A Behavior Marker tool for measurement of the Non-Technical Skills of Software Professionals: An Empirical Investigation

Lisa L. Lacher[1], Gursimran S. Walia[2], Fabian Fagerholm[3], Max Pagels[4], Kendall Nygard[5], Jürgen Münch[6]

Department of Computer Science

University of Houston-Clear Lake[1]; North Dakota State University[2, 5]; University of Helsinki[3, 4, 6]

Lacher@uhcl.edu[1]; {gursimran.walia[2], Kendall.Nygard[5]}@ndsu.edu; {fabian.fagerholm[3], max.pagels[4], juergen.muench[6]}@cs.helsinki.fi

Abstract— **Managers recognize that software development project teams need to be developed and guided. Although technical skills are necessary, non-technical (NT) skills are equally, if not more, necessary for project success. Currently, there are no proven tools to measure the NT skills of software developers or software development teams. Behavioral markers (observable behaviors that have positive or negative impacts on individual or team performance) are beginning to be successfully used by airline and medical industries to measure NT skill performance. The purpose of this research is to develop and validate the behavior marker system tool that can be used by different managers or coaches to measure the NT skills of software development individuals and teams. This paper presents an empirical study conducted at the Software Factory where users of the behavior marker tool rated video clips of software development teams. The initial results show that the behavior marker tool can be reliably used with minimal training.**

Keywords-Non-technical Skills; behavior marker; performance.

I. INTRODUCTION

Most software is developed by teams and the success of a software project depends on the effective performance of the software project team. The PMI and the most recent PMBOK Guide [1] acknowledges that, non-technical (NT) skills in comparison to the technical skills are equally important for project success and team development. Several authors agree that the NT skills are critical to project success [2, 3]; and there are even some that assert that NT skills can have the largest impact on software development [4, 5].

The growing need for an agile workforce is one major factor that is driving the demand for NT skills [6]. Agile Manifesto's [7] first principle - "*individuals and interactions over processes and tools*" clearly points to the importance of NT skills. Agile teams depend greatly on NT skills such as efficient communication, taking responsibility, initiative, time management, and leadership.

While it is obvious that NT skills are important, and that the performance of individuals is very important to creating an effective team, there are no established guidelines for measuring team effectiveness. Different criteria for assessing team effectiveness have been identified by different authors [8, 9]. Generally, these criteria include measurements of task performance as well as the interpersonal skills of the team members. The interpersonal skills include attitudes and behaviors. Although there is extensive literature with respect to different ways to measure task performance for software development (e.g., lines of code) [10], scant research has been performed on the measurement of NT skills, especially for software developers. A couple of notable exceptions can be found in the aviation and health care industries. Both industries have already recognized the importance of NT skills to the success of their teams, and have been using behavioral marker (BM) systems (e.g., LOSA, ANTS) to structure individual and team assessments of these NT skills. We believe that software teams can also draw upon these BM's from the aviation and health care industries. It is often Software Development managers and coaches that are responsible for assessing the performance of their development teams – not HR departments, thus a tool like a BM system needs to be available to them.

As educators and software project development managers, we are concerned with questions such as: how can managers objectively measure the NT skills of their employees to determine if their NT skills need improvement or how would feedback be provided to the team members so that they could improve their performance? This research attempts to begin answering these kinds of questions.

II. BACKGROUND –NT SKILLS, BEHAVIOR MARKERS

Non Technical (NT) Skills: NT skills are the cognitive, personal resource, and social skills that complement a person's technical skills and contribute to overall task performance [11]. Some classic examples of NT skills include communication, cooperation, decision making, leadership, stress management, and workload management. Basically; NT skills cover the cognitive and social sides of a person. In the most recent survey released by the Association of American Colleges and Universities [12], it was found that employers feel that NT skills are more important than a particular major. Several different surveys of U.S. employers have also identified a lack of NT skills as the area where young job-seekers have the largest deficiency [13]**.** Even professional organizations such as Professional Engineering Competence (UKSPEC), IEEE Computer Society state that professionals have an obligation to possess NT skills [14].

