Daily Questionnaire to Assess Self-Reported Well-Being During a Software Development Project

Miikka Kuutila
University of Oulu
Oulu, Finland
miikka.kuutila@oulu.fi

Maëlick Claes
University of Oulu
Oulu, Finland
maelick.claes@oulu.fi

Mika V. Mäntylä
University of Oulu
Oulu, Finland
mika.mantyla@oulu.fi

Marko Elovainio
University of Helsinki
Helsinki, Finland
marko.elovainio@helsinki.fi

ABSTRACT

According to authors best knowledge, this workshop paper makes two novel extensions to software engineering research. First, we create and execute a daily questionnaire monitoring the work well-being of software developers through a period of eight months. Second, we utilize statistical methods developed for discovering psychological dynamics to analyze this data. Our questionnaire includes elements from job satisfaction surveys and one software development specific element. The data were collected every day for a period of 8 months in a single software development project producing 526 answers from eight developers. The preliminary analysis shows the strongest correlations between hurry and interruptions. Additionally, we constructed temporal and contemporaneous network models used for discovering psychological dynamics from the questionnaire responses. In the future, we will try to establish links between the survey responses and the measures collected by conducting software repository mining and sentiment analysis.

KEYWORDS

questionnaire, job well-being, software development, empirical study, strain, stress, gaussian graphical model

1 INTRODUCTION

While well-being is used synonymously with happiness in everyday language, in academic literature it mostly refers to psychological well-being. Characteristics that define well-being are emotional conditions, and its phenomenological and long term nature [21].

That is, a person with good well-being is more likely to experience positive emotions and believe to be well themselves.

The job demands-resources model [1] proposes that work related well-being is controlled by work demands and resources. Demands include emotional demands, time and work pressures, while resources can include available autonomy, time, and tools. The model assumes that job strain is caused by an imbalance between job resources and demands. In other words, demands such as work pressure are antecedents to job strain and stress. Additionally, the model assumes that different demand and job resource variables interact in predicting job strain, e.g. increased autonomy can buffer the effects of time pressure [2, 22].

Extensive overtime work (part of job strain) has been widely reported on the software industry [6, 17]. Possible causes for overtime are incorrect effort estimation and scheduling problems, company culture and tight deadlines. Overtime work in turn has been associated with physical and mental distress in the software industry [16].

In this paper, we develop a daily questionnaire to measure self-reported well-being with questions assessing: hurry, interruptions, stress, sleeping problems, ineffective software development and a buffering variable of autonomy/ independence. This questionnaire was administered on a single software development project for eight months. The acquired data is analyzed with correlations and a gaussian graphical model. Our long term goal is to broaden the understanding of job strain of developers and its relation to development activities with sentiment analysis and repository mining. This paper describes the questionnaire and focuses on the initial research questions covering:

RQ1 Are the answers to the different questionnaire questions correlated with each other?

RQ2 Can possible causal mechanisms be created based on the results?

2 RELATED WORK

A longitudinal study by Fujigaki [5] examined the mental health of software developers and found a statistically significant correlation between job events and an increase in depressive symptoms up to a week after the job event. The study did not differentiate between different job events, but instead they included: "time pressure of a deadline, work-overload, amount of work increase, responsibility increase, and trouble with clients".

Experiment investigating interruptions in information work by Mark, Gudith and Klocke [15] found out, that interruptions actually decreased the amount of time needed for the task, but with a price of increased stress, time pressure and effort.

A study by Laanti [12] investigated reported stress before and after transitioning to agile development methodologies. The study found no increase or decrease of stress by the change of development method, but teams who reported being more empowered reported less stress. Tuomivaara et al. [20] investigated job strain in agile and lean development projects. While the evidence gathered suggests that constant deadlines in agile projects levels the workload more evenly across the whole project, more strain was still felt in the end of investigated projects.

In addition to well-being, the effects of hurry have been studied on the productivity of software developers. Mäntylä and Itkonen [14] found increased performance on testing task with time pressure. Conversely, Land et al. [13] do not find increased performance in code reviews, or Topi et al. on database query tasks [19].

Affective states have been shown to alter the productivity of software developers [8]. Effects of unhappiness while developing software include low cognitive performance, low motivation, work withdrawal, low productivity and low code quality [7].

