The ebb and flow of job engagement: Engagement variability and emotional stability as interactive predictors of job performance.

Article *in* Journal of Applied Psychology · October 2023

DOI: 10.1037/apl0001129

3 authors, including:

Basima Tewfik
Massachusetts Institute of Technology
9 PUBLICATIONS **31** CITATIONS

Journal of Applied Psychology

The Ebb and Flow of Job Engagement: Engagement Variability and Emotional Stability as Interactive Predictors of Job Performance

Basima A. Tewfik, Daniel Kim, and Shefali V. Patil

Online First Publication, October 5, 2023. https://dx.doi.org/10.1037/apl0001129

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Applied Psychology

ISSN: 0021-9010

https://doi.org/10.1037/apl0001129

The Ebb and Flow of Job Engagement: Engagement Variability and Emotional Stability as Interactive Predictors of Job Performance

Basima A. Tewfik[1], Daniel Kim[2], and Shefali V. Patil[3]

[1] Work and Organization Studies, Sloan School of Management, Massachusetts Institute of Technology

[2] Department of Management, Warrington College of Business, University of Florida

[3] Department of Management, McCombs School of Business, University of Texas at Austin

Scholars have long recognized that employees often ebb and flow in how engaged they are in their jobs—what we term "engagement variability." Yet, to date, we have little insight into how an employee's engagement variability—that is, the degree of inconsistency in their engagement—affects job performance. Drawing on and extending habit theory, we hypothesize that, controlling for average engagement, engagement variability is negatively related to job performance. We further hypothesize that emotional stability moderates this relationship: Although engagement variability hinders performance when an employee is higher in emotional stability, this effect weakens when an employee is lower in emotional stability. Finally, we hypothesize that flow mediates the interactive effect of engagement variability and emotional stability on performance. We test our hypotheses across three studies: a multisource, ten-wave field study of 160 cadets across three Army and Air Force divisions of the Reserve Officer Training Corps, an experiment with 600 full-time employees, and a multisource, two-week experience sampling study with 152 full-time employees and their supervisors. We find consistent support for engagement variability's negative relationship with performance and the moderating role of emotional stability, but mixed support for the mediating role of flow. We conclude by discussing the implications of our work.

Keywords: work engagement, job performance, personality, neuroticism

The study of job engagement—or the simultaneous investment of an employee's cognitive, emotional, and physical resources into their work role—has long been a dominant focus among organizational scholars (e.g., Christian et al., 2011; Crawford et al., 2010; Macey & Schneider, 2008; Maslach & Leiter, 2008; Matta et al., 2017; Sonnentag et al., 2012). Over the past two decades, scholars have examined key correlates of engagement (e.g., Sarwar & Abugre, 2013; Schaufeli & Salanova, 2007), how engagement emerges (e.g., Christian et al., 2011), and how engagement shapes important work outcomes (e.g., Hakanen et al., 2006; Rich et al., 2010). Of note, a burgeoning line of work has shown that job engagement uniquely predicts job performance: Those who are highly engaged in general also perform well on the job (Christian et al., 2011; Rich et al., 2010).

Despite tremendous progress, our understanding of how engagement affects performance remains limited, given that individuals not only differ from each other in their average level of engagement in a role, but also in how consistently or inconsistently they engage in their role—which likely also affects performance. Indeed, engagement, as originally theorized, not only reflects the simultaneous investment of one's cognitive, emotional, and physical resources into one's work role performance but also the consistency of such investments over time (Kahn, 1990, 1992). For example, consider two employees who, on average, display the same medium level of engagement, aggregated from their moments of engagement. Whereas one employee (the "low variability"/"more consistently engaged" employee) may reach this medium level of engagement by displaying that same level of engagement over time, the second employee (the "high variability"/"more inconsistently engaged" employee) may reach this medium level by switching between very high and very low levels. As a result of these differing patterns in engagement over time, these two employees may differ drastically in their subsequent performance.

