

Characterising Meeting Practices in Software Engineering Teams: A Mixed-Methods Study of Perceptions, Objective Patterns, and Impacts

Nitish Patkar 

University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Windisch, Switzerland

Xeno Isenegger 

Coop, Bern, Switzerland

Gideon Monterosa 

Swiss Post, Bern, Switzerland

Sebastiano Panichella 

University of Bern, Switzerland

AI4I – The Italian Institute of Artificial Intelligence for Industry, Turin, Italy

Norbert Seyff 

University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Windisch, Switzerland

Abstract

Background. Meetings are an integral part of software practitioners' workday, but many practitioners see them as disruptive to focused work and productivity. Prior research has mainly focused on overall meeting load and broad perceptions; however, teams cannot effectively improve meeting practices without knowing which specific meeting patterns cause disruption.

Aims. Our core aim is to move from general perceptions to an empirical investigation of meeting characteristics that are strongly linked to productivity, flow, and well-being.

Method. We conducted a two-phase mixed-method study. In Phase 1, we surveyed 55 software practitioners to establish a baseline of meeting practices and perceived impact on productivity, flow, and well-being. In Phase 2, a two-week field study with experience sampling collected real-time post-meeting ratings (usefulness, energy, focus disruption) and daily end-of-day reports, complemented by objective meeting characteristics extracted automatically from calendar metadata (135 post-meeting and 40 end-of-day records).

Results. In the baseline survey, 61.8% reported at least moderate flow disruption and 76.4% lower productivity on high-meeting days, with unclear purpose and excessive duration as the top frustrations. The field study revealed a more nuanced pattern: overall usefulness was moderately positive, but varied by type – ad-hoc meetings were among the most useful ($Mdn = 4$), planning meetings were most disruptive ($Mdn = 4$), and one-on-ones were least disruptive ($Mdn = 1$). Type contrasts are exploratory and based on small sample. At the daily level, higher meeting counts coincided with lower productivity and deep-work ratings; energy was lowest on high-meeting days. This daily pattern reflected between-participant differences rather than within-person day-to-day effects. The directional alignment between survey complaints and real-time observations strengthens confidence in these emerging patterns.

Conclusions. This study shifts the conversation from broad dissatisfaction with meetings toward identifiable and actionable meeting patterns that organizations can actively redesign. The findings suggest that not all meetings are equally harmful—or beneficial—and that factors such as meeting purpose clarity, duration, and scheduling context may play a critical role in shaping employees' productivity, ability to sustain flow, and daily energy levels. We treat these patterns as design hypotheses pending larger confirmation. They also provide an empirical starting point for longitudinal research on meeting design.

2012 ACM Subject Classification Human-centered computing → Empirical studies in HCI; Software and its engineering → Requirements analysis; Software and its engineering → Programming teams; General and reference → Empirical studies; General and reference → Experimentation



© Nitish Patkar, Xeno Isenegger, Gideon Monterosa, Sebastiano Panichella, and Norbert Seyff; licensed under Creative Commons License CC-BY 4.0

20th International Symposium on Empirical Software Engineering and Measurement (ESEM 2026).

Editors: Robert Feldt, Maria Paasivaara, Daniel Mendez, Stefan Wagner, and Marvin Muñoz Barón;

Article No. 7; pp. 7:1–7:19



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Keywords and phrases Software engineering, Meeting practices, Productivity, Flow (deep work), Well-being, Experience sampling method (ESM)

Digital Object Identifier 10.4230/LIPIcs.ESEM.2026.7

Category Technical Track Paper

Supplementary Material *Dataset (Survey and field-study data, prototype source code, and analysis scripts):* <https://figshare.com/s/453a1f1a5c920066393f> [14]

1 Introduction

Meetings are one of the most ubiquitous forms of collaborative work in software development. Prior studies show that practitioners spend a substantial share of work time in meetings and informal communication [11], while meeting frequency has risen across knowledge work more broadly [18]. Meetings are essential for coordination [23, 16], however, they are also often experienced as disruptive. Developers frequently report interruptions to deep work [7, 11], and many practitioners perceive current meeting load as excessive [22, 18].

The tension between coordination value and productivity cost has received growing empirical attention. In organisational psychology, meeting load has been linked to well-being and daily fatigue [18, 10]; in software engineering, studies show that perceptions of meeting value vary across context, role, and meeting type [11, 22]. Related work on interruptions and transitions further suggests that meeting cost extends beyond meeting time itself: regaining focus after context switches can be cognitively expensive [7, 1]. At the same time, most prior evidence remains aggregate (e.g., total meeting time/frequency) or retrospective, which limits practical guidance for improving everyday meeting practice.

A key empirical gap is therefore not only whether meetings are experienced as costly, but which specific meeting characteristics are associated with that cost in situ. In particular, objective scheduling properties, such as timing within the day, duration, and back-to-back clustering (meetings with ≤ 5 min gap), have rarely been examined against the within-person momentary experience of productivity impact, flow disruption, and energy change. As a result, teams still have limited evidence for deciding which meeting patterns to modify first and under what conditions changes are likely to help.

To address these gaps, we conduct a two-phase mixed-method study of software practitioners. In Phase 1, a practitioner survey establishes a baseline characterisation of meeting practices and perceived effects on productivity, flow, and well-being. In Phase 2, a two-week field study combines experience sampling with objective calendar metadata to examine how meeting characteristics relate to momentary disruption, energy change, and daily outcomes. Together, these phases allow us to contrast aggregate perceptions with field observations under natural working conditions. This study makes three contributions.

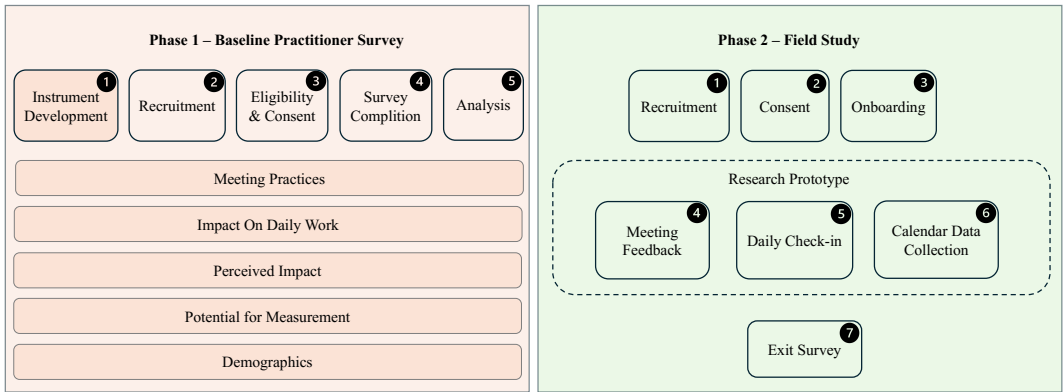
1. **Moving beyond “meeting overload” to actionable meeting patterns.** We provide an empirical characterisation of how meeting practices shape not only productivity, but also employees’ flow, energy, and day-to-day well-being. Rather than treating meetings as uniformly costly, the study identifies which meeting characteristics most strongly predict disruption and usefulness in practice.
2. **Capturing meetings in real work contexts through in-situ analysis.** We provide within-person, real-time evidence on how objective meeting characteristics – timing, duration, and back-to-back clustering – relate to momentary experiences of productivity, focus, and energy, moving the literature beyond retrospective self-report toward observations grounded in practitioners’ actual workdays.