Behavior Markers (BM): Behavioral markers (BM) are defined [15] as "*observable, non-technical behaviors that contribute to superior or substandard performance within a work environment*". They are derived by analyzing data regarding performance that contributes to successful and unsuccessful outcomes. The overall purpose of a BM system is to use markers as a method to assess both team and individual behaviors. These BM systems provide an observation-based

method to capture and assess individual and team performance on data rather than on gut feelings. The BM tool is designed in the form of a structured list of behaviors. The Observers then use this form during a selected work situation to rate performance. This allows an individual's or team's skills to be rated in their real context. BM systems can provide a common language for giving feedback as well as discussing and teaching NT skills.

Behavior Marker (BM) Systems: BM systems have demonstrated value for assessing and providing feedback on these NT skills, for improving training programs, and in the use of building databases to identify norms and prioritize training needs. It is important to recognize that BM systems need to be specific to the domain and culture. A brief description of successful BM systems (airline, medicine) follows:

The first BM system, Line Operation Safety Audit (LOSA) is a very successful BM system that focuses on interpersonal communication, leadership, and decision making in the cockpit. Trained observers ride along in the cockpit and observe the flight crews during normal flight operations. They score the behaviors of the crew using the LOSA tool. LOSA has been endorsed by the International Civil Aviation Organiztion because it has been used so successful in measuring the strengths and weaknesses of flight crews' interpersonal skills [16]. The Anesthetists' NT Skills (ANTS) [17] used in healthcare has proven very useful in assessing the NT skills of anesthetists in simulation training and has provided important performance feedback for the individuals. Another successful healthcare BM system is the Observational Teamwork Assessment of Surgery (OTAS). Many studies have shown that poor communication, coordination, and other aspects of teamwork, rather than technical failures, have been the primary causes of adverse events in surgery. OTAS has been found to be a valid measure of the NT performance of surgical teams [18].

Our goal is to develop and validate a BM system that can improve software professional team member performance by providing feedback in the form of an objective and documented assessment of the NT skills of the team members. We wanted to create a tool that is very usable by practitioners: it requires little or no training to use and does not require unreasonable effort to use. It is a concern of the researcher that if the tool took a lot of training or was too difficult to use, that the potential practitioners, such as project managers and team leads for whom the tool was meant to assist, would not find the tool useful because of the amount of effort required.

III. BEHAVIOR MARKER SYSTEM DEVELOPMENT

The development process for our behavioral marker system for software developers is detailed in our previous work [19]. As a ***first*** step, we performed a systematic literature review to develop NT skill inventory. The high-level question addressed by the review was: *"What are the NT skills required of software professionals performing well in their field and how can we discover what NT skills are valued by employers?"*

Details on the review protocol (sources searched, search execution, inclusion and exclusion criteria, quality assessment,

Fig. 1: Desired NT skills of Software Professionals

data extraction) can be referred to in a report [20]. The output of this step was an initial list of 35 NT skills that were clustered into four major categories: communication, interpersonal, problem solving, and work ethic (see Fig. 1). The detailed desription of each skill can be referred [20].

During the ***second*** step, the initial list of NT skills had their quality assessed and were validated by focus group of experts in industry and academia. Two surveys (and focus groups) were conducted online (using a cross sectional design) to gather NT skill priorities, missing NT skills, description clarifications, and examples of examples of good and poor behaviors for the top rated NT skills of software developers. So that we could prioritize our efforts, focus group ranked the importance of each NT skill to software professionals during the first survey. After the survey analysis, we had a reduced list of 16 skills to focus on. During the second focus group survey, we gather a total of 408 examples of observable actions that indicated good performance and behavior of each NT skill as well as examples of observable actions that indicate poor performance and behavior of each NT skill. These examples were reviewed, clarified, and redundancies were eliminated. The final set of NT skills consisted of: *teamwork*, *initiative/motivation to work*, *listening*, *attitude*, *critical thinking*, *oral communication*, *problem solving*, *attention to detail*, *flexibility*, *integrity/honesty/ethics*, *time management*, and *questioning*. Some behavioral examples, such as "*being a good team player*" and "*body language and persona emitting that you do not enjoy your work*", were too ambiguous and removed. It was also felt that the "*Leadership*" skill did not have enough observable