To date, few insights exist regarding how inconsistency in engagement—or what we term "engagement variability"—impacts performance. Instead, much of the past research has generatively examined employees' general engagement over time or their momentary engagement at specific points in time (e.g., Parke et al.,

Basima A. Tewfik https://orcid.org/0000-0001-9750-3736

The authors are deeply appreciative of the helpful feedback provided by Matthew Bidwell, Jared Curhan, Noah Eisenkraft, Trevor Foulk, Adam Grant, Erin Kelly, Kate Kellogg, Anthony Klotz, Klodiana Lanaj, Michael Parke, Jessica Rodell, Nancy Rothbard, Tyler Sabey, Phil Tetlock, Mo Wang, and Michele Williams. They are also very grateful for the beneficial analytical insights from Joel Koopman and Kristopher Preacher. For her invaluable research assistance, they wish to thank Cara Krupnikoff-Salkin. They also thank Jennifer Duan, Alexis Kim, and Cindy Luo for their assistance on different parts of the project.

An earlier version of this article was presented at the 2017 Annual Academy of Management Meeting and the 2018 International Association for Conflict Management Conference. All analytical output and appendices are available on the Open Science Framework at https://osf.io/7r9mu/?view_only=ee1a674612b948b281f4c5564dbd7540.

Correspondence concerning this article should be addressed to Basima A. Tewfik, Work and Organization Studies, Sloan School of Management, Massachusetts Institute of Technology, 100 Main Street, Cambridge, MA 02142, United States. Email: btewfik@mit.edu

2018; Rich et al., 2010; Vogel et al., 2022; Wang et al., 2019). Without consideration of engagement variability, however, scholars miss developing a holistic understanding of the engagement–performance relationship given that variability constructs are crucial for extending theory and empirical work; in essence, these constructs increase the power scholars have at their disposal for predicting individual behavior like performance (Fleeson & Leicht, 2006). Indeed, because engagement variability, as theorized, captures between-person differences in the consistency of momentary engagement over time, it offers a way to bridge existing within- and between-person perspectives (Dalal et al., 2008), thereby augmenting our understanding of how engagement more comprehensively operates in the workplace. Finally, in an age in which questions around job engagement among scholars and practitioners alike have evolved to consider how to elicit consistent engagement among employees (Knight, Patterson, & Dawson, 2017; Rothbard & Patil, 2013; Yohn, 2019), developing such insight is particularly timely.

Accordingly, in this article, we develop a theory of how engagement variability affects job performance. To build our theoretical model, we turn to habit theory (Ouellette & Wood, 1998; Rothman et al., 2011; Verplanken, 2006; Wood & Rünger, 2016). Habit theory is a particularly relevant framework because it explains the consequences of consistent, or repeated, allocation of personal resources—and engagement variability, at its core, involves the (in)consistent allocation of resources over time. According to habit theory, those individuals who are used to allocating resources in a consistent manner experience efficiencies in the form of greater automaticity, which can have a positive effect on downstream outcomes like performance (Ouellette & Wood, 1998; Rothman et al., 2011). Thus, building on this logic, we first propose that, controlling for average engagement, engagement variability should be negatively related to job performance because it reflects the inconsistent investment of resources. Extending habit theory, we further suggest that emotional stability—an individual difference alluded to but not explicitly discussed in habit theory (Verplanken, 2018)—may affect the relationship between engagement variability and performance. When an employee is higher in emotional stability, there should be a negative relationship between engagement variability and performance because higher emotional stability does not habituate people to fluctuations in resource allocation. But, when an employee is lower in emotional stability, this relationship should weaken because those lower, as opposed to higher, in emotional stability are more accustomed to shifting their resources in an inconsistent manner (Wallace & Newman, 1998). Finally, we hypothesize that the interactive effect of engagement variability and emotional stability on job performance is mediated by flow—a state associated with habituation in which people feel like what they are doing is "almost automatic" (Csikszentmihalyi, 1990, p. 53).