3. **Revealing where perceptions and observed behaviour align – or diverge.** By comparing survey-reported beliefs with field measurements, we identify which commonly reported meeting frustrations are supported by real-time evidence and which are less consistently reflected in observed patterns. This bridges practitioner perceptions with empirical measurement and helps distinguish widely held assumptions from effects that appear in day-to-day work practice.

The remainder of this paper is structured as follows. Section 2 describes the study design, including the baseline survey and in-situ field study, instruments, and analysis approach. Section 3 reports findings for both research questions. Section 4 discusses implications for research and practice. Section 5 outlines limitations and validity considerations, and Section 6 discusses key related literature. Section 7 concludes the paper.

2 Study Design

The study comprises two sequential phases: (i) a practitioner survey that establishes baseline perceptions and quantifies meeting practices, and (ii) a two-week field study with a subset of survey respondents ($n = 3$) as well as additional participants, combining experience sampling with objective calendar data. Figure 1 provides a graphical overview of our study design. Together, the phases allow us to triangulate survey perceptions against self-reports and



■ **Figure 1** Research Design Overview.

objective calendar data. The replication package contains: (1) raw data for the baseline practitioner survey, (2) raw data from the field study, (3) the source code of the research prototype, and (4) analysis scripts [14].

2.1 Phase 1: Practitioner Survey

RQ1

What are software practitioners’ current meeting practices, and how do they perceive the impact of meetings on their productivity, flow, and well-being?

Instrument Design. We followed established guidelines for empirical survey design in software engineering [15, 6, 13]. Two authors drafted the instrument based on the research question, reviewed it internally, and piloted it with two external colleagues. Their feedback

led to revisions in wording and question ordering. The survey ran from late January to late February 2026. The instrument contains 33 questions across eight sections: eligibility and consent, current meeting practices, impact on daily work, reasons for declining or skipping, perceived impact on satisfaction and productivity, potential for measurement, demographics, and follow-up consent. Question types include single-select (16), multiple-select (4), Likert scale (9), one matrix item, and one open-ended item. Likert scales use 5-point ordinal formats, varying by construct (disruption, impact, or agreement). Participation required confirmation of recent meeting involvement and explicit informed consent. The survey was implemented in Google Forms.

Participants and Recruitment. We distributed the survey through professional networks and platforms, such as LinkedIn, targeting software practitioners with recent meeting experience. Participation was voluntary and anonymous. Respondents willing to join a follow-up field study were asked to provide contact details, forming the candidate pool for Phase 2. Sample demographics are reported in Section 3.

Analysis Approach. Given the exploratory nature, we report descriptive statistics throughout: frequencies and percentages for categorical items. For 1–5 field ratings we report medians and IQRs ($Q_3 - Q_1$). Associations among ordinal ratings and between meeting counts and daily ratings use Kendall’s τ_b . Where subgroup differences are discussed, they are interpreted as descriptive patterns rather than confirmatory hypothesis tests.

2.2 Phase 2: Field Study

RQ2

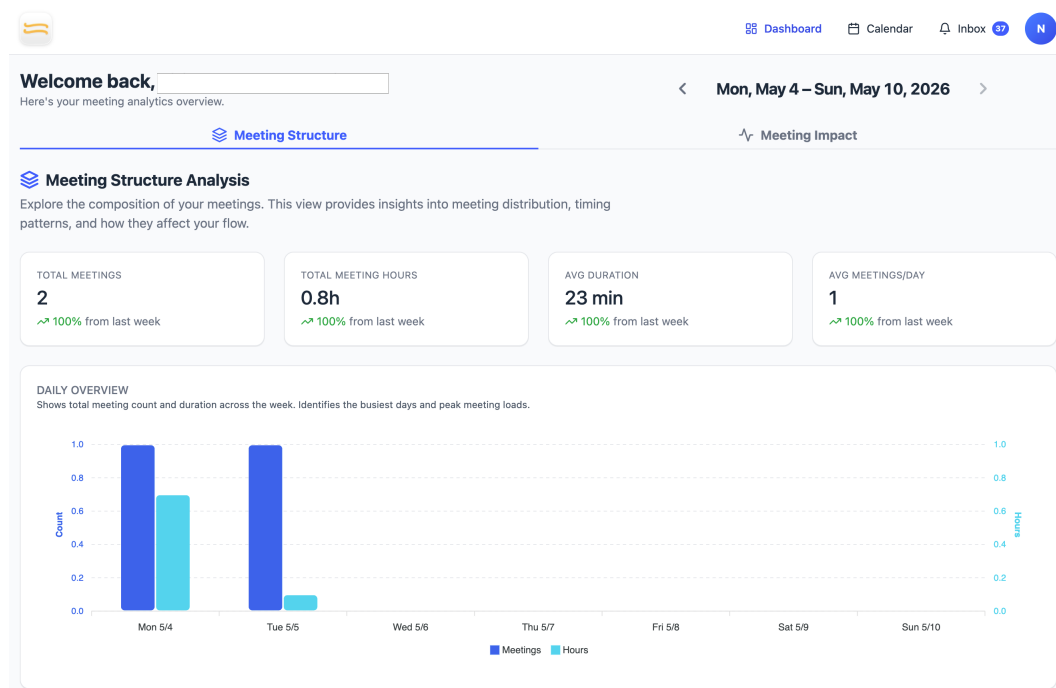
How do objective meeting characteristics – type, duration, timing, and clustering – relate to practitioners’ experience of productivity impact, flow disruption, and energy?

In Phase 2, we used an experience sampling approach [2], where participants reported immediately after meetings and once at the end of each workday. We separate productivity, flow, and well-being following SE work that treats perceived progress, deep focus, and affective state as related but non-identical constructs [11, 12, 3, 5, 4]. We operationalise *productivity* as perceived ability to make progress on work tasks on a given day; *flow* (deep focus) as the capacity for sustained concentration without interruption; and *well-being* as energy level and psychological state (stress, satisfaction) as affected by meetings. These three constructs are conceptually orthogonal: a planning meeting can be highly useful for the team while being highly disruptive and energy-draining for the individual; an ad-hoc sync can be both useful and lightly disruptive precisely because it is short. Several of the findings reported in Section 3 depend on this distinction.