Fig. 2: Example of 'Listening' behaviors (good and bad examples)

behaviors that would be able to be clearly identified, so that NT skill was removed. The result of the second survey was a behavior-based software engineer NT skills taxonomy. Fig. 2 shows the resultant examples of good and poor behavior for the "*Listening*" skill. The same process was used to create examples of good and poor behavior for each NT skill.

During the ***third*** step, the behavior marker systems being used in aviation, health care, rail transport and maritime transport were examined. Each system's structure was examined to select which elements would have the most potential for use in software development and our final tool was a composition of several systems. The NT skills validated by the focus group along with the good and bad behavior examples for those skills were structured into a BM audit tool for software development. For reference, we refer to the BM audit tool as the Non-Technical Skill Assessment for Software Developers (NTSA).

The NTSA is designed to be used by an observer (i.e. manager, team leader, coach) during routine team interactions or meetings. It is intended that each time a behavior is observed, a mark is placed in the appropriate column by placing a tick mark in that column: observed and good, or expected but not observed. Observations can be clarified by placing explanations in the comments section. The observer can see skill definitions and examples of good and poor behavior for a particular behavioral marker by viewing the second page. A manager is allowed to list as many or as few skills as desired in the behavioral marker column. The observer will score the behaviors based on how well the behavior meets the behavioral examples and its definition.

IV. Empirical Validation of Behavior Marker

In order to evaluate our BM tool, an empirical study rated video clips of student software development teams that were working on industrial strength projects within the Software Factory (as shown in Fig. 3 and explained).

1) Software Factory Background

The Software Factory is a software development laboratory created by the University of Helsinki, Department of Computer Science. All research was performed in Finland due to the requirements of international privacy laws. The University of Helsinki is consistently ranked in the top 100 out of world's 15,000 universities, in part because the university promotes science and research together with European's top research-intensive universities. The master's degree programs are taught in English in order to support the large number of international students who study at the university. The Software Factory's primary participants are students, but the businesses provide team members who work with the students, and university faculties oversee the projects, although the faculty involvement is kept to a minimum. Almost all project communication is in English. Faculty involvement consists primarily of project orientation and project intervention if problems cannot be resolved by the students, coach, and customer. The coach is generally an upper level student with Software Factory project experience. University students take on the role of the development team for projects provided by businesses. The customer has company representatives that take on the role of the product owner and represents the interests of the company. Although these representatives are not co-located, they do come by the Software Factory for weekly demos, sometimes for meetings, and are generally available via telephone and email. Researchers are able to observe what happens in the project due to the seven cameras that provide multiple angles of view and four microphones that record activities in the Factory room. In Software Factory projects, the participants take on the core roles of a typical Scrum project. Projects at the Software Factory last for seven to eight weeks; the students work approximately 6 hours per day, 4-5 days per week.

2) Study Design

This study investigates whether the BM system can be used with consistency by different raters to capture a measurement of the NT skills of software developers, thus facilitating objective feedback to software development teams and individuals. This study used a blocked subject-project study. This type of analysis allows the examination of several factors within the framework of one study. Each of the non-technical skills to be studied can be applied to a set of projects by several subjects and each subject applies each of the non-technical skills under study. In this study, raters evaluated the NT skills of project teams using the NTSA tool. The project teams worked together using state-of-the-art tools, modern processes and best practices to prototype and develop software for real business customers in an environment that emulates industry. Video tapes of the projects were evaluated to rate the student team's NT skill performance. The details of the study are provided as follows.

