CHALMERS |  UNIVERSITY OF GOTHENBURG

Group Development and Software Engineering Performance

The Connections Between Group Maturity, Software Development Velocity and Planning Effectiveness

Master of Science Thesis in Software Engineering

KHALED AL-SABBAGH

The University of Gothenburg
Chalmers University of Technology
Department of Computer Science and Engineering
Gothenburg, Sweden, June 2016

Group Development and Software Engineering Performance
Khaled Al-Sabbagh

Supervisor: Lucas Gren, Department of Computer Science and Engineering
Examiner: Eric Knauss, Department of Computer Science and Engineering

Master's Thesis 2016:NN
Department of Computer Science and Engineering
Division of Software Engineering
The University of Gothenburg
Chalmers University of Technology SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2016

Abstract

Background: Empirical evidence regarding the connection between group development (maturity) and the success of software development teams is lacking. Since software development is primarily a product of team effort that involves human interaction, it is important to investigate the influence of group development on the performance of software development teams.

Objective: The purpose of this research is to gain a qualitative and quantitative understanding of how performance of software teams relates to group development. More specifically, the analysis of group maturity and its association with velocity and planning effectiveness is the objective of this research.

Method: The Group Development Questionnaire (GDQ) was given to four participating work groups from company A to assess their group development levels. The work groups' responses to the survey were checked for correlation with development velocity and planning effectiveness. Additionally, semi-structured interviews were conducted with 16 individuals to explore issues about their group maturity and to increase the validity of our findings.

Results: The group maturity measurement had a strong association with the planning effectiveness measurement and showed a significant convergent validity, which means that a more mature team is also a more effective one in planning its requirements. On the other hand, the correlation results between group development and velocity showed no convergent validity, i.e, the group maturity is not related to the velocity of work groups in accomplishing tasks.

Conclusion: We conclude that the dynamics within software development teams relate to their ability to deliver the expected outcome as planned, but does not relate to their ability to work faster.

Acknowledgements

I would like to thank various individuals for their guidance and support throughout the conduct of this thesis. These include

My supervisor, Ph.D. student Lucas Gren, for helping me throughout the process by ensuring a critical review of the work, and for providing the resources needed to complete this research.

The Swedish Institute for providing me the opportunity to study, live, and experience the Swedish culture,

My love, Hanouf, for standing up for me during the difficult and stressful moments,

My mother, father, and sisters for raising me up to become who I am today, and

My friends, especially those in Sweden. We have made it together through two winters in Sweden.

Khaled Al-Sabbagh, Gothenburg, Sweden June 9, 2016

We don't regard any scientific theory as the absolute truth.

— Kenneth R. Miller

Contents

List of Tables

List of Figures

1

Introduction

Groups, like humans, move through successive phases; they tend to advance and regress. A group is defined as three or more members that interact with each other to perform a number of tasks and achieve a set of common goals [17], which means they have neither developed a structure nor created common goals. A team, on the other hand, has developed both the goals and the means to achieve them [38]. The emphasis on the importance of arranging work in a group-form emerged, in part, from the growing awareness of the role of groups in facilitating or blocking individual and organizational effectiveness [9]. This is because when a group works well, more work can be accomplished compared to other work methods [17]. As a result, organizations are counting on teams as the main asset for accomplishing goals [9].

Group development can be defined as the process in which a group navigates through a number of stages until it becomes a mature team. In other words, teams in organizations could actually be groups or work groups rather than mature teams. In fact, 83% of teams that were assessed in a study were found to be work groups [38]. A team, therefore, is one that has successfully navigated the earlier stages of group development and has emerged as a mature, high performing unit capable of achieving common goals [35].

Although group development research began in the 1930 [12], it is the period from the 1950's onwards that exhibited a plethora of research on small groups [12] whereby various systems for behavioural analysis within groups were proposed. The first group development model postulated by Bales and Strodtbeck [7] explains that verbal and non-verbal behaviour of individuals are either an emotional reaction, whether positive or negative, or task questions and answers. Several other models were proposed by other researchers until Tuckman [32] performed a comprehensive analysis on these models in 1965. His

suggested model includes 5 stages of group development, each characterized by a social domain and a task domain. These stages are "Forming", "Storming", "Norming", "Performing", and "Adjouring" [33].

The work of Susan Wheelan on group development research helped determine the common threads amongst group development models and postulate the basis for the Integrated Model of Group Development **(IMGD)**. In this model, a group is believed to go through 5 successive stages of development, namely "Inclusion and Safety", "Counter-dependency and Fight", "Trust and Structure", "Productivity and Work", and "Termination". The IMGD is considered the most acceptable model that describes the group development behavioural patterns [35]. The importance of this model, in our own opinion, lies in the fact that it proposes a statistically validated instrument that measures the maturity of a given group at a given time, namely the Group Development Questionnaire **(GDQ)**. The instrument, developed by Susan Wheelan in 1993, contains four sub-scales bases on the stages from her IMGD. The "Termination" stage is not addressed in the GDQ, since the tool aims at measuring maturity with ongoing groups. Each sub-scale contains 15 items which measure the amount of energy a group is spending on the corresponding stage of IMGD. A comprehensive validation study on the GDQ, performed by Wheelan and Hochberger [37], revealed reliability scores for scales one through four to be 0.74, 0.87, 0.69, and 0.82 respectively, which indicate a good overall reliability of the GDQ items. In this research, we used the IMGD model as the theoretical framework for understanding the group dynamics of the participating work groups and the GDQ was used to assess their group development (maturity) stages.

While team performance is defined as the extent to which a team is able to meet cost, time, and quality objectives, a differentiation between two variables, effectiveness and efficiency, needs to be made in order to gain insights into the actual performance of software teams. Effectiveness refers to the team's adherence to the predetermined quality of a product [18]. In a software context, effectiveness could be the robustness or reliability of functionality in software. Efficiency, on the other hand, is evaluated in terms of team's commitment to schedules [18], like launching software on the target date and within budget. Therefore, effectiveness reflects a comparison of actual versus intended outcomes, whereas efficiency ratings are based on a comparison of actual versus intended inputs [18].

The performance of software teams can be measured using two approaches: objective and subjective [24]. The subjective approach relies on the perception of key stakeholders (e.g., the customer) on the performance of a given team whereas the objective approach relies on a quantitative assessment of team performance [24]. One way to measure team performance is to look at the team's adherence to schedule. In software teams that adopt scrum in their development, planning occurs on a sprint level, where all sizes of completed work items are collected at the end of a sprint to determine the velocity of the team [8]. The value of the completed work is only recognized when the work gets accepted by the product owner at the end of the sprint. In other words, no points are given for

any work done until it gets accepted. Based on this, we used the Schedule Performance Indicator **(SPI)** to measure the effectiveness of the scrum teams in planning their stories and delivering the expected outcome. In this research, we used the term planning effectiveness to describe the teams' ability to deliver the planned work as expected. Also, we measured the velocity of the teams in accomplishing their scrum tasks, at the end of a given sprint, by calculating the number of hours spent on those. As a result, the velocity measurement used in this research reflects the teams efficiency in accomplishing scrum tasks while planning effectiveness reflects their ability to estimate what can be delivered within each sprint as expected.

Several studies that used the IMGD as a theoretical framework have been conducted to examine the effect of group maturity on the productivity of teams in different contexts [9, 36, 39]. This highlights the usefulness and versatility of the GDQ. However, empirical evidence regarding the influence of group maturity on the success of software engineering groups with innovative tasks is lacking. In fact, studies demonstrating a link between teamwork in the field of software engineering has just begun [17]. To the best of our knowledge, only one study investigates the link between group maturity and software performance [17]. In this study, Gren et al. [17] suggested the use of velocity as a factor to further validate their findings since the tool used in their study, Sidky's [31], is not thoroughly validated, which means that it might not even measure agility. Therefore, our research investigates the correlation between group maturity and velocity to help in addressing this gap.

The overall structure of this thesis is organized in the following way. The first Chapter presents the motivation for conducting this research. Chapter 2 provides a review of the literature on the different group development models and software performance measurement approaches. The research objectives and the method used in this research are discussed in Chapter 3. Chapter 4 provides the results of the implemented methodology. Chapter 5 discusses the results and their implications on the research community purport for group development in the software engineering context. Conclusion and future work are presented in Chapter 6.

2

Related Work

This section is organized into three broad categories: a) group development b) performance of software development teams c) correlation between group development and the agility of software development teams. In the first section, various group development theories and their corresponding instruments are discussed. The second section defines productivity in software development teams, investigates the various human factors affecting it, and reviews the different units used to measure performance in this context.

2.1 Group Development

2.1.1 Models of Group Development

Group development research began in the 1930s with the work of Lewin on group climate and conflict between groups. The study of the behavior of small groups was launched with the establishment of a research center for group dynamics in 1946 [12]. The 1950s and onwards was a period of research on small groups [12]. During that time several research groups proposed different systems for analyzing the behavior within groups. The first system, Interaction Process Analysis, describes groups' interaction as a predictable process in which groups display cognitive and affective behaviors [12]. In this system, Bales proposed classifying verbal and non-verbal behavior of individuals within a group as either task questions or answers, or positive or negative emotional reactions [12].

In 1958, Schutz integrated aspects of personality into group development [12]. He theorized that personal needs do not only affect the individuals within a group, but also apply to the group as a whole [12]. In his theory, Fundamental Interpersonal Relations Orientation (FIRO), he proposed that groups go through sequential stages of integration and resolution. These are **a)** inclusion of members also described as *in or out* issues, **b)** control activities labeled as *top or bottom* issues, **c)** affection between members also

known as *near or far*. The order of these phases is reversed at the group's termination.

While Schutz focused on the interpersonal aspects, Bion described the effect of emotional states on group development [12]. The results revealed that two levels of activity are found in groups. One level is geared towards the accomplishment of tasks, known as *work group*, whereas the other, known as *basic assumption group*, interferes with tasks achievement. Dependency, fight-flight, and pairing were identified as the emotional states that deviate a group from its work task, i.e. the basic assumption group, but these are crucial for group cohesion. These are not necessarily sequential and they can occur at anytime during the life of a group. Bion's theory was further expanded by Slater, who postulated that the themes that affect group development are the relationship of members to its leader, its need for order, and its wish for immortality [12].

In 1961, the AGIL model of group development was proposed by Parsons [35]. The model suggests that groups oscillate through sequential phases of development. The first phase is known as *sharing information versus withholding*, after which the group navigates to a phase of *integration*, characterized by group decisions making versus furthering distance. The final phase, *latent pattern maintenance*, focuses on working towards achieving the group's goals versus having individual ones [12].