Participants and Setting. We recruited participants via two channels: (1) Phase 1 respondents who volunteered for a follow-up, and (2) additional outreach through professional networks. All field-study participants underwent the same consent and onboarding procedure before participation. During the two study weeks, participants used their own devices and followed their normal work routines.

Research Prototype Implementation. To enable data collection, we developed a browser-based web application. The research prototype was motivated by the need to combine objective calendar data with subjective ratings in a single instrument, while imposing

minimal burden on participants. The application is implemented with Angular (frontend), a Spring Boot REST API (backend), a PostgreSQL database. Calendar data is ingested automatically through a Google Calendar OAuth 2.0 integration that synchronises events every ten minutes. After each meeting ends, a push notification prompts participants to complete a four-item feedback form covering focus disruption, overall feeling, energy, and perceived usefulness. A separate end-of-day notification, timed one hour before each participant's configured work end, triggers a four-item daily check-in on productivity, deep work, energy, and meeting load. All pending feedback items are also collected in an inbox, ensuring that missed notifications do not result in data loss. Two interactive dashboards allow participants to review their own data: one visualising meeting load (Figure 2), timing heatmaps, type distribution, and a computed daily flow-potential score; and another linking subjective ratings to meeting types and surfacing recurring positive and negative tags.



■ **Figure 2** Prototype tool – dashboard.

Field Study Procedure. Before the study began, participants completed a brief background questionnaire covering role, years of experience, organisation type and size, and work location. They then received an onboarding session (online or in-person) in which we demonstrated the prototype, and individual consent for calendar access and notifications was obtained.

Measurement Instruments. We collected data through four streams.

1. A lightweight feedback prompt was triggered after each calendar event ended. Participants rated: (1) *Usefulness* (“This meeting was worth my time?”; 1 = not at all – 5 = very useful); (2) *Energy impact* (“How did this meeting affect your energy?”; 1 = very draining – 5 = very energizing); and (3) *Flow disruption* (“Did this meeting disrupt your focus?”; 1 = not at all – 5 = very much). All meetings were sampled; no stratified or random subsampling was applied.

2. At the end of each workday, participants reported on four items: (1) *Daily productivity* (“How productive were you today?”; 1 = very unproductive – 5 = very productive); (2) *Daily deep focus* (“How much deep, focused work did you do today?”; 1 = none – 5 = a lot); (3) *Daily energy* (“How energized do you feel right now?”; 1 = very drained – 5 = very energized); and (4) an ordinal assessment of perceived meeting load (“How did today’s meeting load feel?”; 1 = too few – 5 = too many).
3. With participant consent, we extracted objective meeting characteristics from calendar metadata. Per-meeting variables include: start time and end time, duration in minutes, inter-meeting gap (minutes to the next meeting), back-to-back status (gap ≤ 5 min before or after), time-of-day category (early morning: 08:00–10:00; morning: 10:00–12:00; around lunch: 12:00–13:00; afternoon: 13:00–17:00; late afternoon: 17:00+), position in the day (1st, 2nd), and consecutive meeting block duration. From these, we derived daily aggregates: total meeting count, total meeting time in minutes, number of back-to-back pairs, longest consecutive meeting block, and meeting time as a percentage of the workday (assuming an 8-hour, 480-minute day).
4. At the end of the two weeks, participants completed an *Exit Survey* assessing response honesty, perceived behaviour change due to participation, and feedback burden. The exit survey additionally included an open-ended question on participants’ awareness of their own meeting practices, perceived changes in meeting behaviour during the study weeks, and usability of the research prototype. Open-ended responses were analysed manually through inductive thematic analysis; given the small participant count, no inter-rater reliability check was conducted.

3 Results

We report the samples in Subsection 3.1, Phase 1 baseline survey results in Subsection 3.2, the Phase 2 field-study findings in Subsection 3.3, and answers to RQ1 and RQ2 in Subsection 3.4.

3.1 Sample

Survey sample. We received 57 responses; one respondent reported no meeting participation and was ineligible, and one additional respondent declined consent, leaving $N = 55$ consented, eligible responses. The majority of respondents were from Europe (96.4%), while the rest (3.6%) were from Asia. The sample skewed young: 80% of respondents were aged 18–44 (18–29: 40.0%; 30–44: 40.0%; 45–60: 14.5%). The majority of respondents were Software Engineers/Developers (72.7%), followed by Team Lead *or* Engineering Manager (4.5%), and Product Owner *or* Product Manager (5.5%). Respondents were predominantly male (81.8%). Most respondents (72.7%) work in an IT department inside a larger, non-IT organisation – i.e. “embedded IT” in banks, whereas another (9.1%) work in a Software company *or* IT services provider. A majority (69.1%) work in Large companies or organisations with 250 or more employees.

Field-study sample. Eleven participants enrolled in the field study. We excluded five who submitted insufficient data. The remaining six participants form the *meeting-level analysis group*; of these, five additionally completed daily end-of-day check-ins and form the *full analysis group*. Roles included software developer (2), requirements engineer (1), product owner (1), and coach/developer (1), with professional experience ranging from 4 to 29 years. Participants were located in Austria (3; same organisation), Switzerland (2), and India (1).

Across the observation period, the six participants provided feedback on 135 meetings; the five-participant full-analysis subset contributed 112 meeting feedback records and 40 daily check-ins. We summarise observation periods and compliance rates in Table 1.

■ **Table 1** Field study participants included in analysis. Meeting compliance reflects the share of calendar events for which feedback was submitted. P8 did not configure a daily work schedule and submitted no daily check-ins; P10 did not configure a schedule but submitted 7 daily entries voluntarily.

ID	Observation period	Work days	Meetings	Daily Check-in	Mtg. compl.
<i>Full analysis group (meetings + daily, $n = 5$)</i>					
P1	02 Mar – 13 Mar 2026	10	43	10	100%
P2	02 Mar – 13 Mar 2026	10	19	10	95%
P3	03 Mar – 12 Mar 2026	6	15	5	100%
P9	17 Apr – 01 May 2026	11	16	8	94%
P10	21 Apr – 01 May 2026	–	19	7	100%
<i>Meeting-only group (additionally, $n = 1$)</i>					
P8	20 Apr – 30 Apr 2026	–	23	0	92%

Self-reported response quality. Six participants completed the exit survey. All six reported having *always* answered honestly throughout the study. Perceived burden was minimal, ranging from *not at all* to *moderately* burdensome. All six noticed patterns in how meetings affected them during the observation weeks, suggesting that the measurement instruments supported, rather than disrupted, natural reflection. No participants reported technical issues other than the calendar-sync suggestion captured in the quote below:

“This experiment was useful as a checkpoint as to what’s the state of these efforts. Especially looking back at last week made me reflect on which meetings were actually useful, [...]. Overall usability was good. Calendar sync would definitely be useful. [...]”