3 METHODOLOGY

3.1 Developing a questionnaire

Our goal is to construct a repeatedly taken questionnaire, to produce longitudinal data to measure the daily level of job well-being. This meant having relatively few items, which can be answered quickly to achieve a high response rate. Measurements of stress relying on only one item have produced valid data [3]. The questionnaire was constructed by picking relevant items on the survey done by Heponiemi et al. [9], to which we added one software engineering specific item. To make sure respondents can answer quickly, we decided to include only one software specific question in the questionnaire. The questionnaire contains six items in total:

- I can make independent decisions in my work
- I am in a hurry and have too little time to finish the task properly
- I feel interrupted while working
- I experience ineffective software development (poor processes, poorly performing tools or poor communication with the development team)
- I feel stressed (refers to a situation in which the respondent feels tense, restless, nervous or anxious)
- I experience sleeping problems (difficulty in falling asleep or waking up several times during the night)

The respondents were asked to rate items with the question: "How frequently has the following condition occurred since the last time you answered this survey?". These items were then ranked in five point Likert-scale. From 1 to 5, the corresponding textual answers are "Very rarely or never", "Rarely", Once in a while", "Often" and "Frequently of continuously". Before starting the data collection, we met with the project personnel to explain the purpose of the study, as well as the voluntary nature of participation.

3.2 Software project context

The software project used as a case study is developed by a medium-sized software company, for a single customer, with weekly meetings and continuous delivery. The project was originally started in 2014. The developed questionnaire was sent to developers of an ongoing software project from April 10th 2017 to January 12th 2018. We used Webropol[1] to send the questionnaire every working day by email at 8am and to collect the responses. Developers who moved from or to another project, or started working in multiple projects at the same time, stopped answering the questionnaire. Individuals with less than ten responses were discarded from the data analysis. For data analysis, a total of 526 responses were received from eight respondents. Taking into account summer holidays, the total response rate is 37,5% (526 / 1404) for eligible respondents.

3.3 Analysis

For correlation analysis and the produced networks, we handled missing values by carrying back observed values. This is because we asked respondents to answer how frequently the condition had occurred since last time the questionnaire was answered. Carrying back the observed value was implemented with package imputeTS[2] and "na.locf" function. Carrying back values was used for correlation analysis and the creation of Gaussian graphical models.

In time series analysis, in order to investigate correlations between series, it is necessary to take the seasonal and trend components into account. After filling the missing values by carrying back, we removed data from weekends and assumed a weekly seasonality. Then we aggregated the data into time series by daily mean, removed the seasonal and trend components from the time series with the *decompose* function [3] in R. Afterwards, we tested for the stationarity of the residuals with Augmented Dickey-Fuller Test (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. ADF test rejected the null hypothesis of non-stationarity and the possession of unit root for every time series residual while KPSS test could not reject the null hypothesis of stationarity for any of the time series residuals. Hence stationarity is assumed.

3.4 Temporal and Contemporaneous Networks

Gaussian graphical models (GGM) have been used as an exploratory tool for modeling networks between variables by using partial correlation, i.e. correlation between two variables after controlling for the effects of all other variables. Similarly to penalized regression, *e.g.* Ridge [10] or Lasso regression [18], penalized models have been proposed for GGMs as well [23]. GGM can be combined with vector-autoregression (VAR) to address time series where temporal independence cannot be assumed, e.g. stress level of today is not independent of yesterday's stress level.

Combining GGM and VAR with Lasso penalty has been proposed for discovering psychological dynamics collected with Experience Sampling Method (ESM), e.g. daily survey, by Epskamp et al. [4] and by reanalyzing the data from previous studies. Epskamp et al. states that psychological processes should be modeled as complex dynamical systems where different psychological and sociological

---

[1]http://w3.webropol.com/start/
[2]https://cran.r-project.org/web/packages/imputeTS/imputeTS.pdf
[3]https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/decompose

Figure 1: 5-day moving average of all the answers of the aggregated responses.
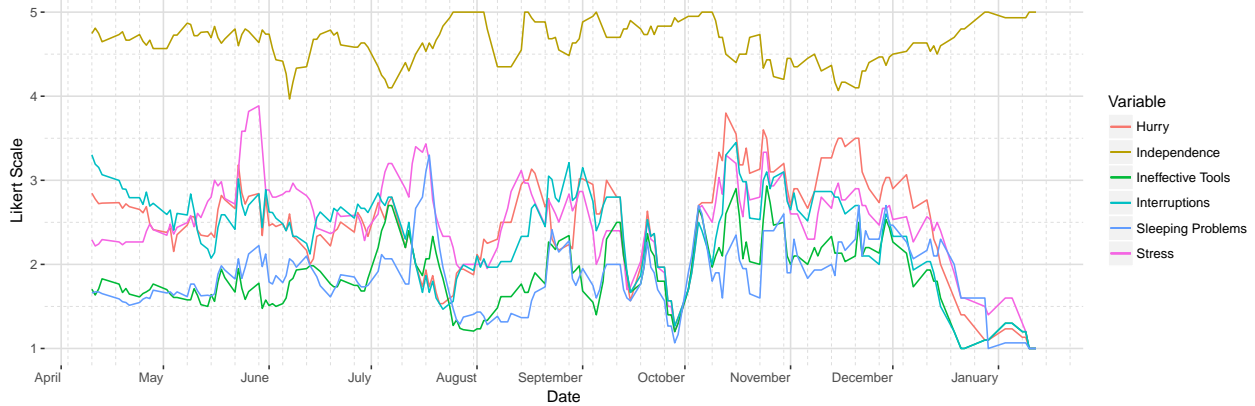