Independent and dependent variables: The experiment manipulated the following independent variable:

a) Behavioral Marker System tool and Example Behaviors: Each non-technical skill has its own set of good and

Fig. 3: Software Factory

poor behavioral examples that are used by the raters to evaluate team performance of each non-technical skill.

The following dependent variable was measured:

b) Rater's Evaluations: The behavioral rating for each non-technical skill by each rater. This measure includes the percent positive for each rater for each non-technical skill.

Participating Subjects: The participant subjects (students in the Computer Science master's degree) were software developers from two different projects. There were two different projects that were evaluated. One project had five team members and the other had seven team members. The students worked together to develop a software solution to a project posed by the business customer.

Artifacts: Although the NTSA tool could be used to evaluate the NT skills of both individuals and teams, it was decided to test for team skills first. Because we were primarily interested in how the team member's NT skills manifested when interacting with others, it was decided that the first clips to be evaluated would be of team meetings, and so standup meetings, impromptu team meetings, and customer demos were targeted. After extracting all of these clips, it was determined that we would focus on standup meetings because of the consistency and quantity of footage. Two raters used the NTSA tool to independently rate each clip. The NTSA was in the form of a spreadsheet on a computer.

Experiment Procedure: Study steps as described below:

Step 1 – Project Selection: We decided to focus on two projects. We selected one project that had gone well and one that had not gone well (as the first project) in the expectation of producing diverse scorings.

Step 2– Video Clip Collection: Video and audio recordings of the entirety of each project were collected. The Software Factory deployed 7 video cameras and 4 microphones. The cameras were situated such that one could not actually view what was on the computer monitors or clearly see any of the paper artifacts, although anything written on the white board or displayed on either of the two projectors could be clearly viewed. Video clips were labeled with the type of meeting along with date and start and end times so if the clip because corrupted and needed to be re-created, the researcher would know exactly what day and time to go retrieve the clip. A spreadsheet was used to store this information along with which cameras and microphone were used in the clip.

Step 3 – Test Rater Understanding of the NT Skill and Behavioral Descriptions: During the initial phase of the empirical evaluation, two researchers from the Software Factory reviewed the NTSA tool to make sure they understood the descriptions of the good and poor behaviors. Each researcher has extensive experience with project teams in the Software Factory. Each of the researchers reviewed the behavioral descriptions independently, and added comments. Then we met as a group to discuss potential changes. Following the discussion, some behavioral descriptions were modified, some eliminated and some added. Ultimately, the group reached a consensus on all descriptions. It was also determined that it

was unrealistic to observe the behaviors for Integrity, Honesty, and Ethics, Attention to Detail, and Time Management and that it would be better to look at other documents and devices, such as Kanban metrics, bug reports and customer feedback to observe and rate those non-technical skills.

Step 4 – Test Usability of the Tool: The Software Factory researchers used the initial NTSA tool to evaluate several clips to test usability. First, each researcher reviewed the descriptions of each behavior and the good and poor behavioral examples. Then, each researcher did independent evaluations of the clips, after which we met for discussion of the evaluations. There was consensual agreement that fine gradations in quality were difficult to determine and the researchers agreed that the tool would only include ratings for good and poor behavioral observations. The final NTSA tool is shown in Fig. 4. The raters also noted that it was very difficult to determine how often to place a mark for exhibition of good and poor behavior because the meetings were continuous. Because the raters are not classifying discreet events or statements, it was decided that the raters would be notified when a minute had passed, which would prompt them to decide if the team exhibited any good behaviors or poor behaviors and to put a mark in the appropriate column. If they did not feel that any good or poor behaviors were exhibited by the team, they did not place a check mark. If they felt that both good and poor behaviors were exhibited, they put a check mark in each column. After the evaluation of the last clip and post discussion, there was consensus that the tool was ready for testing.

