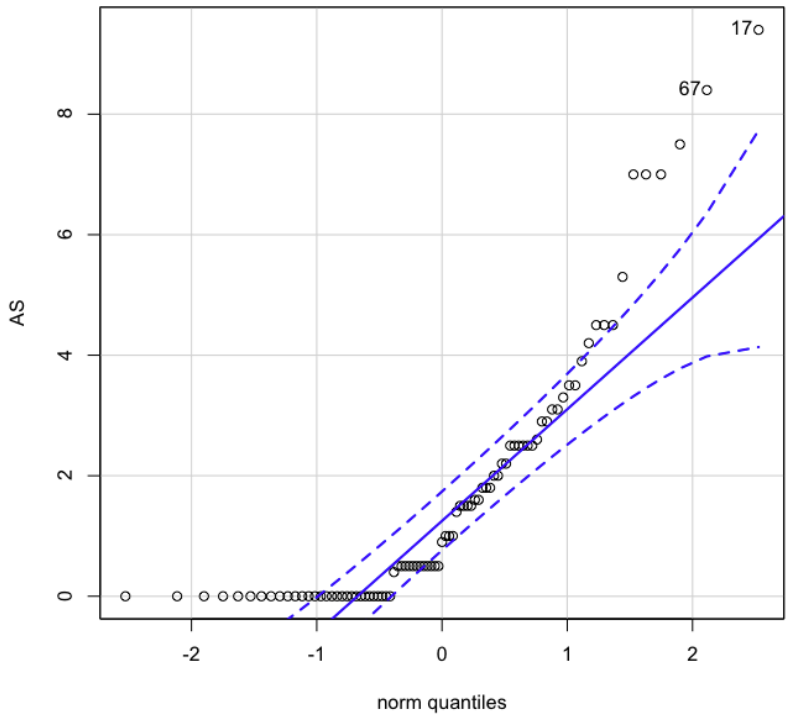


Figure 135 Q-Q Plot Agreeabiliy Std Deviation

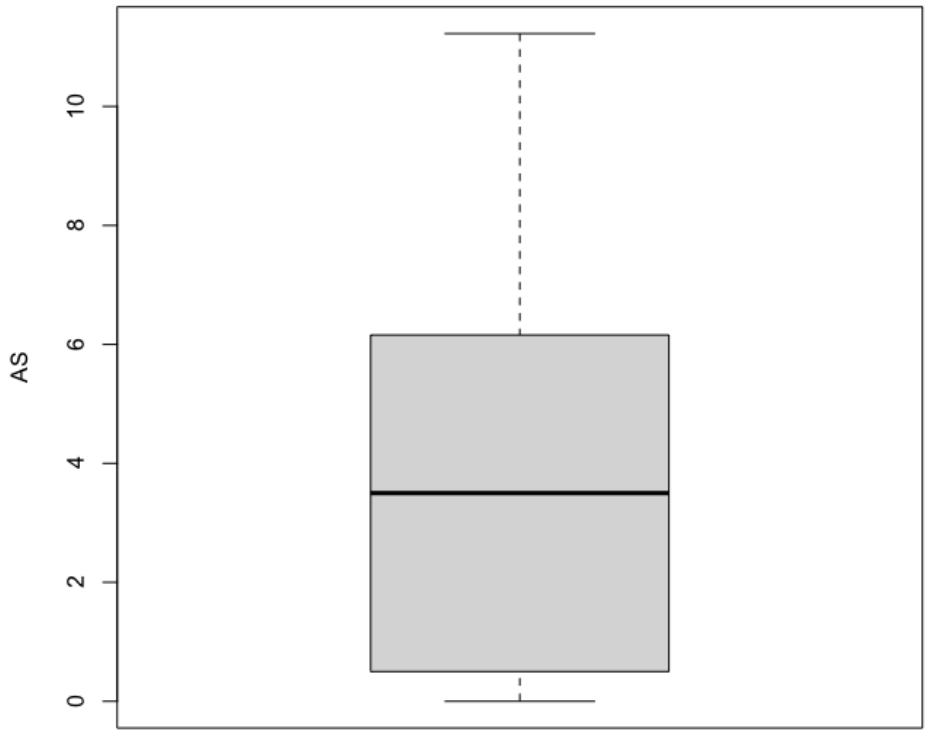


Figure 136 Box Plot Agreeabiliy Std Deviation

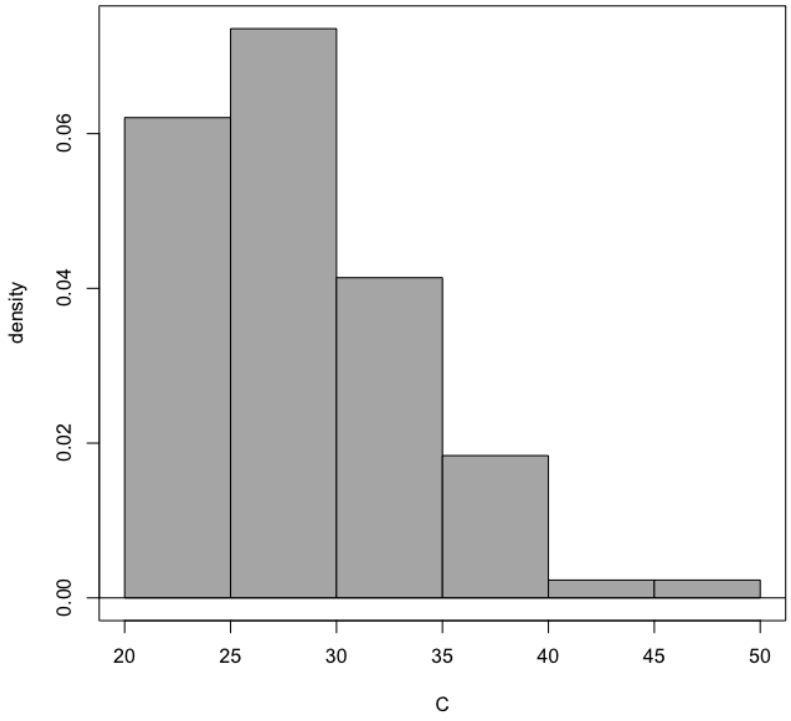


Figure 137 Histogram Consientousness

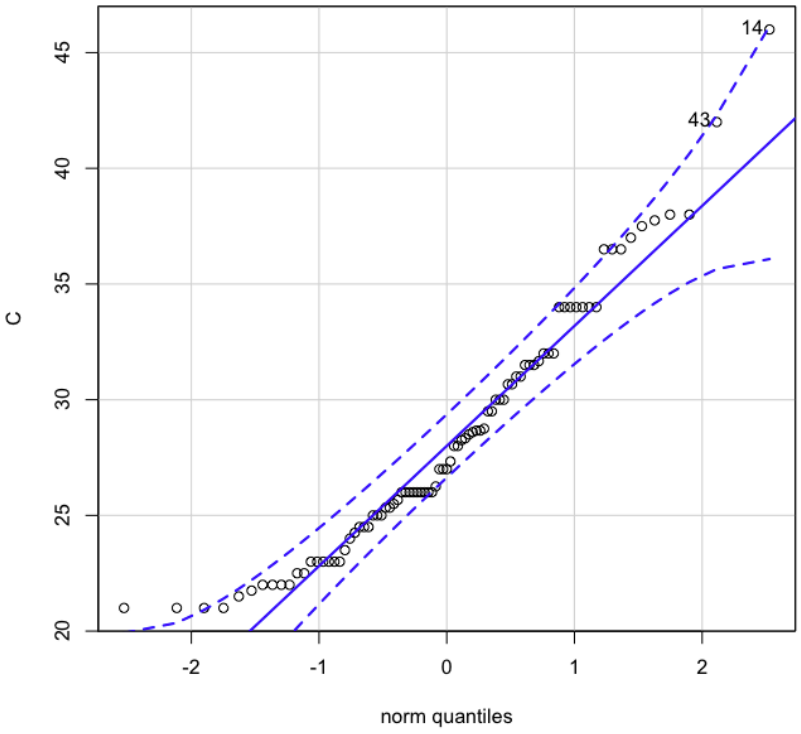


Figure 138 Q-Q Plot Consientousness

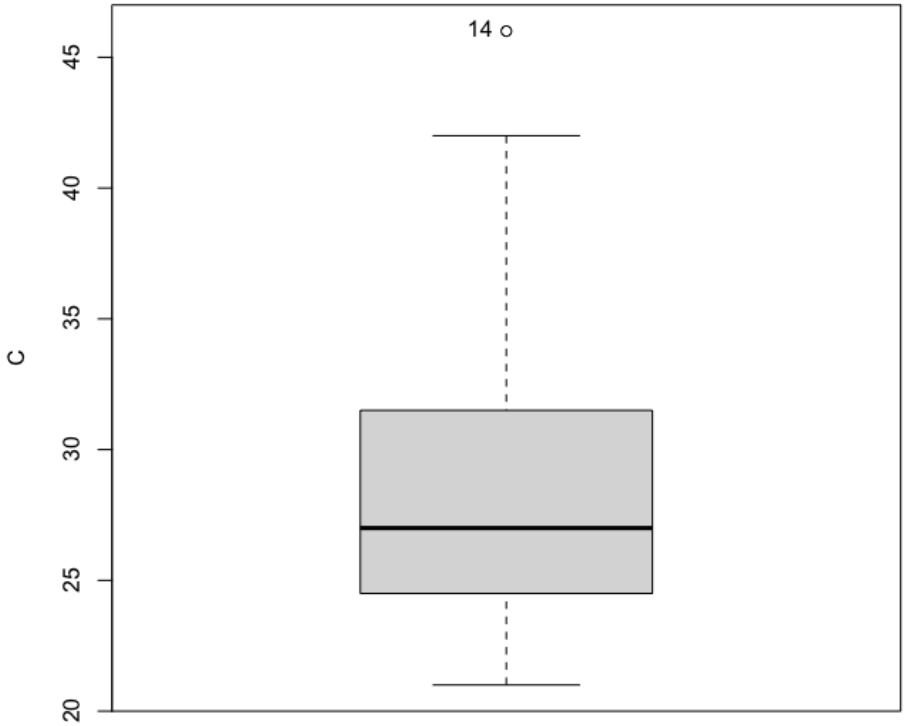


Figure 139 Box Plot Consientousness

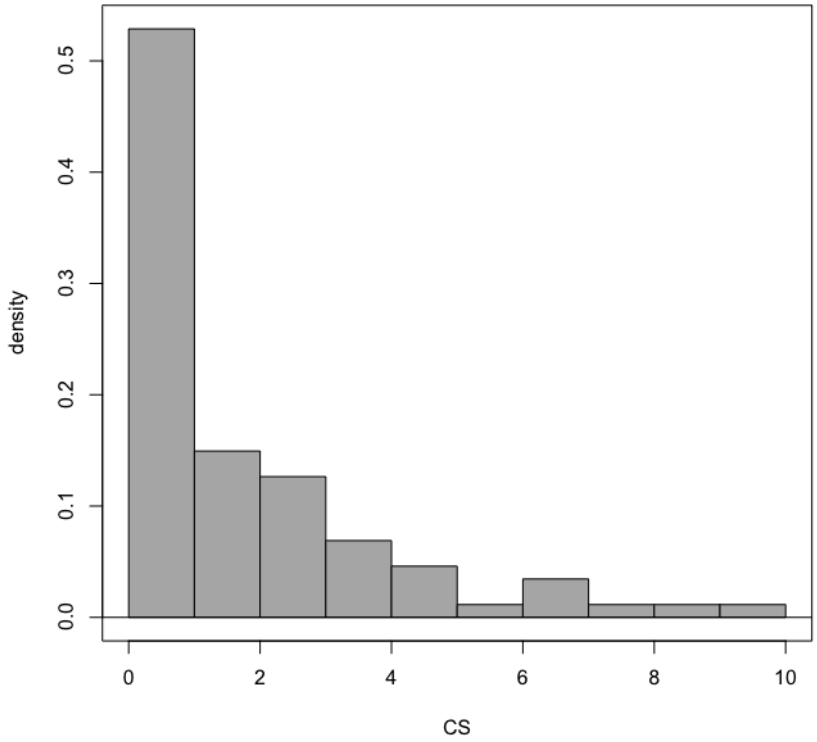


Figure 140 Histogram Consientousness Std Deviation

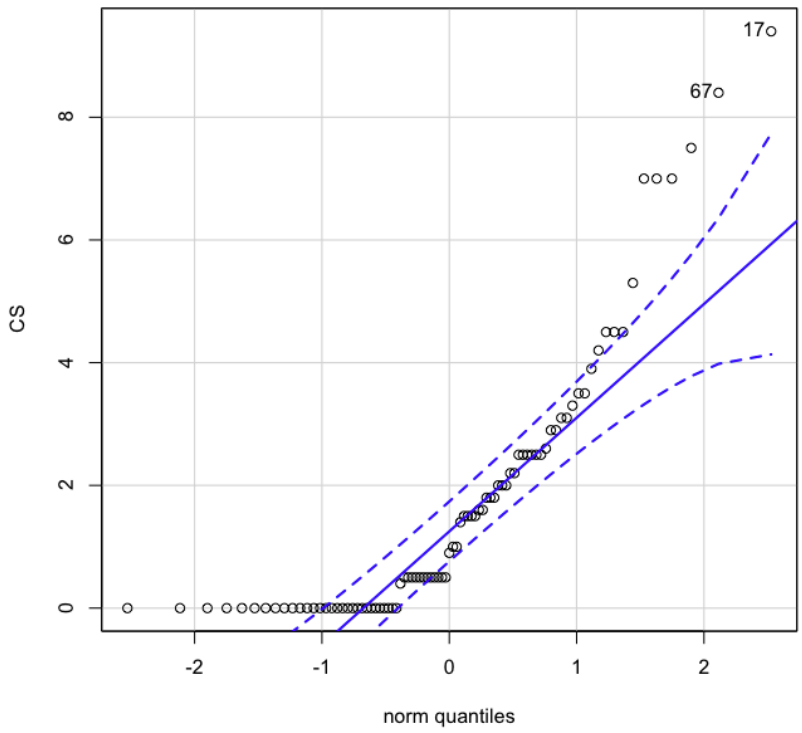


Figure 141 Q-Q Plot Consientousness Std Deviation

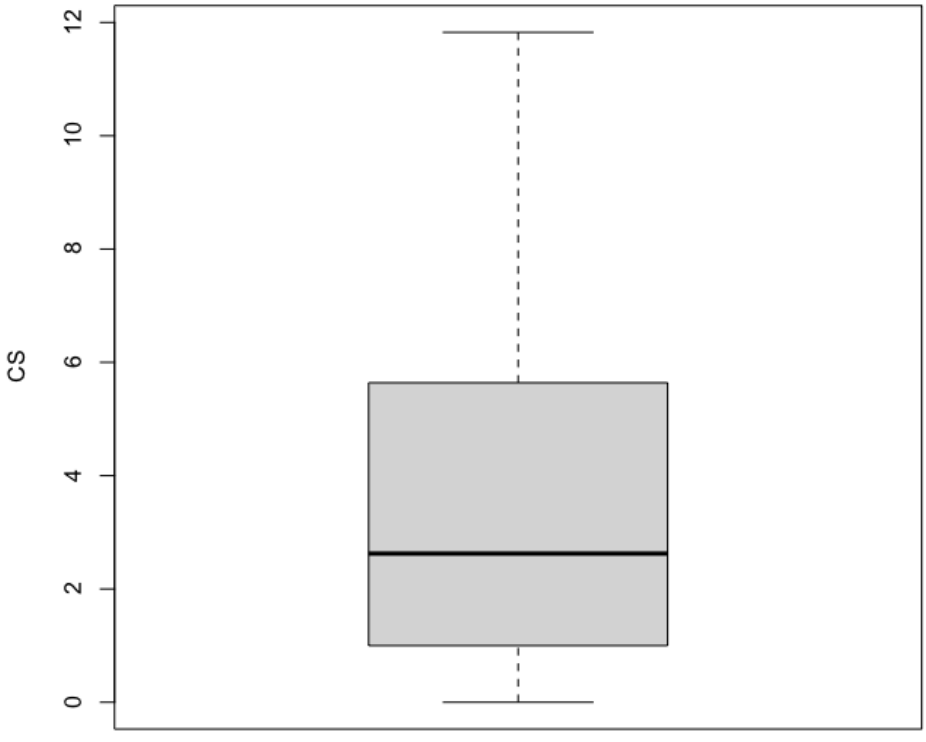


Figure 142 Box Plot Consientousness Std Deviation

C.3 SHAPIRO-WILK NORMALITY TEST

data: PQ1

W = 0.78976, p-value = 2.023e-11

data: PQ2

W = 0.84419, p-value = 0.000000001484

data: PQ3

W = 0.82682, p-value = 3.88e-10

data: PRO_Exp

W = 0.94189, p-value = 0.00009542

data: PRO_Team

W = 0.96264, p-value = 0.003034

C.4 WILCOX SIGNED RANK TEST ON RAW DATA

data: PQ1

V = 6216, p-value < 2.2e-16

data: PQ2

V = 6328, p-value < 2.2e-16

data: PQ3

V = 6328, p-value < 2.2e-16

data: PRO_Exp

V = 6441, p-value < 2.2e-16

data: PRO_Team

V = 6441, p-value < 2.2e-16

C.5 DESCRIPTIVE STATISTICS DATA CONSOLIDATED ON PROJECT LEVEL N=115

	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
E	16	47	24.50	6.193	38.349	1.603	.357
O	17	43	25.63	6.161	37.953	1.091	.361
A	13	41	28.38	5.642	31.831	.060	.354
C	20	46	27.71	6.595	43.488	1.388	.343
N	7	27	19.37	3.687	13.590	-.733	.383
independent	1.7	4.1	3.007	.6516	.425	-.040	.448
egalitarism	1.2	4.0	2.730	.5999	.360	-.245	.448
risk	1.1	4.0	2.804	.7320	.536	-.834	.448
direct	1.2	4.9	2.952	.8059	.650	.039	.448
task	1.2	4.3	2.781	.7716	.595	-.143	.448
birthdate	1949	1993	1977.94	9.503	90.309	-.757	.297
age	23	67	38.06	9.503	90.309	.757	.297
startdate	1988	2016	2009.68	4.773	22.785	-2.139	.297
grade	1	6	3.54	1.105	1.221	.007	.297

