A Follow up Study of the Effect of Personality on the Performance of Software Engineering Teams

John Karn
Dept of Computer Science,
University of Sheffield,
Regent Court,
Sheffield, UK,
(+44) 114 2221958
J.Karn@dcs.shef.ac.uk

Tony Cowling
Dept of Computer Science,
University of Sheffield,
Regent Court,
Sheffield, UK,
(+44) 114 2221823
A.Cowling@dcs.shef.ac.uk

ABSTRACT
This paper describes ethnographic observations and analysis of the performance of student teams working on year-long software projects (2004-2005 UK academic year) for industrial clients. Personality types were measured using an online version of the Myers Briggs Type Indicator (MBTI), as a basis for studying how individuals interacted within the teams, and the effects of disruptive issues on the quality of work produced by the team. The behavior of the observed teams is analyzed and the results compared with those from the previous year's (2003-2004) research, also carried out on student teams. A significant finding in 2003-2004 was that issues which teams did not discuss adequately caused more problems for the quality of work than issues which produced actual disruption within the team; the results from 2004-2005 differ in that actual disruptions proved most damaging to the teams involved.

Categories and Subject Descriptors
D.2.9 [**Management**]: Programming Teams – *personality types, conflict in teams, group-think.*

General Terms: Management, Human Factors

Keywords
Jungian Personality Types, Myers-Briggs Type Indicator, Group Work in Software Engineering, Ethnographic Observations

1. INTRODUCTION
The work described in this paper builds on previous work [9] carried out by the authors, which studied the interactions of personalities in software engineering (SE) teams and how disruptions and lack of debate between individuals affected the overall performance of the team.

The initial expectations when starting the previous study were that a team that experienced a large amount of disruption would exhibit the most serious adverse effects on the quality of the work produced, and that such damage would have repercussions on the kinship and rapport between different team members. The ways in which disruptions led to conflicts within the teams had been observed in earlier studies [8], [7] and were thought to be the most likely source of poor quality products at the outset of the 2003-2004 study. While the initial expectation for the previous study to this [9] was that greater disruption would be more damaging, the opposite proved to be true, in that teams without disruption but who lacked debate and suffered from inadequate communication encountered serious problems, failing to meet project milestones. Such negative outcomes for observed teams without sufficient debate included: limiting discussion to only a few alternatives; or initial solutions never being reconsidered; or alternatives either not being proposed or being ruled out by the majority of the team, resulting in further lack of debate on future issues.

In this previous study the authors had analyzed 5 student teams, each with 5 members, and so the total number of students yielded 25 data points, full details for all teams observed during the previous study 2003-2004 are given in [10]. For the follow up study described here 3 student teams were observed increasing the number of data points to 40. The aim of this research was to build on the previous study by acquiring more data points and adding to the reliability and validity of the earlier study by analyzing results and comparing the effects of no debate and disruptions between the teams from 2003-2004 with those from 2004-2005.

The context for both of these studies is the SE Observatory at the University of Sheffield; this is a facility that aims to allow researchers to carry out studies on student teams working on real software development projects for real clients, and that last for either a semester or an academic year. The specific project in question for this research is the "Maxi Project", as it run in 2004-2005 which is a Master's-level team-based two semester's SE course that involves the analysis, design and implementation of a complete new system for an external client.

The structure of the rest of this paper is therefore as follows: section 2 gives a description of the relevant personality research; section 3 describes the research methods used; section 4 explains how the observed behavior was classified and analyzed; and section 5 summarizes the main results for the teams studied, who were all undertaking the "Maxi project" during 2004-2005. Section 6 then aggregates these results, section 7 discusses the findings of both the 2003-2004 and 2004-2005 research in both

ISESE'06, September 21–22, 2006, Rio de Janeiro, Brazil.

the current and historical context, and finally section 8 draws conclusions from the work done and proposes future work.

2. PERSONALITY RESEARCH

Personality can have an affect on group processes and decision making. It can be measured according to traits and types, however these are not different definitions as types are collections of packaged traits. A trait is nothing more than an enduring characteristic of the individual's behavior in a wide variety of situations. A type is a categorization according to a fixed pattern of traits and provides shorthand for talking about a set of separate traits by making the assumption that generally recognized patterns of personality characteristics exist.

The theory underlying this research is provided by the work of Jung [6], as developed by Myers-Briggs [11] in designing the MBTI (Myers-Briggs Type Indicator), and applied to SE development teams as described in [9], [8] and [7] . Jung believed that people adopted different ways of relating to experience as they attempted to evolve towards self-hood. They adopted different attitudes towards life and utilized different psychological processes or functions to make sense of their individual experiences. Jung became convinced that people did indeed belong to different psychological types and he set about documenting this in his book "Psychological Types" [6] which was first published in Zurich in 1921.

Jung distinguished four functions thinking, feeling, sensation and intuition. These four functions were also associated with an attitude type, either extraversion or introversion. The idea of stimulating and advancing academic research was always at the forefront of Jung's world view. He claimed that his psychological typology was a central tool for researchers; it would provide definite points of view and guidelines to allow them to reduce the chaotic profusion of individual experiences to any kind of order. Essentially Jung's work in this area forced people to realize that individual differences are important and must not be overlooked.