Step 5 – Actualizing Rater's Evaluations: Each rater individually rated forty five standup meetings over the course of ten weeks. The time spread of the ratings simulates the frequency with which a manager, team lead, or coach would use the tool. We also wanted to eliminate the amount of fatigue that could transpire. The raters used the spreadsheet version of the NTSA behavioral marker system tool with the one minute timer. Unlike the trial evaluations, the raters rated all NT skills while viewing the video clip as opposed to only rating one non-technical skill per viewing.

Non-Technical Skills (NTSA)
Assessment Instrument, Final Version

Date: _____ Observer: _____ Segment: _____

Event (place check mark): Stand up Meeting: ____ Customer Demo: ____ Other (describe): _____

Category	Skill	Good	Poor	Comments
Communication	Listening			
	Oral Communication			
	Questioning			
Interpersonal	Attitude			
	Teamwork			
Problem Solving	Critical Thinking			
	Problem Solving			
Work Ethic	Flexibility			
	Initiative/ Motivation to Work			

* Place one mark in the appropriate column each time behavior is observed.

Fig. 4: NT skills assessment instrument

V. RESEARCH RESULTS

Because we were primarily interested in how the team member's NT skills were manifested when interacting with others, it was decided that standup meetings would be the focus of our analysis. We were able to limit the video footage to view based on the schedule that the development team agreed upon. Generally, the team limited their development efforts to Monday through Friday from eight in the morning to five in the afternoon. Thus, for a typical seven to eight week time period, this means that there were approximately 2,205 to 2,520 hours of video footage per project available, with four different audio choices for each hour.

We evaluated the percentage of positive ratings, and developed a binary data set for statistical analyses. By inspecting the distributions of the raters when examining the skills, a critical value (specific to each NT skill) was chosen to separate the 0 or 1. For example, for the Listening NT skill, a critical value of 0.8 was chosen. This value was chosen because it approximately separated the raw data evenly into two parts. Thus, if the good percentage was greater than or equal to 0.8, the rating was assigned to 1, and the rating was assigned to 0 if the good percentage was less than 0.8. Using this information, a 2X2 table containing the good and bad percentages of two raters was created. Next, a McNemar's test was used to evaluate whether or not there are significant differences between the raters. A value of $p < 0.05$ would tell us that there is a significant difference between the raters and p value greater than 0.05 would signify inter-rater reliability.

As mentioned earlier, an analysis of the quantitative data includes the rater's evaluations for good and poor behaviors observed in the standup meetings. It was decided to follow John Uebersax's [21] recommendation to run McNemar's test of marginal homogeneity and calculate the inter-rater reliability between two individuals. Cohen's kappa could not be used because the sample size was not large enough to be reliable.

To analyze the agreement between the two raters, analyses were performed for each of the nine NT skills: *listening*, *oral communication*, *questioning*, *attitude*, *teamwork*, *c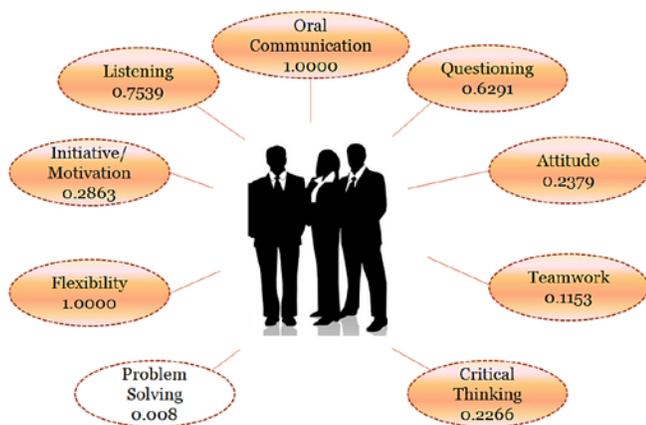ritical thinking*, *problem solving*, *flexibility*, and *initiative and motivation to work*. Figure 2 shows the McNemar test results for each of the NT behaviors evaluated.