Table 22 Descriptive statistic table form raw data

C.6 R_{WG} AND ICC CALCULATIONS FOR PEER DATA

grpid	rwg.j	gsize
Length:105	Min. :0.7914	Min. :0.000
Class :character	1st Qu.:0.9707	1st Qu.:2.000
Mode :character	Median :0.9872	Median :3.000
	Mean :0.9738	Mean :3.267
	3rd Qu.:0.9916	3rd Qu.:4.000
	Max. :1.0000	Max. :9.000
	NA's :10	

Table 23 r_{wg} Results Statistics

	grpid	rwg.j
1	1	0.962
10	10	0.9581
100	100	0.982
101	101	0.979
102	102	0.989
103	103	0.9915
104	104	0.9719
105	105	0.985
106	106	0.991
107	107	0.966
108	108	0.989
109	109	0.985
11	11	0.996
110	110	0.983
111	111	0.994
112	112	0.982
113	113	0.991
114	114	0.989
115	115	0.988
12	12	0.965
13	13	0.9433
14	14	0.982
15	15	0.989
16	16	0.929
17	17	0.995
18	18	0.993
19	19	0.917
2	2	0.991
20	20	0.985
21	21	0.994
22	22	0.791
24	24	0.977
25	25	0.997
26	26	0.982
29	29	0.991
30	30	0.993
31	31	0.991
32	32	0.973
33	33	0.983
35	35	0.996
36	36	0.989
37	37	0.975

	grpid	rwg.j
38	38	0.989
39	39	0.990
40	40	0.993
41	41	0.988
42	42	0.990
43	43	0.945
44	44	0.980
45	45	0.989
46	46	0.923
47	47	0.810
48	48	0.964
49	49	0.975
5	5	0.942
50	50	0.989
53	53	0.974
54	54	0.993
55	55	0.973
56	56	0.954
57	57	0.944
58	58	0.993
59	59	0.958
6	6	0.941
60	60	0.954
61	61	0.941
62	62	0.988
63	63	0.975
64	64	0.992
65	65	0.990
66	66	0.917
67	67	0.983
68	68	0.992
69	69	0.991
70	70	0.981
71	71	0.991
72	72	0.994
73	73	0.98
74	74	0.990
75	75	0.897
76	76	0.980
77	77	0.988
78	78	0.996
79	79	0.993

	grpid	rwg.j
8	8	0.979
80	80	0.998
81	81	0.999
82	82	0.988
83	83	0.917
85	85	0.994
86	86	0.987
87	87	0.987
88	88	0.956
89	89	0.978
9	9	0.975
90	90	0.979
91	91	0.970
92	92	0.958
93	93	0.934
94	94	0.956
95	95	0.992
96	96	0.938
97	97	0.982
98	98	0.923
99	99	0.971

Table 24 r_{wg} Results per project

```
Single Score Intraclass Correlation

  Model: twoway
  Type : agreement

  Subjects = 10
    Raters = 372
  ICC(A,1) = 0.509

F-Test, H0: r0 = 0 ; H1: r0 > 0
 F(9,2732) = 484 , p = 0

95%-Confidence Interval for ICC Population Values:
 0.328 < ICC < 0.776
```

Table 25 ICC Peer results

C.7 REGRESSION CALCULATION N=111

C.7.1. REGRESSION RESULTS USING PO_TEAM AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	58.38**	[26.95, 89.80]						
PO_Exp	0.25	[-0.08, 0.58]	0.15	[-0.05, 0.34]	.02	[-.03, .07]	.16	
Peer	2.23	[-3.22, 7.68]	0.08	[-0.12, 0.29]	.01	[-.02, .03]	.09	
Age	-1.10	[-6.89, 4.70]	-0.04	[-0.24, 0.17]	.00	[-.01, .01]	-.08	
Education	-1.71	[-6.79, 3.37]	-0.07	[-0.29, 0.15]	.00	[-.02, .03]	-.03	
Gender	3.72	[-0.53, 7.98]	0.17	[-0.02, 0.37]	.03	[-.03, .09]	.18	
Residence	-0.16	[-4.37, 4.05]	-0.01	[-0.23, 0.21]	.00	[-.00, .00]	-.01	
								R^2 = .068
								95% CI[.00,.13]

Table 26 Regression calculation PO_Team N=111

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

C.7.1.1. DURBIN-WATSON TEST

DW = 1.8372, p-value = 0.1833

alternative hypothesis: true autocorrelation is greater than 0

C.7.1.2. REGRESSION

Residual standard error: 9.097 on 104 degrees of freedom

Multiple R-squared: 0.06799, Adjusted R-squared: 0.01422

F-statistic: 1.265 on 6 and 104 DF, p-value: 0.2802

C.7.1.3. EFFECT PLOTS

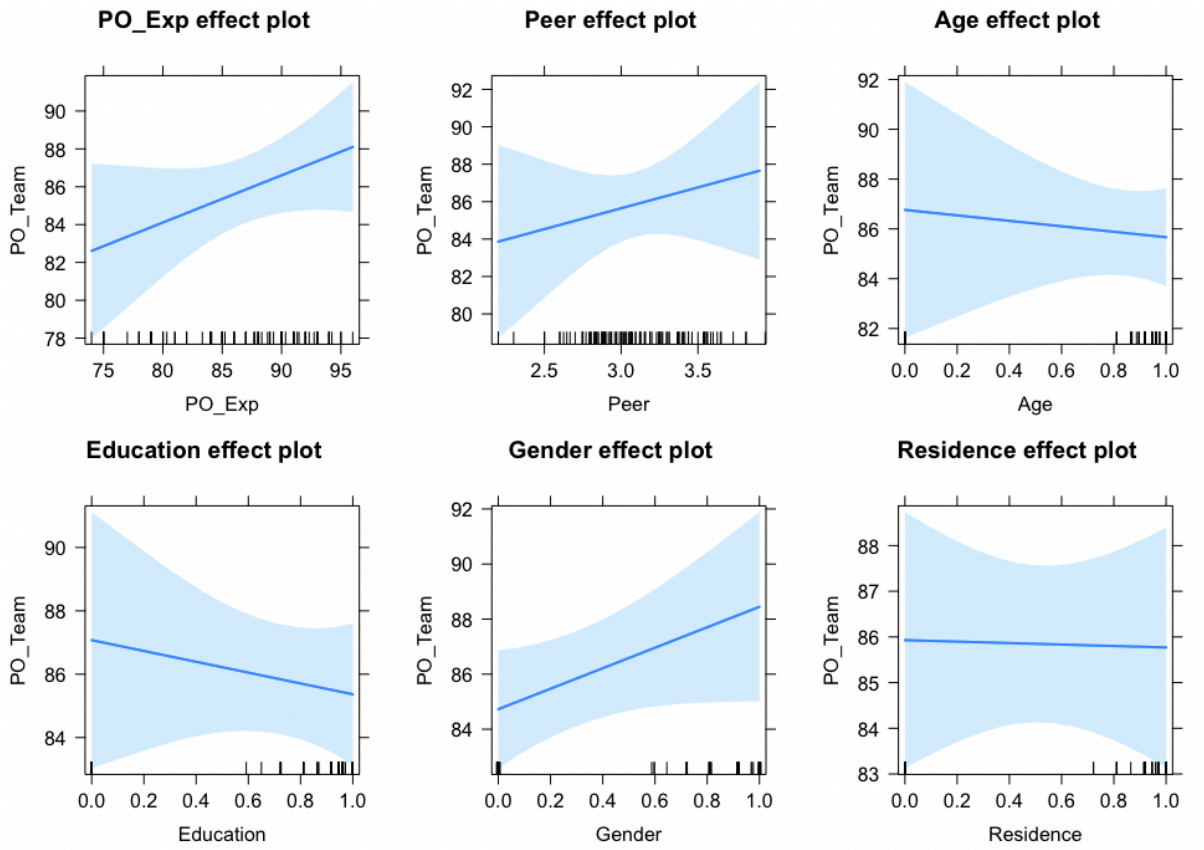


Figure 143 Effect plots regression PO_Team N=111

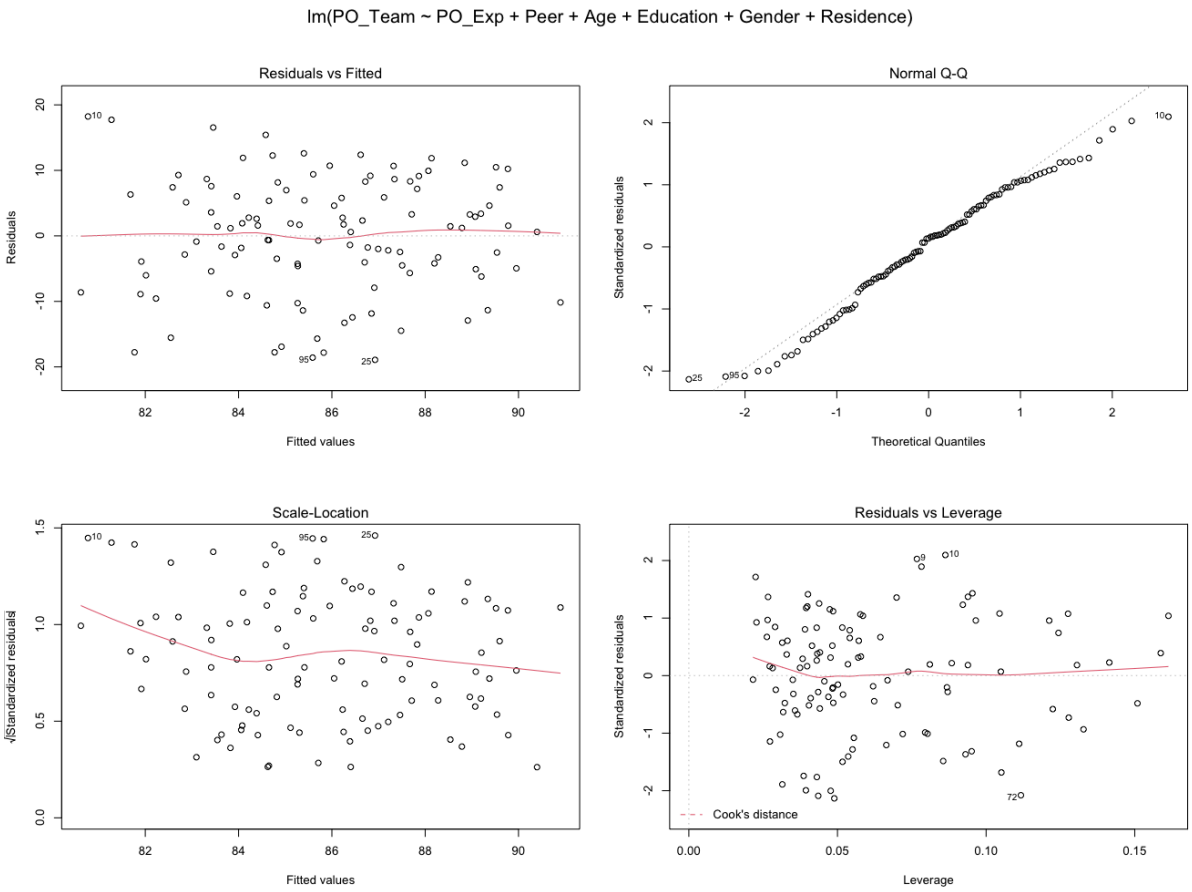


Figure 144 Residual plots regression PO_Team N=111

C.7.2. REGRESSION RESULTS USING PO_EXP AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	74.43**	[61.31, 87.55]						
PO_Team	0.09	[-0.03, 0.20]	0.15	[-0.05, 0.34]	.02	[-.03, .07]	.16	
Peer	1.87	[-1.31, 5.05]	0.12	[-0.08, 0.33]	.01	[-.03, .05]	.10	
Age	-1.95	[-5.32, 1.41]	-0.12	[-0.32, 0.09]	.01	[-.03, .05]	-.08	
Education	0.62	[-2.36, 3.59]	0.05	[-0.18, 0.27]	.00	[-.01, .02]	.07	
Gender	-0.11	[-2.63, 2.41]	-0.01	[-0.21, 0.19]	.00	[-.00, .00]	.07	
Residence	1.23	[-1.22, 3.68]	0.11	[-0.11, 0.33]	.01	[-.03, .04]	.09	
								R^2 = .058
								95% CI[.00,.11]