We test our hypotheses across three studies. In Study 1, a multisource, ten-wave field study of 160 cadets across three Army and Air Force divisions of the Reserve Officer Training Corps (ROTC), we examine the negative effect of engagement variability on performance and the moderating role of emotional stability, controlling for average engagement. In this study, we also rule out alternative explanations, such as workload variability, and substantiate that engagement variability, which reflects a between-person difference in fluctuations in engagement, is a meaningful parameter that complements extant between-person and within-person conceptualizations of engagement. Then, in Study 2, an online experiment with a behavioral performance measure involving 600 full-time employees, we provide causal evidence for our full theoretical model, including the hypothesized mechanism of flow. Finally, in Study 3, a two-week experience sampling study with 152 employees and their supervisors, we test our full model in the field and examine relevant micro-mechanisms implicit in our theory. Across our studies, we also conduct additional supplementary analyses to enhance understanding of our findings, particularly around flow as a mediator, given that we find mixed support for its mediating effect.

In highlighting the interactive effect of engagement variability and emotional stability on job performance, we offer several contributions. First, we respond to scholars' calls to extend our knowledge of engagement by examining it "not as a continuous process filled with constant intensity, but rather as a noncontinuous process with intermittent exhibitions of disengagement" (Rothbard & Patil, 2013, p. 61); this noncontinuous process, in essence, reflects engagement variability, the focus of our article. In doing so, we not only introduce the conceptualization of engagement variability to the literature but we also importantly show that, alongside past research that has found that individual differences in average engagement are related to performance, inconsistently engaging may be problematic for performance even if the average level of engagement is the same. Relatedly, our model allows us to offer insights into the relationship between engagement and flow at work, which is useful given that the conceptual bounds of the two constructs and how they relate remain ill-defined (Bakker, 2011; Schaufeli et al., 2002).

We also contribute to research on emotional stability (i.e., neuroticism) by extending the burgeoning stream of work that shows that neuroticism (i.e., low emotional stability) can have some functional properties (e.g., Bendersky & Shah, 2013; Smillie et al., 2006). Our findings show that whether engagement variability negatively impacts performance crucially depends on an employee's level of emotional stability; perhaps counterintuitively, those who are lower (vs. higher) in emotional stability may not exhibit as strong of a negative relationship between engagement variability and performance. In incorporating emotional stability into our model, we thus also expand the scope of habit theory, which has thus far remained largely agnostic to specific individual factors that may qualify its core premise that inconsistently deploying resources always harms downstream outcomes like performance.

Defining Engagement Variability

Variability constructs (from which we derive engagement variability) have a long history in the field of psychology (e.g., Berdie, 1969; Cattell, 1973; Fiske, 1961; Murray, 1938); but only within the past two decades have organizational behavior scholars begun to rigorously attend to them theoretically and empirically (Eid & Diener, 1999; Matta et al., 2017; Scott et al., 2012). Such constructs are crucial for extending theory and empirical work by increasing the power scholars have at their disposal for predicting individual behavior (Fleeson & Leicht, 2006). In an attempt to situate these constructs theoretically, Fleeson and Leicht (2006) advanced the density distributions approach, which advocates that there are three parameters of interest when

predicting individual behavior. First, one can attend to the moment-to-moment assessments of a given construct (i.e., the within-person differences of a given construct). Second, one can describe an individual by the between-person differences of a given construct, which is the average of one's within-person differences. Third, one can account for the fact that individuals not only differ in terms of the mean level of a particular construct, but also in how much they vary around that mean level (i.e., how consistent or inconsistent they are in their display of a given construct). This third parameter of interest, which attends to the distribution of a particular construct capturing both between- and within-person differences, offers a way to develop a full understanding of a construct of interest that complements the first two parameters.