Myers-Briggs wanted to take Jung's theories out of universities and propel them into mainstream life. Jung saw his theory as an aid to self-understanding and as a subject to be debated by academics in universities. Many of his writings were aimed at academic audiences and as a result of this his theory found scant enthusiasm among the general public interested in personality. The application of Jung's theory extends beyond the point he was content to stop. Myers correctly noted that the theory shed light on the way individuals perceive and judge and on the things that they value most; the type concepts are useful whenever one person must communicate with another or live with another or make decisions that affect another's life. They thus developed the MBTI, which is widely used to assess an individual's personality style on four dimensions: social interaction, information gathering, decision making, and dealing with the external world.

The test used in this research is not the official paper version of the MBTI, but an online test based on the MBTI developed by Human Metrics, a consortium of Israeli psychologists [5], who claim they have found no significant statistical differences between this test and the official paper version of the MBTI. An important aspect of this form of the test is that there is no risk of the participants drawing potentially harmful conclusions from its results, as these can not suggest that they might be psychologically unhealthy.

The official MBTI is a Jungian based inventory that uses a pen and paper self-report format. It is composed of 94 forced choice items that constitute the four bipolar discontinuous scales that are implied in Jung's theory, Introversion-Extraversion, Sensation-iNtuition, Thinking-Feeling and Judging-Perceiving. Respondents are classified into one of 16 personality types based on the largest score obtained for each bi-polar scale (i.e. a person scoring higher on extraversion than introversion, intuition than sensing, thinking than feeling and judging than perceiving would be classified as an Extraverted iNtuitive Thinking Judging type) or ENTJ. On completion of the test a subject will also know the clarity of preference for each function and attitude: an example is I 52% N33% T22% J42%. An indicator for clarity of preference is:

- 40 % or Higher (30 for T/F) - Very clear preference.
- 31-39 % (21-29 T/F) - Clear preference.
- 11-20 % - Moderate Preference.
- 1-10 % -Slight Preference.

Another popular personality test is the NEO-PI model which is based on the five factor model of trait personality [2]. This model is widely used in academic circles and differs from the MBTI in the sense that it seeks to measure traits as opposed to types. Furnham [4] believes that both models would benefit by examining the behavioral and cognitive correlates of the various dimensions of both scales that overlap, but this is outside the scope of this paper. Since there was no online version of this test available at the time it was not appropriate for use in this study

3. RESEARCH METHODS

Ethnographic methods were employed to observe selected student teams. In order to gain an in-depth understanding of the user culture (in this context SE teams), the authors observed, and inquired about the research subjects' normal activities throughout their specific projects. Ethnography is a rigorous research method in which good data collection and analytical skills are vital. This method aims to allow the researcher to arrive at a theoretically comprehensive understanding of a group or culture.

Ethnographers claim that the process of learning behavior is absent from other forms of research. For instance a questionnaire asks questions at a particular time. It is a static causal snap-shot of attitudes, how and why people change is not understood. The design of a questionnaire involves the researchers developing ideas and testing or exploring them using questions. Critics argue that researchers using this method already know what is important. In contrast ethnography is said to make no firm assumptions about what is important. Instead, the method encourages researchers to immerse themselves in the day to day activities of the people whom they are attempting to understand. In contrast to testing ideas (deductive), they may be developed from observations (inductive). Ethnography leads to an empathic understanding of a social scene. It is said to exclude over time the preconceptions that researchers may have and express them to a new social milieu which demands their engagement.

The way in which these methods were applied in this study is that students were initially asked to complete an online test based on the MBTI. This personality data was then collected by the

researchers and the Maxi students agreed to take part in the observations, they were then instructed to inform researchers of any meetings that were to take place, whether on or off campus. This was usually done by e-mail, although on some occasions it was also done by SMS. The students proved to be very helpful and provided details of meetings on a regular basis, even meetings arranged at the last minute. Once a time and a venue had been arranged it was then left to the researcher to make his own way to the meeting, the usual place to sit was at the back of the room out of the way of the main discussion. This allowed one to make field notes with minimal interference in the meeting. Field notes were taken over the course of the academic year and were analyzed at three separate intervals.

The first interval was during the Christmas break, then during Easter and finally at the end of the project towards the end of May. The collection of field notes over the academic year provided one with a detailed picture of the behavior and actions of both the team and the individuals within it.

The other important aspect of this process was that of anonymising results. To ensure complete anonymity of subjects each person was given a randomly generated number within each team, such as A1 and B2 and these were used throughout the field notes and for publication of the results. The subjects showed a high level of acceptance of the observations.

In classical ethnography data collection and analysis and the decision as to when to withdraw from fieldwork take place in what is referred to as 'theoretical sampling' and 'theoretical saturation', the latter referring to the time when observations no longer serve to question or modify the theories generated from earlier observations, thus rendering the theory saturated with data. For the work described in this paper, the 'theoretical saturation' was taken out of the researchers hands, as it took place when the student projects finished.

This did not have an adverse effect on either the quality or the quantity of collected data. Ethnography is not an easy method to use. It requires researchers to spend a great deal of time in surroundings with which they may not be familiar; to secure and maintain relationships with people with whom they may have little personal affinity; to take copious notes on what would normally appear to be everyday happenings; to spending a lot of time on the analysis after the fieldwork.

Ethnography is about engaging in a social scene, experiencing it and seeking to understand and explain it. The researcher is the medium through which this takes place. By listening and experiencing, impressions are formed and theories considered, reflected upon, developed and modified. Thus as demonstrated in this study it is a systematic and disciplined approach which, if performed well greatly assists in understanding human activities and brings with it new ways of viewing the social world.