Table 27 Regression calculation PO_Exp N=111

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

C.7.2.1. DURBIN-WATSON TEST

DW = 1.7726, p-value = 0.1065

alternative hypothesis: true autocorrelation is greater than 0

C.7.2.2. REGRESSION

Residual standard error: 5.319 on 104 degrees of freedom

Multiple R-squared: 0.05765, Adjusted R-squared: 0.003279

F-statistic: 1.06 on 6 and 104 DF, p-value: 0.3913

C.7.2.3. EFFECT PLOTS

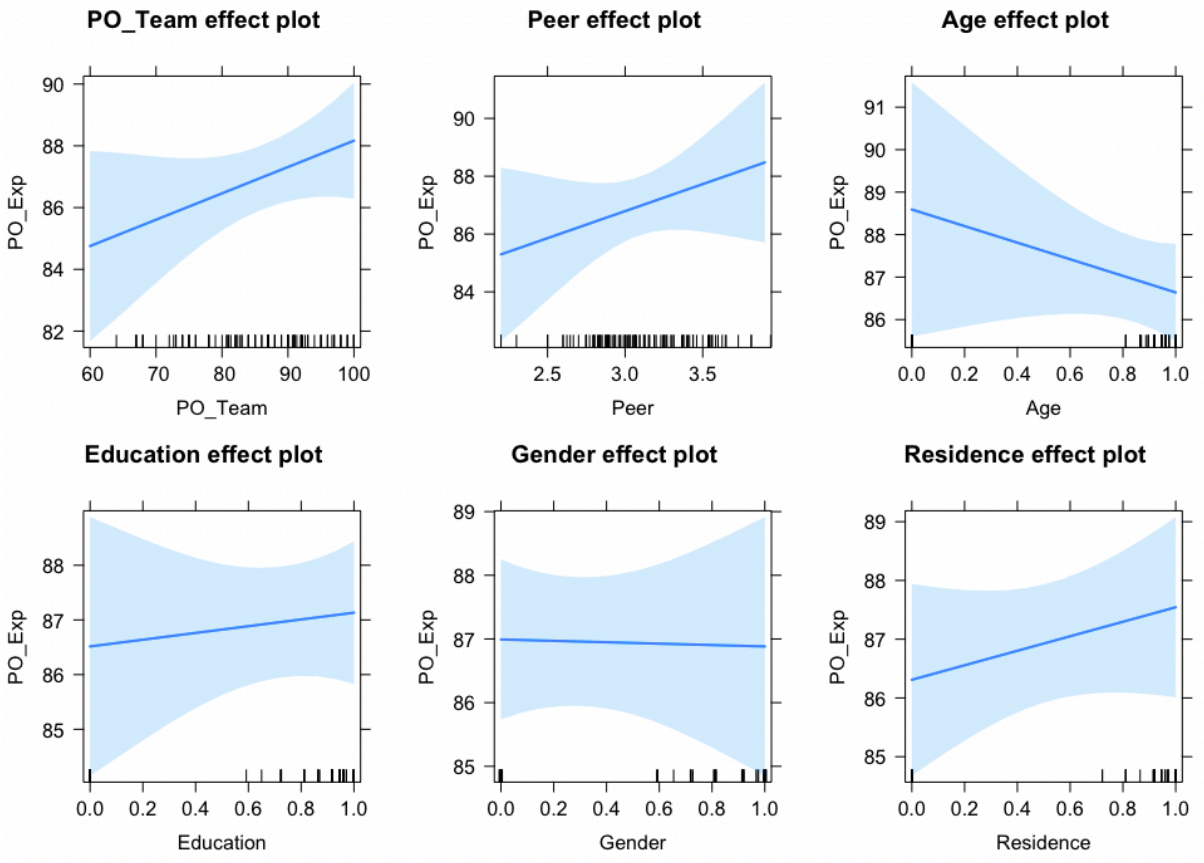


Figure 145 Effect plots regression PO_Exp N=111

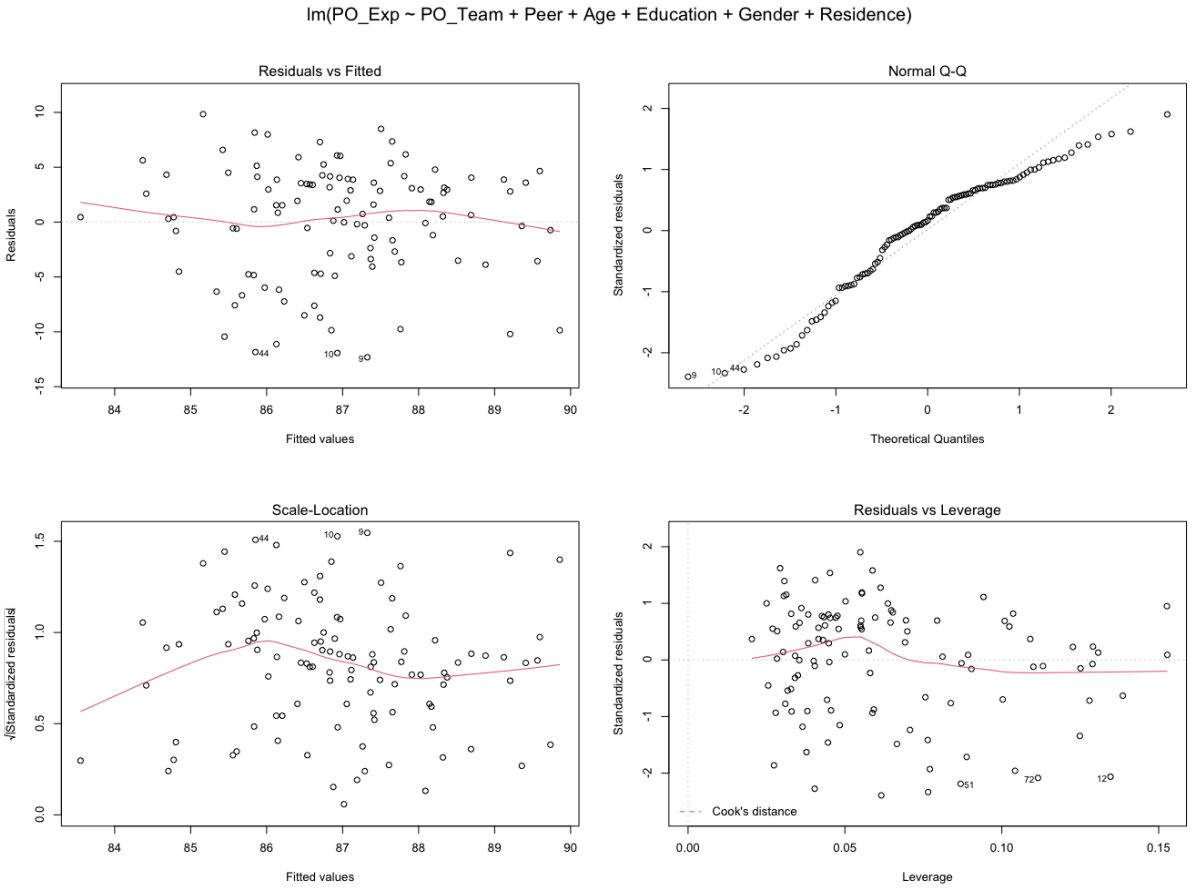


Figure 146 Residual plots regression PO_Exp N=111

C.7.3. REGRESSION RESULTS USING PEER REVIEW AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	2.06**	[0.94, 3.18]						
PO_Exp	0.01	[-0.00, 0.02]	0.11	[-0.07, 0.29]	.01	[-.02, .05]	.10	
PO_Team	0.00	[-0.00, 0.01]	0.07	[-0.11, 0.26]	.01	[-.02, .03]	.09	
Age	0.16	[-0.04, 0.37]	0.15	[-0.04, 0.34]	.02	[-.03, .07]	.17	
Education	0.22*	[0.04, 0.40]	0.25	[0.05, 0.46]	.05	[-.02, .12]	.20*	
Gender	0.09	[-0.06, 0.24]	0.11	[-0.08, 0.30]	.01	[-.02, .05]	.09	
Residence	-0.25**	[-0.39, -0.11]	-0.34	[-0.54, -	.10	[-.00, .20]	-.20*	

				0.15]				
								R^2 = .171**
								95% CI[.02,.26]

Table 28 Regression calculation Peer_Review N=111

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

C.7.3.1. DURBIN-WATSON TEST

DW = 1.725, p-value = 0.06741

alternative hypothesis: true autocorrelation is greater than 0

C.7.3.2. REGRESSION

Residual standard error: 0.3235 on 104 degrees of freedom

Multiple R-squared: 0.1711, Adjusted R-squared: 0.1233

F-statistic: 3.579 on 6 and 104 DF, p-value: 0.002901

C.7.3.3. EFFECT PLOTS

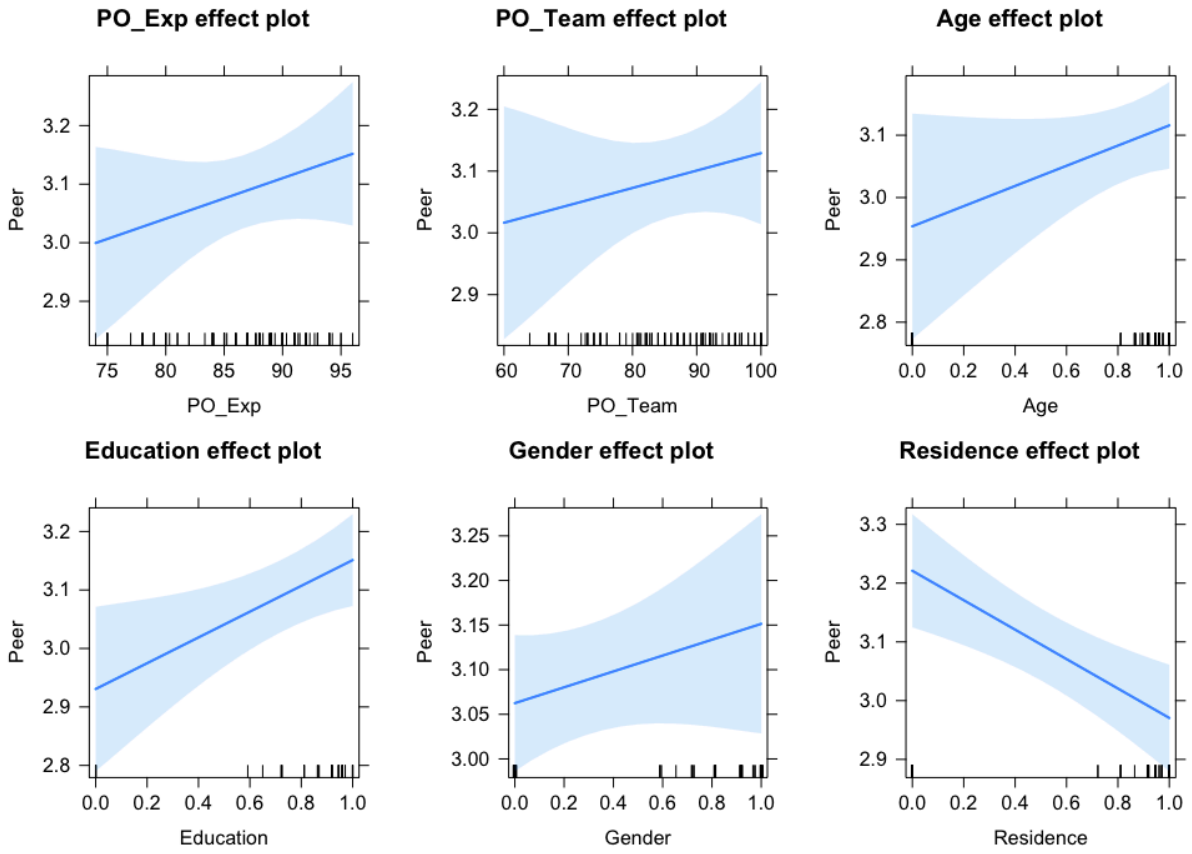


Figure 147 Effect plots regression Peer_Review N=111

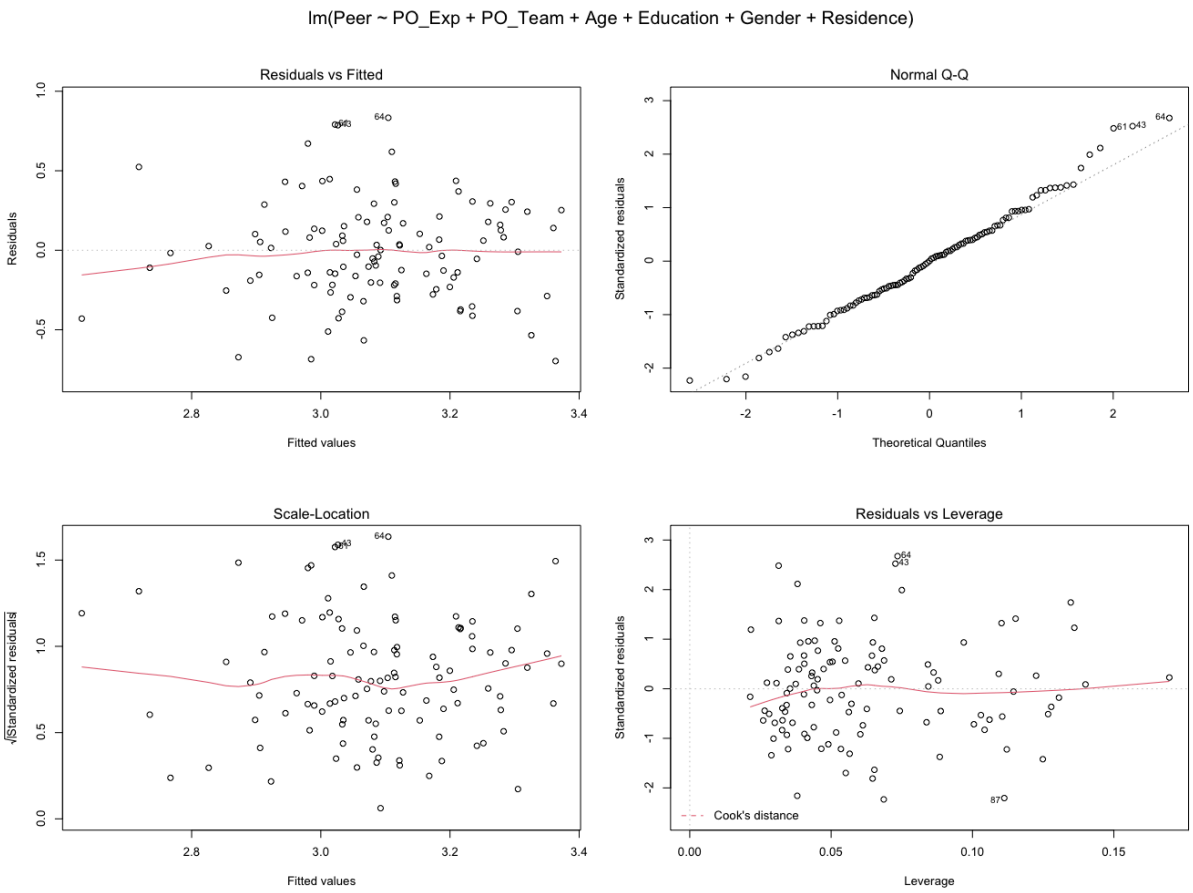


Figure 148 Residual plots regression Peer N=111

C.8 REGRESSION CALCULATION N=109

C.8.1. REGRESSION RESULTS USING PO_TEAM AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	89.46**	[41.46, 137.46]						
PO_Exp	0.21	[-0.14, 0.56]	0.12	[-0.08, 0.33]	.01	[-.03, .05]	.18	
Peer	1.61	[-4.10, 7.32]	0.06	[-0.16, 0.28]	.00	[-.02, .02]	.07	
Age	-2.78	[-8.96, 3.41]	-0.10	[-0.31, 0.12]	.01	[-.02, .04]	-.10	
Education	-1.45	[-6.84, 3.94]	-0.06	[-0.30, 0.17]	.00	[-.02, .02]	-.04	

Gender	3.14	[-1.39, 7.66]	0.15	[-0.07, 0.36]	.02	[-.03, .07]	.16	
Residence	0.56	[-4.05, 5.17]	0.03	[-0.21, 0.27]	.00	[-.01, .01]	-.00	
N	-0.61	[-1.62, 0.41]	-0.14	[-0.36, 0.09]	.01	[-.03, .05]	-.10	
NS	-0.17	[-1.55, 1.21]	-0.04	[-0.34, 0.27]	.00	[-.01, .01]	-.03	
E	-0.15	[-0.96, 0.65]	-0.07	[-0.46, 0.31]	.00	[-.01, .01]	-.08	
ES	0.14	[-0.80, 1.08]	0.04	[-0.25, 0.34]	.00	[-.01, .01]	.01	
O	0.15	[-0.41, 0.72]	0.08	[-0.22, 0.38]	.00	[-.02, .02]	-.01	
OS	-0.57	[-1.46, 0.33]	-0.19	[-0.49, 0.11]	.01	[-.03, .06]	-.11	
A	-0.30	[-0.85, 0.26]	-0.12	[-0.35, 0.11]	.01	[-.03, .05]	-.10	
AS	0.34	[-0.72, 1.40]	0.11	[-0.23, 0.44]	.00	[-.02, .03]	.00	
C	-0.12	[-0.84, 0.61]	-0.06	[-0.47, 0.34]	.00	[-.01, .01]	-.07	
CS	-0.05	[-1.09, 1.00]	-0.02	[-0.37, 0.34]	.00	[-.00, .00]	-.02	
								R^2 = .122
								95% CI[.00,.10]