We treat all field-study estimates in Subsection 3.3 as descriptive and exploratory. With $n = 6$ participants providing meeting-level data ($n = 5$ providing daily data), we report direction and effect size; we include p -values for completeness, but do not treat them as confirmatory.

3.2 Baseline survey results: Reported meeting practices

Survey respondents reported attending a moderate number of meetings (typically 1–3 per day) across a familiar mix of types (stand-ups, planning, ad-hoc syncs), with mostly remote or hybrid attendance. We summarise the descriptive results in Table 2.

Perceived impact of meetings. Table 3 summarises how respondents perceive the impact of meetings on productivity, focus, and well-being. Two patterns stand out and recur in the field study: meetings are rated as useful and as costly by the same respondents – 63.6% perceive them as rather/very useful; 61.8% think that they moderately disrupt flow; while 61.8% need at least 15 minutes to recover focus after a meeting. Most respondents (76.4%) report lower productivity on high-meeting days. Two open responses illustrate reform wishes:

■ **Table 2** Self-reported meeting practices (survey, $N = 55$). Multi-select items (meeting types attended) do not sum to 100%. For “Most disruptive time of day”, smaller categories (early morning, around lunch, late afternoon, n/a) are omitted.

Category	<i>n</i> (%)	Category	<i>n</i> (%)	Category	<i>n</i> (%)
<i>Meetings per typical workday</i>		<i>Total daily meeting time</i>		<i>Modality</i>	
1 meeting	23 (41.8%)	<30 min	9 (16.4%)	Remote video	21 (38.2%)
2–3 meetings	24 (43.6%)	30 min–1 h	17 (30.9%)	Hybrid	21 (38.2%)
4–5 meetings	7 (12.7%)	1–2 h	17 (30.9%)	On-site	13 (23.6%)
>5 meetings	1 (1.8%)	2–4 h	10 (18.2%)		
		>4 h	2 (3.6%)		
<i>Meeting types (multi-select)</i>		<i>Largest share of meeting time</i>		<i>Most disruptive time of day</i>	
Stand-ups	43 (78.2%)	Planning	27 (49.1%)	Morning	12 (21.8%)
Planning	42 (76.4%)	Stand-ups	23 (41.8%)	Afternoon	12 (21.8%)
Ad-hoc syncs	38 (69.1%)	Ad-hoc	19 (34.5%)	No difference	14 (25.5%)

“Meetings should only cover essential topics. [...] I would prefer to participate in meetings only when I have something to contribute, not just to listen.”

“1. Clear agenda/goal, 2. Time-box strictly (15–25 min max), 3. Prep required (read doc before meeting)”

■ **Table 3** Survey perceptions of meeting impact, disruption drivers, and avoidance behaviour ($N = 55$).

<i>Perceived usefulness (%)</i>		<i>Task completion (%)</i>		<i>Decline frequency (%)</i>	
Rather/very useful	63.6	Positive	32.7	Never	18.2
Neutral	18.2	Neutral	38.2	Rarely	60.0
Not useful	16.4	Negative	29.1	Sometimes	18.2
				Often	3.6
<i>Top negative drivers (%)</i>				<i>Top decline reasons (%) (n = 26)</i>	
Unclear purpose	45.5	No clear action items	25.5	Not relevant	29.1
Excessive duration	30.9	Irrelevant/too many attend.	29.1	Deadline overlap	16.4
				Too frequent/long	9.1
				Already informed	7.3
				Focus/async (each)	5.5

3.3 Field-study findings

We organise our analysis around seven findings: (1) ad-hoc meetings were rated most useful; (2) planning meetings were rated most disruptive; (3) disruption increased across the day; (4) one-on-one meetings were the least disruptive and most energising; (5) back-to-back meetings amplified disruption and energy drain; (6) end-of-day productivity dropped as daily meeting count rose; (7) survey complaints and in-situ flags named the same drivers of bad meetings.

Finding 1. Ad-hoc meetings were rated most useful.

The 135 logged meetings spanned six types: *other* ($n = 55$, 40.7%), *stand-up* ($n = 28$, 20.7%), *planning* ($n = 24$, 17.8%), *ad-hoc* ($n = 19$, 14.1%), *one-on-one* ($n = 7$, 5.2%), and *retrospective* ($n = 2$, 1.5%). All type comparisons are descriptive. Meeting durations ranged from 15 to 210 minutes ($M = 47.1$ min, $Mdn = 30$ min, $SD = 31.6$ min). We show median ratings by type in Table 4. Ad-hoc meetings were among the most useful types ($Mdn = 4$, $IQR = 1$; Table 4), despite being unscheduled and the fourth-most-frequent type. Across all 135 meetings, usefulness was positively skewed ($Mdn = 4$, $IQR = 1$), with 58.5% scored 4 or 5. Post-meeting energy was similarly positive ($Mdn = 4$, $IQR = 1$): 51.1% scored 4 or 5, with a large neutral cluster at 3 (34.1%). Usefulness and energy were moderately associated (Kendall $\tau_b = .50$), meaning they are related but not the same judgment. The survey baseline matches the field-study direction (Table 3).

■ **Table 4** Median (IQR) per-meeting ratings by meeting type (meeting-level analysis group, $n = 135$). Scale: 1–5 for all items. Focus disruption: higher = more disruptive. ROTI and energy: higher = more positive. Retrospectives are included for completeness ($n = 2$) but should be interpreted with caution.

Type	n	ROTI		Energy after		Focus disrupt.	
		Mdn	IQR	Mdn	IQR	Mdn	IQR
Ad-hoc	19	4	1	4	1.5	3	2.5
One-on-one	7	4	2	4	0.5	1	1
Planning	24	4	1	3	2	4	2
Stand-up	28	3.5	1	4	1	2	1
Other	55	4	2	4	1	2	2.5
Retrospective	2	2.5	0.5	3.5	0.5	3	2

Finding 2. Planning meetings were rated most disruptive.

Planning meetings have the highest median focus disruption ($Mdn = 4$, $IQR = 2$; Table 4) and the lowest median post-meeting energy among non-retrospective types ($Mdn = 3$). Planning meetings remain useful ($Mdn = 4$), but their cost falls on focus and energy. Across all 135 meetings, focus disruption had $Mdn = 2$ ($IQR = 2$), with 52.6% scored 1 or 2 and 31.1% scored 4 or 5; self-reported post-meeting feeling was positive in 57.8%, neutral in 37.0%, and negative in 5.2% ($n = 7$). The survey baseline names the same pattern at the population level (Table 3): respondents report flow disruption, prolonged focus-recovery time, and perceived meeting waste.

Finding 3. Disruption increased across the day.

We analysed focus disruption by sequential position within each participant-day (Table 5). Median disruption rose from the 1st meeting ($Mdn = 2$) to the 5th ($Mdn = 4$), with a dip at the 4th ($Mdn = 2.5$); only three observations occur beyond the 5th meeting.