An integrative theory of linear and cyclic models was first introduced in 1964 [35]. The theory postulates the existence of four primary elements in group development. *Acceptance*, which focuses on the creation of trust and the reduction of anxiety, and the growth of self-confidence among members of the group. *Data-flow*, involves the ability of a group to make decisions as a result of communicated feelings and data across its members. *Goal Information*, relates to the group's productivity as evidenced by their ability to perform problem solving and decision-making. The final element is referred to as *Control*, the degree by which members of the group are recognized as interdependent and organized [35].

A major comprehensive analysis of various group development models was conducted by Tuckman [32]. In this analysis, 50 articles on group development were reviewed based on a classification system of three elements: 1) setting (such as a laboratory group, natural group, or therapy group) 2) task or social focus 3) stage of development. The result of this analysis was a conceptual model comprising of four stages of group development in which each stage has a social realm and a task realm. The four stages proposed by Tuckman are: *forming* which is categorized by high dependency, orientation, and testing; *storming* during which resistance to both tasks and the influence of a group is apparent; *norming* in which opinions are more freely expressed; and *performing* in which a focus is on tasks' accomplishment after structural issues in a group are resolved. A review of this model was made by Tuckman in which he added a fifth stage of *adjourning* [32, 33]. This was following a review made by Mills [23] on his four-stages model suggesting adding a separation and a conclusion stage. Tuckman's theory gained empirical support by many researchers [35].

The Punctuated Equilibrium Model, designed by Gersick [15] in 1988, suggesting that groups go through an initial phase characterized by the establishment of behavioral patterns in which members of the group approach their project during their first meeting. A midpoint transition halfway through the project occurs afterward. At this point, members try to capitalize on the learning they have gained, consequently strive to make significant work progress. The final phase suggests that members of group reap the consequences of the past choices they made (positive and negative) as evidenced by the degree to which the expectations of external stakeholders were met [15].

Another group development model emerged in 1992. Similar to many previous cyclic models, it allows for circular process and does not presume a specific amount of time for any stage. The model is comprised of five stages: *discontent*, *precipitating event*, *group identification group productivity*, and *decay*. A limitation to this model is that membership in a group should be made by individual preferences rather than assignment [12].

In her review of a number of the studies that did not support the group development stage theory, Cissna [10] pinpointed a number of erroneous approaches in the methodologies adopted in these studies, consequently citing that "every group is like all groups in some respect, and like no groups in other respect". Moreover, there is ample evidence in the body of literature which support the theory of stages in group development [35]. While these models share the same view that groups face a basic set of developmental changes over time, the differences persist in the recognition and labeling of each stage and their sub-components in the group development [35].

According to Zimpfer [40], group development models fit best for groups that do not have shifting in their membership, are small in size, formed for a specific purpose, have relatively unlimited involvement with each other with less leadership structure.

2.1.2 The Integrated Model of Group Development (IMGD)

The IMGD was theorized after consolidating previous theories which proposed a unified group development model for all group types [34]. The overall goal of group development was set to establish an organized unit of members capable of working effectively to achieve specific goals. What follows is a description of the five stages, found in the IMGD, which describe the behavioral pattern of any group type [34].

Stage One The first stage is a period of *Dependency and Inclusion*, where members tend to show significant dependency on the leader in resolving new issues. At this stage, members spend a significant amount of energy to achieve a feeling of safety and inclusion in their group. Since members have not yet interacted enough to establish a cohesive relationship, they cannot rely on each other for support. As a result, members become leader-focused, in a sense that he/she will provide protection and structure for the members. Members are indulged in an exploratory phase for the sake of identifying their

roles, rules, and the structure within the group. Their exploration is characterized by being tentative and polite since they fear being rejected or excluded. Their actions are seldom independent and a reliance on a leader for support is evident. Members live with the assumption that other members may try to secure their own safety at the expense of others. This assumption results in a defensive and polite attitude towards each other [35].

Stage Two The second stage is referred to as *Counterdependency and Fight*. At this stage, members feel freer to express conflict between each other or among members and leaders since some needs for safety have been achieved in the previous stage. The group tries to free itself from being leader focused, and tends to fight about the group's goals and procedures. Coser explained that conflict is an important part for the development of cohesion in the group, as it provides the opportunity for setting the psychological boundaries, which facilitate the establishment of goals, shared values, norms, culture, and structure [12]. The occurrence of conflict is a result of the members' attempt to reach a unified direction out of the many divergent viewpoints. The rise of coalitions between members who share similar values and ideas is very much apparent. In other words, the group becomes unintentionally split into at least two different subgroups. On the other hand, a prolonged period of conflicts is likely to result in the destruction of the group [35].

Stage Three After navigating the inevitable stage of conflicts, the group is now ready to go through civilized and more mature dialogue. Communication is more open and members' trust and cooperation increase. Feedback and information sharing increase rather than being kept as a way to gain power. The aforementioned characteristics consolidate a more solid and positive relationship between members, which allow the group to carry out more mature negotiation processes to set their goals and procedures. Although still occurring, the amount of conflicts at this stage lessens in density. Essentially, the group is at a stage where it is designing and preparing itself to start working effectively. Although work occurs in all the stages of group development, the group's focus on structure and goals at this stage significantly increases the group's capacity to work productively [35].

Stage Four As soon as the goals and structure of the group are set from the previous stage, the group's focus is diverged into getting the work done well at the same time as the group cohesion is maintained, and remains cohesive while engaging in task-related conflict.

Stage Five Most groups, temporary and continuous, experience an ending point at some point in the course of their lives. At the ending point, functional teams tend to give feedback about each other [35]. It has been reported that this type of processing is important for individual members since it enhances their ability to work effectively in

the future. Impending termination of a group alters its structure and is likely to result in the group's regression to earlier stages of the group development [35].

2.1.3 Tools for Measuring Group Development

Various self-reporting instruments have been developed in the last few decades to aid team building and highlight the importance of group development. We reviewed some of those instruments since they reported some evidence of reliability and validity [12]. Below are some of these tools that we have come across in our literature review.

The Group Development Assessment (GDA) The GDA instrument is based on the group development model developed by Jones, which is based on Tuckman's theory [35]. It suggests that groups navigate through four phases: the immature group, the fractionated group, the sharing group, and the effective team. The GDA has two dimensions: one for task behaviour and the other is for process behaviors. The task dimension is comprised of four phases: orientation, organization, open data flow, and problem solving, while the process dimension includes phases that are characterized by dependency, conflict, cohesion, and interdependency. The GDA instrument contains 40 items that attempt to evaluate group development on four scales. According to Wheelan [35] the reliability and validity of this instrument could not be established.

The Team Development Inventory (TDI) Originally developed for use with work groups [35], the TDI instrument requires each member to rank all members of a group on eight dimensions, which are linked to teamwork. These dimensions are: participation, collaboration, flexibility, sensitivity, risk taking, commitment, facilitation, and openness. The instrument was initially developed to help members think about how well their group is performing and to aid in identifying ways to improve the group's productivity. A major shortcoming to the tool is that it only takes into account interpersonal dimensions which are insufficient to measure group development. The developer of the instrument gave no regard to testing its reliability and validity [35].

The Group Development Stage Analysis The tool determines the stage of group development defined as one of the following: orientation, dissatisfaction, resolution, and production [35]. It relies on the eight characteristics that are exhibited by high performing teams and lists four options for each of them, each corresponding to one of the development stages of groups. These eight characteristics are productivity, roles and goals, empowerment, open communication, morale, empathy, flexibility, and recognition. Like the tool before, reliability, validity, and norms were not reported [35].

The Group Attitude Scales The tool, which measures the members' attraction to their group at a given time, was developed in 1986 by Evans and Jarvis [13]. According to its developers, it is based on the principle that the attraction of members to their group changes over time in a way that reflects the development stage of a group. This

means that this tool cannot measure group development directly, but rather it is inferred [35].

Group Development Questionnaire (GDQ) Based on the IMGD, the GDQ was developed after being subjected to a number of statistical tests for reliability and validity [13, 37]. The 60-item instrument contains a total of four scales. Each scale contains fifteen items, which corresponds to a single stage in the IMGD. The instrument does not assess the termination stage since it is meant for use with existing groups only. Items on scale I measure the amount of energy a group is spending in dealing with issues of inclusion and dependency. Items on scale II seek to measure the amount of group focus on issues of counterdependency and conflict. The group's current level of trust and structure is measured by scale III, which corresponds to stage three in the group development model whereas the group's maturity on the "work and productivity" is measured by scale IV [34].

Internal consistency tests for each fifteen-item scale were performed to ensure that all items within each scale were consistent [37]. Furthermore, the instrument was correlated with the Group Attitude Scale to establish concurrent validity [13]. The results indicated a significant concurrent validity between the two measures. Moreover, criterion-related validity was investigated. Results showed that groups who ranked high on productivity had significantly lower scores on the first and second scales of the GDQ. Similarly, groups that ranked high on productivity had significantly scored high scores on the third and fourth GDQ scales [37].

Group Development Observation System (GDOS) Similar to GDQ, the GDOS is a self reporting instrument that is based on the IMGD theory [12]. It relies on observing patterns in communication and independent verbal statements between group members. The method requires at least two raters to observe each communicated verbal statement to increase the accuracy of the analysis. Each complete thought is placed into one of seven categories: Dependency, Counterdependency, Fight, Flight, Pairing, Counterpairing or Work. Incomplete thoughts are placed under the category Unscorable. The raters are required to be trained in using the tool in order to increase reliability [12].

2.1.4 Application of IMGD in Different Contexts

Several studies, adopting the IMGD as a theoretical framework, have been conducted to examine the effect of group maturity using GDQ on the productivity of teams in different contexts, highlighting the usefulness and versatility of this tool. One study looked at the learning outcomes of students in schools as measured by math, reading, and achievement ranks and the maturity level of school administrators as measured by GDQ. The study, which involved 292 faculty members, concluded that there is a significant relationship between the functioning of faculty group and students' learning outcomes [39]. Similarly, another study investigated the relationship between the level of teamwork in the Intensive Care Unit (ICU) and the patients' outcome. Data were analyzed by correlating the ICU

mortality rate (patients' risk of dying in the hospital using a mortality prediction system) and stage of group development of 394 staff members in the participating 17 ICU in nine hospitals. A significant correlation was identified between a unit's stage of group development and that unit's mortality rate [36]. As the staff perception of their level of group development increased, mortality rate in their unit decreased, i.e. the higher the level of group development a group is, the fewer deaths occurred. A third study used the GDQ to plan an appropriate intervention to improve the effectiveness of three work groups in semi-governmental organizations. In this study, the group development scores of the three groups on the four GDQ scales were determined, an appropriate intervention to improve the teams' effectiveness was devised, and a three-months follow-up plan was set to determine whether significant positive changes had occurred. The intervention revolved around the issues revealed from the GDQ data. For example, member discussion was encouraged to focus on the importance of hearing opinions from all team members, reducing the dominance of the leader without creating a hostile environment, etc. Paired samples tests were employed to determine whether the intervention resulted in a positive significance on the fourth GDQ scale and effectiveness ratio within each group from pre to post-tests [9].