Table 29 Regression calculation PO_Team N=109

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

C.8.1.1. DURBIN-WATSON TEST

DW = 1.7307, p-value = 0.06465

alternative hypothesis: true autocorrelation is greater than 0

C.8.1.2. REGRESSION

Residual standard error: 9.128 on 92 degrees of freedom

Multiple R-squared: 0.1219, Adjusted R-squared: -0.03082

F-statistic: 0.7982 on 16 and 92 DF, p-value: 0.684

C.8.1.3. EFFECT PLOTS

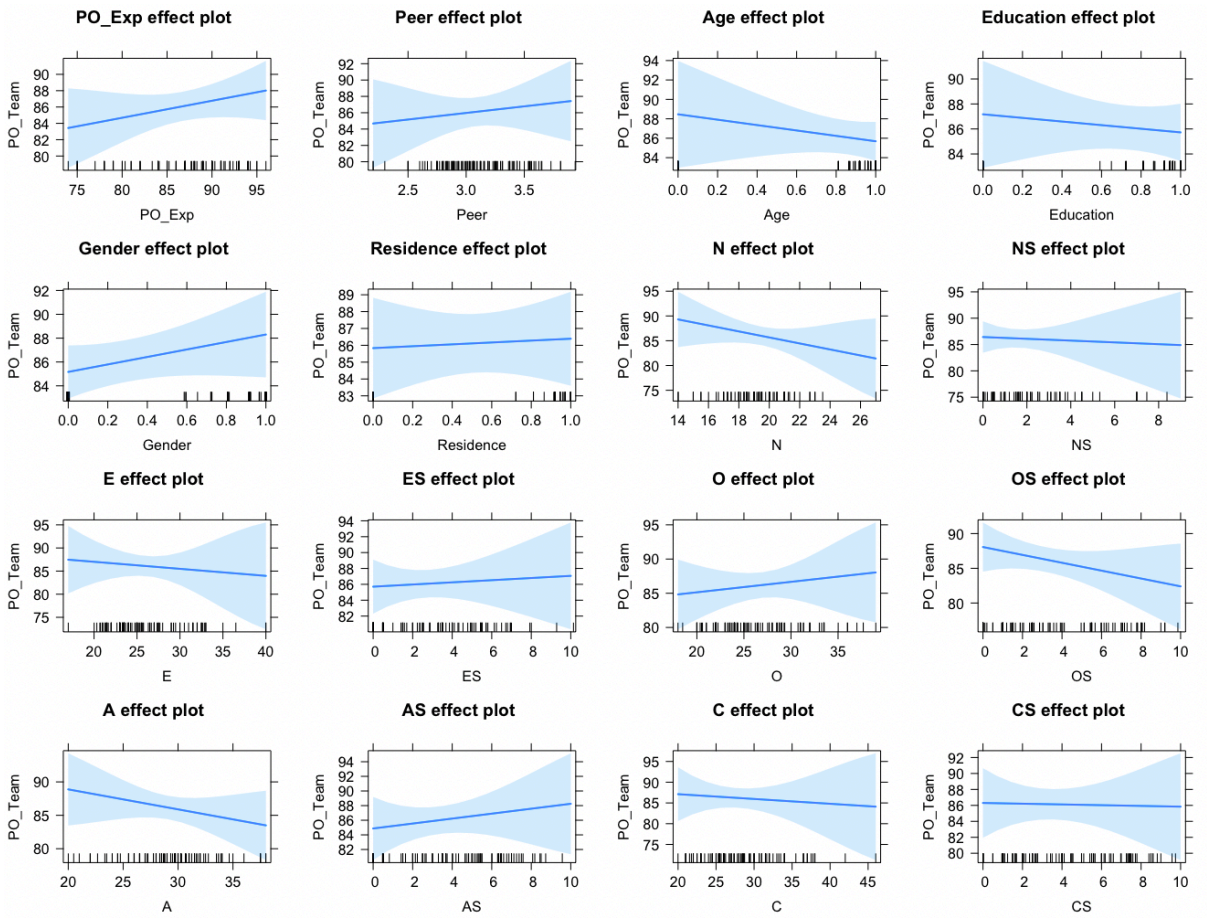


Figure 149 Effect plots regression PO_Team N=109

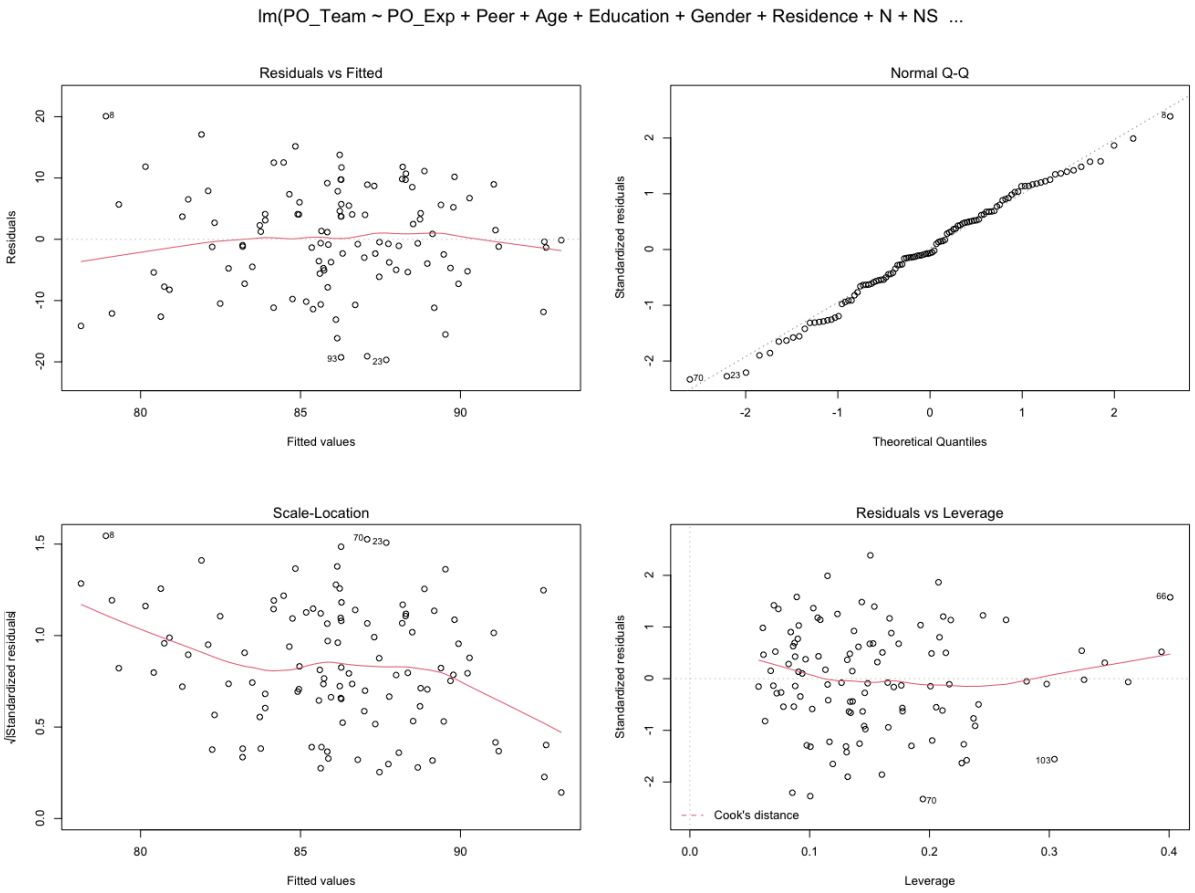


Figure 150 Residual plots regression PO_Team N=109

C.8.2. REGRESSION RESULTS USING PO_EXP AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	90.78**	[67.15, 114.40]						
PO_Team	0.07	[-0.05, 0.19]	0.12	[-0.08, 0.32]	.01	[-.03, .05]	.18	
Peer	1.63	[-1.71, 4.96]	0.10	[-0.11, 0.32]	.01	[-.02, .04]	.10	
Age	-2.02	[-5.64, 1.60]	-0.12	[-0.33, 0.09]	.01	[-.03, .05]	-.10	
Education	1.10	[-2.06, 4.26]	0.08	[-0.15, 0.31]	.00	[-.02, .03]	.06	
Gender	0.24	[-2.45, 2.92]	0.02	[-0.20, 0.23]	.00	[-.01, .01]	.07	
Residence	1.76	[-0.92, 4.44]	0.16	[-0.08, 0.39]	.02	[-.03, .06]	.07	
N	-0.41	[-1.00, 0.18]	-0.15	[-0.38, 0.07]	.02	[-.03, .06]	-.05	
NS	0.27	[-0.54, 1.07]	0.10	[-0.20, 0.40]	.00	[-.02, .03]	.07	
E	0.05	[-0.42, 0.52]	0.04	[-0.34, 0.42]	.00	[-.01, .01]	-.14	

ES	-0.20	[-0.75, 0.35]	-0.11	[-0.40, 0.18]	.01	[-.02, .03]	.04	
O	0.02	[-0.32, 0.35]	0.01	[-0.28, 0.31]	.00	[-.00, .00]	-.08	
OS	-0.33	[-0.86, 0.19]	-0.19	[-0.49, 0.11]	.01	[-.03, .06]	-.04	
A	-0.01	[-0.34, 0.32]	-0.01	[-0.23, 0.22]	.00	[-.00, .00]	-.05	
AS	-0.17	[-0.79, 0.45]	-0.09	[-0.42, 0.24]	.00	[-.02, .02]	.03	
C	-0.31	[-0.73, 0.11]	-0.29	[-0.69, 0.11]	.02	[-.03, .07]	-.14	
CS	0.54	[-0.06, 1.14]	0.31	[-0.03, 0.66]	.03	[-.03, .09]	.11	
								R^2 = .146
								95% CI[.00,.14]

Table 30 Regression calculation PO_Exp N=109

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

C.8.2.1. DURBIN-WATSON TEST

DW = 1.9229, p-value = 0.3026

alternative hypothesis: true autocorrelation is greater than 0

C.8.2.2. REGRESSION

Residual standard error: 5.357 on 92 degrees of freedom

Multiple R-squared: 0.1464, Adjusted R-squared: -0.002089

F-statistic: 0.9859 on 16 and 92 DF, p-value: 0.4786

C.8.2.3. EFFECT PLOTS

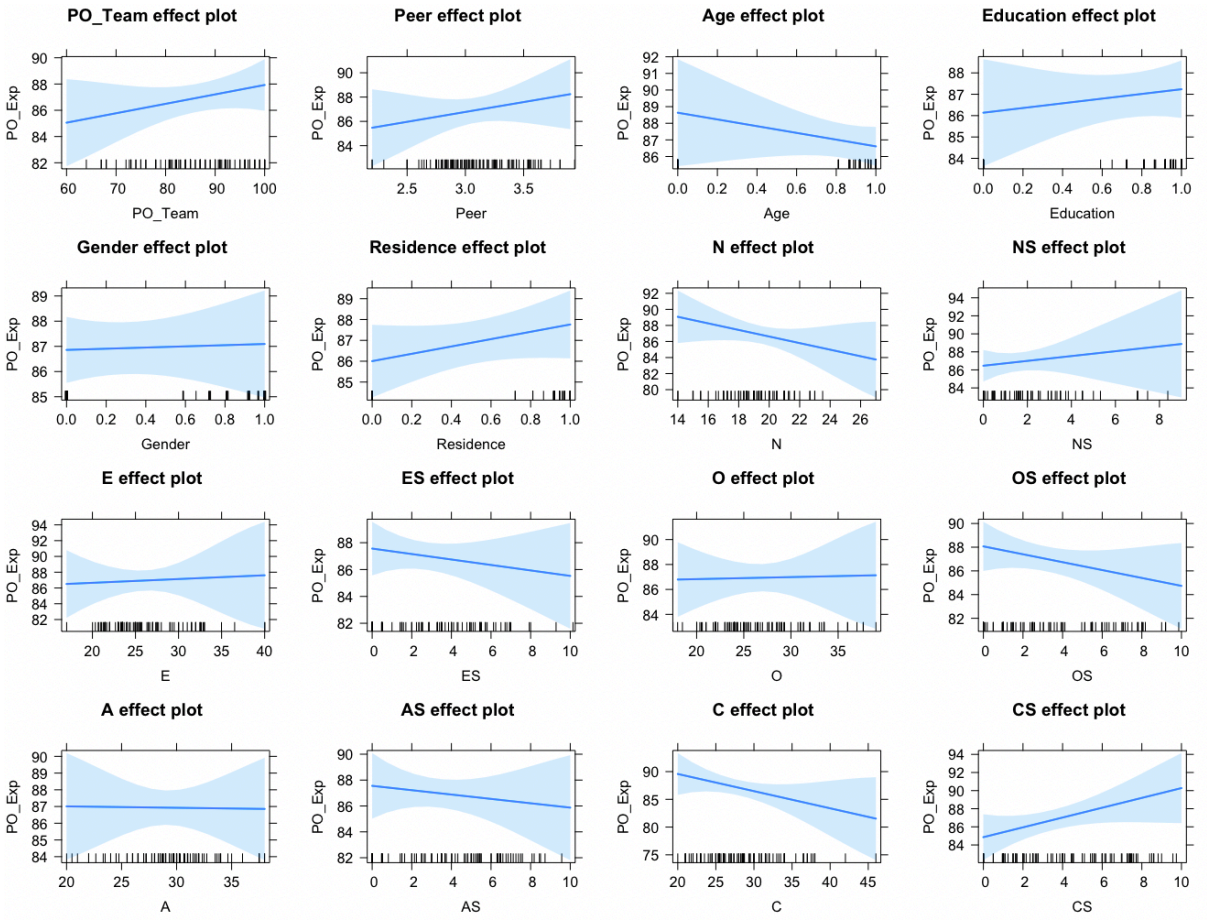


Figure 151 Effect plots regression PO_Exp N=109

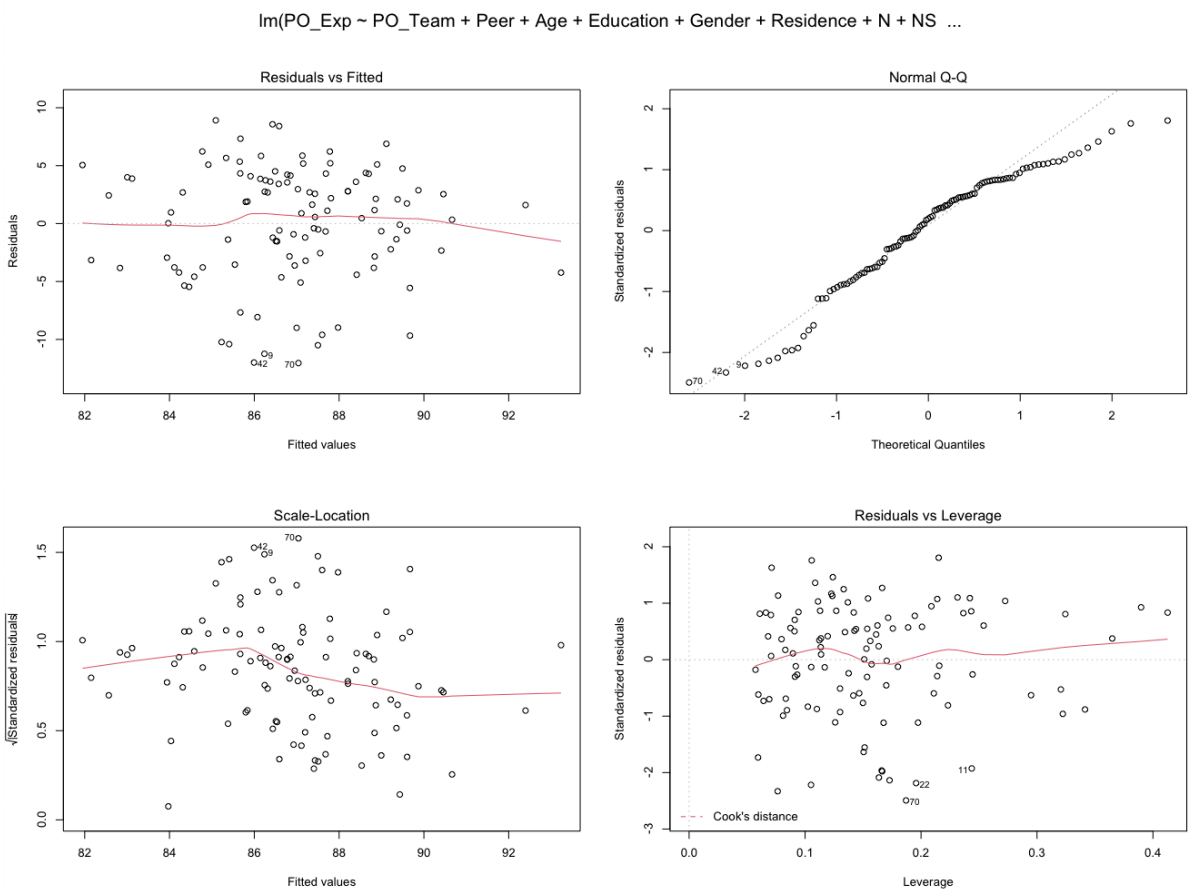


Figure 152 Residual plots regression PO_Exp N=109

C.8.3. REGRESSION RESULTS USING PEER REVIEW AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	2.04*	[0.23, 3.85]						
PO_Exp	0.01	[-0.01, 0.02]	0.10	[-0.10, 0.29]	.01	[-.02, .04]	.10	
PO_Team	0.00	[-0.01, 0.01]	0.06	[-0.14, 0.25]	.00	[-.01, .02]	.07	
Age	0.14	[-0.08, 0.37]	0.13	[-0.07, 0.33]	.01	[-.03, .05]	.14	
Education	0.22*	[0.03, 0.41]	0.25	[0.03, 0.46]	.04	[-.02, .11]	.18	
Gender	0.07	[-0.09, 0.23]	0.09	[-0.12, 0.29]	.01	[-.02, .03]	.08	
Residence	-0.27**	[-0.43, -0.12]	-0.38	[-0.59, -0.16]	.10	[.00, .20]	-.21*	
N	-0.01	[-0.04, 0.03]	-0.03	[-0.25, 0.18]	.00	[-.01, .01]	.02	
NS	0.03	[-0.02, 0.08]	0.17	[-0.12, 0.46]	.01	[-.02, .05]	.11	
E	-0.02	[-0.05, 0.01]	-0.22	[-0.58, 0.14]	.01	[-.02, .05]	-.06	
ES	0.00	[-0.03, 0.04]	0.01	[-0.27, 0.29]	.00	[-.00, .00]	.07	
O	-0.00	[-0.02, 0.02]	-0.01	[-0.29, 0.28]	.00	[-.00, .00]	-.05	

