"*The candle is burning out on its own...*": Modeling Fatigue and Empathy Among Chinese Developers

Damian A. Tamburri[0000−0003−1230−8961]1, Haotian Zhang[0000−0000−0000−0000]2, Kelly Blincoe[0000−0003−4092−9706]3, Rick Kazman[0000−0003−0392−2783]4, Giammaria Giordano[0000−0003−2567−440X]5, Valeria Pontillo[0000−0001−6012−9947]6, and Fabio Palomba[0000−0001−9337−5116]5

[1] University of Sannio, Italy
[2] Politecnico di Milano, Italy
[3] HASEL — University of Auckland, New Zealand
[4] University of Hawaii, USA
[5] Software Engineering (SeSa) Lab — University of Salerno, Fisciano, Italy
[6] Software Languages (SOFT) Lab — Vrije Universiteit Brussel, Belgium
Gran Sasso Science Institute (GSSI), L'Aquila, Italy
datamburri@unisannio.it, haotian.zhang@studenti.polimi.it,
k.blincoe@auckland.ac.nz, kazman@hawaii.edu
{giagiordano,fpalomba}@unisa.it, valeria.pontillo@gssi.it

Abstract. Developer turnover and layoffs are at a historical peak, contributing to increased stress, fatigue, and declining morale among software developers. To investigate this issue, in this paper, we surveyed 178 developers in China and found that over half reported experiencing psychological distress, which is significantly higher than the national average. Using factor analysis and regression modeling, we identified key psychological dimensions of fatigue and empathy and examined their relationship to workplace conditions. We complemented the survey with 17 behavioral metrics from *Azure DevOps* and *Microsoft Viva Insight*, enabling a data-driven assessment of developers' work context. Finally, we developed the *Empathy Catalogues Analysis Model*, a statistical model linking work context metrics to empathy scores, revealing a significant negative correlation between workload burden and perceived empathy. Our findings provide a foundation for scalable, automated monitoring of psychological well-being in teams.

Keywords: Social Software Engineering · Application Lifecycle Management · Organisational Structures.

1 Introduction

As software systems grow increasingly complex and interconnected [17], software developers are facing ever-intensifying demands on their morale and energy. These demands stem from the intricate balancing act between technical problem-solving, team collaboration, and tight delivery schedules leading to persistently high levels of stress, burnout[7] [34], and reduced empathy among developers and their teams [5]. A recent *Stack Overflow* survey[8] confirms the severity of this issue, with 58.3% of developers reporting burnout, and over a quarter experiencing it "sometimes" or "often". These figures highlight the urgent need for solutions that can support the well-being and productivity of software teams through measurable and manageable; hence, potentially data-driven—approaches. We call this research effort *PsyOps*: an initiative to embed psychological operational capacities into DevOps pipelines, aimed at quantifying team morale and developer fatigue[9] through dedicated, automated metrics.

One of the key challenges in developing such analytics lies in assessing the anthropometric [10] and psychological characteristics associated with developer fatigue, with the aim of supporting both well-being and team productivity. Although work-related data can be automatically collected via widely used collaboration and development tools, e.g., *Microsoft Viva Insight*, *Azure DevOps*, and *Jira*, the integration and comprehensive analysis of these data sources remains limited. In particular, current approaches often fall short in capturing indicators of software developers' fatigue, as well as the empathy exercised by Application Lifecycle Management governors, e.g., product owners [31], toward their teams.

In this paper, we develop a data-driven approach to analyse developer fatigue and empathy, focusing on key anthropometric and psychological factors such as stress levels, emotional turmoil, work engagement, and perceived empathy. Building on the need for measurable indicators of well-being in software teams, our mixed-methods study combines survey responses with work context data. As summarized in Figure 1, we first conducted a large-scale questionnaire with 178 software developers in China, focusing on those with 3+ years of experience to represent a typical workforce. We then complemented the survey findings with work context data extracted from *Microsoft Viva Insight* and *Azure DevOps*. Our study addresses the following research questions:

- **$\mathbf{RQ}_1$:** What are the key anthropometric characteristics of software developers' fatigue?
- **$\mathbf{RQ}_2$:** To what extent can work context data be used to evaluate these characteristics?
- **$\mathbf{RQ}_3$:** How effective is this data in identifying mitigating factors towards fatigue?

Our results show that developer fatigue and empathy levels are strongly influenced by work context factors such as workload intensity, overtime, and meeting overload. Through factor and regression analyses, we find that increased work context burden, as measured by metrics like meeting frequency, overtime hours, and defect resolution pressure, is significantly associated with higher psychological strain and reduced empathy. As part of our *PsyOps* initiative, we developed

[7] `https://crackedlabs.org/dl/CrackedLabs_Christl_MobileWork.pdf`
[8] `https://www.theregister.com/2022/05/11/stack_overflow_stress/`
[9] `https://betterprogramming.pub/development-fatigue-fe092f036d4f`

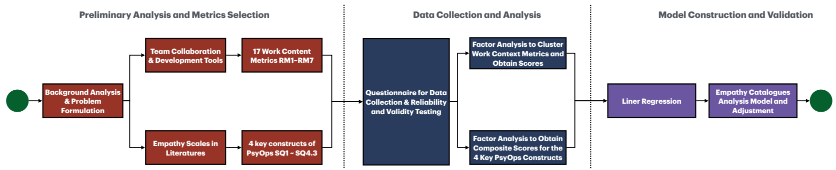

Fig. 1: Overview of Research Framework for *PsyOps*.

the *Empathy Catalogues Analysis Model*, which quantifies the relationship between work context and psychological well-being. This model confirms a negative correlation between work burden and empathy, suggesting that intensified workloads contribute to empathy fatigue and elevated stress. Our findings highlight the importance of monitoring and managing work context factors to support developers' mental health. The *PsyOps* framework demonstrates how workplace data, collected from tools like *Microsoft Viva Insight* and *Azure DevOps* platforms, can be used to assess and predict psychological strain. This data-driven approach offers a scalable path for organizations to proactively identify mental health risks and foster a more empathetic and sustainable work environment.

