

Towards Group Development Stages in Software Engineering Courses Project Work

Davide Fucci

Ehsan Zabardast

SERL Sweden, Blekinge Institute of Technology (Sweden)

Abstract

Wheelan et al. established that a group of individuals working together towards a common goal goes through four different stages—i.e., Dependency & Inclusion, Counter-Dependency & Fighting, Trust & Structure, and Work & Productivity. Such development includes both group structure (e.g., which individual takes which responsibility) and interpersonal processes. Group work projects, spanning several weeks, are common in applied disciplines such as Software Engineering (SE). This study is the first step towards validating whether the four stages of group development take place for SE students working together in groups during a semester-long software development project. Moreover, we focus on factors of interest, namely the students' different backgrounds and the new working conditions they had to adapt to due to the COVID-19 pandemic. We carried out a longitudinal survey with 11 students taking part in the group work of a Master-level course in SE at Blekinge Institute of Technology, Sweden. We used a simplified version of the Likert-scale questionnaire originally proposed by Wheelan et al., as well as open-ended questions. The preliminary analysis of the survey data shows that the four stages of group development do not occur among students working together on an SE project, although later stages are more evident than others are. These results seem to be impacted by the social aspects of unusual working conditions.

Introduction

The ability to work in groups is desirable for university graduates. To develop such skill, university programs have introduced group assignments, which have now become an inseparable part of higher education (Violet and Mansfield 2006). From a pedagogical perspective, group work aims to achieve two goals, intellectual and social. Regarding the former, it promotes active learning, conceptual learning, and creativity; regarding the latter, it increases trust and communication, as well as keeping students involved with the subject (Cohen and Lotan 2014). Other studies have shown that group work has positive effects on students' achievement, self-esteem, outlook on learning, and retention (Cumming 2010; Johnson, Johnson, and Smith 1998). Group work is relevant, especially for university students, as it contributes to develop the soft skills sought by employers—e.g., communication with peers, conflict management and time management (Branine 2008).

There are, however, factors that influence group performance, such as group heterogeneity, individual motivation, type of task, and individuals' culture and language (Springer *et al.* 1999; Popov *et al.* 2012). These factors make university group work challenging for both teachers and students. In some cases, students who experience group activities numerous times during their training are shown to not be satisfied with their group work, nor with how they are integrated with their projects (e.g., Chang and Brickman 2018; Burke 2011). To improve students' experience of group work, teachers should also consider the stage at which the group happens to be at a given time since this may be a possible limiting factor for group work success.

The vast majority of the existing body of literature supports the existence of several stages during which a group develops over time. Wheelan and colleagues show that newly created groups go through several stages before they start to be productive (e.g., Wheelan 2014; Wheelan and Hochberger 1996). In particular, they identified four stages of group development, *Dependency & Inclusion*, *Counter-dependency & Fighting*, *Trust & Structure*, and *Work & Productivity* (Wheelan, Davidson, and Tiliin 2003).

According to the literature supporting this view (e.g., Shaw 1981; Steiner 1986; Sundstrom, De Meuse, and Futrell 1990), the term *team development* includes both group structure, such as the individual skills distribution, as well as interpersonal processes. Temporal patterns that the group processes are tied to effectiveness, even for brief work sessions. For example, Hackman and Morris (1975) examined the effect of social interactions during the team formation process. If the first few minutes of group formation leaves the group with a feeling of having been productive or having achieved something together, this can activate the initial group cohesion effect. This effect strongly determines the relationship between the team members and how they are going to interact with each other over the course of the project (Wheelan 2014). Eventually, those groups that experienced a strong initial cohesion are likely to perform better than those that did not.

Nonetheless, there are a few contrasting theories that do not support group development (Tyastuti and Ariany 2017; Zellmer-Bruhn 2003). One of the most notable examples is the work of Gersick (1988) which distinguishes group development into two phases, a punctuated equilibrium in which little initial progress is made, followed by a phase towards the second half of a project in which goals are actually accomplished. However, the majority of the current research supports Wheelan's theory and counter-theories are based on less solid evidence.

Software Engineering (SE) is not only a technical field—in professional settings it also requires co-operation and co-ordination among people having different roles (e.g., project manager, software developer, interface designer, quality assurance). Students practice such “soft skills” through group work aimed at collectively delivering a project. In this study, we investigate how student groups go through the four Wheelan stages in the context of a course in the Software Engineering curriculum. Moreover, we aim at comparing different student sub-groups based on their experience (i.e., Bachelors' and Masters'), and the impact of their remote working conditions that arose during the second half of the project because of the COVID-19 pandemic.

This paper is organized as follows: Section 2 presents an overview of Wheelan’s theory of group development as well as some of its application to the SE field. In Section 3, we report the research questions guiding this study and its design. In Section 4, we show the results of our study and answer the research questions. Section 5 reports some of the limitations of this work and its implications for teachers. Finally, Section 6 concludes the paper and outlines future work.

Background

In this section, we summarize Wheelan’s work, which this study draws upon, as well as the few studies in the SE field using GDS.

Group Development and Wheelan’s Four Stages

Initial investigations of group dynamics (Tuckman 1965) proposed states or levels of activities. Building on the work of Tuckman—who proposed four stages (*Forming*, *Storming*, *Norming*, and *Performing*)—Wheelan’s group development theory describes groups in later stages as *mature* (Wheelan 2014).

The model of Group Development covers different maturity stages. This model is based on the idea that individual behavior changes according to how well a person knows another member of the group. Wheelan compares the group development level to that of an individual going through childhood (where the focus is on figuring out the environment), adolescence (where the focus is on questioning the environment and social structures), and adulthood (where the focus is on how to mature and improve). Figure 1 shows Wheelan’s model in which groups progress through cyclical stages.