Scholars posit that variability constructs are particularly useful when questions exist as to whether a given construct is best thought to vary within persons or between persons because they offer a way to clarify and extend extant theory and empirical work (Fleeson & Leicht, 2006). Engagement, defined as the investment of one's cognitive, emotional, and physical resources, is one of these constructs. Some scholars have conceptualized engagement as a within-person difference (e.g., Macey & Schneider, 2008), whereas others have conceptualized it as a between-person difference (e.g., Schaufeli & Bakker, 2004). Yet, a third group of scholars has theorized that a full understanding of engagement comes from incorporating both between-person and within-person perspectives (e.g., Dalal et al., 2008; Rothbard & Patil, 2013). Indeed, as originally theorized, Kahn (1990) noted that role engagement ebbed and flowed as individuals pulled away and pushed toward their role memberships. However, such theorizing around individuals' patterns of their ebbs and flows has been limited and remains unexplored empirically. As such, introducing a variability construct is particularly suitable in the domain of engagement.

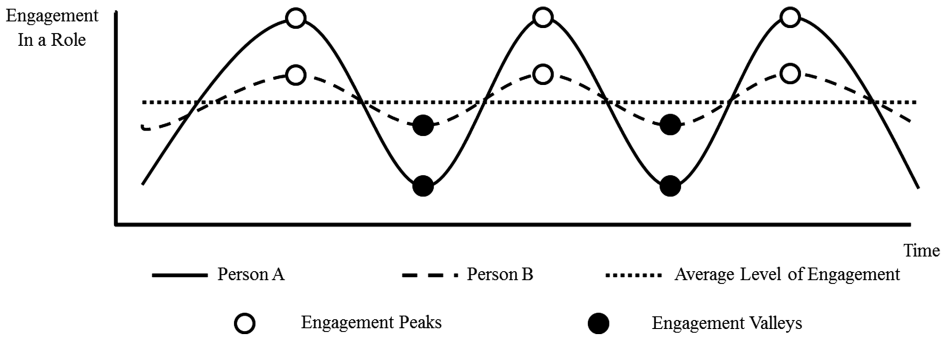
Accordingly, consistent with how other variability constructs have been conceptualized (e.g., Eid & Diener, 1999; Kernis et al., 1993; Matta et al., 2017; Scott et al., 2012), we define engagement variability as an individual's pattern of fluctuation around their average level of engagement. In other words, it reflects an individual's tendency to inconsistently engage in a role over time. Those who are lower in engagement variability fluctuate more tightly around their mean level of engagement in a role, whereas those who are higher in engagement variability fluctuate more wildly. For example, consider two individuals, Person A and Person B, who have the same average level of engagement in a role (represented by the horizontal dotted line in Figure 1). As can be seen in Figure 1, despite their similar average engagement levels, Person A and Person B exhibit differing levels of engagement variability as defined by the two curved lines (the solid line and the dashed line, respectively). Person B (the "low variability"/"consistently engaged" individual) fluctuates more tightly around the mean level of engagement in their role, whereas Person A (the "high variability"/"inconsistently engaged" individual) fluctuates more wildly. The peaks in engagement variability represent greater engagement in one's role, while the valleys represent lesser engagement in one's role.

The imagery of engagement variability as composed of peaks representing high engagement in a role, and valleys representing low engagement in that role, may evoke the possibility that the construct is a mere reflection of inconsistencies in workload, or workload variability. However, this is not necessarily the case. Whereas workload variability consists of objective demands in one's role (e.g., formal deadlines), engagement variability pertains to the psychological experience of the manner of one's investment of resources in one's role. As such, two individuals may have the same workload—and workload demands—but report drastically different degrees of engagement variability. Additionally, the locus of engagement variability and workload variability differs. Whereas workload variability emanates from factors external to the self, engagement variability likely emanates primarily, but perhaps not exclusively, from factors internal to the self.