2 Related Work

In software development, empathy refers to developers' ability to understand and relate to the thoughts, feelings, and perspectives of others, including users, team members, and stakeholders [25,5]. It involves emotional awareness, sensitivity to others' needs, and supportive behaviour.

Empathy spans several dimensions. *User* empathy helps developers design with user needs in mind. *Team* empathy fosters collaboration through communication and mutual support [1]. *Ethical* empathy concerns socially responsible design decisions. *Contextual* empathy reflects awareness of organizational and project constraints. *Self*-empathy involves recognizing and managing one's own stress and emotions [22,36]. As the reader may see, analysing empathy in development is complex, involving emotional states, expressions, and interactions, often supported by technological tools [37]. Gathering input from team members can improve emotional insight and enhance problem-solving [15]. Previous studies showed that empathy contributes to project success [19,26] and can be implemented through behaviours like active listening and emotional support [21,30]. High stress levels, linked to long hours and pressure, can impair empathy, with symptoms such as anxiety and fatigue [20,35,24,18].

Compared to the current body of knowledge on developer well-being and empathy in software engineering—which lacks integrated, operationalizable models—*our paper contributes a novel, data-driven framework* for quantifying developer empathy and fatigue using subjective survey responses and objective workplace metrics, along with an *Empathy Catalogues Analysis Model* that reveals a significant negative correlation between work context burden and empathy.

3 Empirical Study Design

To address our **RQ**s, we gathered developers' perceptions of empathy and fatigue through an online questionnaire and complemented this with work habit data from *Microsoft Viva Insight* and *Azure DevOps*. We then applied factor analysis and linear regression to examine the relationships between the two data sources. The following sections detail our methods.

3.1 Part I: Questionnaire

Design of the study. Based on prior literature in psychology and software engineering [8,6,27], we identified four key constructs to explore in our questionnaire: (1) *empathy*, (2) *stress levels*, (3) *personal or emotional turmoil*, and (4) *work engagement*. These constructs were selected for their strong theoretical and empirical links to developer well-being and team dynamics. Each was operationalized through validated indicators drawn from established scales and mapped to specific questions in our instrument. Table 1 presents the full list of items, which participants answered using a five-point Likert scale.

Empathy. In software teams, empathy plays a critical role in fostering collaboration and responsiveness to the emotional and cognitive needs of colleagues and users [8,33]. It captures the degree to which developers perceive, express, and respond to others' emotions.

Stress level. Stress reflects physiological and behavioural signs of psychological strain. Measuring stress helps identify whether developers experience chronic pressure or burnout due to high job demands [6,35,38].

Work engagement. This construct assesses how motivated, committed, and energized developers feel in their work. It is a key predictor of productivity and job satisfaction [14,15,27].

Personal or emotional turmoil. This dimension includes major life events, mental health conditions, and interpersonal conflicts that may negatively affect well-being. It captures the broader emotional landscape in which developers operate [20,40].

Participants Selection and Recruitment. We focused our study on software developers based in mainland China for both theoretical and practical reasons. China hosts one of the world's fastest-growing software industries, marked by rapid digital transformation, intense work cultures, and concentrated urban tech hubs.[10] Yet, the psychological well-being of developers in this context remains largely understudied [12]. These conditions make Chinese developers a particularly relevant population for examining stress, fatigue, and empathy in high-pressure environments. Additionally, we had direct access to this

[10] `https://www.globenewswire.com/news-release/2024/12/27/3002160/28124/en/The-Digital-Transformation-Market-in-China-Forecast-to-2029-Trends-Demand-Drivers-Challenges-and-Emerging-Opportunities.html`

Table 1: Scale items for the main constructs: empathy, stress level, work engagement, and personal or emotional turmoil.

ID	Item
Empathy [9,33]	
SQ1.1	How well do you think you understand the needs and emotions of your colleagues and users?
SQ1.2	How often do you communicate with your colleagues and users about their emotions and concerns?
SQ1.3	How often do you provide emotional support or encouragement to your colleagues and users?
SQ1.4	How well do you feel supported and understood by your colleagues and superiors?
Stress level [6,35,38]	
SQ2.1 †	Have you ever experienced burnout or mental health issues related to your work as a software developer?
SQ2.2 †	How often do you experience symptoms of stress (e.g., anxiety, insomnia, fatigue, irritability)?
SQ2.3 †	How well does the stress in your professional life affect your personal life?
Work engagement [14,15,27]	
SQ3.1	How satisfied are you with working as a software developer?
SQ3.2	How motivated are you to perform well in your job?
SQ3.3	How often do you feel engaged and energized by your work?
Personal or emotional turmoil [20,40]	
SQ4.1 †	Have you experienced any significant life events (e.g., divorce, bereavement, illness) in the past year that have impacted your well-being?
SQ4.2 †	Do you have some mental health conditions (e.g., depression, anxiety) that impact your work as a software developer?
SQ4.3 †	Have you noticed some physical symptoms or changes that may be related to stress or emotional turmoil (e.g., headaches, fatigue, muscle tension, changes in appetite or weight)?

population through professional and social platforms in the Chinese IT sector, which enabled efficient recruitment. Participants were indeed selected through *convenience sampling* by distributing the questionnaire via WENJUANXING™ (`https://www.wjx.cn`), a popular online survey tool in China. The 10-minute questionnaire, consisting of 12 open and closed questions, was disseminated without incentives to minimize selection bias. In total, we received 178 complete responses. As shown in Figure 2, most respondents were based in major software development regions, including Shanghai, Jiangsu Province, Beijing, and Fujian Province. While the focus on a single national context may limit generalizability, it provides valuable insights into a high-intensity and globally significant segment of the software workforce.