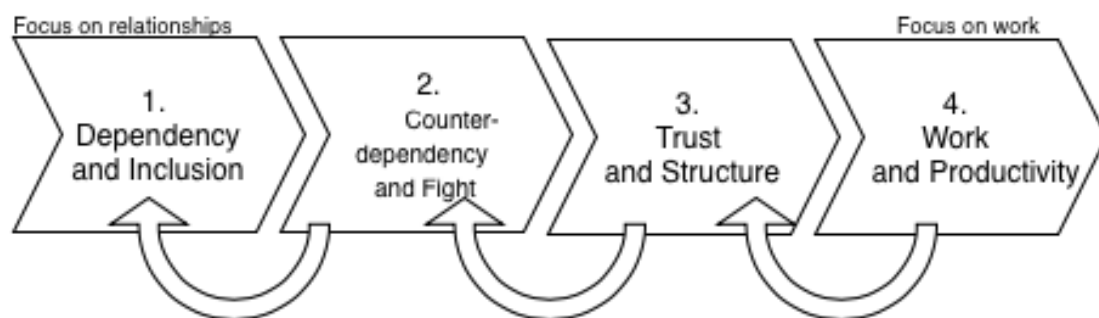


Figure 1: Wheelan’s group development stages.

Stage 1: Dependency and inclusion.

In this stage, the group’s main concerns are safety and inclusion. A leader is designated and there is a wish for structure. At this stage, the desire for becoming productive also arises.

Stage 2: Counter-dependency and fight.

The group has gained a sense of loyalty and the environment is becoming more stable and safe. In turn, this allows members to easily share their opinions. Fights arising in this phase lead to the creation of roles and a higher level of trust.

Stage 3: Trust and structure.

In this stage, the group structure develops further through the establishment of skill-based roles. Accordingly, members communicate in a task-oriented fashion rather than resolving non-work-related issues. More mature processes and organization arise in this stage.

Stage 4: Work and productivity.

In this stage, the main activity of the group is getting work done. Cohesion is maintained, and focus is on decision making in order to progress the work. This is the stage at which the group becomes a team.

Group Development in Software Engineering

Gren and Al-Sabbagh (2017) performed a case study with 19 professional software developers in order to observe how performance based on productivity and effectiveness relate to the maturity level of a development team. Using the answers to the Group Development Questionnaire, the authors performed a correlation analysis. Their results show that maturity is correlated with planning effectiveness, indicating that a relationship between group development and team performance exists in the software engineering domain.

Team self-organization, empowerment, and accountability are essential properties of an Agile software development—an approach that has shown to increase project success thanks to its focus on people (McDonald and Edwards 2007). Gren, Goldman, and Jacobsson (2019) performed a study during which 12 Agile teams (85 members in total) participated in training regarding group development psychology. Thereafter, in a survey, the participants reported positive views regarding the way they think about teamwork and how to deal with team members.

Wheelan's model of Group Development is the basis for a questionnaire used in research (e.g., through surveys) to assess team or group maturity. Such a Group Development Questionnaire (GDQ) consists of 60 items, which may hinder its reliability, for example because of transient measurement error (Schmidt, Le, and Ilies 2003) and in general because of the survey response rate. Gren et al. (2018) proposed a shorter version of GDQ, the GDQS. This includes 13 items and has been developed and validated across four independent and diverse samples—for a total of 2,646 groups—including one from the IT industry. The results from all the four studies show that GDQS is internally consistent with respect to GDQ ($\alpha \geq 65\%$). In this study, we use GDSQ to assess the students' group maturity.

Research Design

In this section, we report the research questions driving this study and the methodology used to answer them.

Research Questions

The aim of this investigation is to observe whether the Group Development Stages (GDS) proposed by Wheelan take place during group work in SE. Accordingly, we formulated our first research question as follows:

RQ_{GroupDev}. To what extent do GDS take place during SE students' group work?

By answering this question, we apply the research methodology and instruments used in Wheelan's model to a new context. Accordingly, the aim of this study is not to establish any causal relationship between a group's maturity and, for example, its productivity or performance. Rather, we want to provide a steppingstone for further researchers and lecturers who can benefit from recognizing in which phase the group is situated. Validating the existence of the different phases is a sensible action towards our goal.

The study involves students working on several group assignments within a project during the Spring Term of 2020 at the Software Engineering Department (DIPT), Blekinge Institute of Technology (BTH), Sweden. Students at DIPT have different backgrounds than solely Software Engineering (e.g., Economics) and are enrolled into different programs (i.e., B.Sc., M.Sc., and Civil Engineering (CIV: *Civilingenjör*)). We formulated our second research question as follows:

RQ_{Background}. To what extent do the students' backgrounds impact their conformance to GDS?

When answering this question, we consider the academic background of the student as a mediating factor. In this study, background—i.e., the study program in which students are enrolled, either Industrial Economics, Masters' in Software Engineering, or Civil Engineering—is a proxy for existing knowledge (e.g., computer programming, experience with a similar project) which can have an impact on the group project success (Hammar Chiriac 2014). By answering this question, we can increase the generalizability of Wheelan's theory and pinpoint the impact that a student's background has on it—e.g., by eventually showing that students of different backgrounds progress differently through the phases.

As the COVID-19 pandemic changed BTH operations in Spring 2020, students needed to collaborate remotely for the entire duration of the project. This condition offered the opportunity to understand how GDS are affected by remote collaboration. Accordingly, we formulated our third research questions as follows:

RQ_{Remote}. How does remote collaboration affect GDS?

RQ_{GroupDev} and *RQ_{Background}* will be answered by conducting and analyzing the answers to a survey based on GDQS (Gren et al. 2018), whereas *RQ_{Remote}* will be answered after performing thematic analysis of students' responses to open-ended questions.

Research Methodology

We selected longitudinal survey as a research method. We originally planned to administer the same survey several times during the project group work of an SE course. However, due to a restructuring of the course, it was possible to administer the survey only twice. The survey aims to assert facts about a population that is compared against an existing theory, i.e., GDS.

Sampling.