Finally, following Suddaby's (2010) guidance on construct clarity, given that engagement variability reflects an individual's tendency to inconsistently engage in a role over time, it is worth considering the temporal boundaries of the construct. Past theory suggests that individuals can meaningfully vary in how engaged they are in role episodes that range from minutes to days to weeks (Bakker & Leiter, 2010; Kahn, 1990; Lanaj et al., 2019; Newton et al., 2020; van Woerkom et al., 2016), making such time horizons appropriate for observing engagement variability (Shipp & Cole, 2015). Indeed, it is not difficult to imagine engagement variability

Figure 1

Visualization of Engagement Variability Concept



Note. Person A can be said to be higher in engagement variability, whereas Person B can be said to be lower in engagement variability. The peaks in an individual's engagement variability represent greater engagement in their role, while the valleys represent lesser engagement in that same role.

manifesting in role episodes that are only a few minutes in duration (e.g., consider attorneys who bill clients in six-minute increments in a quarter-hour) or in those that last several weeks (e.g., consider a soldier on a twelve-week deployment). As such, we posit that engagement variability assessed over minutes, days, and weeks may theoretically operate similarly.

The Effect of Engagement Variability on Job Performance

A central aim of this investigation is to develop a model linking engagement variability to job performance. To do so, we draw on habit theory. Habit theory posits that the manner by which an individual allocates or invests their resources impacts how efficiently these resources can be used, which can have downstream effects on outcomes like performance (Ouellette & Wood, 1998; Rothman et al., 2011; Verplanken, 2006; Wood & Rünger, 2016). Although habits were initially conceived as purely behavioral, scholars have since acknowledged that habits implicate one's cognitive and emotional resources in addition to physical resources associated with behavior (Wood et al., 2002), making it an appropriate framework to draw upon to explain the effect of engagement variability.

Engagement Variability and Performance

According to habit theory, consistent execution of a cognitive, emotional, or behavioral response can facilitate performance because it eliminates the need to consciously switch investments of resources from one target to another (Rothman et al., 2011; Scott et al., 2012; Verplanken, 2006; Wood & Neal, 2007). Consequently, anything that prevents consistency in one's response hinders the automaticity (and thus efficiency) of that response. Engagement variability is one such possible factor that, by definition, impedes consistency. As defined, engagement variability reflects an individual's tendency to inconsistently engage in their role, which means individuals higher in engagement variability inconsistently invest their cognitive, emotional, and physical resources. Because they inconsistently allocate resources toward their role, they fail to develop habitual routines when engaging in their role, resulting in resource inefficiencies. In contrast, those who are lower in engagement variability exhibit greater consistency and thus can more efficiently allocate resources due to habituation. These individuals approach engagement in their role in an even manner, consistently investing in their role. In this vein, those lower in engagement variability reap the benefits associated with cognitive, emotional, and behavioral habit formation, notably the efficient allocation of resources, whereas those higher in engagement variability comparatively fail to do so.

Because engagement variability results in the inefficient allocation of resources, it is therefore likely to be negatively associated with job performance, controlling for average engagement. Indirect evidence suggests this hypothesis may hold. For example, in a mixed-methods study of the U.S. Navy, Stanko and Beckman (2015) found that Navy members who did not maintain consistent cognitive engagement in their work roles, instead shifting their attention toward and away from their jobs, exhibited poor work performance. Taken together, we hypothesize that engagement variability should be negatively associated with job performance, controlling for employees' average levels of engagement (the primary focus of extant work on engagement).

> *Hypothesis 1:* Engagement variability is negatively related to performance, controlling for average level of engagement.