2.1.5 Summary

These various models suggest that interactions within a group display predictable patterns and that human interactions affect work performance within a group. These models have been the result of mainly observation of groups functioning in different settings (laboratory group, natural group, therapy group, etc.). The culmination of these models helped Susan Wheelan formulate the IMGD (Integrated Model for Group Development) which, unlike many other models, developed an instrument, the GDQ, to capture data on how groups behave and progress relative to stages of group development. The GDQ has been studied thoroughly relative to validity and reliability [37], which makes it a desirable choice for investigating the maturity level of work groups.

2.2 Software Development Team Performance

This section addresses the soft factors which influence the performance of software development teams and discusses the different approaches for measuring their performance.

2.2.1 Soft Factors Affecting Software Team Performance

Performance in software development teams, like all other teams, is defined as the extent to which a team is able to comply to cost, time, and quality objectives [25]. The factors that influence the performance of software teams were classified by Purna Sudhakar et al. [25] into technical, non-technical (soft), organizational, and environmental as shown in Figure 2.1. Since our research focused on exploring the relationship between group maturity and software performance, only soft factors were considered. Below are some of

the non-technical factors that positively or negatively affect the performance of software development teams, as we came across in our literature review.

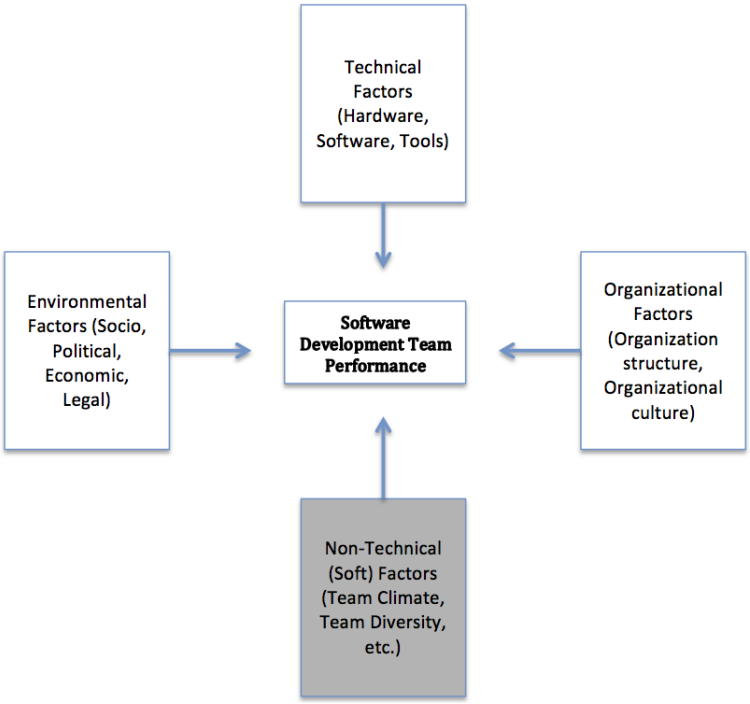


Figure 2.1: Influencing Factors on Software Teams Performance
Taken from [25]

Team climate According to Anderson and West [3], climate is defined in two ways. One that refers to the individual's understanding and cognitive representation of the work environment and this is referred to as "cognitive schema approach". The second refers to the individuals' collective perception of the policies, practices and procedures within an organization. A climate that supports innovation in a team must take four elements into account:

- A vision
- A safe working environment that supports the participation of individuals
- An established accountability system whereby team performance is evaluated and work methods are consequently improved
- A work environment whereby teams are supported to share new ideas

In a study conducted by Ganesh and Gupta [14], a positive team climate is shown to be a vital factor in determining the performance of teams. In another, but related study, held in collaboration with 35 software development teams of students in a university in Spain, team climate was identified as a mediating variable that causes mediation between the team size and team performance [25].

Team diversity Although team diversity stems from a myriad of reasons such as education, experience, ethnicity, culture, skills, age, gender, etc, Liang et al. [21] conducted research on 30 software development teams and found conflicting results in regards to the effects of diversity on the performance of the teams. Namely, knowledge diversity in the teams positively affects team performance whereas value diversity negatively affects the performance of teams [21].

Team innovations The relationship between members of a team can have a profound effect on the team's creativity, which is reflected on the team's performance. In a study carried out by Bain et al. [6] on 54 research and development teams in Australia found that a positive team climate results in better innovation and team performance.

Team member competencies and characteristics Competencies can be classified into two categories: technical and personal competencies [4]. Asproni [4] explained that personal competencies can sometimes outweigh the technical in their influence on team performance. For example, a team of junior programmers with high personal competencies can perform better than a team of senior software developers. Similarly, another research conducted by Huckman et al. [19] in cooperation with a software development company in India confirmed that having team members who have previously worked with each other has a positive influence on team performance regardless of the years of experience of team members.

Team leader's behavior The role that a project manager plays is pivotal in dealing with uncertainty and problems that a team might face. Thus, it contributes positively to team performance and to conformance to budget and schedule [19].

Top management support Team empowerment is expressed as a key element in positively influencing team performance. It is the responsibility of top management to empower the working teams within the organization and also to ensure that their work is well noticed within and even outside the organization [16].

Conflicts in team In a study conducted by Sawyer [30] on 40 software development teams at a hardware and software manufacturing company found that team members' characteristics and the intra-group conflicts explained half of the variance between good and bad performing teams. Different factors might accommodate for intra-group conflicts. These include, but are not limited to, differences between the behaviors of individuals, social pressures, and interdependence in teams. The result of the study concluded that intra-group conflicts have a negative impact on the performance of software teams. Sawyer [30] advised on establishing a constructive conflict management approach within team members in order to improve their performance.

The IMGD model characterized a productive work group as one that has navigated the earlier stages of group development and has become more focused on building trust and structure, and work and productivity. As the IMGD describes some of the behavioral aspects manifested by groups in all the stages of group development, these aspects show similarity with some of the soft factors described above. For example, Wheelan [34] described stage two in group development as a period of fight and counter-dependency where conflicts between members are prominent, which negatively affects the group productivity. Likewise, Sawyer [30] suggested that conflict in teams is a significant factor that yields to a deterioration in team performance. Moreover, Wheelan [35] described a stage three group as one whose members communicate more openly, cooperate more effectively, and share information and feedback. Similarly, Anderson and West [3] proposed that a clarity on the team's goals, a safe working environment that supports idea sharing and the participation of individuals are key factors in poistively affecting the performance of software teams.

2.2.2 Software Team Performance Measurement

Ong et al. [24] identified two approaches in which the performance of software development teams can be measured: objective and perceptual or subjective. The first approach includes measuring function points, object points, use case points, kilo lines of code, and defect rate. Sawyer [30] explained that perceptual measures, such as quality of the product and satisfaction with the product should be taken from external stakeholders in order to account for self-bias. The perceptual or subjective approach relies on the group's perception of their team's performance and is based on items such as "*our group is very productive, we work well as a team, and the quality of our work is very good*" [5]. Table 2.1 classifies the two approaches.

Table 2.1: Approaches for Measuring Software Teams performance
Taken From [25]

M1 - Objective measures	M2 - Subjective/Perceptual Team performance Ratings By:
• Function Point • KLOC • Object Points • Use Case Points • Defect Rates • Defect Density • Quantitative Metric	• Team Members • Management • Customer

Similarly, Ramasubbu and Balan [26] concluded in another study that software teams performance is measured in terms of function points per person hour and conformance to quality. The conformance quality refers to the defect rate claimed by the customer during acceptance testing. Team's adherence to budget and schedule is another measure of performance reported by [8]. According to Purna Sudhakar et al. [25], a team's performance is a function of what individual team members are doing. More specifically, a successful team is one that is characterized by the following: 1) shared leadership roles, 2) specific and clear goals, 3) mutual accountability, 4) collective problem solving.

Albero Pomar et al. [1] proposed two techniques for predicting future performance of scrum software teams. The first approach relies on plotting the accrued velocity for all previous sprints in order to identify the trends (downward or upward) of performance. The second approach depends on calculating a confidence interval to comprehend the probability of future velocities. They also proposed exploiting a traditional (non-agile) project management metric to gauge the amount of completed work over the planned work. The metric is calculated as the ratio of the total earned points over the total points planned in the sprint planning meeting [1].

$$SchedulePerformanceIndicator = \frac{EarnedPoints}{PlannedPoints} * 100$$

2.2.3 Summary

In summary, factors that affect the performance of software development teams include technical, non-technical, organizational and environmental. In this research, only soft factors or non-technical ones are considered since the scope of this research is to investigate the association between the maturity in work groups and their performance. The

means to measure software development performance can be broadly categorized into objective and subjective approaches, which were both employed in this research. First, the performance of the work groups was subjectively measured from the GDQ in which they were asked about their perceptions about their productivity. On the other hand, the planning effectiveness and velocity of the work groups were more objectively measured using the SPI metric and by calculating the sum of hours spent on delivering selected tasks in an iteration, respectively.

2.3 Correlation between Group Maturity and Agility

In a study that involved 45 employees and twelve managers in two large multinational companies, Gren et al. [17] investigated the existence of a connection between group maturity, measured by GDQ, and agility, measured by Sidkey survey [31]. The study found a positive significant correlation between group maturity and agility, highlighting the importance of considering group maturity when the agile practices are introduced. In this study, Gren et al. [17] suggested the use of velocity as a factor to further validate their findings since the tool used in their study, Sidky's, is not thoroughly validated, which means that it might not even measure agility. Therefore, our research, which investigates the correlation between maturity and velocity, can help in addressing this gap.

3

Research Method

This chapter presents the details of a case study conducted at a Swedish-based software company, which will be anonymously named A. It outlines the purpose of this research, addresses the research questions, and provides details of the method used.

3.1 Research Purpose

The objective of this research is to investigate and analyze whether group maturity is related to the performance of software development teams. More specifically, performance is examined by measuring both planning effectiveness and development velocity of four participating work groups from company A.

3.1.1 Research Questions

This study aims to contribute to answering of the following questions.

1. What is the association between group maturity and planning effectiveness?
2. What is the association between group maturity and software development velocity?

Group maturity in the four participating work groups was measured using GDQ. The software development velocity was in turn measured by calculating the number of hours spent on developing scrum tasks for each member in the participating teams whereas planning effectiveness was assessed by using the Schedule Performance Indicator metric.

3.2 Case Study

A combination of qualitative and quantitative data was used in this study. According to Runeson and Höst [28], a case study is a suitable methodology for software engineering research, since it provides a deeper understanding of the phenomena under study. As a result, a case study was selected as the most suitable means for conducting this research. Using both qualitative and quantitative data provides an in-depth understanding to the way the participating groups are functioning and facilitates a better comparison between the groups.