4. CLASSIFICATION SCHEMES

4.1 Ordinal Ratings

Three Maxi teams from 2004-2005 were observed and described in this study. There were several phases involved in going from the field notes stage into quantifiable data. These phases are described in detail in [8]. Previously, the authors had created an ordinal scale to measure what is now being called the 'level of disruption', and this is described in [7] and [8].

Table 1. Level of Disruption

Level	Kind of Disruption
1	Premise uncritically accepted with no interaction between team members
2	Dealt with smoothly and harmoniously after a brief discussion
3	Lengthy period of constructive debate discussing the virtues of an issue
4	Caused slight disruption by forcing people off relevant issues
5	Lengthy period of destructive debate, lengthy disruption
6	Caused complete disruption to the work of the team

The following scheme was developed to aid in deciding which area an issue was related to:

- PSI- Project Specific Issue
- MSI- Methodology Specific Issue
- GSEI- General Software Engineering Issue
- THF- Team Human Factors
- C- Client
- M- Manager

For the 2003-2004 study additional scales were created to deal with issues which were rated 1, 4, 5 and 6, and which were expected to have some impact on the quality of the work produced. It was observed that there were two kinds of effects: an internal impact where the relevant issues were resolved without management intervention and an external impact when management was forced to intervene.

Table 2. Internal Impact of Disruption

Level	Internal Impact
0	No Impact
1	Small changes made to document
2	Large scale restructuring to one document
3	Major restructuring to more than one document
4	Complete re-write of documentation
5	No working system at end of project

Table 3. External Impact of Disruption

Level	Managerial Involvement
0	No Impact
1	Management intervention led to small changes
2	Major large scale restructuring due to management criticism
3	Large scale restructuring to several documents
4	Marks lost for area of project
5	Deadlines missed, marks lost for late work

5. RESULTS

5.1 Team 1

Table 4. MBTI Type Team 1

ID	Type	E-I %	S-N%	T-F%	J-P %
1A	INFP	I 44	N 33	F 44	P 22
1B	INTP	I 44	N 22	T 11	P 33
1C	ENTJ	E 44	N 33	T 44	J 11
1D	ENFJ	E 11	N 22	F 11	J 22
1E	INTJ	I 33	N 67	T 56	J 1

Table 4 shows the MBTI type for each member of Software Hut Team A. The E-I, N-S, F-T and J-P dichotomies indicate the preferred function or attitude for each team member. The number indicates the clarity of preference. Such tables are used for each team.

This team managed to work remarkably well over the entire project, despite great differences in both personality and ethnicity. This team wasn't dominated by one or two individuals, and communication between the members was excellent throughout. As Figure 1 shows, the levels of disruption for all issues for this team give a very positive picture.

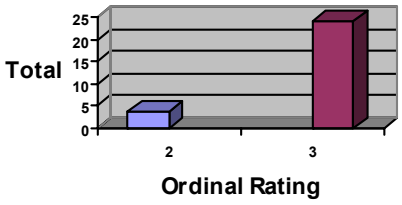

Figure 1. Level of Disruption for all issues Team 1.

1A (an INFP) and 1C (an ENTJ) were the main spokesmen for this team, although they were ably assisted by other team members. 1A in accordance with type theory was a sensitive individual, who promoted positive growth amongst team-mates. 1A was very polite, a good listener and had the ability to put team-mates at ease, 1A was genuinely interested in understanding points from team-mates, and was keen to avoid any conflict. Even if 1A strongly disagreed with someone there was no animosity, instead the focus was on intuitively understanding the other person's perspective and feelings. As an INFP, 1A focused on feelings and human conditions and attempted to avoid rushing to any impersonal judgments.

1C worked very well with 1A. 1C as an ENTJ with clear preferences had a natural tendency to organize and direct the team. 1C was decisive, knowing what needed to be done, and was confident when assigning roles to team-mates. Although 1C was very strong in debate, there were no instances of this character becoming confrontational, despite not being tuned in to people's feelings; a lot of this was down to the personality types of team-mates. 1C was determined not to repeat any mistakes, and if the manager or client commented on a particular piece of work, 1C ensured that these comments were taken into account whilst either amending a previous piece of work or starting a new piece.

The other key member of this team was 1E whose knowledge of technical issues proved to be invaluable. 1E's type is INTJ. 1E was the person most likely to answer any technical query or deal with any problem in this area. 1E was aided and abetted in this by a hardworking though subservient colleague in 1B (INTP). 1E was very insightful and was quickly able to understand and then apply a concept. This came to the fore when technical issues such as XML, Java, Object-Oriented Design, and Software Testing were discussed.

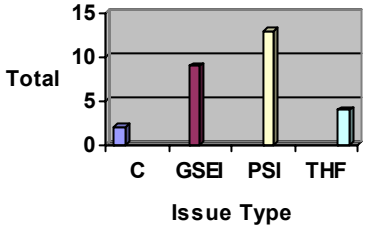

Figure 2. Level of Disruption for all issues Team 1.

The final member of this team 1D is an ENFJ. 1D had very good people skills and sought to bring out the best in team mates. 1D focused on understanding, supporting and encouraging team mates, especially when working on documentation. Although at the same time 1D was always honest when assessing others' work, this honesty was appreciated by team mates. 1D's intuitive awareness of peoples needs coupled with the preference for feeling helped to create a positive human centered environment that was conducive to good team work.