Data Analysis. Prior to performing analysis, the negatively worded items, as indicated by † in Table 1, were reverse-scored. Cronbach's alpha was used to check scale reliability. Scores range from 0 to 1, with higher scores being indicative of greater consistency in the scale items [7]. A score of 0.9 or higher is excellent, 0.8 or higher is good, and 0.7 or higher is acceptable. The result of Cronbach's alpha coefficient was 0.829, which is greater than 0.8, thus indicating good reliability of the questionnaire data.

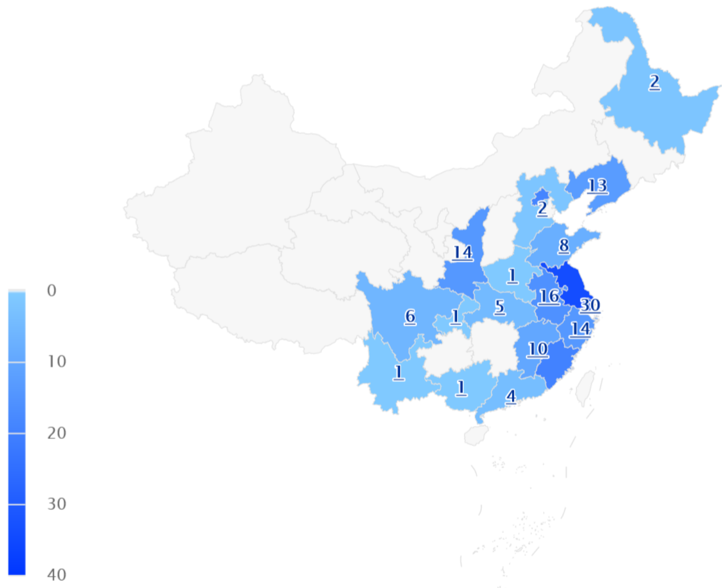

Fig. 2: Distribution of completed questionnaires.

3.2 Part II: Data Pipeline

As the second part of the study, we aimed to complement self-reported perceptions with objective behavioural data from workplace tools. *Microsoft Viva Insight* collects and analyses data from email, calendars, and collaboration platforms to generate metrics on productivity (e.g., focus hours, meeting load) and collaboration (e.g., interaction frequency, team engagement, network centrality). In addition, we considered data from *Azure DevOps*, which offers process-level metrics like code commits, build activity, deployment frequency, lead time, and defect reports. Together, these sources allowed us to examine how work context factors relate to developers' psychological well-being in a data-driven manner.

We collected the 17 metrics listed in Table 2 from *Microsoft Viva Insight* and *Azure DevOps* platforms using standard log mining tools in Python (PM4Py[11]). After removing outliers, we normalized each metric to a 0–1 scale to account for differences in magnitude across indicators. The normalization was performed using min–max scaling, as shown in Equation (1), where X_{ij} is the original value of the j-th metric for the i-th participant, and X_{jmin} and X_{jmax} are the minimum and maximum values of that metric across all participants:

$$\hat{X_{ij}} = \frac{X_{ij} - X_{jmin}}{X_{jmax} - X_{jmin}} \tag{1}$$

[11] `https://pypi.org/project/pm4py/`

Table 2: Representative metrics of work context data.

	Metric	Note
Viva Insight	RM1	Work overtime in a month (%)
	RM2	Meeting arranged at least one day in advance (%)
	RM3	Meeting no overlap (%)
	RM4	Meeting Ended on time (%)
	RM5	Collaboration time within work hours (hours)
	RM6	Collaboration time outside work hours (hours)
	RM7	Available to focus time (%)
	RM8	Meeting time out of collaboration time (%)
	RM9	Emails time out of collaboration time (%)
	RM10	Chats time out of collaboration time (%)
DevOps	RM11	Number of builds per week (times)
	RM12	Number of commits per week (times)
	RM13	Number of tasks assigned per week (pieces)
	RM14	Number of tasks completed per week (pieces)
	RM15	Failed test cases out of planned test cases per week (%)
	RM16	Number of defects reported per week (pieces)
	RM17	Defects resolved out of the total number of defects(%)

We developed data pipeline integrations to automatically collect and transfer relevant metrics from *Microsoft Viva Insight* and *Azure DevOps* platforms into a centralized data repository. This repository supports further analysis of patterns and correlations between work context metrics and the anthropometric indicators described above.

3.3 Part III: Developer Fatigue Indicator Model

Next, we explore the relationship between work context metrics and developers' self-reported perceptions. We apply factor analysis to identify latent constructs and use multivariate linear regression to model their associations. An overview of the combined data sources is shown in Figure 3.

Factor Analysis. We sought to identify the underlying relationships in both sources of data. To do this, we employed an *Explanatory Factor Analysis* with the maximum variance rotation method (varimax), which is the most common rotation that maximizes the differences between the factors [13,11].

Factorability. We assessed factorability using the Kaiser-Meyer-Olkin (KMO) metric and Bartlett's test of sphericity [29]. For both data sources, the KMO exceeds the 0.5 threshold and Bartlett's test is significant ($p < 0.01$), confirming the analysis's suitability (Table 3).