3.3 Subject Selection

3.3.1 Company Description

Company A is a Swedish company with 1,400 employees located in four different countries. The company is active in the fields of software development, quality management, qualified project management and business development, mobility and web development. The increasing growth of the company's market share has stemmed a need for the company to work towards achieving more efficient and effective ways to develop its products. Part of their development effort is spent on developing the group dynamics in their software development teams. This research was conducted in collaboration with the company's staff at their branch located in Gothenburg, Sweden.

3.3.2 Work Groups in Company A

First, we would like to reinforce the distinction, made in section one, between teams and work groups for the purpose of clarifying the terms we used in this research. A team is a structured group of individuals who share well-defined common goals that require coordinated interactions in order to effectively accomplish their tasks. A work group, on the other hand, is one in which members accomplish their tasks successfully, but not necessarily coordinate well and share the same goals [20]. Accordingly, we decided to use the term "work groups" to refer to the participating groups in this research. Additionally, we gave anonymous names to the work groups to keep their identity unknown.

Four software development work groups adopting scrum participated in this research. Members' age ranged from 20 to 60 years. All the development work groups were cross-functional, which means that each team has all the skills needed to build products and that each member is specialized in doing more than just one role. All the work groups receive work packages, analyzed and defined by company B, which acts as the main customer for company A. These work packages shape out the groups' product backlogs, which contain a number of requirements, written in the form of user stories, from which teams select and plan their development cycles (or sprints) respectively. The assignment of work packages to the work groups is done based on the their competence level. This was communicated by the gatekeeper to the researcher during a meeting at company A.

Estimations of user stories are done using planning poker, which is used to estimate complexity in unit of points for either new features or change requests. At the end of each sprint, all the points assigned to finished stories are added together to what is known as velocity. Each work group collaborates closely with a designated product owner assigned by company B to represent the business, prioritize requirements, and conveys the product vision. Participating groups use a web-based project management and issue tracking tool. This allows them to manage their projects and visualize their work progress at any point in time. Stories are located at the leftmost part of the UI and are moved to the right as stories progress towards completion. This UI is divided into seven columns, starting from the far left: new, in progress, needs review, blocked, closed, and rejected. The column in progress indicates scrum tasks that have been assigned to an individual for development. The column "blocked" contains all the stories that are temporary blocked because of other external dependencies or the absence of the assignee. On the other hand, column "closed" refers to the stories that were completed by members.

3.3.3 Data Collection

Permission to begin the data collection was arranged with the company's administration following the signing of a Non Disclosure Agreement. The selection of software development work groups was carried out with help from a "gatekeeper" at the case company. Accordingly, the author did not have any personal relationship with the subjects. Data were collected from the work groups (N=4) at their work site during regularly scheduled meetings, with all members of each respective group present. To guarantee the anonymity of the participating work groups, we avoided stating any information that would indicate their identity. We used multiple data sources in this case study in order to increase the validity of the findings. Below are the data collection steps arranged in chronological order.

Unstructured Interviews Brief interviews of approximately 15 minutes each with the scrum master of each work group were conducted at the onset of the data collection process. These interviews allowed the author to gain a better understanding of the context of the groups' work and to schedule for the GDQ fill-out sessions and semi-structured interviews with the 19 participants from the four work groups. Some scrum masters were interviewed twice over the course of the research as new issues emerged.

Groups Maturity To examine the maturity level of the participating groups, the GDQ, discussed in Section 2, was used to obtain the members' perception about how each group is functioning. Individuals were requested to answer the sixty questions of the GDQ. Additionally, some demographic information and their perception about their group's productivity were requested to be filled out by the participants. The groups' responses to these demographic questions aided investigations into the impact of educational background, time in the company, and members age on the groups' perceptions

about their productivity. All the GDQ fill-out sessions occurred during the last week of the group's ongoing sprints. This time was chosen to give the work groups the longest time possible in the sprint to resolve any issue related to their dynamics.

Development Velocity In this research, the velocity of the four participating work groups in accomplishing scrum tasks, at the end of a Sprint, was measured by calculating the sum of hours spent on implementing a number of scrum tasks that were arbitrary selected. This method of measurement was discussed and approved by Company A. Also, scrum tasks were chosen over user stories since tasks, unlike stories, share similar complexity as each corresponds to a small unit of work planned by a scrum team [11].

To measure the velocity of the work groups, access to their task boards was granted and data about velocity was collected during the same sprint when the GDQs were administered. This typically involved the author visiting the site about three days a week (averaging about 10 hours per week) for a period of one and a half months. For each group, an average of 40 completed tasks were arbitrary selected whereby eight tasks on average were taken per individual. Consequently, the difference between the end and start time (in unit of hours) for each of those tasks was computed. The duration in which a task gets "blocked" is deducted from the time spent on completing tasks. A given task may get blocked if there are external dependencies that temporary prevent the development of the task or if the assignee was on leave. Subsequently, the mean value of tasks accomplishment, for each work group, was calculated and recorded as their velocity.

Planning Effectiveness Since all of the four participating work groups adopt scrum as their development methodology, they decide what can be accomplished in each sprint following the poker planning approach. Accordingly, teams take into consideration the complexity of stories, the group's availability, and their technical competence level in planning what they can commit to in each sprint. The planning effectiveness of the work groups was measured using the Schedule Performance Indicator metric, which calculates the ratio of their total earned points over the total planned points for a given sprint.

$$SchedulePerformanceIndicator = \frac{EarnedPoints}{PlannedPoints} * 100$$

The mean planning effectiveness for all the sprints, which were selected according to two criteria, for each work group was calculated. The first criterion for the sprint selection is that the structure of the work groups remained unchanged, that is, no individuals joined or left the work group. The second is that their maturity level remained stable. This was confirmed by the interviewed group members during the semi-structured interviews when participants were asked "For how long has the team's maturity level been stable?" allowing for an estimation of the duration that their work group maturity has not changed. Table 3.1 shows the total number of planned versus earned story points for each selected sprint for all the participating work groups. As can be seen, the number of sprints from which the planned and earned points were collected varied considerably.

This reflects the difference in the duration that a given group's maturity remained unchanged. For example, the responses of the majority of members from group A during the semi-structured interview, revealed that their maturity has remained unchanged over the past ten sprints, in their opinion. Therefore, data from this period only was collected. On the other hand, the majority of members of work groups C and D agreed that their maturity has remained unchanged over the past four sprints. Therefore, the planned and earned points were collected from those sprints only. For more details on the groups' responses to their maturity levels, refer to section 4.4.

Table 3.1: Planned vs. Earned Points

	Sprint	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Group A	Planned Points	4	3	5	7	4	6	2	4	2	8
	Earned Points	0	3	3	6	2	6	2	4	2	0
Group B	Planned Points	4	4	6	8	6	5	2.5	-	-	-
	Earned Points	2	2	4	6	0	1	0.5	-	-	-
Group C	Planned Points	80	63	40	9	-	-	-	-	-	-
	Earned Points	65	24	21	3	-	-	-	-	-	-
Group D	Planned Points	22	18	14	30	-	-	-	-	-	-
	Earned Points	18	10	11	21	-	-	-	-	-	-

Semi-Structured Interviews A primary source of data collection was semi-structured interviews, a common way of interviewing in case study research [22]. These involve working from an interview guide – a list of prepared questions and topics aimed to ensure systematic and chronological coverage across interviews. However, the interview is flexibly conducted to allow for self-elaboration and exploration of emerging issues [2]. In this research, the main purpose was to explore more issues of group development and to strengthen the validity of responses obtained from the surveys (the GDQ). Also, it was used to validate responses of the work groups to the GDQ surveys. Following the interviewees' approvals to take part in the interviews, the entire interview for each individual was taped, transcribed, and coded. 18 out of the 19 members agreed to have their interviews taped, while one member did not. As a result, the author did not include the latter as part of the data collection. Table 3.2 shows a summary of the total durations of both the unstructured and semi-structured interviews for each of the participating work group.

Table 3.2: Interview summary

Group	Total Interviews time (minutes)
Group A	158
Group B	175
Group C	111
Group D	116

3.3.4 Data Analysis

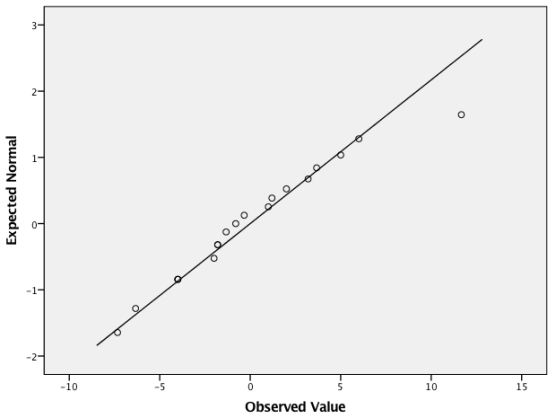
3.3.4.1 Normality Test

A first step to decide which correlation method to use in the data analysis would be to evaluate if the data is normally distributed. We conducted a Shapiro-Wilk analysis for each residual value of the four GDQ scales and the velocity of the four participating work groups. Table 3.3 demonstrates the results of the Shapiro-Wilk test on normality. It would be assumed that the data for the velocity of groups A and C are not normally distributed. We also plotted the frequency of residuals for the velocity and the four GDQ scales for the individuals in the participating groups. The results can be seen in Figure 3.1a, Figure 3.1b, Figure 3.1c, Figure 3.1d, and Figure 3.1e.