To test this study hypothesis, we ran McNemar's on the percentage positive ratings (calculated to produce a binary data set) for each rater and for each NT skill to test for rater agreement in cases where there were enough observation data points. The results showed that, inter-rater reliability of NTSA was found for eight of the nine NT skills in the tool. These results provide initial evidence that NTSA can be a useful tool that could be easily used by managers, team leaders, etc. responsible for the development of these skills, to objectively and consistently measure their employee's NT skills. A tool, such as the NTSA, provides a mechanism to not only improve a team and by extension the software that they produce.

The fundamental finding is that inter-rater reliability of NTSA was found for eight of the nine NT skills in the tool. The "*Problem solving*" NT skill needs further enhancements and subsequent validation before it could be used. In fact, it is possible that "*problem solving*" simply is not observable. The Non-Technical Skills Assessment for Software Developers (NTSA) system can be used reliably by individuals responsible for the NT skills of software development teams, such as educators, managers, team leads, etc. Although the raters did practice rating several video clips with the tool, and this is equivalent to a few meetings, it is also very interesting to note that the raters do not need to be human factors experts, nor did it require extensive initial training for the tool to be used reliably. Although the raters felt that it was very easy to use the tool in its spreadsheet form while working with the form on a computer where the behavioral examples are only a click away, they also noted that they would like to keep the electronic capability if they were rating a live event rather than a video recorded event. The raters also noted that the tool could be customized to only include the NT skills of interest to the rater – not all non-technical skills need to be rated at the same time. This would make the tool even easier to work with. While, these results are encouraging, only two projects and two raters were used. Therefore, more studies need to be performed. A positive aspect of this study is that the raters had different levels of project management experience, and were able to use to tool and get reliable results.

VI. THREAT TO VALIDITY

Although the results of this study are encouraging, there are certain threats to validity that exist. One such threat is that only two projects were evaluated. Like any study, the more a subject is tested, the more empirical studies that are performed, the more one can see if the results are repeatable. Rater agreement testing should continue to be performed on more projects. Another threat is that both projects were rated by the same two judges. More empirical will be performed with different raters using the NTSA tool to ensure the robustness of the tool. One positive aspect about the raters is that each had different levels of software development project management experience. That

Fig. 5 Aggregation of McNemar Test Results

means that the raters do not have to have the same level of experience or backgrounds in order to use the tool and get reliable results. Another potential threat is that both projects were fairly successful, and thus may not have exercised the poor behavior examples enough. Lastly, the projects were performed by student teams and thus many not be generalizable; although this threat was mitigated by the level of professional business-like environment that can be found in the Software Factory and by the fact that both projects were real-world projects.

VII. CONCLUSION AND FUTURE WORK

Our results establish that the NTSA tool can be reliably used with minimal effort. This is valuable knowledge for managers and educators. We recognize that teams need members with the correct technical skill set and knowledge, by using NTSA software development team mangers can identify the areas in which the team's NT skills could use some improvements. Using the same tool on subsequent projects will allow us to determine if there was any improvement in a given skill. Such as tool provides a mechanism with which to improve a team and by extension the software they produce. The NTSA provides a common language with which to understand and communicate about NT skills important to software professionals

In the future, we would plan on repeating this study on other projects. Specifically, we would like to use the tool on more unsuccessful software development project to see if there is a correlation between poor NT skills and an unsuccessful project. This research can be extended to include all of the NT skills deemed important to software developers as identified in the NT skills taxonomy. This would give educators and managers a rich set of NT skills and behaviors that could be evaluated. This tool also needs to be tested on individual software developers within software development teams to see if it can be effectively used to assess the NT skills of the individual as well as the team. This tool should also be tested in industry to verify that it works for professional software developer and teams, as well as student software development teams.

REFERENCES

[1] Project Management Institute. A Guide to the Project Management Body of Knowledge (PMBOK Guide). Newton Square, PA: Project Management Institute, 2008, pp. 215.

[2] S. Acuna, N. Juristo, and A.M. Moreno, "Emphasizing Human Capabilities in Software Development", IEEE Software, vol. 23, 2006, pp. 94-101.