Short meetings (≤ 30 min) and meetings longer than an hour had lower median disruption ($Mdn = 2$) than the default 46–60 min slot ($Mdn = 3$). A small 31–45 min bin ($n = 7$) had $Mdn = 4$ and should not be over-interpreted. One possible explanation is that very short

■ **Table 5** Median (IQR) focus disruption by meeting position within the workday (meeting-level analysis group; positions 6+ omitted, $n = 3$).

Position	1st	2nd	3rd	4th	5th
n	49	37	27	12	7
Mdn	2	3	3	2.5	4
IQR	2	3	2	1.25	1.5

meetings feel cognitively bounded – participants can anticipate the end and return quickly to their prior task – while meetings longer than an hour are typically planned as the primary activity of a time block, so the cognitive transition cost is expected and absorbed. The survey baseline is less unanimous on time-of-day: morning and afternoon were cited as the most disruptive period at equal rates (Table 3). Position in the day is correlated with how many meetings already occurred. The rise in disruption may therefore reflect cumulative meeting load rather than clock time alone; our design cannot fully separate these. Finding 6 speaks to daily load at the person level, but a joint model of position $\times$ prior count was out of scope given n .

Finding 4. One-on-one meetings were the least disruptive and most energising.

One-on-one meetings have the lowest median focus disruption ($Mdn = 1$, $IQR = 1$; Table 4) and high post-meeting energy ($Mdn = 4$). Their usefulness median matches ad-hoc and planning ($Mdn = 4$), but at $n = 7$ they are the rarest non-retrospective type.

Finding 5. Back-to-back meetings amplified disruption and energy drain.

Our ≤ 5 min gap definition yields 55 back-to-back and 80 non-back-to-back meetings. Median focus disruption is higher for back-to-back meetings ($Mdn = 3$, $IQR = 2$) than for non-back-to-back ($Mdn = 2$, $IQR = 3$). Risk-based comparisons sharpen the contrast: high disruption (scores 4–5) occurred in 39.2% of back-to-back meetings versus 27.9% of non-back-to-back meetings ($RR = 1.41$, $OR = 1.67$), and draining energy ratings (1–2) in 21.6% versus 13.1% ($RR = 1.65$, $OR = 1.82$). Clustering amplifies both the focus cost and the energy cost.

Finding 6. End-of-day productivity dropped as daily meeting count rose.

Survey respondents reported that productivity on high-meeting days is lower than on low-meeting days (Table 3); no other survey item shows a comparable one-sided distribution. In the field study, we grouped days by meeting count (0–1, 2–3, 4+) to make the daily pattern interpretable (Table 6). The gradient holds for productivity and deep work; energy is lowest on 4+ meeting days (Table 6).

■ **Table 6** Median end-of-day ratings by daily meeting count (full-analysis group, $n = 5$ participants over 40 participant-days). Scale: 1–5 for all items; higher = more positive.

Meetings per day	Productivity	Deep work	Energy
0–1	4.0	3.5	3.5
2–3	3.5	2.5	4.0
4+	3.0	2.0	2.0

Day-level Kendall associations with meeting count were weakly negative (productivity $\tau_b = -.15$; deep work $\tau_b = -.15$; energy $\tau_b = -.21$). The negative pattern is clearer between participants (productivity $\tau_b = -.20$; deep work $\tau_b = -.40$; energy $\tau_b = -.40$) than within person, where average τ_b was near zero. Perceived meeting load tracked actual meeting count ($\tau_b = .36$) and total meeting time ($\tau_b = .27$).

Finding 7. Survey complaints and in-situ flags named the same drivers of bad meetings.

Participants flagged at least one *positive* aspect in 58.5% of meetings ($n = 79$), most commonly *clear purpose* ($n = 42$), *clarified next steps* ($n = 40$), and *provided useful information* ($n = 33$). Issues were flagged in 20.0% of meetings ($n = 27$), most often *too long*, *no relevance to my tasks*, and *could have been an email* (each $n = 8$). We summarise in Table 7 how the presence of each flag shifts ratings relative to meetings without that flag. Issue flags lower usefulness and energy and raise disruption; positive flags do the opposite. The flag *no clear outcome* is the strongest single negative on usefulness. The factors named by survey respondents as the top frustrations (Table 3) match the field-study flag effects: unclear purpose, excessive duration, irrelevant attendees, and missing action items – the same four factors that the issue flags *no clear outcome*, *too long*, and *no relevance* capture in the in-situ data.

■ **Table 7** Per-meeting rating differences associated with flag presence (Δ = mean for meetings carrying the flag minus mean for meetings without the flag; 1–5 scale). Disruption deltas for the two positive flags were not reported in the source data.

Flag	n	Δ ROTI	Δ Energy	Δ Disruption
<i>Issue flags</i>				
too long	7	−1.77	−1.49	+0.84
no relevance	8	−1.64	−0.85	+1.19
no clear outcome	4	−2.35	−1.07	+0.11
<i>Positive flags</i>				
clear purpose	40	+1.49	+0.68	–
clarified next steps	33	+1.33	+0.66	–

3.4 Summary

Three findings stand out across the two study phases. First, the cost of a meeting is not a single number. Usefulness, focus disruption, and post-meeting energy are three different facets, and a meeting can score high on one and low on another (F1, F2, F4). Treating “too

many meetings” as a single problem overlooks the nuances behind these trade-offs. Second, when and how meetings are scheduled matters. Focus disruption rises with each successive meeting in the day, with back-to-back scheduling, and with meetings booked into the default 60-minute slot (F3, F5). Third, what practitioners say makes meetings good or bad aligns with what we see in real time. Clear purpose and clear next steps appear both as the top survey reform requests and as the strongest positive in-situ flags. On the other hand, unclear purpose, excessive duration, and irrelevant attendees appear both as the top survey frustrations and as the largest negative in-situ tag effects (F7).

Respondents were nearly unanimous that productivity is lower on high-meeting days (76.4%). The field data trend in the same direction at the group level, but within-person – comparing each participant’s own busier and quieter days – the signal was near zero. The group-level pattern reflects who carries heavier schedules, not how individuals respond when their own load increases. We return to this in the Discussion.

Answer to RQ1. Software practitioners attend 1–3 meetings on a typical workday, drawn from a familiar mix of stand-ups, planning sessions, and ad-hoc syncs in mostly remote or hybrid format. They rate these meetings as moderately useful (63.6%) but costly: 61.8% report at least moderate flow disruption, 61.8% need ≥ 15 minutes to refocus afterward, and 76.4% feel less productive on high-meeting days. The complaints behind this negative experience are specific and recurring: unclear purpose, excessive duration, too many or irrelevant attendees, and weak action items.