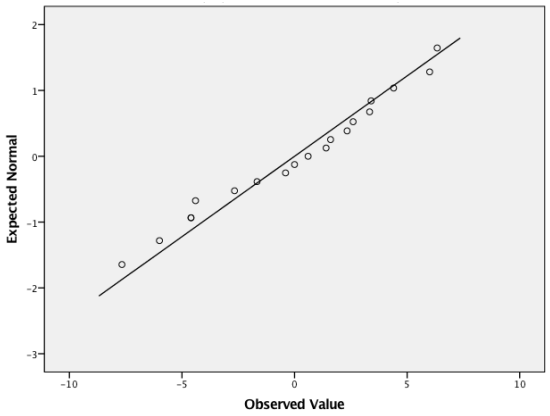
Table 3.3: Shapiro-Wilk Normality Test

Residual	**Group**	**Shapiro-Wilk**	
		Stat	**Sig.**
GDQ1	Group A	0.871	0.272
	Group B	0.931	0.587
	Group C	0.890	0.359
	Group D	0.931	0.587
GDQ2	Group A	0.833	0.145
	Group B	0.972	0.907
	Group C	0.903	0.427
	Group D	1.000	1.000
GDQ3	Group A	0.990	0.294
	Group B	0.849	0.095
	Group C	0.935	0.979
	Group D	0.938	0.948
GDQ4	Group A	0.877	0.294
	Group B	0.824	0.095
	Group C	0.990	0.979
	Group D	0.999	0.948
Velocity	Group A	0.775	0.050
	Group B	0.867	0.216
	Group C	0.773	0.048
	Group D	0.901	0.390

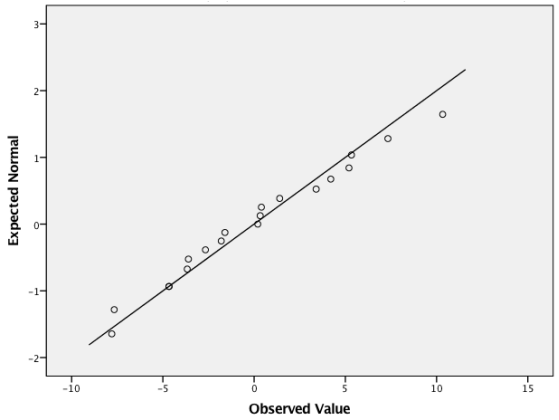
The p values for the velocity of groups A and C indicate statistical significance, were p=.05 for group A and p=.048 for group C. As a result, our normality assumption for our linear regression model is not valid. In addition, the Q-Q plot of residuals for velocity in Figure 3.1e shows a wide scatter in the distribution of residuals across the regression line, which supports our finding from the Shapiro-Wilk analysis that our normality assumption is not valid. Spearman's rank-order correlation analysis was, therefore, selected as the most appropriate method to conduct the correlation for the collected data set.



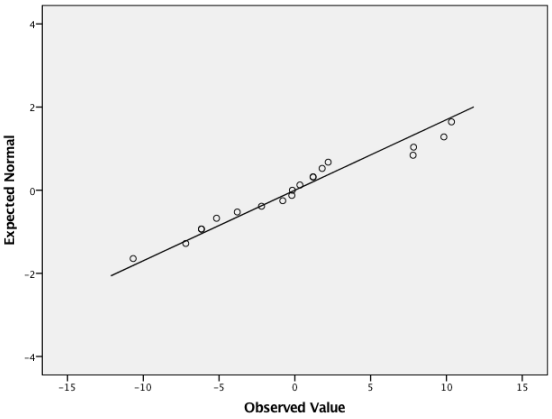
(a) GDQ1



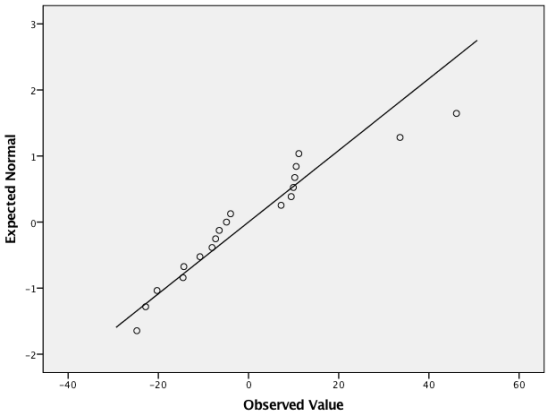
(b) GDQ2



(c) GDQ3



(d) GDQ4



(e) Velocity

Figure 3.1: Q-Q Plot Of Residuals

3.3.4.2 Quantitative Data Analysis

Spearman's rank-order correlation coefficient was used to investigate the connection between group maturity and development velocity, and between group maturity and planning effectiveness. Given the normality analysis check and the small sample size available in this research (four groups), Spearman's correlation was chosen as the most appropriate method to run the analysis, since it does not assume normality in the data. SPSS was used to aid in investigating the aforementioned correlations. For question one, Spearman's correlations were run on both individual and group level, using individual data (19 group members) and then using group data (four groups). For question two, Spearman's correlation was run on the group level only because the planning effectiveness is a group endeavour rather than an individual one. Running the analysis on the group and individual levels will support the idea of the IMGD theory, which states that the dynamics of a particular group constitute the source of individual perceptions of that group. Moreover, it emphasizes the idea that groups, not individuals, should be the key element of any change efforts deemed important. Moreover, some group demographic background collected from the groups' responses on the GDQ were tested for correlation with the four group development scales. Specifically, this was done to examine the impact of the individuals' age, educational background, employment time in company A on the four different maturity scales.

3.3.4.3 Qualitative Data Analysis

Thematic analysis was used to interpret the collected qualitative data. The data from the semi-structured interviews were collated into electronic documents, which made the process of handling, searching and comparing the large volumes of data more convenient and manageable. Data were broadly categorized into seven themes, which were related to the dynamics within the work groups, in order to address some of the issues in the four stages of group development (see Table 3.4).

Table 3.4: Themes Explored in Group Development

Themes	Group Development Stage
Tentativeness and Politeness	I
Participation and Cooperativeness	II
Subgroups or Cliques	II
Goal Clarity	III
Structure	III
Trust	III
Goal Accomplishment	IV

Based on the above themes, a list of seven questions was prepared to address issues

related to the four GDQ scales. Additionally, the last question was asked to help us estimate the number of sprints to consider when calculating the planning effectiveness of each work group (see Section 3.3.3). The Nvivo software was used for transcribing and coding the data.

Table 3.5: Semi-structured Interview Questions

Questions	**Theme**
What are your roles in the team?	Structure
Are members overly polite to each other?	Tentativeness and Politeness
Are members hesitant to ask for support from each other?	Participation and Cooperativeness
Are there cliques or subgroups in the team?	Subgroups or cliques
Is there a high level of trust in the team?	Trust
Are you clear on your team goals? What are they?	Goal Clarity
What is causing delays in accomplishing your sprint goals?	Goal Accomplishment
For how long has the team's maturity level been stable?	

3.4 Ethical Considerations

The importance of ethical standards of conduct for maintaining trust and collaboration with the participants in question has been highlighted by many authors [29]. Participants were spoken to about the objectives of the research, the nature of their involvement, the measures that would be taken to protect their identity, and the right to not participate or to withdraw at any stage. This was first done during scheduled meetings with the scrum masters, then explained to the other group members during the first meeting. Given that this research involves the exposure of the work groups' dynamics to the researcher, it might be sensitive to the participants to openly communicate their views about the functioning of their work groups. As a result, we reiterated the purpose of the research to the participants several times during our presence at company A.

3.5 Chapter Summary

In this chapter, the author presented the research purpose, research questions, research methodology, and the data analysis methods and tools employed in this case study research. The group maturity of the four participating groups were measured by the GDQ and data on the groups' development velocity and planning effectiveness were calculated after collecting the data from the groups' task boards.

4

Results

This chapter presents the results of the case study conducted in company A. It is organized broadly into three sections **a)** maturity level analysis for the participating work groups and the connection between some group demographic background (age, years in the company, and education) and the maturity of the work groups **b)** relationship between group maturity and planning effectiveness which was examined by running Spearman's correlation analysis **c)** relationship between group maturity and development velocity which was also investigated using Spearman's correlation following a normality test. This was examined on both an individual and a group level.

4.1 Maturity Level in The Work Groups

4.1.1 Work Group Profiles

Organizing the collected data from the surveys in a form of profiles, for each of the participating work group, is the first step in getting a general overview of how these work groups are functioning. Each profile contains descriptive statistical information, which include: the scale total, which is the sum of all members' responses to the four scales; the percentage, which reflects how much energy is exhibited on each GDQ scale; and the range difference, which reveals the disparity of viewpoints within each group and for every scale. The group profiles of the four participating work groups can be seen in the Tables 4.1 to 4.4. As these profiles provide a general overview of the work groups in question, it is more insightful to perform a sub-scale analysis to provide recommendations on the specific issues that each work group experiences. A sub-scale analysis looks at the individual scores in all the items related to each scale. If the majority of members scored a given item three or greater than three on scales I and II, this item is checked on the sub-scale analysis form. On the other hand, if the majority of members scored a given item three or less than three on scales III and IV, this item is also checked on the

sub-scale analysis form.

By examining the differences in the range of scores on the group development scales presented in tables 4.1 to 4.4, it can be concluded that work groups B and D are split into at least two subgroups sharing different viewpoints about the way their work groups are functioning because of the large range difference in their scores on scales I, III, and IV (above 15).

Table 4.1: Work Group A – Profile

Scale	I	II	III	IV
Scale Totals	204	158	313	319
Mean	40.8	31.6	62.6	63.8
Percentage	20.523%	15.895%	31.488%	32.092%
Range	39 to 44	27 to 36	59 to 66	60 to 65
Range Difference	5	9	7	5

Table 4.2: Work Group B – Profile

Scale	I	II	III	IV
Scale Totals	242	226	318	320
Mean	40.333	37.666	53	53.333
Percentage	21.88%	20.433%	28.752%	28.933%
Range	33 to 52	30 to 44	45 to 61	47 to 63
Range Difference	19	14	16	16

Table 4.3: Work Group C – Profile

Scale	I	II	III	IV
Scale Totals	126	96	167	194
Mean	42	32	55.66	64.66
Percentage	21.6%	16.4%	28.64%	33.27%
Range	38 to 48	26 to 38	48 to 66	54 to 75
Range Difference	10	12	18	21

Table 4.4: Work Group D – Profile

Scale	I	II	III	IV
Scale Totals	220	147	284	301
Mean	44	29.4	56.8	60.2
Percentage	23.109%	15.441%	29.831%	31.617%
Range	40 to 49	25 to 32	49 to 62	53 to 68
Range Difference	9	7	13	15

4.1.2 Group Demography and Maturity

In order to see if there is a connection between some group demographic information (age, years in company, and educational background) and group development within the four participating work groups, a Spearman's correlation analysis on the individual level was performed. Overall, the results suggest that age relates to the group perceptions about "trust and structure", i.e. the older the members were in the work groups, the higher the work group perception about "trust and structure" was. On the contrary, the number of employment years within the company is negatively correlated, on a moderate level, with the members' perception about their productivity. In other words, the more years the members spent in company A, the less productive they viewed their work groups to be. In this correlation analysis, the education background played no role in the members' group development, according to their views.

Table 4.5: Spearman's Correlations between Group Demography and Perception

Demography	**Statistic**	**GDQ1**	**GDQ2**	**GDQ3**	**GDQ4**	**Productivity**
Age	Coefficient	-0.352	0.068	0.455	0.368	-0.248
	Sig.	0.139	0.783	0.050	0.121	0.305
	N	19	19	19	19	19
Years In Company	Coefficient	-0.239	0.203	0.341	0.220	-0.512
	Sig.	0.325	0.405	0.153	0.366	0.025
	N	19	19	19	19	19
Education	Coefficient	0.315	0.039	0.000*	-0.102	0.090
	Sig.	0.190	0.872	1.000	0.678	0.715
	N	19	19	19	19	19

4.2 Maturity and Planning Effectiveness

A normality test was not performed on the residuals for the planning effectiveness data set since the sample size was too small (N=4). Therefore, Spearman's correlation was run to determine the connection between planning effectiveness and group development since it does not assume normality of the data.