The Moderating Role of Emotional Stability

To date, habit theory has largely remained agnostic about specific individual moderators that may qualify the relationships among consistency, resource efficiency, and subsequent performance. Accordingly, extending habit theory, we posit that emotional stability—an individual difference alluded to but not explicitly discussed in habit theory (Verplanken, 2018)—may moderate the negative relationship between engagement variability and performance. By definition, those who are higher in emotional stability are less reactive, experiencing fewer mood swings and lower levels of negative affect (McCrae & Costa, 1997; Wallace & Newman, 1998). In contrast, those lower in emotional stability exhibit "intense emotional and physiological reactivity to stress" (Connor-Smith & Flachsbart, 2007, p. 1081). As such, those who are higher, versus lower, in emotional stability are less reactive to a given stimulus.

Although higher emotional stability, in and of itself, may appear appealing because it implies that individuals are less reactive, a close look at its cognitive properties suggests that being lower in emotional stability may be useful under certain conditions. Compared to those higher in emotional stability, those lower are predisposed to automatically redirect their attention from one stimulus to another, in what has been termed as the "automatic orienting of attention" (Smillie et al., 2006; Wallace & Newman, 1997, 1998). As we argue, this automatic orienting of attention may be valuable in the presence of engagement variability such that the negative effect of engagement variability on performance may be weaker for those lower in emotional stability (Wallace & Newman, 1998). In other words, because those lower in emotional stability are more habituated to redirecting resources from one stimulus to another, they should be less likely to face the full extent of the resource inefficiencies that emanate from engagement variability, thereby experiencing less of a negative impact on performance.

In contrast, when individuals are higher in emotional stability, the negative relationship between engagement variability and performance should be stronger. This is because those higher in emotional stability are less predisposed to automatically redirecting their attention (Smillie et al., 2006; Wallace & Newman, 1997, 1998). As such, with greater engagement variability, they should be more likely to experience the full extent of the inefficiencies that come with the inconsistent allocation of resources, thereby decreasing their performance.

> *Hypothesis 2:* Emotional stability moderates the relationship between engagement variability and performance such that the negative relationship between engagement and performance will be weaker for those lower in emotional stability and stronger for those higher.

The Mediating Role of Flow

Finally, further extending habit theory, we hypothesize that the concept of flow explains the negative interactive effect of

engagement variability and emotional stability on performance. Flow, a mechanism implicated, but not elaborated upon, in habit theory (Verplanken, 2018), reflects a harmonious, self-reported subjective experience in which an individual feels that their actions have merged perfectly with their awareness such that the activity seems almost automatic (Csikszentmihalyi, 1990, 1997; Debus et al., 2014; Fullagar & Kelloway, 2009; Quinn, 2005). In other words, flow captures a state in which one feels that one has an innate understanding of how to adequately deploy resources to match challenging role demands without experiencing resource inefficiencies (Csikszentmihalyi, 1990, 1997). Those who report being in a high state of flow at work often comment that they "are in the zone," "understand the rhythm of the work," or "are in the groove" (Leybourne & Cook, 2015). Although flow may appear similar to engagement, the two constructs are distinct (Bakker, 2011; Schaufeli et al., 2002). Whereas engagement reflects holistic mobilization and investment of one's cognitive, emotional, and physical resources into one's role, flow reflects a peak, primarily cognitive, state defined by a perception of almost effortless, automatic action (Britt et al., 2007; Csikszentmihalyi, 1990, 1997; Hallberg & Schaufeli, 2006; May et al., 2004).

We posit that engagement variability should be less negatively associated with performance through flow for those lower, as opposed to higher, in emotional stability. As noted previously, the inconsistent allocation of resources that defines engagement variability impedes the development of habitual routines. Without such habitual routines, it should be harder to reach flow, given that flow reflects an optimal state in which one effortlessly understands how to deploy their resources to match demands (Csikszentmihalyi, 1990, 1997; Debus et al., 2014; Fullagar & Kelloway, 2009; Quinn, 2005). Yet, for those lower in emotional stability, the inconsistent allocation of resources is likely to seem more habitual and thus automatic, because they are also those who exhibit greater automatic orienting of attention (Wallace & Newman, 1998). As a result, for these individuals who are lower in emotional stability, there may be little interruption in flow emanating from different levels of engagement variability. In contrast, for those higher in emotional stability, engagement variability should be negatively related to flow, as engagement variability is likely to seem less habitual given that they exhibit less automatic orienting of attention in contrast to their more emotionally unstable counterparts (Wallace & Newman, 1998). This decreased flow, for those higher but not lower in emotional stability, should, in turn, be negatively associated with performance as decreased flow signals that what one is working on no longer feels instinctive or automatic (Csikszentmihalyi, 1990). Indeed, because flow reflects a harmonious experience with one's work, a number of studies point to a positive relationship between flow and performance (Engeser & Rheinberg, 2008; Jackson et al., 2001; Schüler, 2007).