In accordance with Wheelan and Hochberger (1996), the unit of analysis is the individual participant in a group, rather than the group itself. In this study, we are not interested in observing correlations between the different stages and an outcome of interest, such as planning effectiveness in software development projects (Gren, Goldman, and Jacobsson 2019) or mortality rate in health professions (Tyastuti and Ariany 2017). Conversely, we are interested in checking whether SE students participating in group work progress through the stages in the first place. Accordingly, higher-education students participating in SE group work projects represent the target population for this study. The sample was non-probabilistic, as we sampled participants by convenience among students enrolled in the PA1450 Software Engineering¹ course offered at BTH during the 2020 Spring Term. As a result, the theoretical sample size ($n = 77$) coincides with the number of students enrolled in the course when the survey was announced on January 18, 2020.

Instrument.

The instrument for the survey is a questionnaire consisting of three parts: i) a question regarding respondents' study backgrounds, ii) 13 Likert items, where each item maps on one the four stages proposed by Wheelan as shown in Table 1, and iii) two additional questions about the remote work conditions due to the COVID-19 pandemic. We assigned to each respondent a random unique identifier which, while preserving anonymity, enabled us to associate answers to the two iterations of the surveys with an individual subject.

Table 1: Statements from the GDQS (Gren et al. 2018)

ID	Statement
a	Individuals seem to go along with whatever the group suggests
b	When members talk, they address the same few people.
c	Members seem very concerned about feeling safe in the group.
d	People seem hesitant to express a different point of view
e	People express frustration with how this group is functioning

¹ <https://bth.azure-api.net/plan/api/kursplan?kurskod=PA1450&revision=5&lang=en>

f	The group seems split over a number of issues
g	There is quite a bit of tension in the group at this time
h	Trust is high in this group
i	Members are satisfied with the group's progress
j	We can rely on each other. We work as a team
k	The group acts on its decisions
l	The group encourages innovation and creativity
m	Commitment to the group's task is high in this group

The Group Development Questionnaire (GDQ) originally proposed by Wheelan (2014) contains 60 items. During a pilot involving a Ph.D. student in SE, and after consultation with colleagues from the Pedagogy Department, we selected an alternative, less demanding instrument with a reduced number of questions—GDQS (Gren et al. 2018). GDSQ—*Group Development Questionnaire Short*—is a 13-item version of GDQ developed and validated across several studies. GDQS includes four items for Stage I and three items for Stages II–IV of Group Development. It had consistent internal validity across studies, which makes it acceptable and useful for our longitudinal study. All items in Table 1 are answered using a 5-point Likert scale to express agreement.

As the COVID-19 pandemic forced new working conditions among the participants, we included the following two questions:

- n) Rate the complexity of working with your group-mates remotely.
- o) How did working remotely influence your group work?

Question n is answered on a 5-point scale expressing complexity, whereas question o is open-ended, asking the students to reflect on their experience (e.g., challenges, reactions) in collaborating remotely with the rest of their group.

Settings.

The course introduces software development methods and stresses aspects of collaboration over time (i.e., developing software in a team rather than individually). The groups go through the main phases of software development projects and use models, approaches, and tools specific to each phase. The groups (18 in total) consisted, on average, of four students.

The project itself consisted of the development of software for a robot vehicle to perform several tasks (e.g., remote control over the Internet, self-driving around a pre-defined route). The hardware was provided to the students, who could decide to work on it in the lab or at home. The main programming language was Python, although no limitations on the usage of other frameworks was imposed. The project accounted for 2.5 credits of the total six credits planned for the course. Although being familiar with the content and the setup of the course, none of the authors of this paper took an active part in it (e.g., as lecturer, teaching assistant).

Operations.

After the start of the course, which touched on theoretical aspects, we administered the first survey in March 2020—i.e., roughly a month after the group members started working together. The survey was open until March 31, 2020. We informed the students about the survey at the beginning of the course (January 2020) and reminded them to fill it in throughout the first data collection period via announcements on the course e-learning platform.

The second iteration of the survey, planned for April 2020, was postponed because of closing of the university premises and the shift to remote collaboration. Accordingly, the second data collection took place at the beginning of May 2020, and the survey was closed on May 19 after three reminders were sent to the students (via email, e-learning platform, and during one of the online classes with the course lecturer). As the course ended on June 6, 2020 the planned third survey iteration did not happen.

When filling in the surveys, the respondents used a randomly generated unique ID, which we used to trace answers between the two iterations. The survey was administered online using the Questback tool.² Figure 2 summarizes the main activities taking place during the study.

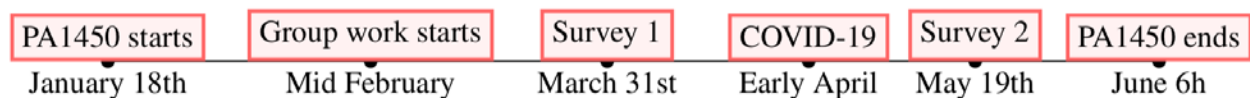


Figure 2: Main activities during the study execution.

Ethics.

The students gave their consent to be part of both iterations of the survey. We clarified the purpose of the survey so that they could make an informed judgment about whether they wished to participate. Moreover, we disclosed any sponsorship. We provided a 110 SEK voucher for a meal to all participants. We kept the results confidential and anonymous and treated the data according to the GDPR regulation.

Data analysis.

We analyzed the questionnaire Likert items (Questions *a–n*) using descriptive statistics and, where possible, inferential statistics. The former method summarizes and organizes data using tabular and graphical representations, whereas, the latter generalizes the sample parameters to the population. The data were analyzed using the R statistical

² <https://www.questback.com>

package (R Core Team 2020). The results were used to answer $RQ_{GroupDev}$ and $RQ_{Background}$.

We applied thematic analysis (Braun and Clarke 2012) to Question *o*. In thematic analysis, relevant themes are identified from qualitative data. The resulting themes were used to answer RQ_{Remote} .

Results

In this section, we present the results of the survey analysis, including the questions related to the remote collaboration conditions. We also answers our research questions.