Answer to RQ2. The same picture appears in real time, with scheduling factors explaining most of the variation. Meeting *type* produces the largest contrasts: ad-hoc syncs are rated most useful, planning sessions most disruptive and draining, one-on-ones appeared least disruptive in our data ($n=7$). Disruption rises with *position* in the day and with *back-to-back* scheduling. *Duration* is non-monotonic, with the default 46–60 minute slot as the disruption peak. *Daily meeting count* trends in the expected negative direction with end-of-day outcomes, but does not reach statistical significance in our small sample. The four survey complaints from RQ1 – unclear purpose, excessive duration, low attendee relevance, weak actionability – also appear as the strongest negative tags in the in-situ data, providing the cross-phase confirmation.

4 Discussion

Meeting experience breaks down into two independent dimensions: perceived usefulness and affective impact on focus and energy. These do not reduce to a single judgment: a meeting can be worthwhile and still leave a participant drained, or feel unproblematic and still consume time without return. One participant captured this in the exit survey: “*That meetings can keep me with good energy but still not be worth my time.*” Planning meetings illustrate the point: rated as useful, yet the most disruptive and draining type observed. The implications below are structured around both dimensions.

4.1 Implications For Practitioners

Cutting meetings does not always help; the problem is often how a meeting is run, not that it happens. A useful-but-draining meeting is not a candidate for removal – it is a candidate for redesign. For high-value, high-drain meetings such as planning sessions, the practical solutions are shorter time boxes, stricter agendas, and scheduling them to end near

a natural break rather than mid-block. The field study identifies the indicators that work: clear purpose, defined next steps, and durations matched to the agenda rather than to the default calendar slot.

Short, on-demand syncs were rated most useful in our data. Policies that suppress ad-hoc communication to reduce interruptions may remove a productive coordination channel and push the same conversations into longer scheduled meetings.

Time of day was not a consistent predictor in either phase. What mattered was sequence position, back-to-back scheduling, and duration. Three scheduling patterns are promising design hypotheses for teams to trial and for future studies to validate:

- **Later-position meetings.** Disruption climbed from the first to the fifth meeting of the day; consider limiting meetings stacked late in a person's meeting sequence.
- **Back-to-back clustering.** Gaps of five minutes or less were associated with higher disruption and more draining energy ratings; consider leaving a recoverable gap between meetings.
- **Default 46–60 minute slots.** Short meetings and those longer than an hour were both less disruptive; consider avoiding the default hour slot when a shorter one will do.

Most respondents reported needing at least 15 minutes to regain focus after a meeting. A day with four 30-minute meetings therefore carries not only 120 minutes of meeting time but up to an hour of recovery – and recovery compounds when meetings are clustered.

The most common reason for declining a meeting was irrelevance to the participant's tasks. Stating explicitly who is a required participant and who is optional shifts that relevance judgment from social pressure to meeting design, and directly addresses the most frequently cited source of wasted meeting time.

The daily-load finding carries an additional organisational implication. The negative relationship between meeting count and end-of-day outcomes was driven by differences between participants, not by day-to-day variation within the same person. The productivity cost of heavy meeting schedules appears concentrated in specific roles, not distributed across the team on busy days. Targeted load reduction for those roles is a more precise intervention than blanket meeting-reduction policies. These implications rest on a small Phase-2 sample ($n=6$ / $n=5$) and single-item measures of productivity, flow, and well-being; they should be read as exploratory.

4.2 Implications For Researchers

Meeting-triggered prompts were answered at high rates by every participant who completed onboarding; end-of-day check-ins were frequently missed. The meeting acts as a natural reminder, and the prompt arrives close in time to the experience being rated. Future field studies should anchor prompts to observable events – meetings, task completions, build failures – rather than fixed clock times.

The study contributes a validated event-triggered ESM instrument for meeting research: calendar-synced prompts, per-meeting ratings on three constructs, and a daily check-in, implemented as a deployable web application. Dropout was entirely among participants who never completed onboarding; every participant who did reached high compliance. The instrument and replication package are available for reuse. For field studies of this kind, live onboarding is not optional – it determines how much usable data the study collects.

The between/within-person decomposition in F6 illustrates a methodological risk common in field studies. A between-person pattern – participants with more meetings report lower productivity – resembles a causal daily effect but is not one: the same participants showed

no such pattern on their own busier days. Field studies on meeting load, interruptions, or communication overhead should report both levels and treat between-person patterns as descriptive context rather than evidence of within-person mechanisms.

Treating each meeting as its own unit of analysis – with post-meeting ratings and calendar metadata for type, duration, position, back-to-back status, and time of day – surfaces variation that meeting-count metrics miss. Studies focused on developer productivity, flow, or well-being more broadly can benefit from treating individual meetings as a unit, since meeting structure explains variance that aggregate load metrics do not capture.

Longer observation windows (four to six weeks) would allow within-person analyses to stabilize and capture sprint-cycle variation. Role-stratified samples are needed before meeting-design recommendations can be made role-specific: managers and individual contributors likely experience planning sessions and back-to-back schedules differently.

Single Likert items kept participant burden low but cannot separate construct components (e.g., speed versus quality for productivity). Longer studies can incorporate short multi-item scales and, where feasible, behavioural traces such as IDE telemetry as an objective check on self-report.

5 Threats to Validity

5.1 Construct Validity

Construct validity concerns whether the operationalizations of our key constructs—productivity, flow (deep focus), and well-being—truly reflect the underlying phenomena of interest.

Key constructs are operationalized with single Likert-type items in both the survey and field study. Single-item measures are vulnerable to random response error, and they prevent internal-consistency estimation. They also compress multidimensional constructs: productivity (e.g., speed vs. quality), flow (depth vs. continuity), and well-being (energy vs. affect) cannot be separated analytically. We mitigated this by grounding items in prior SE and HCI operationalizations and by providing explicit construct definitions during survey completion and app onboarding. Convergence between survey and field responses provides only partial support for construct validity.

We treat calendar event data such as duration, start/end times, and computed gap-based back-to-back status as objective proxies for meeting characteristics. Calendar data may not reflect actual attendance, late starts, or early departures; a meeting recorded as 60 minutes may have lasted 45 minutes in practice. Compliance rates suggest that participants generally engaged with the meetings for which feedback was submitted, but scheduled duration remains an approximation.

Meeting-type labels were AI-assisted because many participants could not share exact titles or detailed descriptions for privacy and organisational confidentiality reasons and instead used masked or custom titles. This improves feasibility in realistic workplace settings but introduces potential misclassification risk for meeting types. Any type-specific contrasts (e.g., between planning and ad-hoc meetings) should therefore be interpreted cautiously.

Both the survey and field study instruments use ordinal Likert-type scales. We summarise 1–5 ratings with medians and IQRs and report associations with Kendall's τ_b .