Number of Factors. To identify the number of factors, we retained those with eigenvalues greater than 1. This commonly used criterion ensures that each

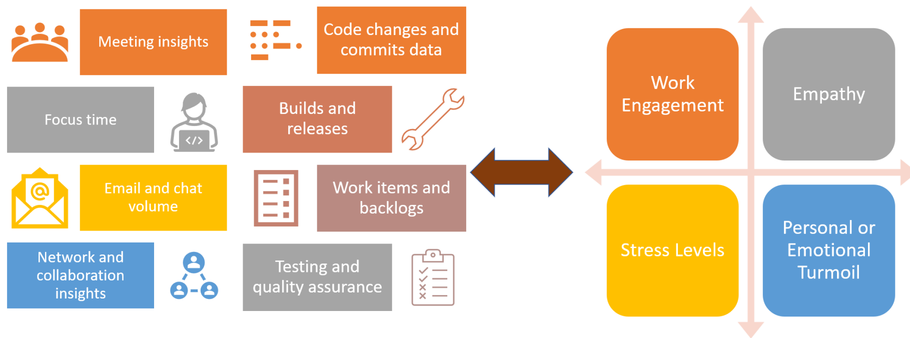

Fig. 3: Overview of data from both sources.

Table 3: Factorability and Factor Analysis Model Fit.

	KMO	Bartlett	RMSR	RMSEA	TLI
Questionnaire Cata	0.821	χ^2=1300.798; df=78; p<0.001	0.17	0.198	0.672
Work Context Data	0.937	χ^2=1179.26; df=136; p<0.001	0.171	0.041	0.96

retained factor explains a significant portion of the variance, thereby representing the underlying structure of the data effectively [16].

Model fit. Model fit was assessed using RMSR, RMSEA, and TLI, with standard thresholds: lower is better for RMSR (0 is ideal), RMSEA < 0.06, and TLI > 0.90 (adequate), > 0.95 (good) [3] [28]. As shown in Table 3, the work context model meets these criteria (RMSEA $= 0.041$, TLI $= 0.960$), though RMSR is relatively high. In contrast, the questionnaire model shows poor fit across all metrics (RMSR $= 0.170$, RMSEA $= 0.198$, TLI $= 0.672$), suggesting it does not adequately capture the data's structure.

Factor loading. We used a 0.3 cutoff for factor loading [16,32]. Variables below this threshold have been excluded.

Factor scores. Finally, to obtain the overall factor score for each dimension of empathy, we compute the weighted sum of the individual factor scores. **E** represents the vector of factor scores for empathy dimensions, with dimensions corresponding to the identified empathy dimensions. The calculation of the overall factor score **p** for each dimension is expressed as:

$$\mathrm{p} = \sum_{j=1}^{n} (E_j \times w_j) \tag{2}$$

where E_j represents the factor score for the survey samples on principal component j, and w_j represents the weight assigned to principal component j. The weights reflect the importance of each principal component in contributing to the respective empathy dimension.

Correlations between items. Figure 4 shows the correlation matrix of the questionnaire responses. A strong correlation is observed among items within each construct. Notably, SQ1.2 and SQ1.4 are negatively correlated with the other variables and were therefore positively normalized.

We also performed a correlation analysis of the work context data (RM1–17). As shown in Figure 5, these metrics likewise exhibit strong correlations.

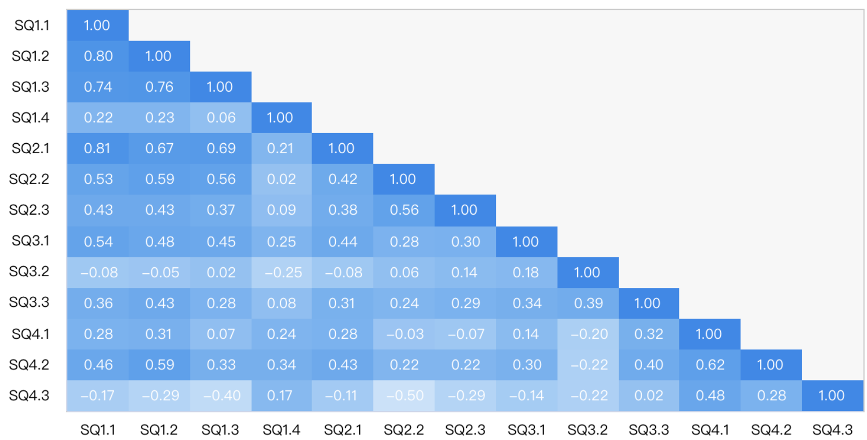

Fig. 4: Correlation matrix of SQ1.1-SQ4.3.

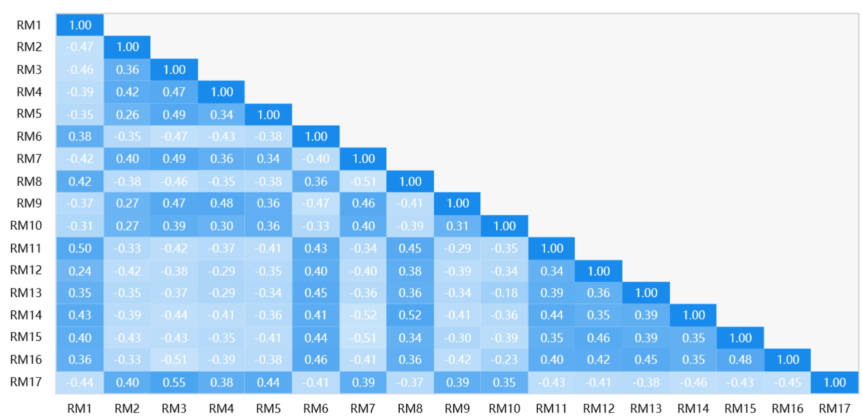

Fig. 5: Correlation matrix of RM1-RM17.