Data Quality

Figure 3 reports an overview of the response rate for both surveys. Survey 1 received 29 impressions (i.e., individual participants who opened the link); this represents our adjusted total sample. From this sample, 17 participants started filling in the survey, but only 11 completed it.

Survey 2 received 14 impressions and was started and completed by 12 respondents. We removed one data point from Survey 2 as the respondent did not take part in Survey 1. Finally, we retained answers to both surveys from the same respondents (i.e., having matching IDs).

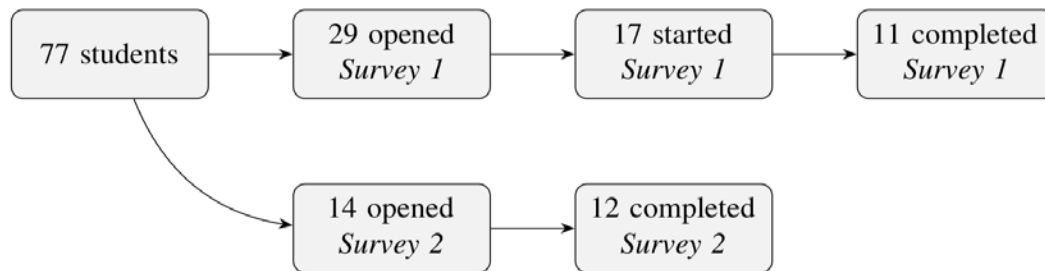


Figure 3: *Surveys completion rates statistics.*

Group Development

The descriptive statistics summarizing the answers to the GDQS are reported in Table 2. Regarding the first stage, there is a visible increment between the first and second iteration for statement *b*. This result was not expected as the group should have been more inclusive already in the second iteration. The second stage shows mixed results. In general, the groups seems to be quite tense (i.e., statement *g*), and frustration has increased (i.e., statement *e*). However, comparing the answers to statement *f*, it seems that the environment is becoming more stable. The previous unexpected results are also supported by the results from the third stage. Trust has decreased—although from the upper-end of the agreement scale—and the group members perceive that they are not making progress towards working as teams. However, there is an increased satisfaction regarding the work

towards solving the project task. The results observed for the final stage are also surprising. The groups already show mature traits during the first iteration (i.e., statements *k* and *l*). However, as predicted by the theory, commitment to work and productivity increase with time.

Figure 4 shows the aggregated results for the four stages over the two survey iterations. Visual inspection of the answer distribution does not support Wheelan's theory. Although we expect a right-skewed distribution for Iteration 1 and a left-skewed distribution for Iteration 2, we observed that there was no difference between the two regarding the first three stages, and only a small increment in Iteration 2 for Stage 4, as predicted by the theory.

Item	Iteration	min	max	mean	median	std. dev.
a	1	3	5	3,90	4,0	0,56
	2	2	2	5	3,91	4,0
b	1	2	4	2,70	2,0	0,94
	2	2	1	4	3,08	3,0
c	1	1	3	1,50	1,0	0,70
	2	2	1	4	2,16	2,0
d	1	1	3	2,00	2,0	0,81
	2	2	1	4	2,58	2,0
e	1	1	4	1,90	2,0	0,87
	2	2	1	4	2,50	2,5
f	1	1	4	2,40	2,0	0,96
	2	2	1	3	2,00	2,0
g	2	1	4	1,90	1,5	1,10
	2	2	1	3	1,75	2,0
h	1	3	5	4,30	4,5	0,82
	2	2	3	5	4,25	4,5
i	1	3	5	4,10	4,0	0,56
	2	2	3	5	4,00	4,0
j	1	4	5	4,50	4,5	0,52
	2	2	3	5	4,16	4,0
k	1	4	5	4,50	4,5	0,52
	2	2	3	5	4,25	4,0
l	1	3	5	4,10	4,0	0,73
	2	2	2	5	3,58	4,0
m	1	3	5	4,20	4,0	0,78

	2	2	3	5	4,25	4,0
--	---	---	---	---	------	-----

From visual inspection of the results, it seems that GDS do not take place during an SE students' group work. There is some evidence that the groups have matured only within Stage 4.