[3] E. Amengual, and A. Mas, "Software Process Improvement through Teamwork Management," in Proceedings of the 8th International Conference on Product-Focused Software Process Improvement, 2007, pp. 108-117.

[4] A. Cockburn, and J. Highsmith, "Agile software development: The people factor", Computer, vol. 34, 2001, pp. 131-133.

[5] N. Gorla, and Y. Wah Lam, "Who Should Work With Whom?" Communications of the ACM, vol. 47 No. 6, pp. 79–82, Jun. 2004.

[6] Abell, Angela, Information World Review; Dec 2002; 186; ABI/INFORM Complete pg. 56.

[7] http://agilemanifesto.org/

[8] S.G. Cohen, and D.E. Bailey, "What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite", Journal of Management, vol. 23, 1997, pp. 239-290.

[9] J.J. Jiang, J. Motwani, and S.T. Margulis, "IS team projects: IS professionals rate six criteria for assessing effectiveness", Team Performance Management, vol. 3, 1997, pp. 236-242.

[10] O. Hazzan and I. Hadar, "Why and how can human-related measures support software development processes?" The Journal of Systems and Software 81, 2008. Pp/ 1248-1252.

[11] R. Flin, P. O'Connor, and M. Crichton., "Safety at the sharp end: A guide to non-technical skills", 2008, Burlington, VT: Ashgate Publishing Company. Pg. 264

[12] Higher Ed News, "Survey Finds Business Executives Aren't Focused on Majors They Hire" accessed Mar. 14, 2014,

[13] http://business.time.com/2013/11/10/the-real-reason-new-college-grads-cant-get-hired/

[14] UKSPEC,"UK-SPEC UK Standard for Professional Engineering Competence," accessed Mar. 14, 2014, www.engc.org.uk/ecukdocuments/internet/document library/UK-SPEC third edition.pdf

[15] B. F. Klampfer, R. L. Helmreich, B. Hausler, B. Sexton, G. Fletcher, P. Field, S. Staender, K. Lauche, P. Dieckmann, and A. Amacher. "Enhancing performance in high risk environments: Recommendations for the use of behavioral markers." Behavioral Markers Workshop, 2001, pp. 10.

[16] B. F. Klampfer, R. L. Helmreich, B. Hausler, B. Sexton, G. Fletcher, P. Field, S. Staender, K. Lauche, P. Dieckmann, and A. Amacher. "Enhancing performance in high risk environments: Recommendations for the use of behavioral markers." Behavioral Markers Workshop, 2001, pp. 10.

[17] G. Fletcher, R. Flin, P. McGeorge, R. Glavin, N. Maran and R. Patey, "Development of a Prototype Behavioural marker System for Anaesthetists' Non-Technical Skills (ANTS)," Workpackage 5 Report, Version 1.1. (2003)

[18] G. Fletcher, R. Flin, P. McGeorge, R. Glavin, N. Maran and R. Patey, "Development of a Prototype Behavioural marker System for Anaesthetists' Non-Technical Skills (ANTS)," Workpackage 5 Report, Version 1.1. (2003)

[19] L.L. Bender, G.S. Walia, F. Fagerholm, M. Pagels, K.E. Nygard, and J. Münch, "Measurement of Non-Technical Skills of Software Professionals: An Empirical Investigation", Proceedings of the 26th IEEE International Conference on Software Engineering and Knowledge Engineering. July 1- 3, SEKE 2014 Vancouver, Canada.

[20] L.L. Bender and G.S. Walia, "Measurement of Non-Technical Skills of Software Development Teams", Department of Computer Science, North Dakota State University, Fargo, ND, Tech. Rep. NDSU-CS-TR-14-001, Mar. 2014.

[21] J. Uebersax. "Statistical Methods for Rater and Diagnostic Agreement" Internet: http://www.john-uebersax.com/stat/agree.htm [Apr. 14, 2013]