Figure 2 shows the kind of issues that were discussed by this team. One THF issue caused slight concern, but was still discussed in a rational manner. This was the matter of 1A checking work for grammatical errors, this was a large extra burden but was deemed necessary after client and managerial criticism over the layout of stage plans, feasibility reports and requirements documents. Due to 1A's diligence and willingness to help those who were not so proficient in terms of written English these initial problems were overcome.

Due to the fact that this team did not suffer from any disruptions or lack of debate they did not contribute any data points to the levels of internal or external disruption. Analysis of the marks awarded to the students carrying out this project showed that this team performed better than any of the other Maxi teams studied in either 2003-2004 or 2004-2005. While these marks can not be regarded as scientifically precise measurements of achievement, even though the marking process tries to get as near to this as possible, the result that a team without any kind of disruption or lack of debate was also the highest performing one is certainly interesting.

5.2 Team 2

Table 5. MBTI Type Team 2

ID	Type	E-I %	S-N%	T-F%	J-P %
2A	INTP	I 78	N 78	T 44	P 33
2B	INTJ	I 11	N 11	T 78	J 44
2C	INTJ	I 56	N 33	T 33	J 44
2D	INTJ	I 11	N 33	T 33	J 56
2E	ENTJ	E 1	N 33	T 11	J 11

This team did not suffer from any problems relating to lack of debate, as Figure 3 illustrates not a single issue came under the ordinal rating 1. On the minus side the large number of issues rated 4 and one rated 6 shows that there was a lot of disruption.

Many of the disruptions occurring in this team were caused by witty responses or pedantry. As shown in Table 5 2A is an INTP (the only team member with a preference for P over J) with very clear preferences for functions and attitudes. 2A was very analytical and often caused annoyance by correcting others in a sharp manner. Pedantry was also a problem as 2A often displayed open-endedness when discussing solutions due to the preference for perceiving coupled with the NT characteristics, so we are left with a character that has an open-mind, but is also driven by a need to be precise and competent. This served the team well on several occasions: when writing documentation, as 2A insisted that the team addresses every point mentioned by the manager; with stage plans.

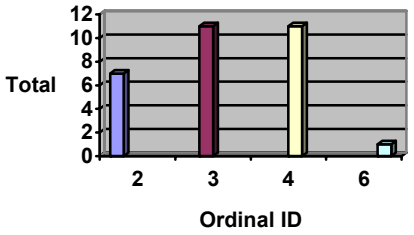

Figure 3. Level of Disruption for all issues Team 2.

2A was also careful to warn the team about overkill, putting in too much information with regards to design documents, especially relating to people and procedures, where he insisted on being precise and to the point, and only mentioning something if necessary. This contradicted the views of other team members and the manager, and indeed earlier views expressed by 2A himself, for 2A it was extremely important that ideas and facts were expressed correctly and succinctly. 2A was typical of an INTP in the sense that they spend considerable time second-guessing themselves; this could be confusing for team-mates. This was not done with any malice or intent to offend as 2A was a generally laid back character and came across as having a good relationship with team-mates, none the less disruptions were caused and discussion was sometimes killed. Another team member 2D admitted that "2A was good with ideas, and proved to be very useful to the team". 2C also stated that "at times the team seemed to be like a one man band", and said 2A's contribution was "crucial".

Figures 4 and 5 show that disruptions had a serious detrimental impact on the performance of this team. This was the case for internal and external impact.

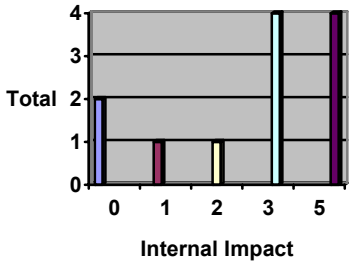

Figure 4. Internal Impact of Issues Team 2.

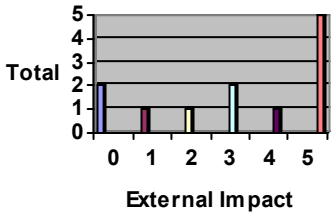

Figure 5. External Impact of Issues Team 2.

2A was the dominant member, and concerns about his heavy workload were justified. These concerns came from 2A himself, from other members of the team and from the manager, who felt that 2A was taking on an unreasonable amount of work, the pressure of this heavy workload took its toll on 2A's health, as the project went on 2A started to complain about feeling weak and tired and was susceptible to viruses, on a couple of occasions 2A was absent due to a heavy cold. Another telling factor was that 2A was the only native English speaker in the team, so perhaps other members were not so confident when informing or coming out with ideas, although they were happy to discuss points once they had been raised.

Because 2A was such a dominant member of this team the remaining members can only be mentioned in passing. All of these members were NTJ's, with three INTJ's (2B, 2C, and 2D) and one ENTJ (2E). 2C in particular had a very clear preference for introversion and seemed very reserved in meetings. This team provided a classic example of what can happen when there is too much reliance on one member; this over reliance came about primarily due to a failure to organize work properly. The over reliance on 2A was due in part to the personality type of the other members of the team, as the three INTJs were naturally cautious when coming forward with new ideas and leading the discussion, add to this the fact that English was not the first language for any of these students and one can have a clearer picture of how the team came to be so reliant on one member, who according to his personality profile was ill suited to play such a role.