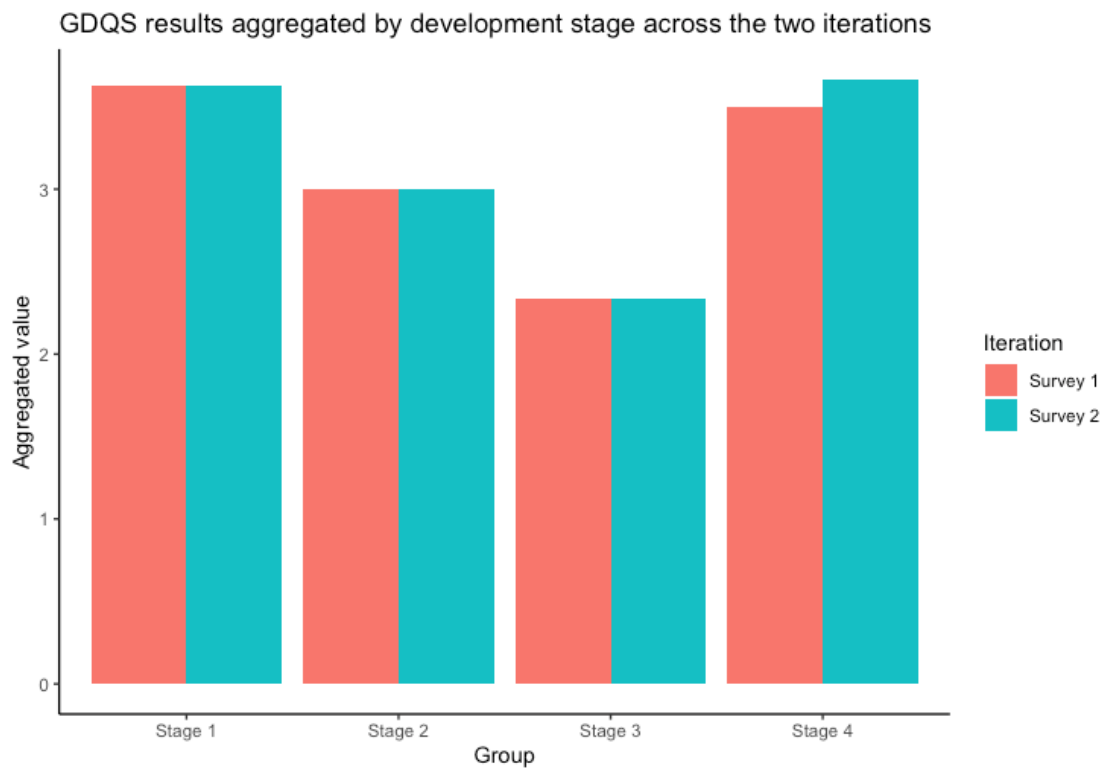


Figure 4: Group maturity over the four stages.

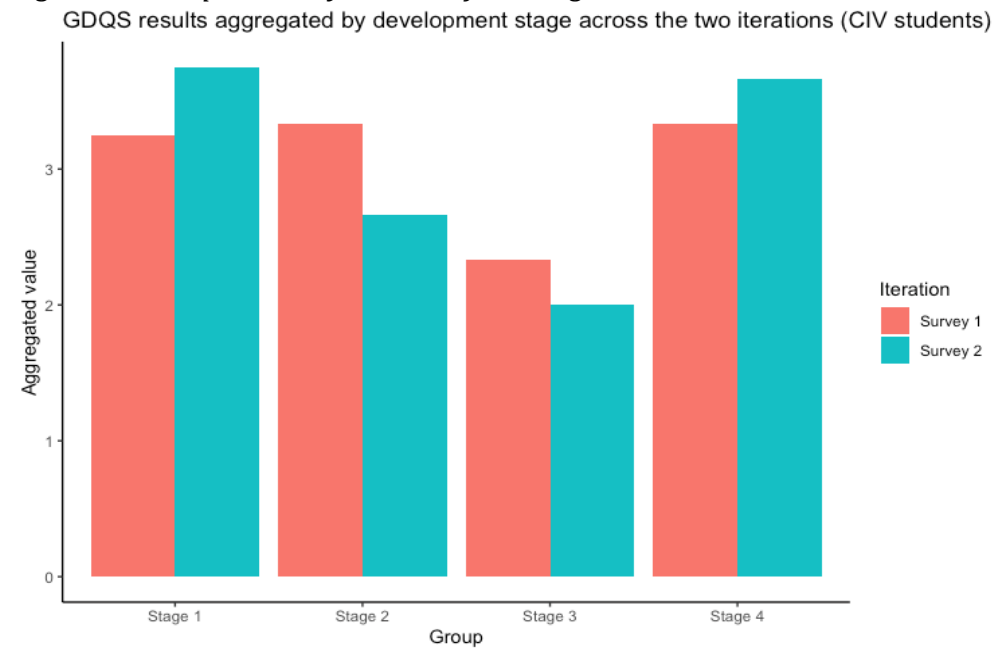


Figure 5: Group maturity over the four stages for CIV students.

Students' Background

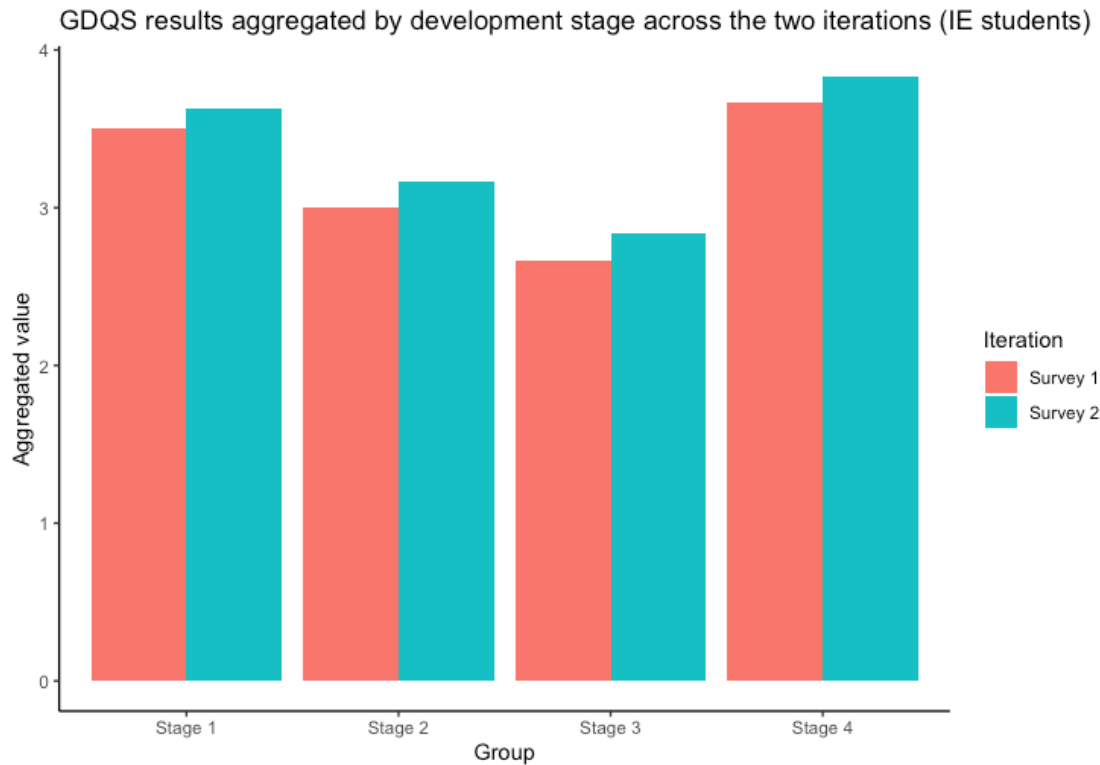


Figure 6: Group maturity over the four stages for IE students.