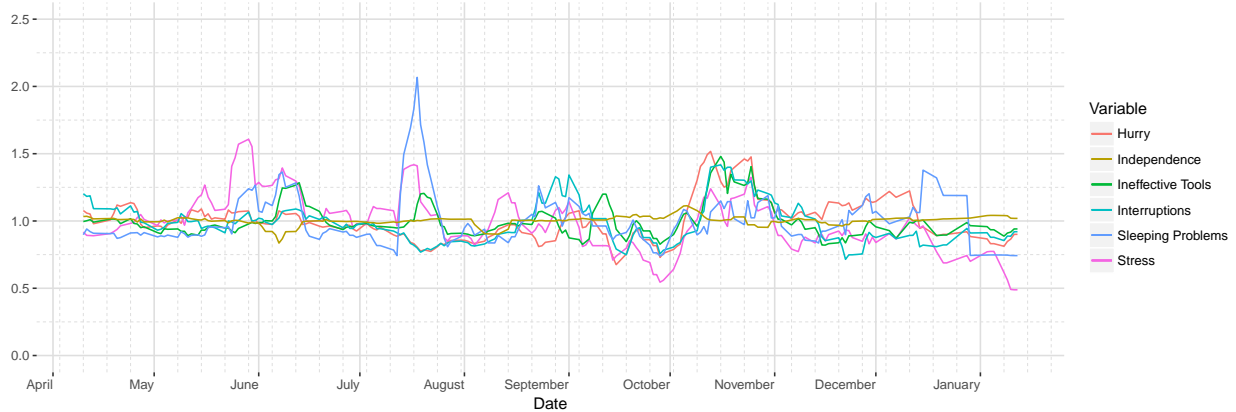Figure 2: Normalized 5-day moving average of the aggregated responses.

Table 1: Spearman correlation analysis, significant p-values on the bottom. Mean and standard deviation of actual responses.

	SD	Mean	Independence	Hurry	Interruptions	Ineffective Software Dev.	Stress	Sleeping Problems
Independence	0.698	4.648	1	-0.039	-0.085	0.001	0.139	0.006
Hurry	1.265	2.492		1	0.513	0.307	0.326	0.300
Interruptions	1.248	2.475		0.001	1	0.522	0.342	0.306
Ineffective Software Dev.	1.074	1.825		0.001	0.001	1	0.350	0.249
Stress	1.054	2.496		0.001	0.001	0.001	1	0.248
Sleeping Problems	0.995	1.838		0.001	0.001	0.001	0.001	1

components interact with each other. Since these interactions are often not known, a probabilistic network model should be used to model the causal relationships. We used the R package graphical-VAR[4] to model our data.

Temporal networks show whether a certain variable in time (t) predicts another variable at later time ($t + 1$). In Temporal networks it is typical to see loops from a variable to itself as the previous state of that variable predicts its next state. Contemporaneous networks show relationships at the same moment in time. They are

needed as "there will likely be many causal relationships that occur much faster than the lag interval of a typical ESM study; in which case, these pathways will be captured in the contemporaneous networks" [4].

4 RESULTS

Before answering our two research questions, we look at the evolution of the aggregated result of the daily questionnaire. For each day, we compute the aggregated answer of all respondents by taking the mean of all responses. Figure 1 shows a 5-day moving average of the aggregated results. In Figure 2, before aggregating the results,