Linear Regression Model. To examine the relationship between empathy levels and work context burden, we used a linear regression model estimated via Ordinary Least Squares (OLS). This approach identifies how changes in work context factors predict variations in empathy, with coefficients representing the effect of each factor while holding others constant.

The *Empathy Catalogues Analysis Model* combines factor analysis and linear regression to quantify empathy dimensions from questionnaire responses and link them to work context metrics. This enables the identification of significant predictors and the estimation of their impact on empathy levels.

4 Analysis of the Results

In this section, we first analyse and observe the frequency counts and descriptive statistics reflective of the data to understand the empathy and workload of the software practitioners. Then, we analyse the data with factor analysis and regression analysis to exercise our proposed fatigue indicator model.

4.1 Frequency Analysis

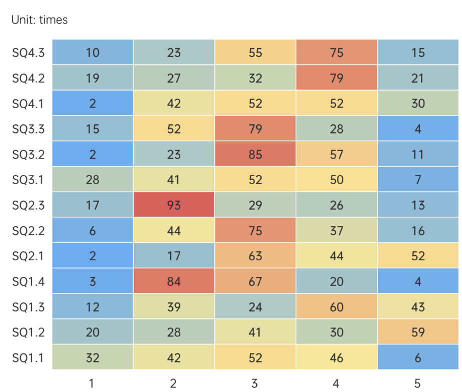

Fig. 6: Answer distribution of SQ1.1-SQ4.3, in which 1 and 5 refer respectively to the least frequent or compliant and the most frequent or compliant.

Empathy Analytics Data. Figure 6 shows the frequency distribution of SQ1.1-SQ4.3. It is noteworthy that 57% of the respondents chose 4 and 5 for SQ1.3, 50% of the respondents chose 1 and 2 for SQ1.4, which suggests that **software developers feel that they provide considerable emotional support to their colleagues and users, but they are supported and understood by the colleagues and superiors not so well.**

Only 10% of the respondents chose 1 and 2 in SQ2.1, but 50% of the respondents chose 3 and 4, indicating that **most software developers have**

experienced burnout or mental health issues related to their work. Moreover, 62% of the respondents chose 1 and 2 in SQ2.3, indicating that **most software developers believe that the stress of their professional life significantly affects their personal life**.

For the work engagement, i.e., questions SQ3.1, SQ3.2, and SQ3.3, most of the respondents gave neutral choices, **suggesting uncertainty about their connection to the work.**

For SQ4.2, and SQ4.3, 55% and 50% of the respondents chose 4 and 5, which **proves that more respondents believe that they have some mental health conditions** (e.g., depression, anxiety), and that **their mental health has caused some physical symptoms** (e.g., headaches, fatigue, muscle tension, changes in appetite or weight). This percentage is much higher than that reported in the *China National Mental Health Development Report (2021-2022)* published by the Institute of Psychology of the Chinese Academy of Sciences and the Social Science Literature Publishing House, where only 13.9% of the respondents chose a high level [39].

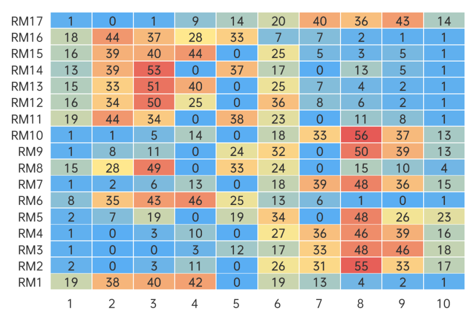

Fig. 7: Answer distribution of RM1-RM17.

Work Context Data. Figure 7 shows the distribution of the answers for the metrics related to the work context. On the one hand, we can observe that for RM2, 3, 4, 5, 7, most of the respondents filled in larger values, proving **that software practitioners in China have better meeting arrangements (scheduled in advance, no overlap, end on time, collaborate during working hours, have more focus time).** Additionally, the answers received for RM8, 9, and 10 prove that **most software developers spend more collaboration time on email and chat than on meetings.**

On the other hand, the comparison of RM11, 12, 13, 14, 15, 16 with RM17, proves that the number of defects resolved still falls in the larger range when most of the people are at the workload, which suggests that **software developers are facing a greater challenge to resolve the defects.**

Table 4: Total Variance Explained for SQ1.1-SQ4.3.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.021	38.620	38.620	5.021	38.620	39.006
2	2.350	18.074	56.694	2.350	18.074	56.694
3	1.370	10.540	67.234	1.370	10.540	67.234
4	.897	6.898	74.132			
5	.816	6.277	80.410			
6	.571	4.391	84.801			
7	.479	3.681	88.482			
8	.367	2.820	91.302			
9	.321	2.471	93.773			
10	.286	2.199	95.972			
11	.222	1.704	97.676			
12	.189	1.452	99.128			
13	.113	.872	100.000			

4.2 Factor Analysis

Empathy Analytics Data. Table 4 presents the results of the factor extraction, including the amount of information captured by the extracted factors. A total of three factors were extracted, each with an eigenvalue greater than 1. After rotation, these three factors explained 38.6%, 18%, 10.5%, respectively. The cumulative variance explained after rotation is 67.2%, denoted as w_j.