5.2 Internal Validity

Internal validity concerns whether observed relationships reflect true causal or associative effects, or are confounded by other factors.

Participants were aware of being observed and were asked to reflect on each meeting immediately after it ended. This may have altered their meeting behaviour (e.g., being more selective about which meetings to attend, or being more consciously critical). We measured reactivity in the exit survey; most respondents reported that participation changed their behaviour *not at all* or only *a little*. However, at least one open-text response indicates active behavioural adaptation (“This experiment was useful as a checkpoint...”), which we acknowledge as a limitation.

Participants knew the study focused on the impact of meetings on productivity. This awareness may have biased some ratings toward confirming the expected direction (i.e., that meetings are disruptive). We cannot fully rule this out, but the open-ended exit responses include both positive and negative reflections, which somewhat mitigates this concern.

Several key relationships rely on self-reports from the same individuals (e.g., usefulness, disruption, energy, daily productivity, and deep focus), which may introduce common-method variance. Transient affect, response style, or demand characteristics could inflate associations among subjective measures beyond what would be observed if different data sources were used. In addition, social-desirability pressures may bias some responses, for example, by discouraging participants from labelling recurring or management-led meetings as low-value, or from reporting persistently low productivity.

Only four participants provided both meeting-level and daily check-in data suitable for multilevel analysis. Two additional participants provided meeting-level data only. This limits daily-level and within-person inference; we therefore interpret these analyses cautiously and as descriptive rather than confirmatory. At the daily level, many factors beyond meetings influence productivity, focus, and energy – personal circumstances, task complexity, deadlines, health, and team dynamics. We do not control for these, and they represent unobserved heterogeneity in the daily check-in data.

5.3 External Validity

External validity concerns the generalizability of our findings beyond the specific sample and setting of this study.

Survey participants were recruited through convenience channels (e.g., professional networks, social media, research listservs). Respondents who chose to participate may systematically differ from non-participants, for example by having stronger opinions about meetings or more available time to complete a survey. This limits the representativeness of our sample and may bias estimates of meeting frequency, usefulness, and perceived impact.

The field study participants were geographically concentrated (Switzerland, Austria, India) and skewed toward senior roles (associate director, product owner, lead researcher, coaches/developers with 10–29 years of experience). Findings, particularly around meeting autonomy, the ability to decline meetings, and the types of meetings experienced, may not generalize to junior practitioners or to practitioners in regions with different organisational cultures. Organisational and national culture may also shape meeting norms (e.g., whether it is acceptable to decline invitations or to voice dissatisfaction about meeting practices), further limiting transferability.

Participants were asked to use the tool across two calendar weeks (i.e., roughly two working weeks), but usable observations covered only a subset of days within that window for some participants. This period may still be unrepresentative of longer-term meeting patterns, which can vary across sprint phases, project milestones, or organisational cycles. Longer longitudinal windows would be needed to assess temporal stability of the observed patterns.

With $n = 6$ participants providing meeting-level data and $n = 4$ providing full meeting+daily data, person-level and daily-level results are based on very few individuals. Effect sizes and directions should therefore be treated as indicative rather than definitive, and replication with a larger, more diverse sample is needed before strong conclusions can be drawn.

5.4 Conclusion Validity

Conclusion validity concerns whether the statistical procedures and their interpretation are sound, given the study's design and sample sizes.

The field study's small n substantially limits statistical power for daily-level multilevel models and person-level correlations. We may fail to detect real effects (Type II error), and effect sizes estimated from small samples tend to be inflated. Where applicable, we report effect sizes alongside significance and flag field study statistical results as descriptive and directional.

Given the sample sizes, we treat field study quantitative findings as illustrative rather than conclusive. We use agreement between survey perceptions and field-study real-time data as convergent evidence across methods. However, with $n = 5$ usable full-data participants for person-level alignment, this comparison is illustrative rather than statistically conclusive.

6 Related Work

6.1 Meetings in Software Engineering and Knowledge Work

Prior work in organisational psychology links meeting load to well-being and fatigue [18, 10], while software engineering studies document that meetings account for a meaningful share of developers' workdays and that planned meetings are associated with lower perceived productivity [11]. In parallel, research on stand-up meetings has clarified their communicative and coordination functions [23, 21], with recent work further linking daily stand-up practices to team satisfaction and performance perceptions [17]. Interruption and recovery research has shown that communication demands fragment developers' time [7] and that transitioning back to focused work after a meeting carries real cognitive cost [1]. This literature establishes that meetings are both necessary coordination mechanisms and potential sources of cognitive overhead. Yet most studies model meeting exposure at the aggregate level (e.g., total meeting time/frequency) or rely on retrospective assessments of quality. Even when identifying strong design factors for effectiveness [9, 20], prior work rarely links objective scheduling characteristics (timing, duration, clustering, sequence) to real-time, within-person experience.

6.2 Productivity, Flow, and Well-Being in Software Engineering

Software engineering research treats productivity as multidimensional and strongly shaped by interruptions and context switching rather than output volume alone [11, 12, 3], with affective state emerging as a further independent influence on performance and process outcomes [5, 4]. This body of work also supports separating perceived performance from experienced cognitive state: developers may report progress and strain simultaneously, and productivity is in large part a personal matter [11, 4]. At the team level, individual productivity perception has been found to predict team productivity more strongly than time spent in meetings or amount of team interaction [19], suggesting that aggregate meeting metrics are a poor proxy for team-level outcomes. This framing motivates our construct space (productivity, flow, well-being)

and our choice to observe it repeatedly in situ. If these constructs are context-sensitive and person-dependent, single retrospective judgments risk collapsing meaningful within-person variation across days and meeting contexts.

6.3 Experience Sampling and Field Methods in Software Engineering

Experience sampling has become a practical way to reduce recall bias in software engineering field research and has been successfully combined with behavioural logs and instrumentation, demonstrating both the feasibility of repeated ecological measurement and the value of pairing subjective reports with objective traces [8, 24]. However, prior field work typically treats meetings as one activity class among many or focuses on interruptions broadly. Our study extends this line by instrumenting the individual meeting as the unit of analysis and combining event-contingent ratings with calendar-derived objective features (timing, duration, and clustering), enabling finer-grained links between meeting structure and real-time experience.

7 Conclusion

This study combined a practitioner survey with a two-week field study to examine how meetings shape productivity, focus, and energy in software engineering. The two phases paint a consistent picture. Meetings carry two costs that do not always move together: whether a meeting was worth attending, and what it did to focus and energy. The association between daily meeting count and end-of-day outcomes was between-person, not evidence that adding meetings on a given day reduces that person's productivity. Meeting type, position in the day, back-to-back scheduling, and duration shape those costs more than meeting volume alone. The complaints most often raised in the survey – unclear purpose, excessive duration, irrelevant attendees, and weak actionability – reappear as the strongest negative signals in real-time meeting ratings.