4.2.1 Correlation Analysis

Since planning is a group endeavour, this correlation analysis was run on group level only. The results revealed a positive correlation between the fourth stage of group development and planning effectiveness and showed a significant convergent validity, i.e. the more mature a team is, the more effective they plan their sprints' stories thus deliver the expected outcome. While significant correlations were not found with scales I, II, and III, correlations on scale II and III are going in the right direction (see table 4.6). The correlation coefficient and significance ($r = 1$ and $p = .0000$) describe the strength of the association between the two variables, the "GDQ4" and "Planning effectiveness", which is a perfect positive one.

Table 4.6: Spearman's Correlations for GDQ Perceptions and Planning Effectiveness

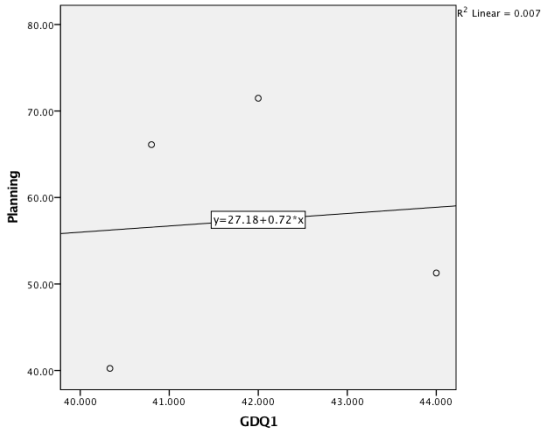
Scale	GDQ1	GDQ2	GDQ3	GDQ4
Planning	0.400	-0.2000	0.4	1.000
Sig. (2-tailed)	0.6000	0.8000	0.6	.
N	4	4	4	4

4.2.2 Planning Effectiveness Comparison

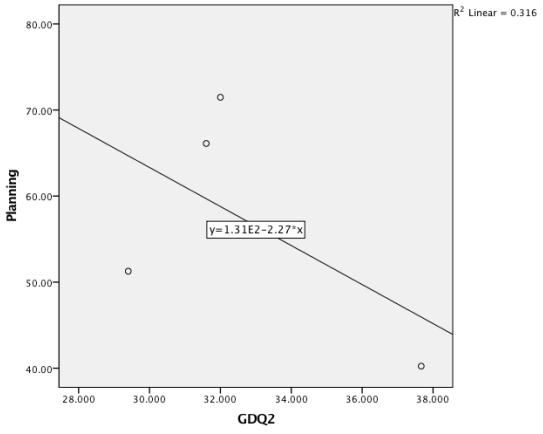
Table 4.7 shows the planning effectiveness and the group development mean values of the four participating work groups. The evidence showed that work groups which scored higher in GDQ4 also scored higher in planning effectiveness. As can be seen from the table, Group D scored the highest GDQ4 score, compared to the other work groups, with a mean value of 64.67. It also outperformed the other work groups in planning effectiveness with a mean value of 71.48. On the other hand, the lowest GDQ4 mean value, 53.17, was scored by work group B, which exhibited the minimum planning effectiveness with a mean value of 40.2.

Table 4.7: Plannng Effectiveness and Group Development Mean Values

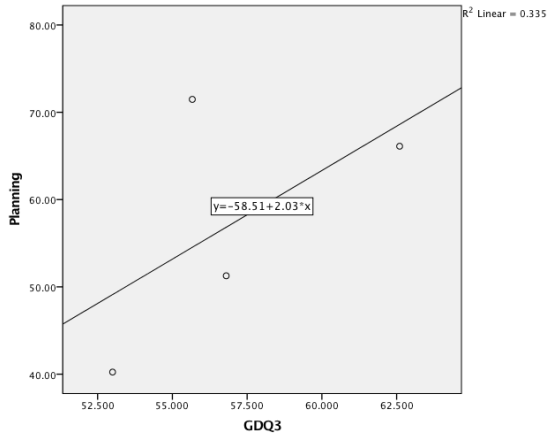
Group	Planning	GDQ1	GDQ2	GDQ3	GDQ4
Group A	66.11	40.80	31.60	62.60	63.80
Group B	40.2	40.33	37.67	54.67	53.17
Group C	51.27	44	29.4	56.8	60.20
Group D	71.48	42	32	55.67	64.67



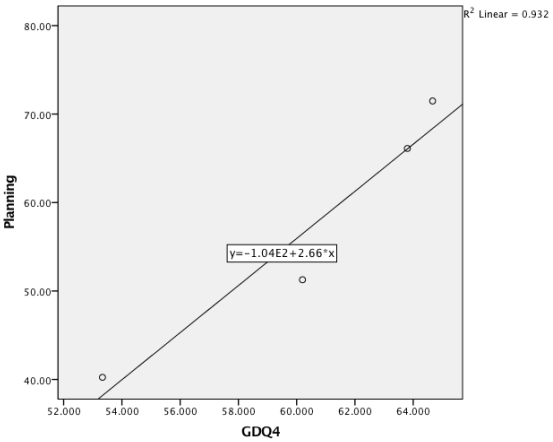
(a) Scatterplot Planning Effectiveness and GDQ1



(b) Scatterplot Planning Effectiveness and GDQ2



(c) Scatterplot Planning Effectiveness and GDQ3



(d) Scatterplot Planning Effectiveness and GDQ4

Figure 4.1: Planning Effectiveness and Group Development Mean Values

Figures 4.2a to 4.2d show the scatter plots for the planning effectiveness as the dependent variable and each of the four group development scales as an independent variable. Each dot represents one of the four participating groups with the x coordinate as the group development mean and the y coordinate as the planning effectiveness mean value. Figure 4.2d shows that $R^2 = 0.93$, which means that 93.2% of the variance in the planning effectiveness can be explained by the fourth scale of the group development (GDQ4). This conclusion is built on the assumption that our model is linear.

4.3 Maturity and Development Velocity

The analysis performed on a group and an individual level helped reinforce the notion that dynamics of a particular group constitute the source of individual perceptions of that group. Tables 4.8 and 4.9 clearly demonstrate that the findings are the same, i.e. no correlation exists between development velocity and maturity on both the individual and group levels.

4.3.1 Correlation Analysis

A second Spearman's correlation was conducted to determine the association between the work groups' perception about their maturity, on all the GDQ scales, and their development velocity. On both the group and individual levels, no significant relationship was identified (see tables 4.8 and 4.9).

Table 4.8: Correlations between GDQ scales and Velocity – Group Level

Scale	GDQ1	GDQ2	GDQ3	GDQ4
Velocity	0.2000	-0.4000	0.8000	0.800
Sig. (2-tailed)	0.8000	0.6000	0.2000	0.2000
N	4	4	4	4

Table 4.9: Correlations between GDQ scales and Velocity – Individual Level

Scale	GDQ1	GDQ2	GDQ3	GDQ4
Velocity	0.310	-2.16	0.236	0.204
Sig. (2-tailed)	0.196	0.374	0.330	0.402
N	19	19	19	19

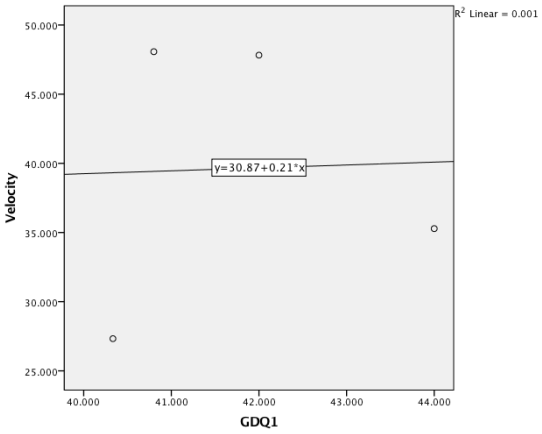
4.3.2 Velocity Comparison

Table 4.10 shows the velocity and the group development mean values of the four participating work groups, which were both measured during the same sprint. As can be

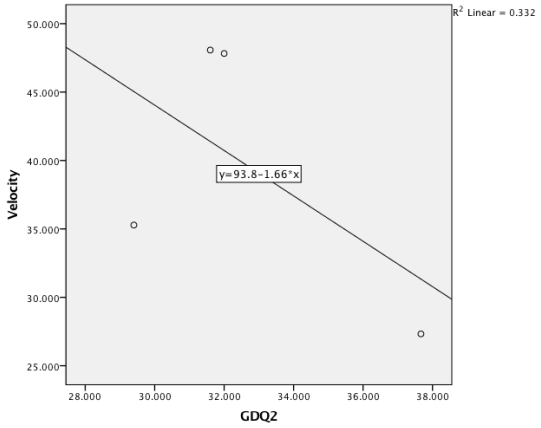
seen from the table, work group B has the minimum velocity mean value of 27.33. On the contrary, the highest mean velocity value was scored by work group A with a mean value of 48.07. As shown in table 4.8, no connection between the groups' mean velocity and the group development scales was found.

Table 4.10: Velocity Mean Values and Group Development Mean Values

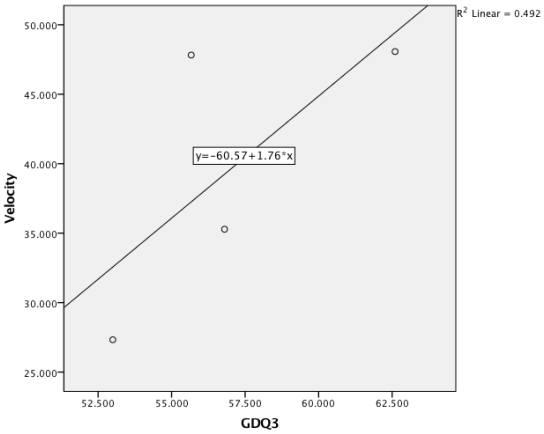
Group	**Velocity**	**GDQ1**	**GDQ2**	**GDQ3**	**GDQ4**
Group A	48.07	40.80	31.60	62.60	63.80
Group B	27.33	40.33	37.67	54.67	53.17
Group C	35.29	44	29.4	56.8	60.20
Group D	47.82	42	32	55.67	64.67



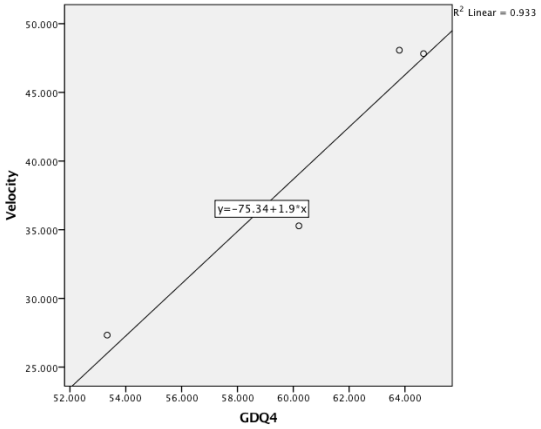
(e) Scatterplot Velocity and GDQ1



(f) Scatterplot Velocity and GDQ2



(g) Scatterplot Velocity and GDQ3



(h) Scatterplot Velocity and GDQ4

Figure 4.2: Development Velocity and Group Development Mean Values

4.3.3 Maturity and Subjective Productivity Correlation

This section presents the results of the correlation run between the measurements of group development and the subjective assessment of productivity for all the 19 individuals. The productivity measurement was collected from the participants' responses to a question in the survey (GDQ), which asked them to rate their group productivity on a scale from one to four. Table 4.11 illustrates the results of this correlation.