Table 5(a) shows the information extraction of the factors for the questions and the correspondence between the factors and the questions—we sorted the coefficients of the questions in order of magnitude. SQ1.4 and SQ3.1 are empty because the value is less than 0.3. Items originally designed to capture empathy (SQ1.1, SQ1.2, and SQ1.3) and stress levels (SQ2.1, SQ2.2, and SQ2.3) loaded strongly on the same factor. This suggests that respondents perceive emotional support and psychological strain as closely linked rather than distinct dimensions. We interpret this as a unified **Empathy–Stress Levels** factor. The second factor, formed by SQ4.1 to SQ4.3, corresponds to **Personal or Emotional Turmoil**, in line with prior literature [20,40]. The third factor, derived from SQ3.2 and SQ3.3, captures **Work Engagement**, which also aligns with previous literature [14,15,27].

Although the questionnaire was designed around four constructs, the analysis revealed a three-factor structure. Based on the rotation variance, we constructed Figure 8 to illustrate the contribution of each factor, where empathy and stress levels had the strongest influence, and work engagement the weakest. These factor scores were then used to compute the overall composite score p.

Work Context Data. For the RM1–RM17 metrics, only one factor was extracted, as shown in Table 5(b). Based on the Rotated Component Matrix and

Table 5: Rotated Component Matrices

(a) Matrix for SQ1.1–SQ4.3

Item	Comp. 1	Comp. 2	Comp. 3
SQ1.3	.862		
SQ1.2	.857		
SQ1.1	.851		
SQ2.2	.777		
SQ2.1	.766		
SQ2.3	.623		
SQ4.1		.846	
SQ4.2		.760	
SQ4.3		.698	
SQ3.2			.855
SQ3.3			.744
SQ1.4			
SQ3.1			

(b) Matrix for RM1–RM17

Item	Component
RM3	−.740
RM7	−.700
RM17	−.693
RM14	.686
RM6	.681
RM15	.671
RM8	.670
RM16	.668
RM1	.658
RM11	.651
RM9	−.645
RM4	−.630
RM5	−.624
RM12	.621
RM2	−.610
RM13	.604
RM10	

the correlation matrix, we developed a data correlation model describing the relationships between key aspects of the work context. The analysis reveals several meaningful groupings.

RM2, RM3, and RM4 form the **Meeting on Time** cluster, which negatively contributes to the work context burden score—indicating that more timely meetings are associated with reduced perceived burden. Similarly, RM5, RM7, and RM9 define the **Available to Focus** cluster, which also shows a negative contribution, suggesting that increased focus time alleviates burden.

In contrast, RM1, RM6, and RM8 belong to the **Work Overtime** cluster, which contributes positively, reflecting that extended working hours increase perceived workload. RM12, RM13, and RM14 form the **Development Load** cluster, also with a positive contribution, indicating that heavier development workloads correlate with greater burden.

Additionally, RM11, RM15, and RM16 compose the **Failed & Defect** cluster, which further increases the burden score, emphasizing the impact of defects and failures on developers. Finally, RM10 and RM17 define the **Defect Resolved** cluster, which negatively affects the burden score, suggesting that resolving defects helps reduce the overall burden.

The relationship between these matrices and their respective contributions to the work context burden score is illustrated in Figure 9. Based on these patterns, we derive a comprehensive score z to represent overall work context burden.

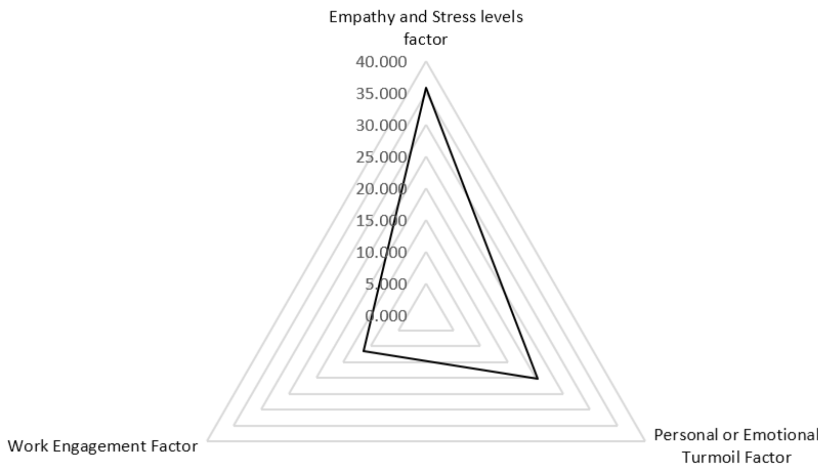

Fig. 8: Anthropometric Factor Contribution Model.

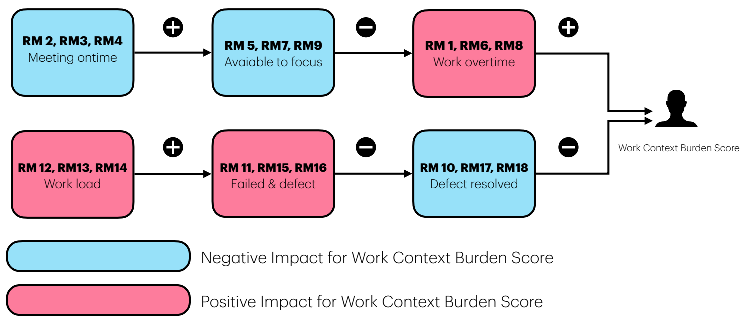

Fig. 9: Developer fatigue indicator model.

4.3 Linear Regression

To build the *Empathy Catalogues Analysis Model* and examine the relationship between the overall factor score **p** (developer empathy) and the work context burden score **z**, we applied linear regression.

As shown in Table 6, the R^2 value is 0.540, indicating that the RM final score accounts for 54.0% of the variance in the composite empathy score. The F-test confirms the model's significance ($F = 206.679$, $p = 0.000 < 0.05$), demonstrating that the RM score has a statistically significant impact. The regression coefficient for the RM final score is -0.371 ($t = -14.376$, $p = 0.000 < 0.01$), indicating a significant negative linear relationship. In other words, **the greater the burden of the work context, the lower the empathy score**.