---

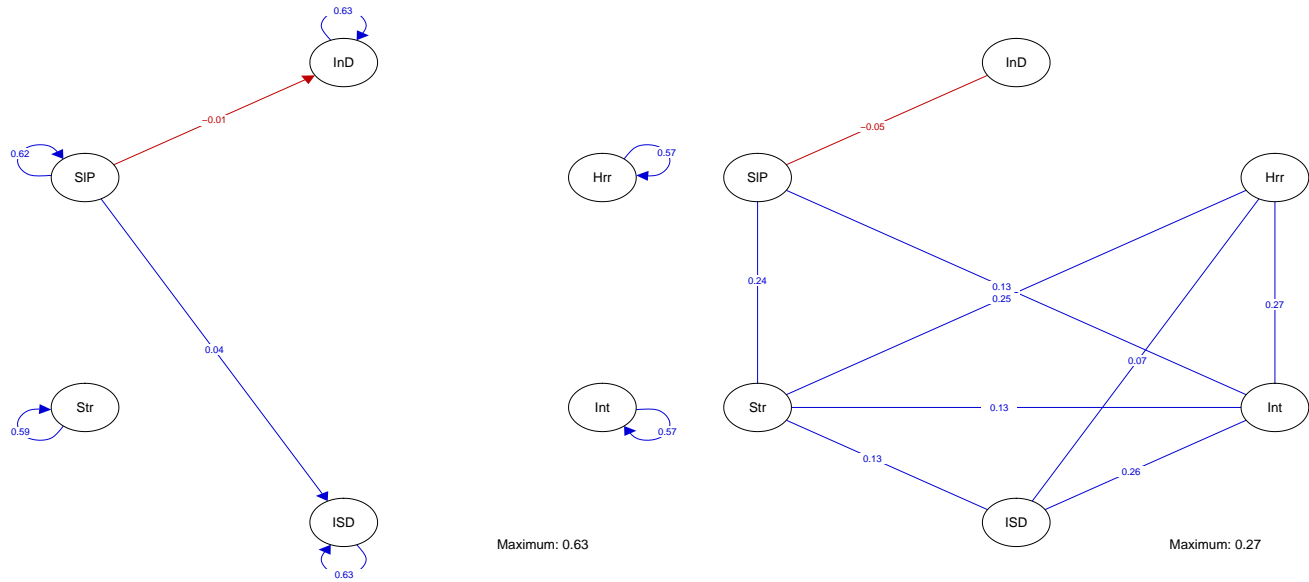[4]https://cran.r-project.org/web/packages/graphicalVAR/graphicalVAR.pdf

Figure 3: Temporal and contemporaneous network. InD corresponding to Independence, SlP to sleeping problems, Str to stress, ISD to ineffective software development, Int to interruptions and Hrr to hurry.

we normalize each respondent answer by dividing all individuals responses by the individual mean response.

Figure 1 shows a decrease in independence variable and increase in all other variables at the start of October. This trend is also captured in the normalized Figure 2. Based on informal talks with the developers, the cause for the spike is increased amount of customer feedback coupled with the start developing a new feature.

RQ1 assessed the correlation between responses. After removing the trend and seasonal component of each time series, we computed correlations between the different time series. We report standard all of these correlations in Table 1. We find statistically significant positive correlations between all variables other than independence. The strongest association (0.522) is between interruptions and ineffective software development. The second strongest association is between hurry and interruptions (0.513).

RQ2 assessed the possible causal mechanisms in responses. Both temporal and contemporaneous networks are presented in Figure 3. In the contemporaneous network, strongest positive associations are between job demands and interruptions, interruptions and ineffective software development, hurry and stress and finally stress and sleeping problems. Autonomy is negatively associated with sleeping problems in the contemporaneous network. The temporal network shows that sleeping problems predict slightly negatively job control and slightly positively ineffective software development. In the temporal model, by far the best predictor for all variables is the prior state of the same variable.

Our results are convergently valid with the job resources model [1], which assumes job strain to be caused by imbalance between resources and demands. The produced temporal network shows the effects of job strain (sleeping problems) to be slightly predictive of job control.

5   CONCLUSION, THREATS AND FUTURE WORK

Our goal was to measure stress from an ongoing software project. Our measurement of stress associates positively with job demands, interruptions and sleeping problems.

The study by Nishikitani et al. [16] found associations between overtime work, physical and mental complaints in software development context. However, sleep duration and job strain were better indicators for physical and mental distress. In our study sleeping problems were associated with interruptions.

A limiting factor for our study is generalizability; the study is made in a single software project of one software company. Stress and sleeping problems are affected by factors outside of work. Among other things, different ways of imputing missing values to time series greatly influence the produced network models.

In the future, we plan to further investigate job well-being by including additional variables in the analysis. In the particular case study here, we have access to instant messaging logs, a work time monitoring system, continuous integration service logs and a version control system. By correlating questionnaire data with software repository metrics, the underlying associations could be better understood, predictive variables for stress identified and causal hypotheses formulated. Some possible testable hypotheses are included in our prior work [11].