2D engaged in several useful debates with 2A and apart from 2A was the most active team member, 2D enjoyed asking questions and was keen to gain an understanding of different areas of the project, this was apparent during discussions with 2A. 2D also admitted to suffering from depression due to the project and other factors and also expressed deep dissatisfaction at the end of the project. 2D was also at the forefront of mentioning concerns relating to work that needed to be carried out and whether the team was ready to undertake certain tasks.

Later on in the project 2C mentioned that the team were putting things off, that they should have done the database before the Easter break, 2B also mentioned that they should have had a meaningful discussion with the demonstrator before the Easter break. Despite being clearly aware of impending problems the team did not act. This inaction did not pass without consequence, the team were late handing in the Stage 1 statistics, they had to be posted to the manager, this resulted in loss of marks, they were well over budget at the end of Stage 2, 2A attempted to shift the blame to the manager for the team being over budget, there was a bug in the final system that prevented users from entering a new artifact. 2A's admitted that one person was responsible for the lion's share of the code and documentation, the login wasn't working, this was directly related to earlier concerns from team mates about being uncomfortable with certain tasks, there were also obscene messages in the system when it was demonstrated, this was wholly unprofessional, and the final documentation was incomplete. The end result of all of this was that there was no working system at the end of the project, this was made worse by the fact that important documentation was incomplete (i.e. requirements and testing), an absolute worse case scenario for a Maxi team.

Whilst having many issues that rated 4 on the ordinal scale, this team did not have any serious screaming matches or confrontations, the disruptions were more related to sardonic responses and idiosyncratic drollery. Nor was lack of debate a problem for this team, they were often aware of problems yet instead of acting on them preferred to leave things to 2A. This tactic could only work on a short-term basis, relying on it over the entire project proved to be a very costly error.

Misunderstandings led to over reliance on 2A which led to poor quality or missed deadlines. Lack of debate was not a problem for this team; they were often aware of problems yet instead of acting on them preferred to leave things to 2A. This tactic could only work on a short-term basis, relying on it over the entire project proved to be a very costly error. The fact that both the documentation and the system were incomplete at the end of the project speaks for itself.

5.3 Team 3

Table 6. MBTI Type Team 3

ID	Type	E-I %	S-N %	T-F %	J-P %
3A	ENFJ	E 67	N 11	F 11	J 1
3B	INTP	I 56	N 44	T 44	P 22
3C	INTJ	I 11	N 33	T 22	J 33
3D	ENTJ	E 33	N 33	T 1	J 22
3E	INTJ	I 78	N 56	T 11	J 11

Observing this team throughout the project was very interesting, for the first half of the project they were dominated by two individuals (3D and 3E), and for the second half dominated by a different one (3B). The other two members had minimal influence. Four different nationalities were represented within the team, without any British students. Another aspect of this team was that for a while they were fractured along both ethnic lines (two members of the same nationality appeared at times to be working as a separate splinter group) and gender lines (the two female members, 3A and 3C were largely frozen out of the discussion or were unwilling to put points forward, as a result they had little impact on the project). Figure 6 shows the level of disruption for all issues involving this team.

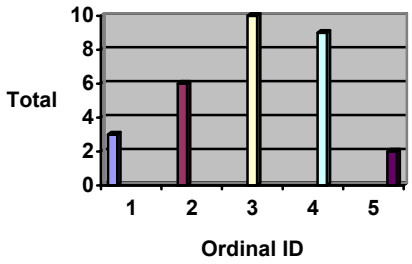

Figure 6. Level of Disruption for all issues Team 3.

The THF problems had a great effect on the team, as they persisted for the entire project. As mentioned earlier there was an ethnic division in this team, so that at times 3D and 3E were almost working as a separate team. Both of these characters were of the same nationality, but different from any of the rest of the team. 3E remarked in meetings that together with 3D they could form a sub-group. 3D and 3E both complained on several occasions about having to do most of the work. 3D said that there was a danger that the team could be seen to be dominated by 3D and 3E. 3E admitted that they were running the show. 3D argued that one or two members were being overloaded with work and putting in a lot more hours, according to 3D only two people were working on project reports and documentation. 3E added that the DB and class designs were also done by the same people and that only two people were working hard in the team.

3B did not agree entirely with the argument that 3D and 3E were doing most of the work and others were deliberately avoiding work, 3B claimed that there were many factors influencing why the team was being dominated by two people, the main reason was that there was a team within a team including having conversations in a foreign language. This was also apparent by 3D's admission that work was going on with 3E on 3D's own personal machine at home. 3E said that everyone seemed to be working on their own machine. 3C also said that 3D and 3E should try to involve others more, 3D responded by stating that the others should show more initiative.

3B admitted that 3D and 3E had worked very hard during the first half of the project, to which 3E responded that there were serious communication problems and problems relating to submitting things at the last minute and not having the chance to check documents prior to submission.

3D disliked seeing mistakes repeated, and had no patience with inefficiency. 3D was not naturally tuned in to other people's feelings, hence the harsh complaints about others not pulling their weight, and leaving work to 3D and 3E. At times 3D could be harsh, and often had difficulty seeing things from outside of a personal perspective and could often lack patience with those who saw things from a different perspective. 3D needed to work on recognizing the value of other opinions, and to take other people's feelings into account. This could be one of the main reasons why the two female members of the team were reluctant to express opinions. 3D could have appeared to have been an overbearing individual with a forceful, abrasive nature.