This study shifts the conversation from broad dissatisfaction with meetings toward candidate meeting patterns and design hypotheses for teams to trial, pending larger confirmation. The findings suggest that not all meetings are equally costly or beneficial, and that purpose clarity, duration, and scheduling context may matter for productivity, focus, and energy. We frame observed meeting patterns as design hypotheses pending larger confirmation, not as prescriptions.

Data Availability Statement

An anonymised replication package containing the baseline survey and field study data, prototype source code, and analysis scripts supporting this study is available via Figshare [14].

References

- 1 Joseph A. Allen, Matt S. Thiese, Emilee Eden, and Sarah E. Knowles. Why am i so exhausted? exploring meeting-to-work transition time and recovery from virtual meeting fatigue. *Journal of Occupational and Environmental Medicine*, 64(12):1053–1058, 2022. doi:10.1097/JOM.0000000000002641.
- 2 Mihaly Csikszentmihalyi and Reed Larson. Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, 175(9):526–536, 1987. doi:10.1097/00005053-198709000-00004.

- 3 Nicole Forsgren, Margaret-Anne Storey, Chandra Maddila, Thomas Zimmermann, Brian Houck, and Jenna Butler. The SPACE of developer productivity: There's more to it than you think. *Queue*, 19(1):20–48, 2021. doi:10.1145/3454122.3454124.
- 4 Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. What happens when software developers are (un)happy. *Journal of Systems and Software*, 140:32–47, 2018. doi:10.1016/j.jss.2018.02.041.
- 5 Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. Happy software developers solve problems better: Psychological measurements in empirical software engineering. *PeerJ*, 2:e289, 2014. doi:10.7717/peerj.289.
- 6 Barbara A. Kitchenham and Shari L. Pfleeger. Personal opinion surveys. In *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer, 2008. doi:10.1007/978-1-84800-044-5_3.
- 7 Amy J. Ko, Robert DeLine, and Gina Venolia. Information needs in collocated software development teams. In *Proceedings of the 29th International Conference on Software Engineering (ICSE'07)*, pages 344–353. IEEE, 2007. doi:10.1109/ICSE.2007.45.
- 8 Miikka Kuuttila, Mika V. Mäntylä, Maëlick Claes, Marko Elovainio, and Bram Adams. Using experience sampling to link software repositories with emotions and work well-being. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 1–10, 2018. doi:10.1145/3239235.3239245.
- 9 Desmond J. Leach, Steven G. Rogelberg, Peter B. Warr, and Jennifer L. Burnfield. Perceived meeting effectiveness: The role of design characteristics. *Journal of Business and Psychology*, 24(1):65–76, 2009. doi:10.1007/s10869-009-9092-6.
- 10 Alexandra Luong and Steven G. Rogelberg. Meetings and more meetings: The relationship between meeting load and the daily well-being of employees. *Group Dynamics: Theory, Research, and Practice*, 9(1):58–67, 2005. doi:10.1037/1089-2699.9.1.58.
- 11 André N. Meyer, Laura E. Barton, Gail C. Murphy, Thomas Zimmermann, and Thomas Fritz. The work life of developers: Activities, switches and perceived productivity. *IEEE Transactions on Software Engineering*, 43(12):1178–1193, 2017. doi:10.1109/TSE.2017.2656886.
- 12 André N. Meyer, Thomas Zimmermann, and Thomas Fritz. Characterizing software developers by perceptions of productivity. In *Proceedings of the 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 105–110. IEEE, 2017. doi:10.1109/ESEM.2017.17.
- 13 Jefferson Seide Molléri, Kai Petersen, and Emilia Mendes. Survey guidelines in software engineering: An annotated review. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 1–6, 2016. doi:10.1145/2961111.2962619.
- 14 Nitish Patkar, Xeno Isenegger, Gideon Monterosa, Sebastiano Panichella, and Norbert Seyff. Replication package for “characterising meeting practices in software engineering teams”. figshare, 2026. URL: <https://figshare.com/s/453a1f1a5c920066393f>.
- 15 Michael Quinn Patton. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Sage Publications, 2014.
- 16 Minna Pikkarainen, Jukka Haikara, Outi Salo, Pekka Abrahamsson, and Jari Still. The impact of agile practices on communication in software development. *Empirical Software Engineering*, 13(3):303–337, 2008. doi:10.1007/s10664-008-9065-9.
- 17 Sarah Rietze and Hannes Zacher. Relations between daily stand-up meetings, work satisfaction, and team performance perceptions: The role of psychological safety. *European Journal of Work and Organizational Psychology*, 34(5):565–582, 2025. doi:10.1080/1359432X.2025.2508178.
- 18 Steven G. Rogelberg, Desmond J. Leach, Peter B. Warr, and Jennifer L. Burnfield. “not another meeting!” are meeting time demands related to employee well-being? *Journal of Applied Psychology*, 91(1):83–96, 2006. doi:10.1037/0021-9010.91.1.83.

- 19 Anastasia Ruvimova, Alexander Lill, Jan Gugler, Lauren Howe, Elaine Huang, Gail Murphy, and Thomas Fritz. An exploratory study of productivity perceptions in software teams. In *Proceedings of the 44th International Conference on Software Engineering*, pages 99–111, 2022. doi:10.1145/3510003.3510081.
- 20 Linda R. Shanock, Joseph A. Allen, Alexandra M. Dunn, Benjamin E. Baran, Cliff W. Scott, and Steven G. Rogelberg. Less acting, more doing: How surface acting relates to perceived meeting effectiveness and other employee outcomes. *Journal of Occupational and Organizational Psychology*, 86(4):457–476, 2013. doi:10.1111/joop.12037.
- 21 Adriana Meza Soria, Taylor Lopez, Elizabeth Seero, Negin Mashhadi, Emily Evans, Janet Burge, and André van der Hoek. Characterizing software maintenance meetings: Information shared, discussion outcomes, and information captured. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, pages 1–13, 2024. doi:10.1145/3597503.3623330.
- 22 Viktoria Stray, Nils Brede Moe, and Gunnar R. Bergersen. Are daily stand-up meetings valuable? a survey of developers in software teams. In *Proceedings of the International Conference on Agile Software Development*, pages 274–281. Springer, 2017. doi:10.1007/978-3-319-57633-6_20.
- 23 Viktoria Stray, Dag I. K. Sjøberg, and Tore Dybå. The daily stand-up meeting: A grounded theory study. *Journal of Systems and Software*, 114:101–124, 2016. doi:10.1016/j.jss.2016.01.004.
- 24 Manuela Züger, Christopher Corley, André N. Meyer, Boyang Li, Thomas Fritz, David Shepherd, Vinay Augustine, Patrick Francis, Nicholas Kraft, and Will Snipes. Reducing interruptions at work: A large-scale field study of flowlight. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 61–72, 2017. doi:10.1145/3025453.3025662.