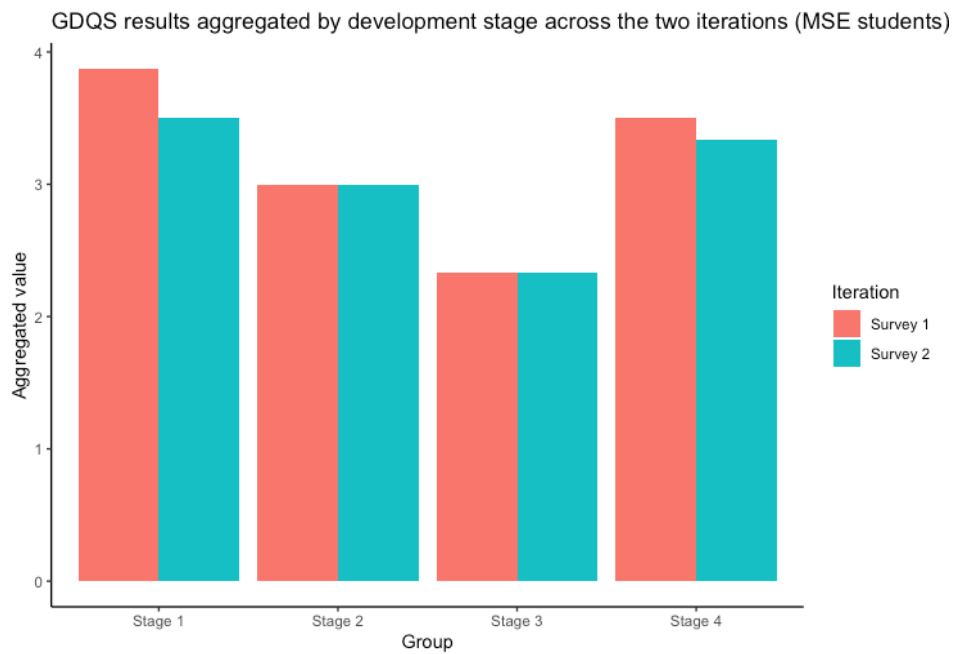


Figure 7: Group maturity over the four stages for MSE students.

Figures 5–7 show the aggregated results for students from the different backgrounds and their evolution over the GSD. None of the student segments follows the expected manner in which their groups should develop over the different stages according to Wheelan's theory. However, there are differences between the different student groups. CIV students evolved with respect to inclusion of members in the group and working productively, whereas they devolved in the middle stages (i.e., fight and structure). IE students seem to be the ones to which Wheelan's theory applies the most with respect to students with other backgrounds—they matured, albeit only to a limited degree, throughout all the stages. Finally, MSE students show results that deviate most from the theory as they did not mature in any of the four stages.

Regarding Stage 1, the results for Civil Engineering students (Figure 5) were also unexpected, and likewise for Stage 2—although with higher variation with respect to the overall results (e.g., statement *e*). Stage 3 shows contrasting results—statement *i* is in the direction predicted by the theory, while *j* is not. For this segment of students, Stage 4 seemed to have occurred earlier.

For Industrial Economics (IE) students (Figure 6), there seems to be a maturation within Stage 1 (i.e., statements *a* and *b*). Stages 2 and 3 also show contrasting results that do not seem to agree with Wheelan's theory, whereas Stage 4 seems to also show a development for such students.

Finally, for Masters' Software Engineering (MSE) students (Figure 7) there are contrasting results for Stage 1 and Stage 2. This segment of students did not seem to mature within Stage 3, which also shows a high level of dissatisfaction and reliance on peers. For this segment, the results for Stage 4 are mixed and do not seem to be aligned with what the theory predicts.

From visual inspection, it seems that groups with different backgrounds experience slightly different development over the four stages. In particular, IE students seem to mature according to Wheelan's theory, whereas that is not the case for CIV and MSE students.

Remote Collaboration

The answer to the last item in the questionnaire, related to the experience of collaborating remotely, is reported in Figure 8. The majority of the students seemed not to have a problem with remote collaboration and evaluated it as *easy* or *very easy*, whereas only a small fraction perceived it as complex.

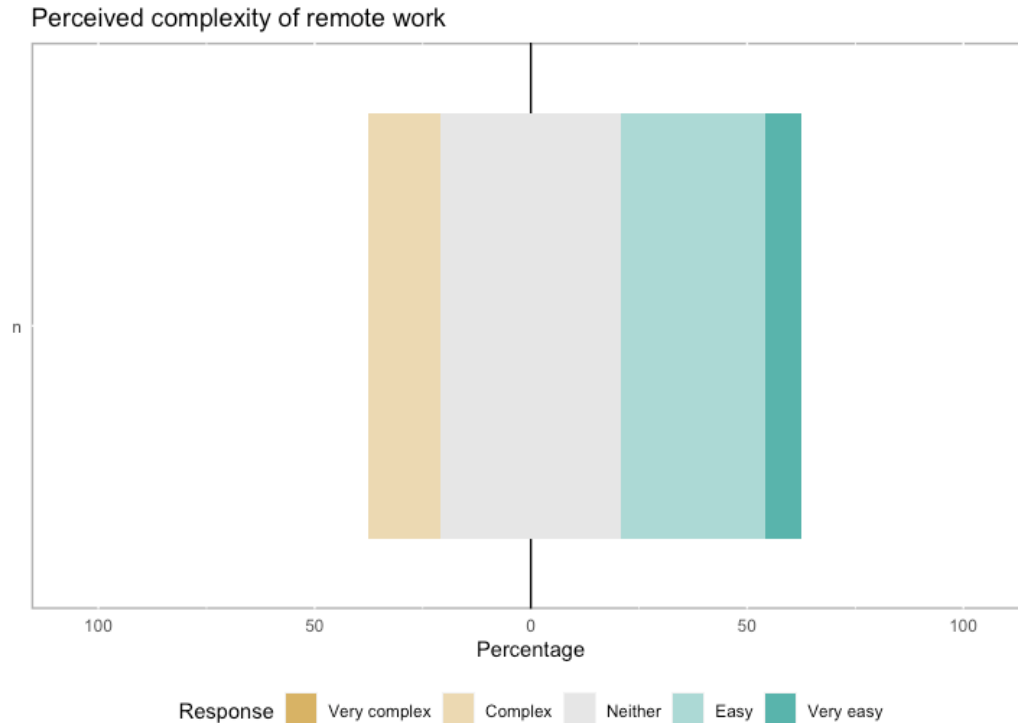


Figure 8: Answers to question “Rate the complexity of working with your group-mates remotely.”