Table 4.11: Spearman's Correlations between Group Development and Productivity

Scale	GDQ1	GDQ2	GDQ3	GDQ4
Productivity	0.134	-0.506	0.159	0.359
Sig. (2-tailed)	0.584	0.027	0.515	0.131
N	19	19	19	19

The Spearman's correlation showed a strong negative association between GDQ2 and productivity, which means that, an increases in GDQ2 leads to a decrease in the participants' self-assessment of their group productivity and vice versa. As to the other three independent variables (GDQ1, GDQ3, and GDQ4), no correlation with productivity was identified.

4.4 Semi-Structured Interviews

Responses from interviewees were thematically analyzed. Below are the main results of this analysis.

Roles in The Work Groups
All members from the participating groups were able to mention their roles with ease, which means they are all clear on their responsibilities.

Politeness and Tentativeness
The majority of members in work group A (75%) did not consider over politeness evident in their work group, whereas 25% considered politeness to be a "*rare occasion*" in the work group. Members of work group B explained that over politeness depends on the situation and the member's personality rather than a general trait for the work group. All of the interviewed members from work group C agreed that members were overly polite with each other. 75% of them linked this over politeness to the nature of engineers and their cultural backgrounds whereas 25% said that members exhibited over politeness only in process related situations rather than technical ones. Finally, members from work group D perceived each other as being overly polite and associated this to the fact that they are newly formed and have not yet built a cohesive relationship with each other. The exception to this is one work group member who has been in the team for the longest time, perceived the work group not extremely polite.

Cooperativeness and Support
All of the interviewees from work group A reported that they are not hesitant to reach out each other when needed. 50% of members from work group B suggested that people are not hesitant to ask for support from any one in the work group. One of those linked this to the fact that members do not want to take responsibilities on their behalf "*they might ask anyone so that they do not take the responsibility*". On the other hand, the remaining 50% of members suggested that members are sometimes hesitant and linked this to the personality and topic type. 25% of the interviewees from work group C shared a consensus that members are hesitant to seek for support from each other "*People are very concerned with each other and are reluctant to ask for support*" whereas 75% suggested that members tend to seek support from knowledgeable people in the work group, regardless of who they are. Finally, the majority of members in work group D explained that members are a bit hesitant to seek for support and explained that it is related to the fact that they have not yet had enough time to interact with each other and create a cohesive relationship. On the contrary, one member explained that members reach each other and are not scared to point out problems.

Subgroups and Cliques
75% of the interviewed members from work group A believed that there are no subwork groups whereas 25% agreed that there are subgroups. work group B explained that their

work group is divided into two subgroups, over half of those linked this to the "age range" factor "there are two subgroups, and they are about the same age and stage in life". All of the interviewed members from work group C explained that people with more technical knowledge or similar interests formed cliques in this work group. In addition, a third of the members in work group D believed that there are no cliques in their work group, the second third suggested the occurrence of subgroups, and the last third were not able to tell.

Trust
All interviewed members from work group A and C believed that trust is high in their work groups. Opinions in work group B is divided, whereby 40% believed there is no trust amongst members, 40% believed that there is trust, while the remaining mentioned that trust is a relative issue in the work group. Members in work group D distinguished between internal and external trust. They referred to external trust as the way in which external teams in company A perceive their work group; whereas internal trust was defined as how much members in the work group trust each other. All members in work group D believed that the external trust is high: "*From outside the team, trust is pretty solid. If something should be done, people trust us*". On the other hand, a third of the members explained that internal trust is high but not satisfactory yet: "*Maybe in the long run we will probably build higher internal trust*", while two thirds of the members suggested a lack of internal trust (within the work group) "*I don't have trust nor members have it to others*".

Goal Clarity
75% of members in work group A mentioned some of their goals on a sprint level only: "*we only look at the sprint goals*" whereas 25% of them did not know any goals, whether on a sprint level or not. The majority of members from work group B could not mention any long or short term goals for their work group "*There are no common team goals, but rather some individual goals to reach*". 50% of them linked this to the poorly defined customer specifications while almost 20% believed that members are too focused on the development that they forget the work group goals: "*Most of them would remember them if you remind them but not everyone realizes that they know them*". On the contrary, all members from work group C recited their short term goals and not the long term ones. There was a 50% overlap in their answers. Finally, third of the members from work group D were able to recite some of their work group's goals, while two third of the individuals could not.

Delays in Goal Accomplishment
The majority of members of work group A agreed that their lack of knowledge in one particular software engineering discipline (kept anonymous here) is negatively affecting their commitment to achieving their goals. Half of the members from work group B explained that the main reason for their delay was the unclarity of requirements received from their customer company: "*We don't know what we are doing. We need clear re-*

quirements". The remaining 50% had a common view that the external dependencies, the underestimation of workload, and the lack of knowledge in the domain of work are the reasons for their delay. 25% of members from work group C attributed the delay to not knowing how to work well enough as a new team, whereas 75% of them gave a common explanation, which persisted in their lack of experience to estimate the time needed for code review. 25% of those gave additional reasons such as external dependencies and sick leaves. The other 25% suggested that delays were the result of lack of knowledge in coding and the product, and underestimation of refactoring. Finally, member of work group D had extremely different explanations for their delay, which mainly persists in their lack of knowledge in the product and the lack of norms within the work group, and the variations in the level of technical competencies within the work group.

Stability of Maturity in The Work Groups
Finally, all of the interviewees from work group A said that their maturity has not changed for ten months. The majority of members from work group B suggested that their maturity has been stable for six months. 75% of members in work group C believed that their maturity level has been stable for three months, whereas 25% could not proximate any specific period. Members from work group D expressed different views about the period of stability. 75% agreed that their group maturity remained unchanged for two months, whereas 25% suggested that the group maturity has continuously progressed and it stopped progressing one month ago.

4.4.1 Summary of the Semi-Structured Interviews

The qualitative analysis revealed that work groups B and D experienced the highest number of group development issues explored in the semi-structured interviews whereas work groups A and C showed the lowest number of these issues. Also, a disparity of viewpoints was evident from the opinions of members in work groups B and D where individuals perceived their work groups' to be functioning differently. The major issues that emerged in work groups B and D seemed to relate to the different technical knowledge or age range of members. Some members with high technical knowledge (more experienced) tend to prefer working collaboratively with each other rather than working with individuals who had less technical experience. This may explain the split in the views, demonstrated in the groups' profiles (see Tables 4.1 to 4.4). Similarly, the former may also explain the lack of trust and goal clarity between individuals in both group B and D.

4.5 Chapter Summary

This chapter presented the results of the analysis performed on both the qualitative and quantitative data. Specifically, the results drawn from the two correlation analysis run against the measurements of group development and planning effectiveness on one hand, and group development and velocity on the other hand. The results showed a perfect

positive correlation between stage four and planning effectiveness, and no correlation between group development and velocity. The results of this analysis revealed a strong negative correlation between productivity and GDQ2.

5

Discussion

This section gives answers to the two research questions presented in section three and concludes with the presentation of validity threats and how we tried to mitigate them during this case study.

5.1 Reflection on Efficiency and Effectiveness

We emphasize on interpreting the results in light of the distinction between efficiency and effectiveness. Our velocity measurement only reflects the efficiency of work groups in accomplishing scrum tasks, with no indication on how effective they were implemented. On the other hand, the measurements of planning effectiveness reveals the work groups' ability to deliver the expected outcome within the planned time frame.

5.2 Answers to Research Questions

5.2.1 RQ1 - What is the association between group development and planning effectiveness?

In this research, we investigated the relationship between four independent variables (the group development stages) and the planning effectiveness. The results showed a perfect positive correlation (+1.0) between the fourth GDQ scale and the planning effectiveness among the four participating work groups, which means that both variables move in a strong tandem with each other and are positive in 100% of the time. In other words, when the group development increases, the planning effectiveness also increases and vice versa. This supports the findings of other studies which confirm that task performance and work activity occur at higher levels later in a group's development [9, 36, 39]. The significance of this research is that it provided evidence to support a relationship between group development and team performance in software engineering context.

Furthermore, the overall conclusion drawn from the qualitative analysis supported those concluded from the quantitative ones. For example the thematic analysis revealed that members from work group B had the highest number of group development issues, compared to the other work groups. Contrary to work group B, members of work group A had the lowest number of issues, which might be an indicator that this work group is at the higher levels in group development. The planning effectiveness results showed that work group B had the lowest rates while work group A had the second highest rates. Although work group D showed the highest GDQ4 score, drawn from the survey, and the highest planning effectiveness scores, their profile (see Table 4.4) and responses to the interviews suggested a disparity of viewpoints in the way they perceive the functioning of their work group. This leads us to suggest that there may be other unknown variables that contribute to the improvement of the groups' planning effectiveness despite their maturity level.

5.2.2 RQ2 - What is the association between group development and Velocity?

We investigated the connection between the two variables group development and velocity. The motivation for investigating this research question was to address a gap recited by Gren et al. [17] in which a positive correlation between maturity and velocity would have added the much needed aspect to support their findings about the connection between agility and group maturity. The results drawn from our analysis to this question were not completely in concordance with what Gren et al. [17] suggested, since we could not provide an empirical evidence to support a significant convergent validity between group maturity and development velocity. The analysis of the qualitative data revealed that the majority of participants linked their tasks development delays to technical and process related aspects rather than issues pertinent to the dynamics and norms within their work groups. However, this needs to be interpreted in light of the fact that members perceived themselves as not good at judging the dynamics within their work groups and linked this to the differences in the cultural backgrounds and the nature of engineers.