Table 6: Linear Regression Analysis Results.

	Regr. Coefficient	**95% CI**	Cov. Diagnostic	
			VIF	Tolerance
Constant	0.000 (0.000)	-0.136 ∼-0.136	-	-
Z	-0.371** (-14.376)	-0.422 ∼-0.321	1	1
Sample Size	178			
R^2	0.540			
Adjusted R^2	0.537			
F value	*F (1,176)=206.679, p=0.000*			

5 Discussion and Lessons Learned

This research addresses our three research questions by identifying and quantifying key work context factors that influence developers' psychological well-being and empathy. Regarding $\mathbf{RQ}_1$, we found that over half of surveyed developers report psychological distress—far above national averages—underscoring the urgency of addressing fatigue in software development, where cognitive demands elevate mental health risks. For $\mathbf{RQ}_2$, factor analysis identified four key matrices—Meeting on Time, Available to Focus, Work Overtime, and Development Load—forming the *Work Context Burden Score*. While overtime and workload increase burden, structured meetings and focus time alleviate it. These constructs, integrated in our model, offer an empirical basis for assessing work-related strain. In response to $\mathbf{RQ}_3$, regression analysis via the *Empathy Catalogues Analysis Model* confirms a strong negative correlation between burden and empathy, indicating that heavier workloads reduce empathetic capacity.

Empathy emerges as context-sensitive rather than fixed. Among the three related constructs—Empathy and Stress Levels, Personal or Emotional Turmoil, and Work Engagement—the first most strongly shapes overall empathy. This suggests that improving work context can directly benefit developers' psychological and interpersonal functioning. Our findings carry implications for both practitioners and researchers. Organizations can reduce burnout and enhance team cohesion by minimizing overtime, managing workloads, and promoting time for structured focus. Since the metrics used are available in common development and collaboration tools, they can support automated monitoring and timely managerial action. For researchers, our results support treating empathy and mental health as operationalizable constructs in software engineering. The *PsyOps* framework opens avenues for longitudinal studies and predictive tools for developer well-being based on passively collected behavioral data.

6 Threats to Validity

This study is subject to several limitations that may affect the generalizability and robustness of its findings. First, the sample was restricted to a specific demographic and geographic region within China, which may limit the applicability of results to other cultural or organizational contexts [4,2]. Cultural nuances unique to the Chinese setting might have influenced the observed correlations between empathy and work context, suggesting the need for cross-cultural studies to validate these results.

Second, reliance on self-reported data for empathy and psychological measures may introduce biases such as social desirability [23], potentially affecting data accuracy. While objective work context data was obtained from *Microsoft Viva Insight* and *Azure DevOps*, these sources may miss interpersonal or unquantified environmental factors relevant to developers' psychological well-being.

Finally, the use of linear regression in the *Empathy Catalogues Analysis Model* may not fully capture the complex or non-linear relationships between empathy and work context burden. Future research could apply advanced models, including machine learning, to identify subtler patterns. Longitudinal data could also provide insights into the temporal dynamics of work conditions and mental health, enhancing both the precision and predictive power of the model.

7 Conclusion and Future Work

In this paper, we investigated the mental health and empathy of Chinese software developers through a questionnaire combining 17 work context items and 13 empathy-related questions. Over half of the participants reported psychological distress, far exceeding the national average of 13.9%, highlighting the urgency of the issue. We identified four key empathy-related dimensions and analyzed 17 objective workplace metrics collected from tools like *Microsoft Viva Insights* and *Azure DevOps*. Our *Empathy Catalogues Analysis Model* revealed a significant negative correlation between work context burden and empathy, enabling automated, data-driven assessments of developers' psychological well-being. This study offers a novel, scalable approach for monitoring empathy and mental health in software teams using workplace data. Future work will aim to refine the model and validate it in broader organizational and cultural contexts.

References

1. Akgün, A.E., Keskin, H., Cebecioglu, A.Y., Dogan, D.: Antecedents and consequences of collective empathy in software development project teams. Information & Management **52**(2), 247–259 (2015)
2. Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B.: Guidelines for the process of cross-cultural adaptation of self-report measures. Spine **25**(24)
3. Bentler, P.M.: Comparative fit indexes in structural models. Psychological bulletin **107**(2), 238 (1990)