We analyzed the answers to the open question “How did working remotely influence your group?”

We found three themes in the data, *Technology*, *Scheduling*, and *Social aspects*. In particular, issues with Technology are associated with the sub-theme of *Distraction* as shown in the following quote.

People start doing other things on their computer that are not a part of the project.

The distraction caused by remote work may have caused delay in Stage 4 of the GDS by impeding teams members’ productivity. On the other hand, the data show positive statements regarding the possibility of scheduling meetings more easily and of increased flexibility for the members in when to accomplish their work. This is possibly due to the time gained by avoiding commuting, and scaling down on other campus activities.

For Social aspects, we found three sub-themes. The first, *Limited interaction*, was perceived negatively as physical distance imposed more structured communication and interactions. In this regard, one student says:

Not being able to point at each other’s screens. Not being able to sketch stuff on a whiteboard or a piece of paper.

This could explain why we did not observe a development in the early stages, for which communication is fundamental to achieve members' inclusion and resolve fights.

Another sub-theme referring to the social aspect is *Peer support*. Working remotely seems to have a negative impact on this aspect, as indicated by the following quote:

Bringing the best out of each other is important if we are going to work together and efficient[ly]. This is impossible when we are working remotely.

This aspect also seems to be related to the group members' struggle to be inclusive and build reliance on peers.

Finally, the last aspect related to social interaction, which seems to be negatively impacted, is *Peer personality*. It seems that the lack of co-located work and the fact that the work interaction happened online revealed differences in terms of members' characters that had a negative impact. For example, one student says:

It is more difficult with these [sic] type of people with strong opinions who think they know best but don't always do [sic].

Similarly, this aspect also negatively influenced the development of the groups in the early stages as a sense of loyalty was either developed late or never developed at all, thus hindering the open sharing of ideas.

Although the students perceived that collaborating with the group members was not complex, it seems that these circumstances had an impact on their development through the different stages. In particular, early-stage development was hindered.

Discussion and Limitations

Group work can be an effective method to motivate students and promote active learning and critical-thinking, as well as to improve communication and decision-making skills. However, it requires teachers to plan and monitor the groups to not thwart the students. Teachers can benefit from understanding at which stage the students are situated in order to suggest (or enforce) corrective actions specific to that stage and accelerate the group progression towards more productive stages. For example, in early stages, the teacher may address issues with leadership of the group.

In the Results section, we showed that there is not a strong indication of group development taking place in our sample of SE students. However, a common misconception is that a group exists only at one stage at a time. In reality, a group should obtain higher values on one scale, while getting some result across *all* scales (Wheelan 2014). In particular, this seems to be the case for our results—all four stages are equally covered with the exception of Stage 4 (i.e., the only stage that consistently shows some maturation between the two survey iterations). One reason for such a result is that the groups matured quickly and our observations simply missed their development. In such a case, our results are based on insufficient observations—i.e., observations taken several weeks between each other. Therefore, we may have missed the complete maturation process because of

the low number of survey iterations. On the other hand, we cannot exclude that a complete maturation took place over the four months of the study, and that being able to observe a full development requires a longer time frame.

Another explanation for our results is that in our analysis we considered an aggregate sample of students working in groups, rather than considering individual groups. This was necessary to maintain student anonymity. Our assumption was that groups should be maturing at roughly the same pace, and we conjectured that the main factor that would eventually explain a different progression through the stages is the background of the individual students in the groups, as investigated in $RQ_{Background}$, rather than the group itself. An early maturation (e.g., as observed for Stage 4 of the first survey iteration) can be explained by students working with the same peers from past projects. We did not control for this factor since we did not look at single groups, and we may be not be able to do so in the future for ethical reasons (e.g., asking students to avoid being in a group with peers from previous projects).

These preliminary results do not provide a diagnosis for all stages, but offer at least some indication that a progression towards more mature stages takes place, and it is perceived by the individuals participating in group work. This result is reinforced by the observation that at least one sub-group of students (IE students) followed a development closer to the one postulated by Wheelan's theory. We observed a different pattern for CIV and MSE students, which we currently cannot explain using the data available. Albeit with healthy skepticism, we recommend teachers using group work in their SE courses to take into account that students may mature to a productive level (Stage 4) faster than expected. Moreover, student background needs to be taken into account when assessing group maturity level, before taking action.

Finally, our investigation of how students develop in their groups during remote work showed that several social aspects impact collaboration. This could hinder groups reaching full productivity. There is evidence from investigations in the field of professional software development showing that individuals perceive a lower productivity when working from home compared to normal settings (Ralph et al. 2020). Therefore, we recommend teachers dealing with collaboration within remote groups to expect lower group productivity, as the first stages of GDS are the ones where most discussion and resolution will take place. Shifting such discussion to an online medium will naturally take more time to arrange, and may need more time to perform.