OS	-0.02	[-0.05, 0.01]	-0.16	[-0.44, 0.13]	.01	[-.02, .04]	-.10	
A	0.01	[-0.01, 0.03]	0.12	[-0.10, 0.34]	.01	[-.02, .04]	.02	
AS	-0.00	[-0.04, 0.04]	-0.02	[-0.33, 0.30]	.00	[-.00, .00]	.03	
C	0.02	[-0.01, 0.04]	0.22	[-0.16, 0.61]	.01	[-.02, .05]	.01	
CS	0.01	[-0.03, 0.04]	0.05	[-0.29, 0.38]	.00	[-.01, .01]	.05	
								R^2 = .217
								95% CI[.00,.23]

Table 31 Regression calculation Peer_Review N=109

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

C.8.3.1. DURBIN-WATSON TEST

DW = 1.7551, p-value = 0.08213

alternative hypothesis: true autocorrelation is greater than 0

C.8.3.2. REGRESSION

Residual standard error: 0.3305 on 92 degrees of freedom

Multiple R-squared: 0.2173, Adjusted R-squared: 0.08122

F-statistic: 1.597 on 16 and 92 DF, p-value: 0.08525

C.8.3.3. EFFECT PLOTS

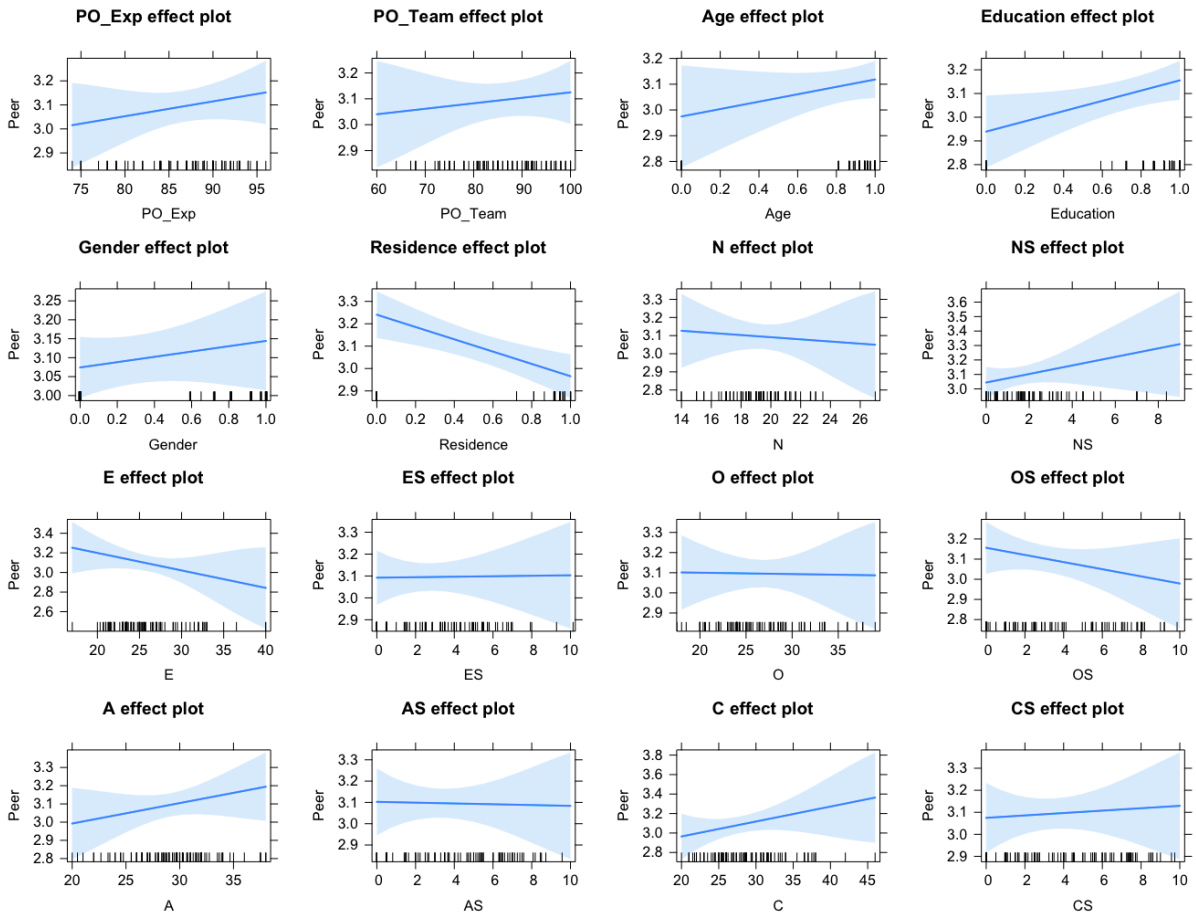


Figure 153 Effect plots regression Peer_Review N=109

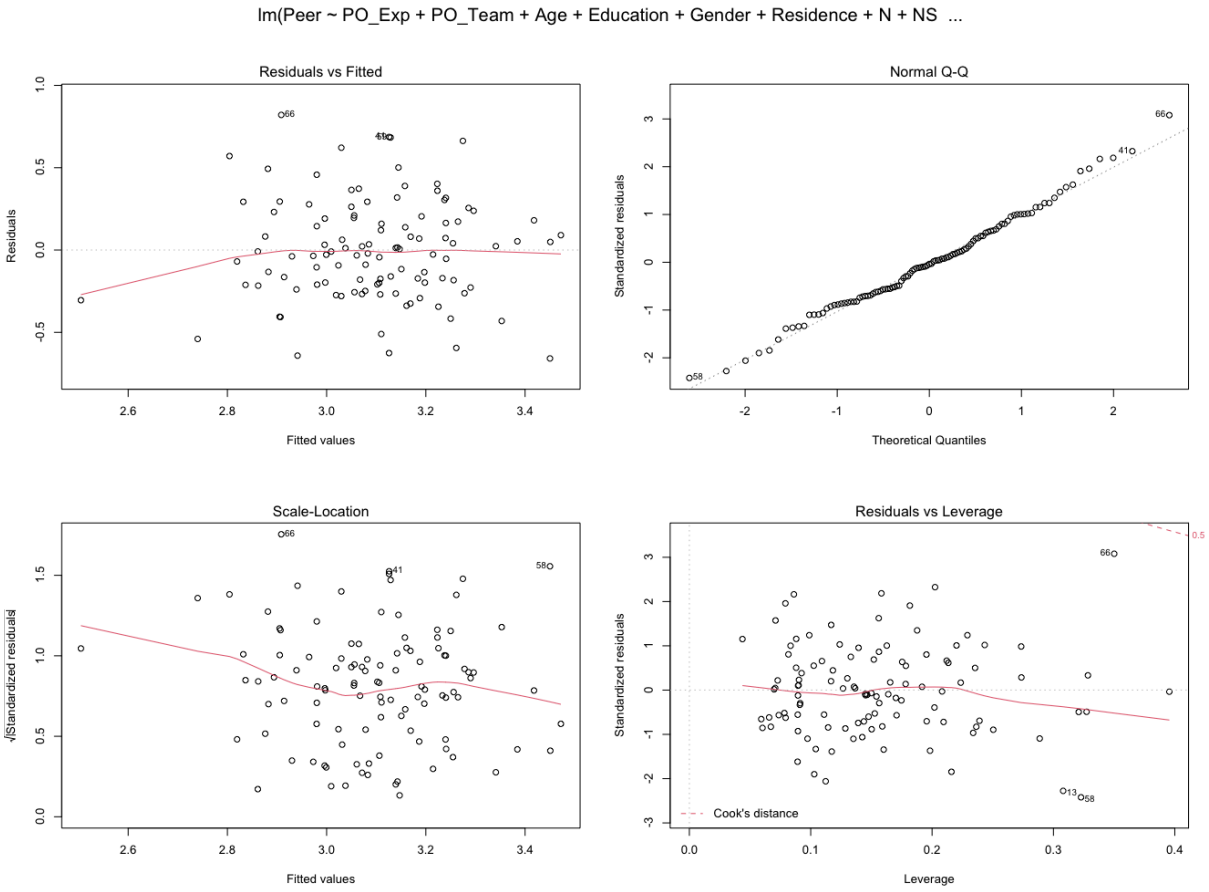


Figure 154 Residual plots regression Peer N=109

C.9 REGRESSION CALCULATION N=87

C.9.1. REGRESSION RESULTS USING PO_TEAM AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	72.66*	[9.40, 135.93]						
PO_Exp	0.31	[-0.10, 0.72]	0.18	[-0.06, 0.42]	.02	[-.03, .07]	.23*	
Peer	1.54	[-5.14, 8.22]	0.06	[-0.19, 0.31]	.00	[-.01, .02]	.07	
Age	0.22	[-8.25, 8.69]	0.01	[-0.24, 0.26]	.00	[-.00, .00]	-.16	
Education	6.58	[-1.84, 15.00]	0.22	[-0.06, 0.51]	.02	[-.03, .08]	.09	
Gender	2.64	[-2.88, 8.16]	0.13	[-0.14, 0.40]	.01	[-.02, .04]	.22*	

Residence	-0.65	[-6.11, 4.81]	-0.03	[-0.32, 0.25]	.00	[-.01, .01]	.07	
I	0.69	[-5.66, 7.05]	0.04	[-0.33, 0.42]	.00	[-.01, .01]	-.06	
IS	6.75	[-4.74, 18.24]	0.21	[-0.14, 0.56]	.01	[-.02, .05]	.19	
EG	-0.87	[-7.56, 5.82]	-0.07	[-0.57, 0.44]	.00	[-.01, .01]	.01	
EGS	17.24**	[6.02, 28.47]	0.54	[0.19, 0.89]	.10	[-.00, .20]	.28**	
R	-0.27	[-4.27, 3.73]	-0.02	[-0.29, 0.26]	.00	[-.00, .00]	.02	
RS	-13.42*	[-23.53, -3.32]	-0.49	[-0.86, -0.12]	.07	[-.01, .16]	.07	
D	-0.12	[-5.68, 5.43]	-0.01	[-0.39, 0.37]	.00	[-.00, .00]	-.03	
DS	-3.41	[-11.74, 4.91]	-0.14	[-0.47, 0.20]	.01	[-.02, .03]	.15	
T	0.37	[-3.64, 4.38]	0.03	[-0.25, 0.31]	.00	[-.01, .01]	-.04	
TS	-0.23	[-6.30, 5.83]	-0.01	[-0.25, 0.23]	.00	[-.00, .00]	.09	
N	-0.84	[-2.34, 0.66]	-0.16	[-0.45, 0.13]	.01	[-.02, .05]	-.11	
NS	2.25**	[0.72, 3.77]	0.53	[0.17, 0.89]	.09	[-.01, .18]	.26*	
E	-0.19	[-1.02, 0.64]	-0.09	[-0.48, 0.30]	.00	[-.01, .02]	.01	
ES	-0.55	[-1.61, 0.51]	-0.19	[-0.54, 0.17]	.01	[-.02, .05]	.11	
O	0.09	[-0.48, 0.66]	0.05	[-0.24, 0.34]	.00	[-.01, .01]	.07	
OS	-1.33*	[-2.44, -0.22]	-0.46	[-0.85, -0.08]	.06	[-.02, .14]	-.05	
A	-0.20	[-1.06, 0.66]	-0.07	[-0.37, 0.23]	.00	[-.01, .02]	.05	
AS	-0.33	[-1.44, 0.78]	-0.11	[-0.48, 0.26]	.00	[-.02, .02]	.07	
C	0.00	[-0.39, 0.39]	0.00	[-0.23, 0.23]	.00	[-.00, .00]	.02	
CS	0.92	[-0.34, 2.18]	0.31	[-0.12, 0.74]	.02	[-.03, .07]	.08	
								R^2 = .390
								95% CI[.00,.33]