3E as an INTJ was drawn to 3D primarily because of shared language and culture, but personality also played a role. INTJ's value intelligence and competence, and 3D had the presence of an intelligent and confident person. Introverted intuition dominated 3E's personality; a lot of energy was focused on generating ideas and possibilities. In the presence of 3D, 3E was encouraged to share ideas and ask questions, particularly as many of the early meetings resembled a private conversation between these two individuals and sometimes involved flitting between different languages. 3E was happy to let 3D take the lead role, due to 3D's greater communication skills, yet was effective at objectively looking at specific situations and often realized that the team were heading towards problems, particularly due to communication barriers. 3E was also realistic with the realization that unless the individuals started to work as a team and end the practice of having splinter groups within the team, there was a good chance that they would sink together. Despite this 3E seemed to lack interest in other people's feelings when discussing human issues, with the exception of 3D. 3E came across as being an aloof and reserved character to team mates, and rarely expressed positive opinions (or any opinion) about other peoples' work or plans. This wasn't down to rigidity or ignorance, people with such a clear preference for INT often have difficulty expressing themselves, particularly when discussing emotional or human centered points. In the absence of adequate communication abilities, 3E ran the danger of appearing blunt to team mates.

3B felt compelled to act in light of complaints from 3D/3E. 3D and 3E made it clear that they were not interested in socializing with the rest of the team, and said they should be too busy working to worry about going out. What happened during the second half of the project was very interesting, because it was almost as if the first half of the project had been turned on its head. Personality played a role in this. 3B was a similar character to 2A in the 04-05 Maxi Team 2 both are INTPs, and both ended up carrying a very heavy workload, the difference being that 2A from Team 2 was the driving force for the entire project and made more of an attempt to involve others, whereas 2B in Team 3 was the driving force during the second half of the project and worked almost totally in isolation. By a strange sequence of events 3B became the dominant member and spokesman of the team. This was remarkable given that 3B played a strictly supporting role in early meetings, and was happy to let 3D and 3E drive the team. This changed when 3D and 3E started complaining about the lack of effort and initiative from the others. 3B certainly responded to this criticism and carried out a lot of research relating to the project, such as reading extensively on user interface design, and SQL. 3B said three members of the team had baled out for the Easter holidays and almost single handedly designed the interface, managed to do most of the DB connectivity work, wrote most of the SQL queries for the DB, tested the interface and the DB, and worked hard on the user manual. This hard work came at a large personal cost to 3B in terms of health and lifestyle. At one point 3B complained about sleep deprivation, and mentioned only having 4hrs sleep over a period of 3 days, 3B was also dogged by burn-out during the last few weeks and admitted to suffering from burn-out in one meeting.

Another side issue was that 3A had managed to insult 3B by using foul and abusive language in response to a legitimate question about the design document. 3A tried to claim that it was done in humor but could barely conceal the simmering anger. 3B said such insults and language were bad for team morale; this followed a threat by 3D to walk out due to dissatisfaction with the way things were progressing. 3E also remarked that they were running over budget.

The other two members 3C and 3A were largely in the background for the duration of the project. They did seek solace in each other and kept out of internal team politics apart from on the occasion when 3A insulted 3B with foul language, this soured relationships between 3B and 3A and ensured that 3B would not seek collaboration with 3A and her ally 3C later on in the project. 3A had a preference for extraversion over introversion; it would have been interesting to have seen their interactions in an all female team, but here with 3A and 3C being the only females, they were largely frozen out of the discussions. When looking for reasons as to why they were frozen out one can state that personality and language played a role in the case of 3C, in that as an INTJ she did not start many discussions but the other team members made little effort to bring her into discussions once they had started. Due to the nature of the team and the combination of 3D and 3E and the solo contributions of 3B, the remaining two members found themselves left on the periphery. 3A did not help matters by alienating 3B after a sharp exchange of insults; 3A wanted to get involved more but was unable to join the sub-group of 3D and 3E and was also unable to create a harmonious relationship with the other key member 3B.

There was another problem facing this team in addition to the fascinating interactions and politics going on. This was the issue of having more than one client. 3A remarked that the initial client they were seeing was not the final user, yet they had to meet this person in order to gain access to the actual client. 3A said this could be a problem as the final client may have very different requirements. 3D also noted that the initial client was changing requirements and messing the team around, 3D's impatience and anger towards the client increased as the project went on. 3D felt the client was wasting the teams' time with pointless meetings and by changing the requirements. 3D added that the client had created a separate list of non-functional requirements which went against what the project manager suggested. 3B later said that "having two clients is a nightmare; they make changes and are not synchronized". 3B was forced to concede late on in the project that any new changes could not be accommodated and the system would have to be frozen.

6. AGGREGATION OF RESULTS

Figures 7 and 8 present the overall results from all five Genesys and Maxi teams from 2003-2004 for the relationship between level of disruption and impact as given in [9] and [10]. The data points in the graphs represent individual issues, and in principle should all have integer co-ordinates, but in practice their positions have been perturbed to enable the numbers of points at each location to be visualized.