Conclusion and Future Work

In this paper, we presented a study based on Wheelan's Group Development Stages in the context of SE group work. We surveyed 11 students on two different occasions using a simplified version of the Group Development Stage Questionnaire. We did not observe a full group development taking place over the four stages presented in Wheelan's work, as the students in our group seemed to have matured quickly towards the last stage, while earlier stages were also covered. We showed that students' background can have an impact on such a result. Finally, as the group were forced to collaborate remotely during the COVID-

19 pandemic, we showed that some social aspects of this new way of working can also impact the development of the groups.

In the future, we will replicate this study, under the same working conditions, but with a larger sample and over a longer time span to get more precise observations. We will also consider how single groups evolve over time and whether such maturation is also perceived by the groups' supervisors and teachers. Finally, we will correlate the evolution of the groups with tangible metrics that measure the groups' productivity.

References

- Branine, Mohamed. 2008. "Graduate Recruitment and Selection in the UK: A Study of the Recent Changes in Methods and Expectations." *Career Development International* 13: 497–513. doi:10.1108/13620430810901660
- Braun, Virginia, and Victoria Clarke. 2012. *Thematic Analysis*. American Psychological Association.
- Burke, Alison. 2011. "Group Work: How to Use Groups Effectively." *Journal of Effective Teaching* 11 (2): 87–95.
- Chang, Yunjeong, and Peggy Brickman. 2018. "When Group Work Doesn't Work: Insights from Students." *CBE—Life Sciences Education* 17 (3): ar52.
- Cohen, Elizabeth G. and Lotan, Rachel A., 2014. *Designing groupwork: Strategies for the heterogeneous classroom Third Edition*. Teachers College Press.
- Cumming, J. 2010. "Student-Initiated Group Management Strategies for More Effective and Enjoyable Groupwork Experiences." *Journal of Hospitality, Leisure, Sport, & Tourism Education* 9. doi:10.3794/johlste.92.284.
- Gersick, Connie JG. 1988. "Time and Transition in Work Teams: Toward a New Model of Group Development." *Academy of Management Journal* 31 (1): 9–41.
- Gren, Lucas, and Khaled Al-Sabbagh. 2017. "Group Developmental Psychology and Software Development Performance." *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*, 232–34.
- Gren, Lucas, Alfredo Goldman, and Christian Jacobsson. 2019. "The Perceived Effects of Group Developmental Psychology Training on Agile Software Development Teams." *IEEE Software* PP (99): 1–10.
- Gren, Lucas, Christian Jacobsson, Niklas Rydbo, and Per Lenberg. 2018. "The Group Development Questionnaire Short (GDQS) Scales: Tiny-Yet-Effective Measures of Team/Small Group Development." <https://doi.org/10.31234/osf.io/u3p8c>.
- Hackman, J Richard, and Charles G Morris. 1975. "Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration." *Advances in Experimental Social Psychology* 8: 45–99.
- Hammar Chiriac, Eva, 2014. Group work as an incentive for learning—students' experiences of group work. *Frontiers in psychology*, 5, p.558.
- Johnson, D. W., R. T. Johnson, and K. A. Smith. 1998. "Cooperative Learning Returns to College: What Evidence is There That It Works?" *Change* 27–35. doi:10.1080/00091389809602629.

McDonald, Sharon, and Helen M Edwards. 2007. "Who Should Test Whom?" *Communications of the ACM* 50 (1): 66–71.

Popov, Vitaly., Brinkman, Dine, Biemans, Harm.J., Mulder, Martin, Kuznetsov, Andrei and Noroozi, Omid, 2012. Multicultural student group work in higher education: An explorative case study on challenges as perceived by students. *International Journal of Intercultural Relations*, 36(2), pp.302-317.

Ralph, Paul, Sebastian Baltes, Gianisa Adisaputri, Richard Torkar, Vladimir Kovalenko, Marcos Kalinowski, Nicole Novielli, et al. 2020. "Pandemic Programming: How Covid-19 Affects Software Developers and How Their Organizations Can Help." *arXiv Preprint arXiv:2005.01127*.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Schmidt, Frank L, Huy Le, and Remus Ilies. 2003. "Beyond Alpha: An Empirical Examination of the Effects of Different Sources of Measurement Error on Reliability Estimates for Measures of Individual-Differences Constructs." *Psychological Methods* 8 (2): 206.

Shaw, Marvin E. 1981. *Group Dynamics: The Psychology of Small Group Behavior*. McGraw-Hill College.

Springer, Leonard., Stanne, Mary-Elizabeth. and Donovan, Samuel. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: A meta-analysis. *Review of Educational Research*, 69(1): 21–51.

Steiner, Ivan D. 1986. "Paradigms and Groups." In *Advances in Experimental Social Psychology*, 19:251–89. Elsevier.

Sundstrom, Eric, Kenneth P De Meuse, and David Futrell. 1990. "Work Teams: Applications and Effectiveness." *American Psychologist* 45 (2): 120.

Tuckman, Bruce W. 1965. "Developmental Sequence in Small Groups." *Psychological Bulletin* 63 (6): 384.

Tyastuti, Dwi, and Devy Ariany. 2017. "The Pattern of Intragroup Conflict and Group Performance During Interprofessional Teamwork Process Among Students of Medical and Health Professionals." In *1st International Integrative Conference on Health, Life and Social Sciences (Ichlas 2017)*. Atlantis Press.

Volet, Simone. and Mansfield, Caroline., 2006. Group work at university: Significance of personal goals in the regulation strategies of students with positive and negative appraisals. *Higher Education Research & Development*, 25(4), pp.341-356.

Wheelan, Susan A. 2014. *Creating Effective Teams: A Guide for Members and Leaders*. Sage Publications.

Wheelan, Susan A, Barbara Davidson, and Felice Tilin. 2003. "Group Development Across Time: Reality or Illusion?" *Small Group Research* 34 (2): 223–45.

Wheelan, Susan A, and Judith M Hochberger. 1996. "Validation Studies of the Group Development Questionnaire." *Small Group Research* 27 (1): 143–70.

Zellmer-Bruhn, Mary E. 2003. "Interruptive Events and Team Knowledge Acquisition." *Management Science* 49 (4): 514–28.