Table 32 Regression calculation PO_Team N=87

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

C.9.1.1. DURBIN-WATSON TEST

DW = 1.5541, p-value = 0.01518

alternative hypothesis: true autocorrelation is greater than 0

C.9.1.2. REGRESSION

Residual standard error: 9.128 on 92 degrees of freedom

Multiple R-squared: 0.1219, Adjusted R-squared: -0.03082

F-statistic: 0.7982 on 16 and 92 DF, p-value: 0.684

C.9.1.3. EFFECT PLOTS

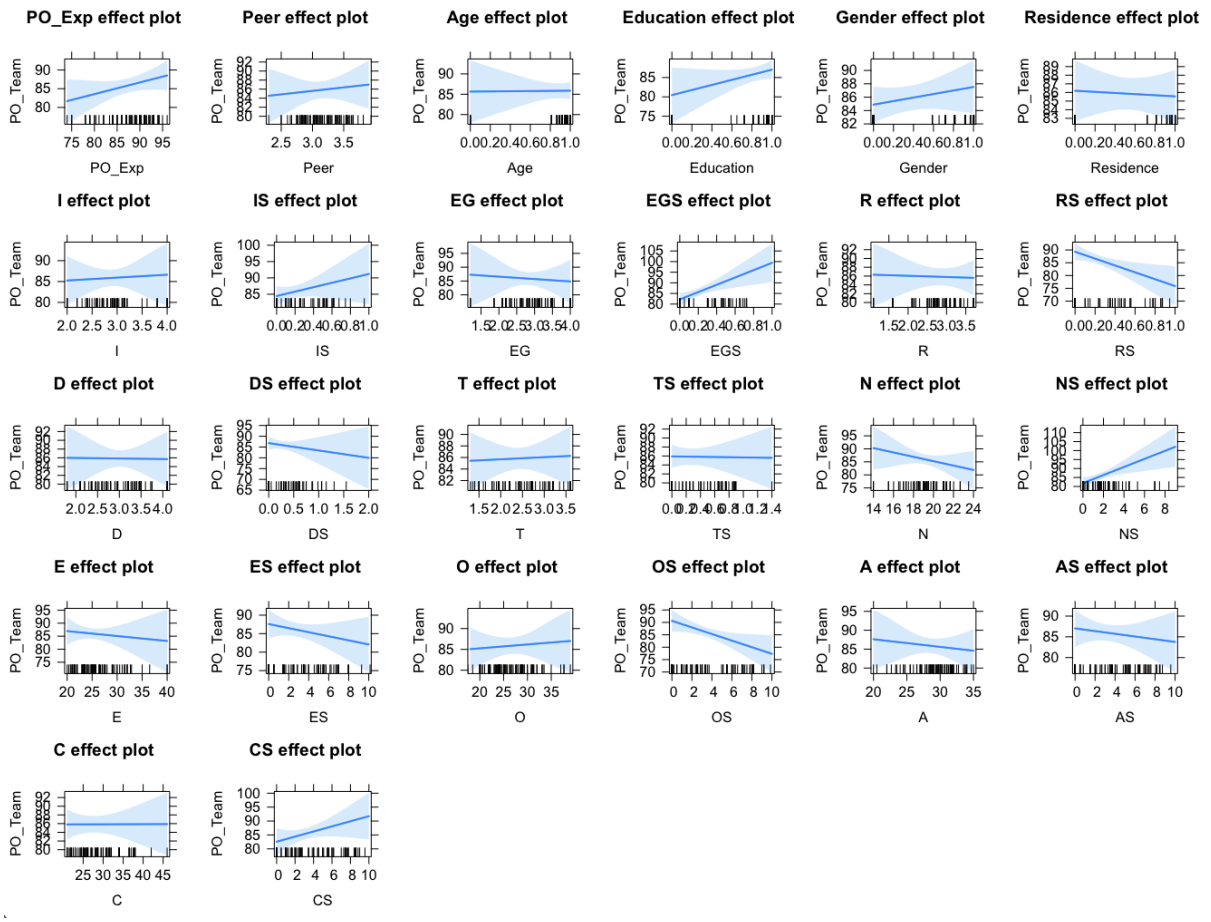


Figure 155 Effect plots regression PO_Team N=87

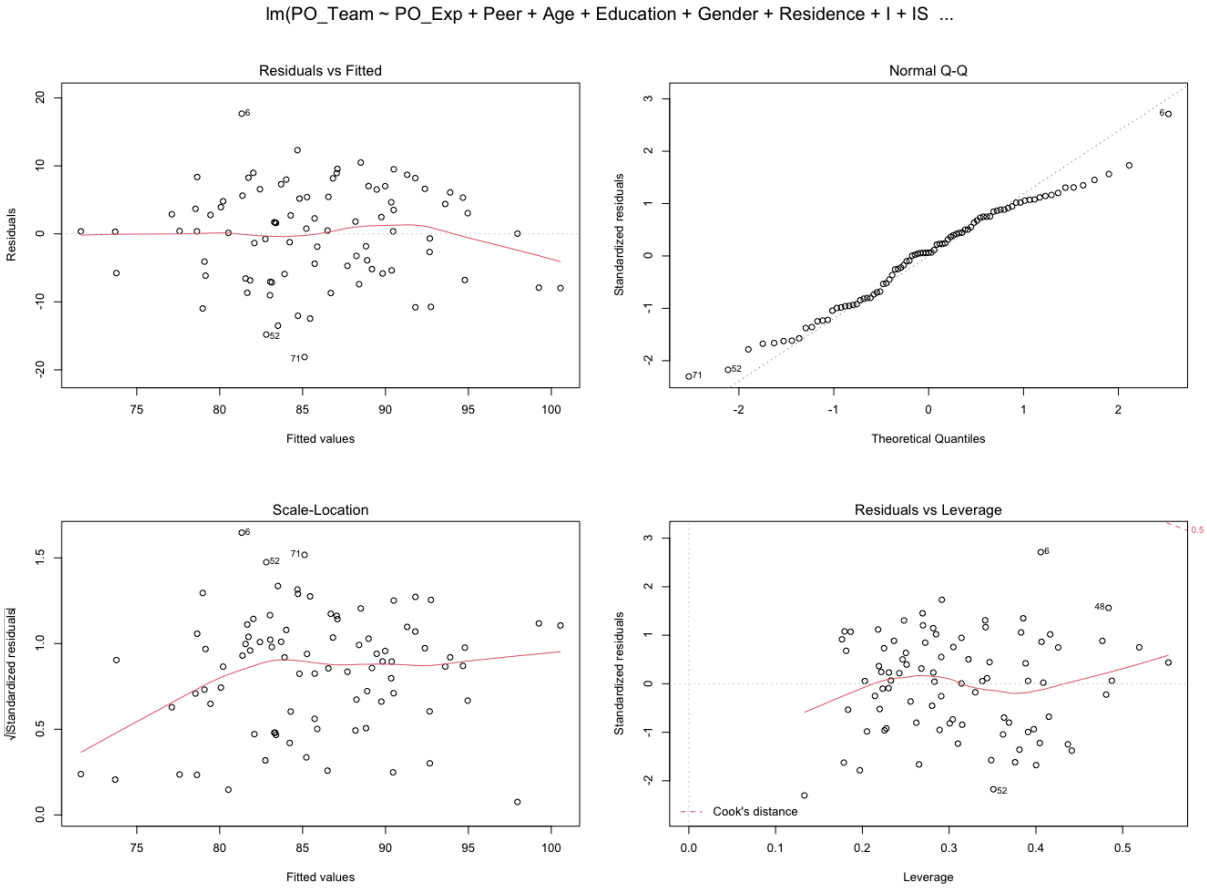


Figure 156 Residual plots regression PO_Team N=87

C.9.2. REGRESSION RESULTS USING PO_EXP AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	72.19**	[35.96, 108.43]						
PO_Team	0.12	[-0.04, 0.28]	0.20	[-0.06, 0.47]	.03	[-.03, .08]	.23*	
Peer	1.05	[-3.07, 5.17]	0.07	[-0.20, 0.33]	.00	[-.02, .02]	.08	
Age	-1.25	[-6.47, 3.97]	-0.06	[-0.33, 0.20]	.00	[-.02, .02]	-.09	
Education	-1.04	[-6.34, 4.25]	-0.06	[-0.37, 0.25]	.00	[-.01, .02]	-.06	
Gender	0.87	[-2.55, 4.30]	0.07	[-0.22, 0.36]	.00	[-.02, .02]	.03	
Residence	3.84*	[0.62, 7.07]	0.34	[0.06, 0.63]	.07	[-.02, .15]	.17	
I	1.02	[-2.89, 4.94]	0.10	[-0.29, 0.50]	.00	[-.02, .02]	-.04	
IS	2.21	[-4.95, 9.36]	0.12	[-0.26, 0.49]	.00	[-.02, .03]	.23*	
EG	2.72	[-1.35, 6.80]	0.36	[-0.18, 0.89]	.02	[-.03, .07]	.17	
EGS	-0.96	[-8.42, 6.49]	-0.05	[-0.45, 0.35]	.00	[-.01, .01]	.15	
R	0.72	[-1.74, 3.19]	0.09	[-0.21, 0.38]	.00	[-.02, .03]	.15	
RS	0.81	[-5.78, 7.41]	0.05	[-0.37, 0.47]	.00	[-.01, .01]	.08	

D	-1.16	[-4.58, 2.26]	-0.14	[-0.54, 0.27]	.01	[-.02, .03]	.12	
DS	-0.80	[-5.97, 4.36]	-0.06	[-0.41, 0.30]	.00	[-.01, .01]	.09	
T	1.21	[-1.25, 3.67]	0.14	[-0.15, 0.44]	.01	[-.03, .05]	.18	
TS	1.66	[-2.06, 5.38]	0.11	[-0.14, 0.36]	.01	[-.02, .04]	.17	
N	-0.40	[-1.33, 0.53]	-0.13	[-0.45, 0.18]	.01	[-.02, .04]	-.06	
NS	-0.05	[-1.06, 0.96]	-0.02	[-0.43, 0.39]	.00	[-.00, .00]	.16	
E	-0.04	[-0.55, 0.48]	-0.03	[-0.45, 0.39]	.00	[-.01, .01]	-.08	
ES	-0.02	[-0.68, 0.64]	-0.01	[-0.40, 0.37]	.00	[-.00, .00]	.05	
O	-0.02	[-0.37, 0.33]	-0.02	[-0.33, 0.29]	.00	[-.00, .00]	.00	
OS	0.63	[-0.07, 1.33]	0.38	[-0.04, 0.80]	.04	[-.03, .10]	.21*	
A	-0.11	[-0.64, 0.42]	-0.07	[-0.39, 0.25]	.00	[-.01, .02]	-.05	
AS	-0.15	[-0.84, 0.53]	-0.09	[-0.48, 0.31]	.00	[-.01, .02]	.10	
C	0.01	[-0.23, 0.25]	0.01	[-0.23, 0.25]	.00	[-.00, .00]	.03	
CS	-0.32	[-1.11, 0.47]	-0.19	[-0.65, 0.28]	.01	[-.02, .04]	.05	
								R^2 = .310
								95% CI[.00,.22]

Table 33 Regression calculation PO_Exp N=87

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

C.9.2.1. DURBIN-WATSON TEST

DW = 1.9241, p-value = 0.3323

alternative hypothesis: true autocorrelation is greater than 0

C.9.2.2. REGRESSION

Residual standard error: 5.357 on 92 degrees of freedom

Multiple R-squared: 0.1464, Adjusted R-squared: -0.002089

F-statistic: 0.9859 on 16 and 92 DF, p-value: 0.4786

C.9.2.3. EFFECT PLOTS

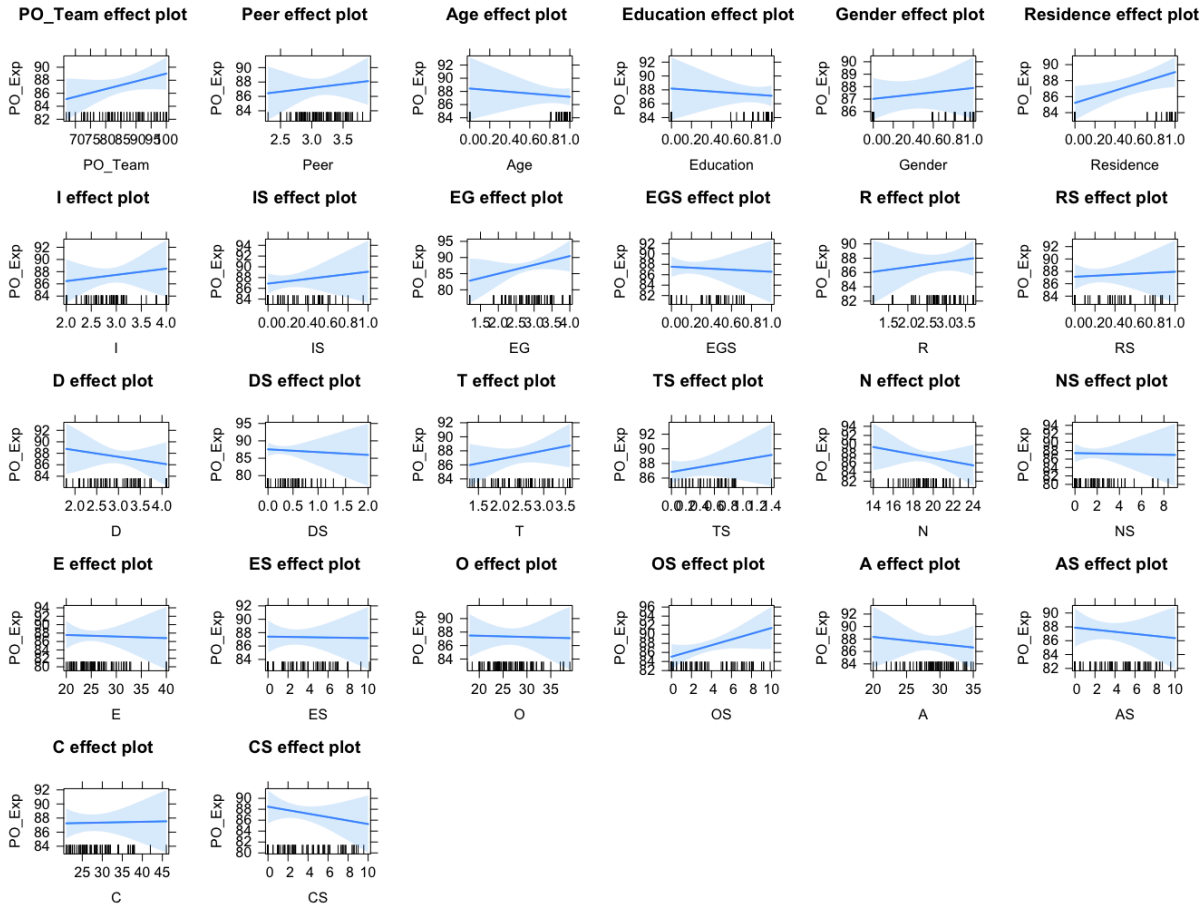


Figure 157 Effect plots regression PO_Exp N=87

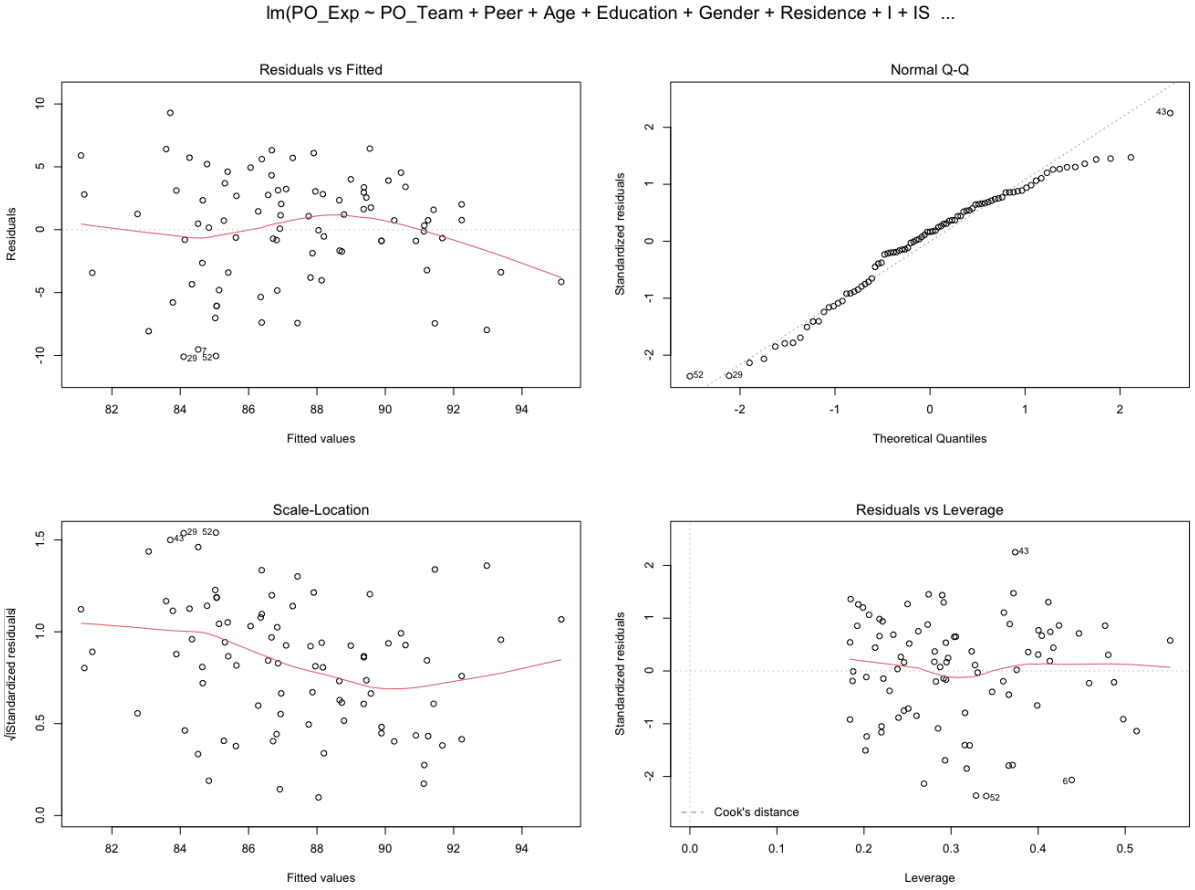


Figure 158 Residual plots regression PO_Exp N=87

C.9.3. REGRESSION RESULTS USING PEER REVIEW AS THE CRITERION

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	2.83*	[0.40, 5.27]						
PO_Exp	0.00	[-0.01, 0.02]	0.06	[-0.19, 0.32]	.00	[-.02, .02]	.08	
PO_Team	0.00	[-0.01, 0.01]	0.06	[-0.21, 0.33]	.00	[-.01, .02]	.07	
Age	0.03	[-0.30, 0.36]	0.02	[-0.23, 0.28]	.00	[-.01, .01]	-.01	
Education	0.11	[-0.22, 0.44]	0.10	[-0.20, 0.40]	.00	[-.02, .03]	.01	
Gender	-0.02	[-0.23, 0.20]	-0.02	[-0.31, 0.26]	.00	[-.01, .01]	.03	
Residence	-0.11	[-0.31, 0.10]	-0.15	[-0.44, 0.14]	.01	[-.02, .05]	-.22*	
I	0.18	[-0.06, 0.42]	0.28	[-0.10, 0.67]	.02	[-.03, .08]	.10	

IS	0.16	[-0.29, 0.61]	0.13	[-0.24, 0.50]	.01	[-.02, .03]	.02	
EG	0.10	[-0.15, 0.36]	0.21	[-0.32, 0.73]	.01	[-.02, .04]	.02	
EGS	-0.32	[-0.78, 0.14]	-0.27	[-0.65, 0.12]	.02	[-.03, .07]	-.08	
R	0.11	[-0.04, 0.27]	0.21	[-0.07, 0.49]	.03	[-.03, .08]	.21	
RS	0.07	[-0.35, 0.48]	0.07	[-0.34, 0.47]	.00	[-.01, .01]	-.02	
D	-0.13	[-0.34, 0.09]	-0.23	[-0.62, 0.16]	.02	[-.03, .06]	-.10	
DS	-0.04	[-0.36, 0.29]	-0.04	[-0.39, 0.31]	.00	[-.01, .01]	.05	
T	-0.03	[-0.18, 0.13]	-0.05	[-0.34, 0.24]	.00	[-.01, .01]	.03	
TS	-0.25*	[-0.47, -0.02]	-0.26	[-0.50, -0.03]	.05	[-.02, .13]	-.13	
N	-0.05	[-0.11, 0.01]	-0.25	[-0.55, 0.05]	.03	[-.03, .09]	-.08	
NS	0.02	[-0.04, 0.08]	0.13	[-0.27, 0.53]	.00	[-.02, .03]	.08	
E	-0.03	[-0.06, 0.00]	-0.34	[-0.74, 0.05]	.03	[-.03, .09]	-.15	
ES	0.02	[-0.02, 0.06]	0.22	[-0.15, 0.59]	.01	[-.03, .06]	.06	
O	-0.00	[-0.02, 0.02]	-0.02	[-0.32, 0.28]	.00	[-.00, .00]	-.09	
OS	0.01	[-0.04, 0.05]	0.06	[-0.36, 0.48]	.00	[-.01, .01]	.01	
A	0.03	[-0.00, 0.06]	0.29	[-0.02, 0.59]	.04	[-.03, .11]	.04	
AS	0.02	[-0.03, 0.06]	0.14	[-0.24, 0.53]	.01	[-.02, .03]	.02	
C	-0.01	[-0.02, 0.01]	-0.08	[-0.32, 0.16]	.00	[-.02, .03]	-.06	
CS	-0.03	[-0.08, 0.01]	-0.31	[-0.76, 0.13]	.02	[-.03, .07]	-.10	
								R^2 = .342
								95% CI[.00,.27]