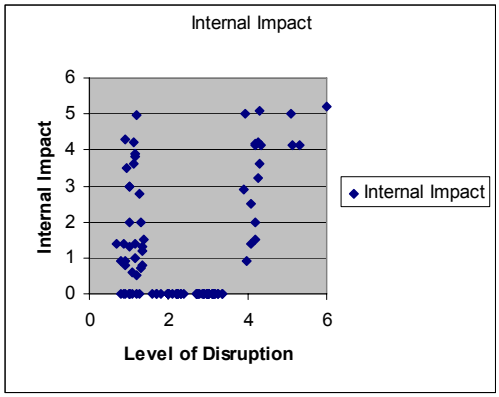

Figure 7. Internal Impact versus level of Disruption 2003-2004.

Figures 9 and 10 present the overall results from the Maxi teams from 2004-2005 for the relationship between level of disruption and impact, but using smaller perturbations than figures 7 and 8. For these teams it was issues that were actually classified as disruptions (i.e. rated 4, 5 or 6 on the level of disruption ordinal scale) that tended to have a serious internal and external impact.

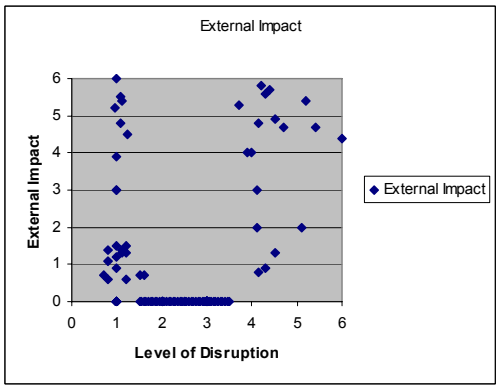

Figure 8. External Impact versus level of Disruption 2003-2004.

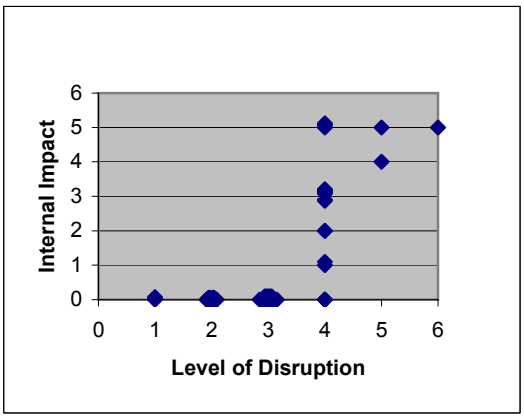

Figure 9. Internal Impact versus level of Disruption 2004-2005.

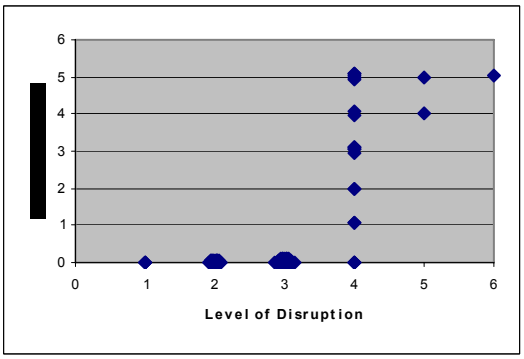

Figure 10. External Impact versus level of Disruption 2004-2005.

7. DISCUSSION

The results gathered during the 2004-2005 research period add another dimension when compared with the previous year's results. Some problems were uncovered during 2004-2005 which confirmed previous findings, typically they related to the idiosyncrasies of the individual personality types of the team members causing problems. Examples of this include 3D's attitude towards team-mates. 3D interacted primarily with 3E and had little patience for mistakes. This contributed to the alienation of 3A and 3C.

Another interesting factor which came to the fore during 2004-2005 was that of individual personality types being forced to play unfamiliar roles. This was apparent when analyzing the relationship between 2A and the rest of his team, although 2A was clearly the dominant member his personality type did not make him naturally predisposed to play such a role. It was noted that other team members had to play a more active role and not allow 2A to do most of the talking: having such a clear preference for introversion meant that 2A was not a natural spokesman, and wasn't totally confident when discussing with his team-mates, clients and managers, since INTP's do not like to lead or control people as it generally goes against their nature. This wasn't so much a clash of personalities; it was more a case of a person of a particular personality being forced to play an unfamiliar role. This phenomenon was also witnessed in Team 3, when 3B came to fore in the second half of the project and failed to consult other team members about important work. 3B started to interpret project issues in a very subjective way, such as diving right into the user interface design on the back of personal research without discussing with others. In many ways the behavior of 3B was typical of that of an INTP with clear preferences acting under pressure, or in the face of what they perceive to be an injustice.

A positive result was gained in relation to Team 1; this team did not suffer from any disruptions or lack of debate and were in fact the only team observed in two years to have such an impeccable record. What this suggests is that a team without disruptions who are able to debate effectively will find it easier to synchronize their combined efforts and ensure that whatever work is produced is consistent with the work of other team members, this is especially vital when different team members are producing separate parts of the same piece of documentation.

By contrast, as was seen in the descriptions of the other teams these disruptive issues had an adverse impact on both internal

interpersonal communication and on the actual work produced. One team (team 3) narrowly avoided complete disaster due to the efforts of one member in the final weeks; another team was not so lucky and failed to produce the required work by the end of the project time-scale.