ACKNOWLEDGMENTS

REFERENCES

[1] Arnold B Bakker and Evangelia Demerouti. 2007. The job demands-resources model: State of the art. *Journal of managerial psychology* 22, 3 (2007), 309–328.

[2] Arnold B Bakker, Evangelia Demerouti, and Martin C Euwema. 2005. Job resources buffer the impact of job demands on burnout. *Journal of occupational health psychology* 10, 2 (2005), 170.

[3] Anna-Liisa Elo, Anneli Leppänen, and Antti Jahkola. 2003. Validity of a single-item measure of stress symptoms. *Scandinavian journal of work, environment & health* (2003), 444–451.

[4] Sacha Epskamp, Lourens J Waldorp, René Mõttus, and Denny Borsboom. 2016. Discovering psychological dynamics: The Gaussian graphical model in cross-sectional and time-series data. *arXiv preprint arXiv:1609.04156* (2016).

[5] Yuko Fujigaki. 1996. Time series investigation of job-events and depression in computer software engineers. *Industrial health* 34, 2 (1996), 71–79.

[6] Sharon Gaudin. 2015. Silicon Valley's 'pressure cooker:' Thrive of get out. (August 2015). https://www.computerworld.com/article/2972723/it-careers/silicon-valleys-pressure-cooker-thrive-or-get-out.html [Online; posted 18-August-2015].

[7] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. 2017. Consequences of unhappiness while developing software. In *Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering*. IEEE Press, 42–47.

[8] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2013. Are happy developers more productive?. In *International Conference on Product Focused Software Process Improvement*. Springer, 50–64.

[9] Tarja Heponiemi, Hannele Hyppönen, Tuulikki Vehko, Sari Kujala, Anna-Mari Aalto, Jukka Vänskä, and Marko Elovainio. 2017. Finnish physicians? stress related to information systems keeps increasing: a longitudinal three-wave survey study. *BMC medical informatics and decision making* 17, 1 (2017), 147.

[10] Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1 (1970), 55–67.

[11] Miikka Kuutila, Mika V Mäntylä, Maëlick Claes, and Marko Elovainio. 2017. Reviewing literature on time pressure in software engineering and related professions: computer assisted interdisciplinary literature review. In *Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering*. IEEE Press, 54–59.

[12] Maarit Laanti. 2013. Agile and Wellbeing–Stress, Empowerment, and Performance in Scrum and Kanban Teams. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE, 4761–4770.

[13] Lesley Pek Wee Land, Graham Low, and Oi Suen Chan. 2003. The Impact Of Time Pressure On Software Inspection Performance: A Pilot Study. In *Seventh IASTED International Conference on Software Engineering and Applications*. 574–579.

[14] Mika V Mäntylä and Juha Itkonen. 2013. More testers–The effect of crowd size and time restriction in software testing. *Information and Software Technology* 55, 6 (2013), 986–1003.

[15] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 107–110.

[16] Mariko Nishikitani, Mutsuhiro Nakao, Kanae Karita, Kyoko Nomura, and Eiji Yano. 2005. Influence of overtime work, sleep duration, and perceived job characteristics on the physical and mental status of software engineers. *Industrial Health* 43, 4 (2005), 623–629.

[17] Jason Schreier. 2016. The Horrible World of Video Game Crunch. (September 2016). https://kotaku.com/crunch-time-why-game-developers-work-such-insane-hours-1704744577 [Online; posted 26-September-2016].

[18] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

[19] Heikki Topi, Joseph S Valacich, and Jeffrey A Hoffer. 2005. The effects of task complexity and time availability limitations on human performance in database query tasks. *International Journal of Human-Computer Studies* 62, 3 (2005), 349–379.

[20] Seppo Tuomivaara, Harri Lindholm, and Marja Känsälä. 2017. Short-term physiological strain and recovery among employees working with agile and lean methods in software and embedded ICT systems. *International Journal of Human–Computer Interaction* 33, 11 (2017), 857–867.

[21] Thomas A Wright and Russell Cropanzano. 2000. Psychological well-being and job satisfaction as predictors of job performance. *Journal of occupational health psychology* 5, 1 (2000), 84.

[22] Despoina Xanthopoulou, Arnold B Bakker, Maureen F Dollard, Evangelia Demerouti, Wilmar B Schaufeli, Toon W Taris, and Paul JG Schreurs. 2007. When do job demands particularly predict burnout? The moderating role of job resources. *Journal of managerial psychology* 22, 8 (2007), 766–786.

[23] Ming Yuan and Yi Lin. 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* 94, 1 (2007), 19–35.