4. Borsa, J.C., Damásio, B.F., Bandeira, D.R.: Cross-cultural adaptation and validation of psychological instruments: Some considerations. Paidéia (Ribeirão Preto) **22**, 423–432 (2012)
5. Cerqueira, L., Freire, S., Neves, D.F., Bastos, J.P.S., Santana, B., Spínola, R., Mendonça, M., Santos, J.A.M.: Empathy and its effects on software practitioners' well-being and mental health. IEEE Software (2024)
6. Cohen, S., Kamarck, T., Mermelstein, R.: A global measure of perceived stress. Journal of health and social behavior pp. 385–396 (1983)
7. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. psychometrika **16**(3), 297–334 (1951). `https://doi.org/10.1007/BF02310555`
8. Davis, M.H.: Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of personality and social psychology **44**(1), 113
9. Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.w., Oishi, S., Biswas-Diener, R.: New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social indicators research **97**, 143–156 (2010)
10. Espitia-Contreras, A., Sanchez-Caiman, P., Uribe-Quevedo, A.: Development of a kinect-based anthropometric measurement application. In: 2014 IEEE Virtual Reality (VR). pp. 71–72 (2014). `https://doi.org/10.1109/VR.2014.6802056`
11. Forina, M., Armanino, C., Lanteri, S., Leardi, R.: Methods of varimax rotation in factor analysis with applications in clinical and food chemistry. Journal of Chemometrics **3**(S1), 115–125 (1989)
12. Godliauskas, P., Šmite, D.: The well-being of software engineers: a systematic literature review and a theory. Empirical Software Engineering **30**(1), 1–42 (2025)
13. Gorsuch, R.L.: Exploratory factor analysis. In: Handbook of multivariate experimental psychology, pp. 231–258. Springer (1988)
14. Graziotin, D., Fagerholm, F.: Happiness and the productivity of software engineers. In: Rethinking Productivity in Software Engineering, pp. 109–124. Springer (2019)
15. Graziotin, D., Wang, X., Abrahamsson, P.: Happy software developers solve problems better: psychological measurements in empirical software engineering. PeerJ
16. Hair Jr, J.F., Anderson, R.E., Tatham, R.L.: Multivariate data analysis with readings. Macmillan Publishing Co., Inc. (1986)
17. Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., Wang, H.: Large language models for software engineering: A systematic literature review (2024), `https://arxiv.org/abs/2308.10620`
18. van der Hulst, M.: Long workhours and health. Scandinavian journal of work, environment & health **29 3**, 171–88 (2003)
19. Judge, T.A., Heller, D.K., Mount, M.K.: Five-factor model of personality and job satisfaction: a meta-analysis. The Journal of applied psychology (2002)
20. Kawakami, N., Haratani, T.: Epidemiology of job stress and health in japan: review of current evidence and future direction. Industrial health **37**(2), 174–186 (1999)
21. King Jr, S., Holosko, M.J.: The development and initial validation of the empathy scale for social workers. Research on Social Work Practice **22**(2), 174–185 (2012)
22. Levy, M.: Educating for empathy in software engineering course. In: REFSQ
23. Van de Mortel, T.F.: Faking it: social desirability response bias in self-report research. Australian Journal of Advanced Nursing, The **25**(4), 40–48 (2008)
24. Noboru, Iwata, Norito, Kawakamiz, Takashi, Haratani, Katsuyuki, Murata, Shunichi, Araki: Job stressor-mental of japanese working negative questions ? health adults associations in a sample : Artifacts of positive and (2007)
25. Pérez-Acosta, A.M.: Vandenbos, gr (ed.)(2007). apa dictionary of psychology. Revista Latinoamericana de Psicología **40**(2), 381–384 (2008)

26. Saleh, H.G.A., Hu, W., Hassan, H.M.A., Khudaykulova, M.: Dark leadership impact on psychological well-being and work-family conflict: Implications for project success of bahrain companies. Journal of Int'l. business research **3**, 32–39 (2018)
27. Schaufeli, W.: Work engagement: What do we know and where do we go? Romanian journal of applied psychology **14**(1), 3–10 (2012)
28. Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J.: Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of Educational Research **99**(6), 323–338 (2006). `https://doi.org/10.3200/JOER.99.6.323-338`, `https://doi.org/10.3200/JOER.99.6.323-338`
29. Shrestha, N.: Factor analysis as a tool for survey analysis. American journal of Applied Mathematics and statistics **9**(1), 4–11 (2021)
30. Skovholt, T.M.: The cycle of caring: A model of expertise in the helping professions. Journal of Mental Health Counseling **27**(1), 82–93 (2005)
31. de Ste-Croix, A., Easton, A.: Proceedings of the agile development conf. 2008. In: AGILE. IEEE, Washington, DC, USA (2008)
32. Stevens, J.P.: Applied multivariate statistics for the social sciences. Routledge
33. Tian, Q., Robertson, J.L.: How and when does perceived csr affect employees' engagement in voluntary pro-environmental behavior? Journal of Business Ethics
34. Tulili, T.R., Capiluppi, A., Rastogi, A.: Burnout in software engineering: A systematic mapping study. Information and Software Technology **155**, 107116 (2023). `https://doi.org/https://doi.org/10.1016/j.infsof.2022.107116`, `https://www.sciencedirect.com/science/article/pii/S0950584922002257`
35. Virtanen, M., Ferrie, J.E., Singh-Manoux, A., Shipley, M.J., Stansfeld, S.A., Marmot, M., Ahola, K., Vahtera, J., Kivimäki, M.: Long working hours and symptoms of anxiety and depression: a 5-year follow-up of the whitehall ii study. Psychological Medicine **41**, 2485 – 2494 (2011)
36. Wallin, A., Nokelainen, P., Kira, M.: From thriving developers to stagnant self-doubters: An identity-centered approach to exploring the relationship between digitalization and professional development. Vocations and Learning
37. Werder, K.: The evolution of emotional displays in open source software development teams: An individual growth curve analysis. 2018 IEEE/ACM 3rd Int'l. Workshop on Emotion Awareness in Software Engineering (SEmotion) pp. 1–6 (2018), `https://api.semanticscholar.org/CorpusID:49864772`
38. Wong, N., Jackson, V., Van Der Hoek, A., Ahmed, I., Schueller, S.M., Reddy, M.: Mental wellbeing at work: Perspectives of software engineers. In: Proceedings of the 2023 CHI Conf. on Human Factors in Computing Systems. pp. 1–15 (2023)
39. Xiaolan, F., Kan, Z., Xuefeng, C., Zhiyan, C.: China National Mental Health Development Report (2021-2022). Institute of Psychology of the Chinese Academy of Sciences and the Social Science Literature Publishing House (2022)
40. Zapf, D., Seifert, C., Schmutte, B., Mertini, H., Holz, M.: Emotion work and job stressors and their effects on burnout. Psychology & health **16**(5), 527–545 (2001)