Previous work has raised some interesting questions for those interested in the nature of the personality make up of a team and the question of homogeneity versus heterogeneity in psychological type. The majority view is that a mix of personality types is required, as described by Teague [13], Capretz [3], Kaiser and Bostrom [1], but by contrast Rutherford [12] suggests that a heterogeneous science/engineering mix is better. Overall, our results gathered over the two years support the widely accepted view that it takes a variety of skills and personalities to solve the myriad of problems related to SE. In other words previous research has tended to support the idea that better software will result from the combined efforts of a variety of mental processes, outlooks and values, and the results described here support this view.

8. CONCLUSIONS

The results of this research reinforce the notion that the personality profile of an SE team is an important factor to consider and that there exists a whole plethora of potential problems within this domain. A positive result was that a team who were very heterogeneous in terms of personality and ethnicity managed to work so well together throughout the academic year. This demonstrated that teams can work well as an effective unit despite significant ethnic, religious and personality differences between individual members. This balances some results of the previous year's research, which found some serious communication problems in ethnically mixed teams.

Another positive result was that individuals were prepared to carry out roles not usually associated with people of their personality type. The sheer tenacity and single mindedness shown in these cases was admirable. The problem of one member attempting to carry a team for an entire project proved to be debilitating, the result was a team failure, which is another important result as it highlights the dangers of over-reliance on one person within a team and how this situation can arise based on the personality and ethnic mix within a given team.

During the previous year the problem of no debate and keeping discussion focused on relevant areas were the main problems which dogged the Maxi teams. During this year the main problem was instances of personality clashes, which contributed to stifling debate and members working alone for long periods. Another clash had both personality and cultural undertones as team members seemed to be unwilling or unable to make significant contributions to the debates going on within the team.

Overall the results from 2004-2005 add substance to existing results whilst offering new insights. The strong performance of a team diverse in personality and ethnicity, the almost total over-reliance on one member within another team, and the relationship of a sub-group who were initially dominant but were eventually superseded by another member were all new phenomena that were uncovered and are welcome additions to the existing knowledge base dealing with the interactions of personalities in SE teams.

The other important conclusion to be drawn from these results, and particularly from the comparison between successful and unsuccessful teams, is that it begins to suggest a hypothesis for how disruptions or lack of debate may be causing the adverse effects on the quality of work that have been observed. This hypothesis is that successful completion of a project does not merely require each member of a team contribute to the work, but also that these individual contributions fit together properly. Discussion of the work is therefore essential to ensure that each member of a team understands how their contribution must fit with the others, and anything that obstructs or distracts from this discussion – whether lack of debate or disruption to the debate – will reduce the extent of mutual understanding, and hence mean that individual contributions will not fit together as well as they should.

Future work should be aimed at exploring this hypothesis, and could involve looking specifically at the effects of factors such as ethnicity, gender and how the teams resolve conflict. In doing this it may be possible to combine qualitative observations with a quantitative approach, and to compare the observed levels with actual quantitative measures such as the quality of the delivered product. This could also be related to the productivity of the team and measure things such as the number of XP story-cards they produce.

9. ACKNOWLEDGMENTS

We would like to thank the following colleagues for their help and assistance throughout the project: Chris Thomson, Marian Gheorghe, Mike Holcombe and Stan Price for allowing us to observe the teams in manager meetings; and for allowing and encouraging the students to take the online test which enabled us to work with them. This research would not have been possible without their help and cooperation so we are grateful to them for their past and continued support and interest.

10. REFERENCES

[1] Bostrom, R.P and Kaiser, K.M, Personality Differences within Systems Project Teams: Implications for Designing Solving Centers *Proceedings of the 18th Annual Computer Personnel Research Conference*, 1981, 248-285.

[2] Costa, P. and McCrae, R. Four ways, five factors are basic. *Personality and Individual Differences*, ***13***. 653-665.

[3] Fernando-Capretz, L. Personality Types in Software Engineering. *International Journal of Human-Computer Studies*, ***58***. 207-214.

[4] Furnham, A. The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, ***21***. 303-307.

[5] HumanMetrics. http://www.humanmetrics.com/cgi-win/JTypes2.asp.

[6] Jung, C.G. Psychological Types *The Collected Works of C.G Jung*, Harcourt Press, 1923.

[7] Karn, J.S. and Cowling, A.J. An Initial Observational Study of the Effects of Personality Type on Software Engineering Teams. *Proceedings of the 8th International Conference on Empirical Assessment in Software Engineering (EASE 2004)*. 155-165

[8] Karn, J.S. and Cowling, A.J. An initial study of the effect of personality on group projects in software engineering. *Department of Computer Science Research Report CS-04-01, University of Sheffield*.

[9] Karn, J.S. and Cowling, A.J. A Study into the Effect of Disruptions on the Performance of Software Engineering Teams. *Proceedings of the 4th International Symposium on Empirical Software Engineering, (ISESE 2005)*. 417-427.

[10] Karn, J.S. and Cowling, A.J. A study into the effect of disruptions on the performance of software engineering teams. *Department of Computer Science Research Report CS-04-17, University of Sheffield*.

[11] Myers, I.B. and Myers, P.B. Gift's Differing: Understanding Personality Type, Davis Black Publishing, California, 1987.

[12] Rutherford, R.H. Using Personality Inventories to Help Form Teams in Software Engineering. *Proceedings of the 6th annual conference on Innovation and technology in computer science education*. 73-76.

[13] Teague, J. Personality Type, Career Preference and Implications for Computer Science Recruitment and Teaching *The Proceedings of the third Australasian Conference on Computer Science Education (ACSE 98)*, 1998, 155-163.