Table 34 Regression calculation Peer_Review N=87

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

C.9.3.1. DURBIN-WATSON TEST

DW = 1.8549, p-value = 0.2249

alternative hypothesis: true autocorrelation is greater than 0

C.9.3.2. REGRESSION

Residual standard error: 0.3305 on 92 degrees of freedom

Multiple R-squared: 0.2173, Adjusted R-squared: 0.08122

F-statistic: 1.597 on 16 and 92 DF, p-value: 0.08525

C.9.3.3. EFFECT PLOTS

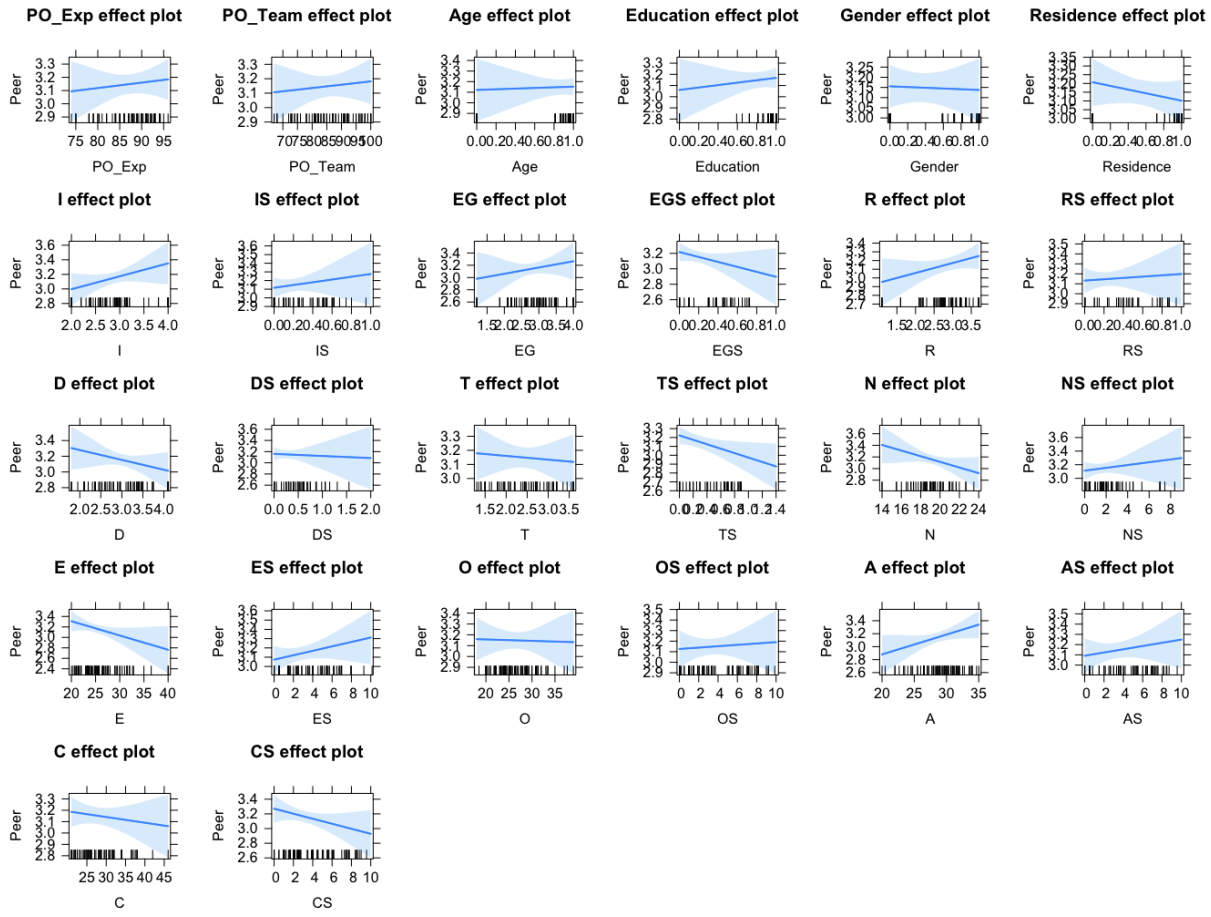


Figure 159 Effect plots regression Peer_Review N=87

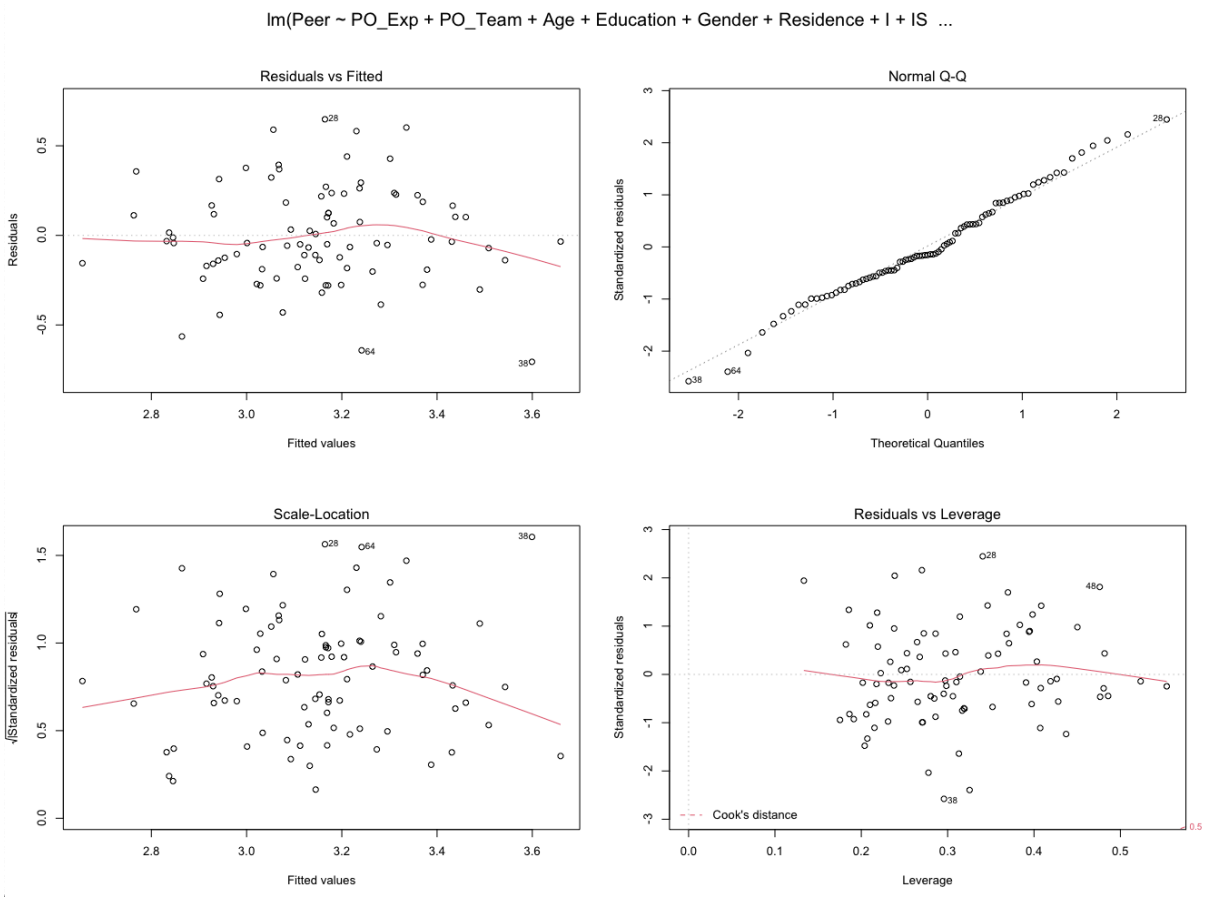


Figure 160 Residual plots regression Peer N=87

C.9.4. MEDIATION FOR PEER IN MULTI REGRESSION PO_TEAM AS THE CRITERION

```
Total effect(c) of Residence on PO_Team = 0.29 S.E. = 2.21 t = 0.13 df= 82 with p = 0.9
Direct effect (c') of Residence on PO_Team removing Peer = 0.61 S.E. = 2.29 t = 0.27 df= 81 with p = 0.79
Indirect effect (ab) of Residence on PO_Team through Peer = -0.32
Mean bootstrapped indirect effect = -0.39 with standard error = 0.7 Lower CI = -2 Upper CI = 0.85

Total effect(c) of Age on PO_Team = -4.47 S.E. = 3.91 t = -1.14 df= 82 with p = 0.26
Direct effect (c') of Age on PO_Team removing Peer = -4.5 S.E. = 3.92 t = -1.15 df= 81 with p = 0.26
Indirect effect (ab) of Age on PO_Team through Peer = 0.03
Mean bootstrapped indirect effect = 0.03 with standard error = 0.55 Lower CI = -1.02 Upper CI = 1.24

Total effect(c) of Gender on PO_Team = 3.53 S.E. = 2.33 t = 1.52 df= 82 with p = 0.13
Direct effect (c') of Gender on PO_Team removing Peer = 3.43 S.E. = 2.34 t = 1.46 df= 81 with p = 0.15
Indirect effect (ab) of Gender on PO_Team through Peer = 0.1
Mean bootstrapped indirect effect = 0.15 with standard error = 0.41 Lower CI = -0.48 Upper CI = 1.21

Total effect(c) of Education on PO_Team = 3.08 S.E. = 3.39 t = 0.91 df= 82 with p = 0.37
Direct effect (c') of Education on PO_Team removing Peer = 2.92 S.E. = 3.42 t = 0.85 df= 81 with p = 0.4
Indirect effect (ab) of Education on PO_Team through Peer = 0.17
Mean bootstrapped indirect effect = 0.19 with standard error = 0.6 Lower CI = -0.9 Upper CI = 1.65
R = 0.27 R2 = 0.07 F = 1.27 on 5 and 81 DF p-value: 0.281
```

```
Total effect(c) of  O  on  PO_Team  =  0.21   S.E. =  0.28  t  =  0.77  df=  81   with p =  0.44
Direct effect (c') of  O  on  PO_Team  removing  Peer  =  0.22   S.E. =  0.28  t  =  0.77  df=  80   with p =  0.44
Indirect effect (ab) of  O  on  PO_Team  through  Peer   =  0
Mean bootstrapped indirect effect =  0  with standard error =  0.04  Lower CI =  -0.07    Upper CI =  0.09

Total effect(c) of  C  on  PO_Team  =  0.07   S.E. =  0.19  t  =  0.37  df=  81   with p =  0.71
Direct effect (c') of  C  on  PO_Team  removing  Peer  =  0.07   S.E. =  0.19  t  =  0.38  df=  80   with p =  0.7
Indirect effect (ab) of  C  on  PO_Team  through  Peer   =  0
Mean bootstrapped indirect effect =  0  with standard error =  0.02  Lower CI =  -0.05    Upper CI =  0.05

Total effect(c) of  E  on  PO_Team  =  -0.35   S.E. =  0.34  t  =  -1.02  df=  81   with p =  0.31
Direct effect (c') of  E  on  PO_Team  removing  Peer  =  -0.32   S.E. =  0.35  t  =  -0.92  df=  80   with p =  0.36
Indirect effect (ab) of  E  on  PO_Team  through  Peer   =  -0.03
Mean bootstrapped indirect effect =  -0.02  with standard error =  0.1  Lower CI =  -0.22    Upper CI =  0.17

Total effect(c) of  A  on  PO_Team  =  0.16   S.E. =  0.34  t  =  0.47  df=  81   with p =  0.64
Direct effect (c') of  A  on  PO_Team  removing  Peer  =  0.15   S.E. =  0.34  t  =  0.42  df=  80   with p =  0.67
Indirect effect (ab) of  A  on  PO_Team  through  Peer   =  0.01
Mean bootstrapped indirect effect =  -0.01  with standard error =  0.07  Lower CI =  -0.19    Upper CI =  0.13

Total effect(c) of  N  on  PO_Team  =  -0.89   S.E. =  0.68  t  =  -1.3  df=  81   with p =  0.2
Direct effect (c') of  N  on  PO_Team  removing  Peer  =  -0.84   S.E. =  0.7  t  =  -1.2  df=  80   with p =  0.23
Indirect effect (ab) of  N  on  PO_Team  through  Peer   =  -0.05
Mean bootstrapped indirect effect =  -0.03  with standard error =  0.17  Lower CI =  -0.37    Upper CI =  0.32
R = 0.17 R2 = 0.03   F = 0.41 on 6 and 80 DF   p-value:  0.891

Total effect(c) of  OS  on  PO_Team  =  -0.95   S.E. =  0.46  t  =  -2.07  df=  81   with p =  0.042
Direct effect (c') of  OS  on  PO_Team  removing  Peer  =  -0.97   S.E. =  0.47  t  =  -2.08  df=  80   with p =  0.04
Indirect effect (ab) of  OS  on  PO_Team  through  Peer   =  0.02
Mean bootstrapped indirect effect =  0.02  with standard error =  0.07  Lower CI =  -0.12    Upper CI =  0.17

Total effect(c) of  CS  on  PO_Team  =  0.34   S.E. =  0.58  t  =  0.59  df=  81   with p =  0.55
Direct effect (c') of  CS  on  PO_Team  removing  Peer  =  0.41   S.E. =  0.6  t  =  0.69  df=  80   with p =  0.49
Indirect effect (ab) of  CS  on  PO_Team  through  Peer   =  -0.07
Mean bootstrapped indirect effect =  -0.07  with standard error =  0.17  Lower CI =  -0.46    Upper CI =  0.25

Total effect(c) of  ES  on  PO_Team  =  -0.28   S.E. =  0.5  t  =  -0.56  df=  81   with p =  0.57
Direct effect (c') of  ES  on  PO_Team  removing  Peer  =  -0.32   S.E. =  0.51  t  =  -0.63  df=  80   with p =  0.53
Indirect effect (ab) of  ES  on  PO_Team  through  Peer   =  0.04
Mean bootstrapped indirect effect =  0.04  with standard error =  0.11  Lower CI =  -0.15    Upper CI =  0.3

Total effect(c) of  AS  on  PO_Team  =  -0.14   S.E. =  0.5  t  =  -0.28  df=  81   with p =  0.78
Direct effect (c') of  AS  on  PO_Team  removing  Peer  =  -0.15   S.E. =  0.5  t  =  -0.29  df=  80   with p =  0.77
Indirect effect (ab) of  AS  on  PO_Team  through  Peer   =  0.01
Mean bootstrapped indirect effect =  0.01  with standard error =  0.05  Lower CI =  -0.08    Upper CI =  0.14

Total effect(c) of  NS  on  PO_Team  =  2.03   S.E. =  0.66  t  =  3.08  df=  81   with p =  0.0029
Direct effect (c') of  NS  on  PO_Team  removing  Peer  =  2   S.E. =  0.66  t  =  3.02  df=  80   with p =  0.0034
Indirect effect (ab) of  NS  on  PO_Team  through  Peer   =  0.02
Mean bootstrapped indirect effect =  0.02  with standard error =  0.09  Lower CI =  -0.16    Upper CI =  0.24
R = 0.37 R2 = 0.14   F = 2.12 on 6 and 80 DF   p-value:  0.0506
```

```
Total effect(c) of I on PO_Team = -1.06 S.E. = 2.61 t = -0.41 df= 81 with p = 0.68
Direct effect (c') of I on PO_Team removing Peer = -1.34 S.E. = 2.66 t = -0.51 df= 80 with p = 0.61
Indirect effect (ab) of I on PO_Team through Peer = 0.28
Mean bootstrapped indirect effect = 0.28 with standard error = 0.58 Lower CI = -0.76 Upper CI = 1.62

Total effect(c) of EG on PO_Team = 0.44 S.E. = 2.89 t = 0.15 df= 81 with p = 0.88
Direct effect (c') of EG on PO_Team removing Peer = 0.21 S.E. = 2.93 t = 0.07 df= 80 with p = 0.94
Indirect effect (ab) of EG on PO_Team through Peer = 0.23
Mean bootstrapped indirect effect = 0.29 with standard error = 0.58 Lower CI = -0.64 Upper CI = 1.8

Total effect(c) of R on PO_Team = -0.04 S.E. = 1.87 t = -0.02 df= 81 with p = 0.98
Direct effect (c') of R on PO_Team removing Peer = -0.25 S.E. = 1.91 t = -0.13 df= 80 with p = 0.9
Indirect effect (ab) of R on PO_Team through Peer = 0.2
Mean bootstrapped indirect effect = 0.25 with standard error = 0.5 Lower CI = -0.55 Upper CI = 1.45

Total effect(c) of D on PO_Team = -1.09 S.E. = 2.69 t = -0.4 df= 81 with p = 0.69
Direct effect (c') of D on PO_Team removing Peer = -0.83 S.E. = 2.73 t = -0.31 df= 80 with p = 0.76
Indirect effect (ab) of D on PO_Team through Peer = -0.25
Mean bootstrapped indirect effect = -0.31 with standard error = 0.62 Lower CI = -1.83 Upper CI = 0.66

Total effect(c) of T on PO_Team = 0.05 S.E. = 1.89 t = 0.03 df= 81 with p = 0.98
Direct effect (c') of T on PO_Team removing Peer = 0.08 S.E. = 1.9 t = 0.04 df= 80 with p = 0.97
Indirect effect (ab) of T on PO_Team through Peer = -0.03
Mean bootstrapped indirect effect = -0.04 with standard error = 0.26 Lower CI = -0.66 Upper CI = 0.47
R = 0.1 R2 = 0.01 F = 0.14 on 6 and 80 DF p-value: 0.994


Total effect(c) of IS on PO_Team = 4.06 S.E. = 4.56 t = 0.89 df= 81 with p = 0.38
Direct effect (c') of IS on PO_Team removing Peer = 3.89 S.E. = 4.57 t = 0.85 df= 80 with p = 0.4
Indirect effect (ab) of IS on PO_Team through Peer = 0.17
Mean bootstrapped indirect effect = 0.14 with standard error = 0.62 Lower CI = -1.13 Upper CI = 1.52

Total effect(c) of EGS on PO_Team = 15.07 S.E. = 5.48 t = 2.75 df= 81 with p = 0.0073
Direct effect (c') of EGS on PO_Team removing Peer = 15.68 S.E. = 5.52 t = 2.84 df= 80 with p = 0.0057
Indirect effect (ab) of EGS on PO_Team through Peer = -0.62
Mean bootstrapped indirect effect = -0.58 with standard error = 0.9 Lower CI = -2.76 Upper CI = 0.87

Total effect(c) of RS on PO_Team = -9.46 S.E. = 4.57 t = -2.07 df= 81 with p = 0.042
Direct effect (c') of RS on PO_Team removing Peer = -9.45 S.E. = 4.57 t = -2.07 df= 80 with p = 0.042
Indirect effect (ab) of RS on PO_Team through Peer = -0.02
Mean bootstrapped indirect effect = -0.01 with standard error = 0.55 Lower CI = -1.21 Upper CI = 1.17

Total effect(c) of DS on PO_Team = -0.13 S.E. = 3.54 t = -0.04 df= 81 with p = 0.97
Direct effect (c') of DS on PO_Team removing Peer = -0.56 S.E. = 3.57 t = -0.16 df= 80 with p = 0.88
Indirect effect (ab) of DS on PO_Team through Peer = 0.43
Mean bootstrapped indirect effect = 0.44 with standard error = 0.63 Lower CI = -0.59 Upper CI = 1.98

Total effect(c) of TS on PO_Team = 0.01 S.E. = 2.77 t = 0 df= 81 with p = 1
Direct effect (c') of TS on PO_Team removing Peer = 0.34 S.E. = 2.79 t = 0.12 df= 80 with p = 0.9
Indirect effect (ab) of TS on PO_Team through Peer = -0.33
Mean bootstrapped indirect effect = -0.4 with standard error = 0.64 Lower CI = -2.06 Upper CI = 0.52
R = 0.37 R2 = 0.14 F = 2.15 on 6 and 80 DF p-value: 0.0477
```