5.3 Implications for Research and Practice

We will now present some of the possible improvements that would increase the software development team performance. The first would be to motivate software developers to focus more on discussing and clarifying their work group goals. By this we mean that members should work more on achieving their group goals rather than focusing on the individual ones only. Our research suggests that the more effective work groups know their group goals, which is in alignment with stage III of the IMGD model which suggests that clarity of goals contribute to the development of more productive and work-focused groups. The second would be to train the software developers in managing group conflicts and to encourage seeing conflicts as an inevitable and positive stage that would eventually

lead to an increase in their performance. This is based on our results which showed that intra-group conflict in software teams results in a decrease of the teams' perceptual assessment of their performance. The third would be to motivate software developers to freely discuss and communicate process-related issues rather than only discussing the technical ones because members reported a tendency of being hesitant to ask for support in process-related issues. The fourth would be to consider having team members of diverse backgrounds working together in order to allow building more trust and structure within teams. This is supported by research done by Roberge and van Dick [27] who attempted to address when and how diversity in teams leads to better performance by conceptualizing a multi-level model that identifies the psychological mechanisms that explain how diversity can have a positive impact on the performance of teams. On the group level, these psychological mechanisms were identified as communication, group involvement, and group trust [27]. Although our research only included one aspect of diversity, which is age, our qualitative and quantitative analysis clearly show that the age of software development team members relates to their perceptions of *trust and structure*.

5.4 Validity Threats

This section discusses the threats to the validity of this research.

It is not possible to generalize the findings of this research outside this specific case because only four participating groups from the same company were studied, which is a small sample. However, the combination of the data collection methods we used in this research, qualitative interviews and quantitative surveys would triangulate our findings, thus would strengthen the validity of this research.

This area of research could be intimidating and sensitive to the participating members since it involves the disclosure of the dynamics within their groups to the researcher, which may influence the validity of the groups' responses to the quantitative survey and the qualitative interviews. At the onset of this research, an attempt to mitigate this was made by explaining the research purpose to the participants and confirming the anonymity of their responses. However, it is not possible to refute the presence of bias in the participants' responses on both the surveys and the interviews. To reinforce the anonymity of the participating work groups, we avoided stating any information that would indicate their identity. In addition, it is not possible to guarantee that the gatekeeper did not share some of the research findings with the work groups before the conduct of the semi-structured interviews, which may have impacted their responses. To minimize this threat, a confirmation to keep the participants' identity anonymous was reiterated and the purpose of the research was clarified several times. Moreover, a self bias in the coding process of the semi-structured interviews cannot be ruled out and requires a second coder to validate the responses and therefore minimize the self bias.

The disparity of viewpoints drawn from the analysis of both the qualitative and quantitative data for work groups B and D indicate that none of the formers work as one functioning unit, but rather rely on the experience and technical knowledge of individuals to get the work done. This analysis implies that the conclusions drawn from the correlation analysis should be interpreted in light of this information. In addition, our measurement of the velocity relied on measuring the time spent by each work group on tasks accomplishment, which means that the amount of teamwork required to accomplish those tasks may not be significant, and an individual endeavor on each task may be sufficient to get the work done. This may provide an explanation to the absence of correlation between velocity and group maturity. Finally, although the majority of members explained that they instantly close their tasks after finishing their implementations, we can not guarantee that all of the tasks we selected for analysis were closed this way. This may have had an effect on the validity of our velocity measurement.

6

Conclusions and Future Work

6.1 Conclusion

In the course of this research, we aimed at investigating how development velocity and planning effectiveness of software teams relate to their group development. The empirical results of this research showed that group development is significantly related to planning effectiveness whereas no evidence was provided to conclude a similar relationship with development velocity. In other words, a more mature team is also a more effective one in its commitment to plans. Moreover, it indicates that there are considerable differences as to how group development relates to the effectiveness and efficiency of software teams performance. That is, a more mature team is possibly a more effective one but not a more efficient one. We believe that this research provided additional knowledge to the prominence of the human interactions within software development teams. Particularly by providing empirical evidence about the association between group maturity and planning effectiveness, and the negative connection between conflicts and software development performance. We believe that the knowledge provided is sufficient to trigger organizations to drive more focus on those aspects, since they may provide benefits to software development teams.

6.2 Future Work

We would like to see the results of similar studies conducted with larger sample sizes from different companies. Also, we would like to encourage further studies to expand upon the connection between group development aspects and team performance in software development. For example by measuring, function points, defects rate, and kilo lines of codes to assess team performance. Moreover, future work is encouraged to use more data collection methods to achieve higher validity in their findings. Particularly, the use of observation systems, such as GDOS, in longitudinal case studies to analyze the verbal

behavioural patterns from the daily meetings of the work groups in question. Finally, we would like to see the results of studies that combine several objective and subjective methods to assess the performance of software development teams and highlight how each relates to group maturity.

Bibliography

[1] Albero Pomar, F., Calvo-Manzano, J. A., Caballero, E., Arcilla-Cobián, M., 2014. Journal of Software: Evolution and Process 26 (9), 776–783.

[2] Allison, I., Merali, Y., 2007. Software process improvement as emergent change: A structurational analysis. Information and software technology 49 (6), 668–681.

[3] Anderson, N. R., 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.

[4] Asproni, G., 2004. Motivation, teamwork, and agile development. Agile Times 4 (1), 8–15.

[5] Bahli, B., Buyukkurt, M. D., 2005. Group performance in information systems project groups: An empirical study. Journal of Information Technology Education 4 (1), 97–113.

[6] Bain, P. G., Mann, L., Pirola-Merlo, A., 2001. The innovation imperative the relationships between team climate, innovation, and performance in research and development teams. Small Group Research 32 (1), 55–73.

[7] Bales, R. F., Strodtbeck, F. L., 1951. Phases in group problem-solving. The Journal of Abnormal and Social Psychology 46 (4), 485.

[8] Boehm, B. W., et al., 1981. Software engineering economics. Vol. 197. Prentice-hall Englewood Cliffs (NJ).

[9] Buzaglo, G., Wheelan, S. A., 1999. Facilitating work team effectiveness case studies from central america. Small Group Research 30 (1), 108–129.

[10] Cissna, K. N., 1984. Phases in group development the negative evidence. Small Group Research 15 (1), 3–32.

[11] Cohn, M., 2015. The Difference Between a Story and a Task. `https://www.mountaingoatsoftware.com/blog/the-difference-between-a-story-and-a-task`., [Online; accessed: 11-April-2016].

[12] Davidson, B. L., 2001. Verbal behavior patterns in groups of different ages. Ph.D. thesis, Temple University.

[13] Evans, N. J., Jarvis, P. A., 1986. The group attitude scale a measure of attraction to group. Small Group Research 17 (2), 203–216.

[14] Ganesh, M., Gupta, M., 2006. Study of virtualness, task interdependence, extra-role performance and team climate in indian software development teams. In: Proceedings of the 20th Australian New Zealand Academy of Management (ANZAM) Conference on Management: Pragmatism, Philosophy, Priorities, Central Queensland University, Rockhampton. pp. 6–9.

[15] Gersick, C. J., 1988. Time and transition in work teams: Toward a new model of group development. Academy of Management Journal 31 (1), 9–41.

[16] Gondal, A. M., Khan, A., 2008. Impact of team empowerment on team performance: Case of the telecommunications industry in islamabad. International Review of Business Research Papers 4 (5), 138–146.

[17] Gren, L., Torkar, R., Feldt, R., 2015. Group maturity and agility, are they connected? – a survey study. In: 41st EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA). IEEE, pp. 1–8.

[18] Hoegl, M., Gemuenden, H. G., 2001. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. Organization science 12 (4), 435–449.

[19] Huckman, R. S., Staats, B. R., Upton, D. M., 2009. Team familiarity, role experience, and performance: Evidence from indian software services. Management science 55 (1), 85–100.

[20] Karn, J., Syed-Abdullah, S., Cowling, A. J., Holcombe, M., 2007. A study into the effects of personality type and methodology on cohesion in software engineering teams. Behaviour & Information Technology 26 (2), 99–111.

[21] Liang, T.-P., Liu, C.-C., Lin, T.-M., Lin, B., 2007. Effect of team diversity on software project performance. Industrial Management & Data Systems 107 (5), 636–653.

[22] McLeod, L., MacDonell, S. G., Doolin, B., 2011. Qualitative research on software development: a longitudinal case study methodology. Empirical Software Engineering 16 (4), 430–459.

[23] Mills, T. M., 1967. The sociology of small groups. Prentice-Hall Englewood Cliffs, NJ.

[24] Ong, A., Tan, G. W., Kankanhalli, A., 2005. Team expertise and performance in information systems development projects. In: Proceedings of the 9th Asia Pacific Conference on Information Systems, Bangkok, Thailand, July. pp. 7–10.

[25] Purna Sudhakar, G., Farooq, A., Patnaik, S., 2011. Soft factors affecting the performance of software development teams. Team Performance Management: An International Journal 17 (3/4), 187–205.

[26] Ramasubbu, N., Balan, R. K., 2007. Globally distributed software development project performance: An empirical analysis. In: Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on the foundations of software engineering. ACM, pp. 125–134.

[27] Roberge, M., van Dick, R., 2010. Recognizing the benefits of diversity: when and how does diversity increase group performance? Human Resource Management Review 20 (4), 295–308.

[28] Runeson, P., Höst, M., 2009. Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering 14 (2), 131–164.

[29] Runeson, P., Höst, M., 2009. Tutorial: Case studies in software engineering. In: Product-Focused Software Process Improvement. Springer, pp. 441–442.

[30] Sawyer, S., 2001. Effects of intra-group conflict on packaged software development team performance. Information Systems Journal 11 (2), 155–178.

[31] Sidky, A., 2007. A structured approach to adopting agile practices: The agile adoption framework. Ph.D. thesis, Virginia Polytechnic Institute and State University.

[32] Tuckman, B. W., 1965. Developmental sequence in small groups. Psychological bulletin 63 (6), 384.

[33] Tuckman, B. W., Jensen, M. A. C., 1977. Stages of small-group development revisited. Group & Organization Management 2 (4), 419–427.

[34] Wheelan, S., 1994. The group development questionnaire: A manual for professionals. Provincetown, MA: GDQ Associates.

[35] Wheelan, S. A., 2005. Group processes: A developmental perspective. Pearson Education.

[36] Wheelan, S. A., Burchill, C. N., Tilin, F., 2003. The link between teamwork and patients' outcomes in intensive care units. American Journal of Critical Care 12 (6), 527–534.

[37] Wheelan, S. A., Hochberger, J. M., 1996. Assessing the functional level of rehabilitation teams and facilitating team development. Rehabilitation Nursing 21 (2), 75–81.

[38] Wheelan, S. A., Hochberger, J. M., 1996. Validation studies of the group development questionnaire. Small Group Research 27 (1), 143–170.

[39] Wheelan, S. A., Tilin, F., 1999. The relationship between faculty group development and school productivity. Small Group Research 30 (1), 59–81.

[40] Zimpfer, D. G., 1986. Planning for groups based on their developmental phases. Journal for Specialists in Group Work 11 (3), 180–187.