C.9.5. MEDIATION FOR PEER IN MULTI REGRESSION PO_EXP AS THE CRITERION

```
Total effect(c) of  Residence  on  PO_Exp  =  2.36   S.E. =  1.3  t  =  1.82  df=  82   with p =  0.073
Direct effect (c') of  Residence  on  PO_Exp  removing  Peer  =  2.75   S.E. =  1.33  t  =  2.06  df=  81   with p =  0.042
Indirect effect (ab) of  Residence  on  PO_Exp  through  Peer   =  -0.39
Mean bootstrapped indirect effect =  -0.44  with standard error =  0.43  Lower CI =  -1.48    Upper CI =  0.18

Total effect(c) of  Age  on  PO_Exp  =  -1.68   S.E. =  2.29  t  =  -0.73  df=  82   with p =  0.47
Direct effect (c') of  Age  on  PO_Exp  removing  Peer  =  -1.71   S.E. =  2.29  t  =  -0.75  df=  81   with p =  0.46
Indirect effect (ab) of  Age  on  PO_Exp  through  Peer   =  0.03
Mean bootstrapped indirect effect =  0.01  with standard error =  0.41  Lower CI =  -0.86    Upper CI =  0.9

Total effect(c) of  Gender  on  PO_Exp  =  -0.3   S.E. =  1.37  t  =  -0.22  df=  82   with p =  0.82
Direct effect (c') of  Gender  on  PO_Exp  removing  Peer  =  -0.43   S.E. =  1.37  t  =  -0.32  df=  81   with p =  0.75
Indirect effect (ab) of  Gender  on  PO_Exp  through  Peer   =  0.13
Mean bootstrapped indirect effect =  0.12  with standard error =  0.27  Lower CI =  -0.36    Upper CI =  0.75

Total effect(c) of  Education  on  PO_Exp  =  -1.83   S.E. =  1.99  t  =  -0.92  df=  82   with p =  0.36
Direct effect (c') of  Education  on  PO_Exp  removing  Peer  =  -2.03   S.E. =  1.99  t  =  -1.02  df=  81   with p =  0.31
Indirect effect (ab) of  Education  on  PO_Exp  through  Peer   =  0.2
Mean bootstrapped indirect effect =  0.2  with standard error =  0.44  Lower CI =  -0.55    Upper CI =  1.25
R = 0.26 R2 = 0.07   F = 1.13 on 5 and 81 DF   p-value:  0.353


Total effect(c) of  O  on  PO_Exp  =  0.12   S.E. =  0.16  t  =  0.76  df=  81   with p =  0.45
Direct effect (c') of  O  on  PO_Exp  removing  Peer  =  0.12   S.E. =  0.16  t  =  0.77  df=  80   with p =  0.44
Indirect effect (ab) of  O  on  PO_Exp  through  Peer   =  0
Mean bootstrapped indirect effect =  0  with standard error =  0.02  Lower CI =  -0.05    Upper CI =  0.03

Total effect(c) of  C  on  PO_Exp  =  0.06   S.E. =  0.11  t  =  0.5  df=  81   with p =  0.62
Direct effect (c') of  C  on  PO_Exp  removing  Peer  =  0.06   S.E. =  0.11  t  =  0.52  df=  80   with p =  0.6
Indirect effect (ab) of  C  on  PO_Exp  through  Peer   =  0
Mean bootstrapped indirect effect =  0  with standard error =  0.01  Lower CI =  -0.03    Upper CI =  0.02

Total effect(c) of  E  on  PO_Exp  =  -0.27   S.E. =  0.2  t  =  -1.35  df=  81   with p =  0.18
Direct effect (c') of  E  on  PO_Exp  removing  Peer  =  -0.25   S.E. =  0.2  t  =  -1.21  df=  80   with p =  0.23
Indirect effect (ab) of  E  on  PO_Exp  through  Peer   =  -0.02
Mean bootstrapped indirect effect =  -0.02  with standard error =  0.05  Lower CI =  -0.11    Upper CI =  0.08

Total effect(c) of  A  on  PO_Exp  =  -0.05   S.E. =  0.2  t  =  -0.27  df=  81   with p =  0.79
Direct effect (c') of  A  on  PO_Exp  removing  Peer  =  -0.06   S.E. =  0.2  t  =  -0.32  df=  80   with p =  0.75
Indirect effect (ab) of  A  on  PO_Exp  through  Peer   =  0.01
Mean bootstrapped indirect effect =  0  with standard error =  0.04  Lower CI =  -0.07    Upper CI =  0.08

Total effect(c) of  N  on  PO_Exp  =  -0.46   S.E. =  0.39  t  =  -1.16  df=  81   with p =  0.25
Direct effect (c') of  N  on  PO_Exp  removing  Peer  =  -0.42   S.E. =  0.41  t  =  -1.03  df=  80   with p =  0.3
Indirect effect (ab) of  N  on  PO_Exp  through  Peer   =  -0.04
Mean bootstrapped indirect effect =  -0.03  with standard error =  0.08  Lower CI =  -0.21    Upper CI =  0.12
R = 0.18 R2 = 0.03   F = 0.44 on 6 and 80 DF   p-value:  0.876


Total effect(c) of  OS  on  PO_Exp  =  0.54   S.E. =  0.28  t  =  1.95  df=  81   with p =  0.054
Direct effect (c') of  OS  on  PO_Exp  removing  Peer  =  0.53   S.E. =  0.28  t  =  1.91  df=  80   with p =  0.06
Indirect effect (ab) of  OS  on  PO_Exp  through  Peer   =  0.01
Mean bootstrapped indirect effect =  0.01  with standard error =  0.04  Lower CI =  -0.07    Upper CI =  0.09

Total effect(c) of  CS  on  PO_Exp  =  -0.41   S.E. =  0.35  t  =  -1.18  df=  81   with p =  0.24
Direct effect (c') of  CS  on  PO_Exp  removing  Peer  =  -0.37   S.E. =  0.36  t  =  -1.03  df=  80   with p =  0.3
Indirect effect (ab) of  CS  on  PO_Exp  through  Peer   =  -0.04
Mean bootstrapped indirect effect =  -0.04  with standard error =  0.1  Lower CI =  -0.24    Upper CI =  0.16

Total effect(c) of  ES  on  PO_Exp  =  0.02   S.E. =  0.3  t  =  0.07  df=  81   with p =  0.95
Direct effect (c') of  ES  on  PO_Exp  removing  Peer  =  0   S.E. =  0.3  t  =  0  df=  80   with p =  1
Indirect effect (ab) of  ES  on  PO_Exp  through  Peer   =  0.02
Mean bootstrapped indirect effect =  0.02  with standard error =  0.06  Lower CI =  -0.09    Upper CI =  0.16

Total effect(c) of  AS  on  PO_Exp  =  -0.09   S.E. =  0.3  t  =  -0.31  df=  81   with p =  0.76
Direct effect (c') of  AS  on  PO_Exp  removing  Peer  =  -0.1   S.E. =  0.3  t  =  -0.32  df=  80   with p =  0.75
Indirect effect (ab) of  AS  on  PO_Exp  through  Peer   =  0
Mean bootstrapped indirect effect =  0.01  with standard error =  0.03  Lower CI =  -0.04    Upper CI =  0.08

Total effect(c) of  NS  on  PO_Exp  =  0.35   S.E. =  0.4  t  =  0.89  df=  81   with p =  0.37
Direct effect (c') of  NS  on  PO_Exp  removing  Peer  =  0.34   S.E. =  0.4  t  =  0.85  df=  80   with p =  0.4
Indirect effect (ab) of  NS  on  PO_Exp  through  Peer   =  0.01
Mean bootstrapped indirect effect =  0.01  with standard error =  0.06  Lower CI =  -0.09    Upper CI =  0.14
R = 0.28 R2 = 0.08   F = 1.11 on 6 and 80 DF   p-value:  0.362
```

```
Total effect(c) of I on PO_Exp = -0.3 S.E. = 1.46 t = -0.2 df= 81 with p = 0.84
Direct effect (c') of I on PO_Exp removing Peer = -0.43 S.E. = 1.49 t = -0.29 df= 80 with p = 0.77
Indirect effect (ab) of I on PO_Exp through Peer = 0.14
Mean bootstrapped indirect effect = 0.15 with standard error = 0.28 Lower CI = -0.3 Upper CI = 0.81

Total effect(c) of EG on PO_Exp = 1.64 S.E. = 1.62 t = 1.01 df= 81 with p = 0.31
Direct effect (c') of EG on PO_Exp removing Peer = 1.53 S.E. = 1.64 t = 0.93 df= 80 with p = 0.35
Indirect effect (ab) of EG on PO_Exp through Peer = 0.11
Mean bootstrapped indirect effect = 0.09 with standard error = 0.25 Lower CI = -0.39 Upper CI = 0.64

Total effect(c) of R on PO_Exp = 0.5 S.E. = 1.05 t = 0.48 df= 81 with p = 0.63
Direct effect (c') of R on PO_Exp removing Peer = 0.41 S.E. = 1.07 t = 0.38 df= 80 with p = 0.7
Indirect effect (ab) of R on PO_Exp through Peer = 0.1
Mean bootstrapped indirect effect = 0.12 with standard error = 0.23 Lower CI = -0.24 Upper CI = 0.69

Total effect(c) of D on PO_Exp = -0.63 S.E. = 1.5 t = -0.42 df= 81 with p = 0.68
Direct effect (c') of D on PO_Exp removing Peer = -0.51 S.E. = 1.53 t = -0.33 df= 80 with p = 0.74
Indirect effect (ab) of D on PO_Exp through Peer = -0.12
Mean bootstrapped indirect effect = -0.08 with standard error = 0.26 Lower CI = -0.65 Upper CI = 0.45

Total effect(c) of T on PO_Exp = 1.87 S.E. = 1.06 t = 1.76 df= 81 with p = 0.081
Direct effect (c') of T on PO_Exp removing Peer = 1.88 S.E. = 1.06 t = 1.77 df= 80 with p = 0.08
Indirect effect (ab) of T on PO_Exp through Peer = -0.02
Mean bootstrapped indirect effect = -0.03 with standard error = 0.12 Lower CI = -0.31 Upper CI = 0.2
R = 0.28 R2 = 0.08 F = 1.18 on 6 and 80 DF p-value: 0.326


Total effect(c) of IS on PO_Exp = 5.15 S.E. = 2.72 t = 1.89 df= 81 with p = 0.062
Direct effect (c') of IS on PO_Exp removing Peer = 5.04 S.E. = 2.72 t = 1.85 df= 80 with p = 0.068
Indirect effect (ab) of IS on PO_Exp through Peer = 0.11
Mean bootstrapped indirect effect = 0.07 with standard error = 0.36 Lower CI = -0.68 Upper CI = 0.87

Total effect(c) of EGS on PO_Exp = 1.71 S.E. = 3.26 t = 0.52 df= 81 with p = 0.6
Direct effect (c') of EGS on PO_Exp removing Peer = 2.1 S.E. = 3.29 t = 0.64 df= 80 with p = 0.53
Indirect effect (ab) of EGS on PO_Exp through Peer = -0.38
Mean bootstrapped indirect effect = -0.34 with standard error = 0.53 Lower CI = -1.58 Upper CI = 0.59

Total effect(c) of RS on PO_Exp = -1.7 S.E. = 2.72 t = -0.62 df= 81 with p = 0.54
Direct effect (c') of RS on PO_Exp removing Peer = -1.69 S.E. = 2.72 t = -0.62 df= 80 with p = 0.54
Indirect effect (ab) of RS on PO_Exp through Peer = -0.01
Mean bootstrapped indirect effect = 0 with standard error = 0.32 Lower CI = -0.67 Upper CI = 0.68

Total effect(c) of DS on PO_Exp = -1.21 S.E. = 2.11 t = -0.58 df= 81 with p = 0.57
Direct effect (c') of DS on PO_Exp removing Peer = -1.48 S.E. = 2.13 t = -0.7 df= 80 with p = 0.49
Indirect effect (ab) of DS on PO_Exp through Peer = 0.27
Mean bootstrapped indirect effect = 0.27 with standard error = 0.37 Lower CI = -0.31 Upper CI = 1.13

Total effect(c) of TS on PO_Exp = 2.08 S.E. = 1.65 t = 1.26 df= 81 with p = 0.21
Direct effect (c') of TS on PO_Exp removing Peer = 2.28 S.E. = 1.66 t = 1.37 df= 80 with p = 0.17
Indirect effect (ab) of TS on PO_Exp through Peer = -0.21
Mean bootstrapped indirect effect = -0.22 with standard error = 0.35 Lower CI = -1.1 Upper CI = 0.31
R = 0.31 R2 = 0.09 F = 1.38 on 6 and 80 DF p-value: 0.226
```