Hypothesis 3: Flow mediates the negative interactive effect of engagement variability and emotional stability on performance.

Overview of Studies

To test our hypotheses, we conducted three studies. In Study 1, a field study involving ROTC cadets, we tested Hypotheses 1 and 2 with multisource, time-lagged (ten time periods) data. In Study 2, an online experiment, which involved full-time employees as participants and a behavioral measure of performance, we tested the full model. In Study 3, a multisource, two-week experience sampling study, we again tested the full model and examined implicated micro-mechanisms and alternative mechanisms. Together, the three studies enhance both external and internal validity, offset weaknesses that may be associated with any particular method (McGrath, 1981), and allow for constructive replication—which is often thought to be the strongest test of hypothesized relations—given that they employ different measures/operationalizations of our constructs (Lykken, 1968).

Transparency and Openness

Across our studies, we describe our sampling plans, all data exclusions (if any), and all manipulations. We report our measures of all three studies in full in Appendix A, and we adhered to the *Journal of Applied Psychology's* methodological checklist. Data for Study 1 and Study 3 (our field studies) are not available due to the data agreement with our field site and institutional review board (IRB) agreement, respectively. We post all appendices and provide the analytical output for all studies at https://osf.io/7r9mu/?view_only=ee1a674612b948b281f4c5564dbd7540. Data were analyzed using SPSS V27, Stata V17, and Mplus V8.5. Studies were not preregistered. Study 1 received IRB approval from the University of Texas at Austin (2017-01-0012). Study 2 received IRB approval from the Massachusetts Institute of Technology (E-2485). Study 3 received IRB approval from the University of Florida (202200961).

Study 1

Sample and Procedure

In Study 1, we collected data from three branches of the ROTC, which train college students, referred to as cadets, to become commissioned officers of the U.S. military. Specifically, we engaged in a multisource, time-lagged data collection with these ROTC cadets, with the goal of connecting engagement variability to performance assessed at the end of the study period. Cadets completed one electronic survey sent out at the same time each week for nine consecutive weeks, for a total of nine surveys. The first survey, administered at Time 1, captured individual differences that served as moderating or control variables in our theoretical model (emotional stability, self-esteem, and self-rated performance). The second through ninth surveys, administered at Time 2 through Time 9, respectively, measured cadets' levels of engagement, such that we could form a measure of our independent variable, engagement variability. At Time 10 (survey 10), we asked ROTC leadership to provide performance ratings for each of the cadets. Because we obtained data on the dependent variable, independent variable, and moderating variable at different time periods and from different sources, we were able to mitigate concerns related to the common method–common source bias (Ostroff et al., 2002; Podsakoff et al., 2003).

After listening to a brief presentation describing the purpose of our time-lagged study as one meant to gain a better understanding of how cadets handle their ROTC responsibilities, 176 out of 177 cadets opted into Survey 1 at Time 1 (a response rate of 99%). To encourage high response rates at subsequent time periods, we implemented a monetary incentive structure that would support the

time-lagged data collection process: $10 for Survey 1, $2 each for Surveys 2 through 6, $4 each for Surveys 7 and 8, and $6 for Survey 9. Additionally, a lottery prize of $200 was offered as an incentive for completing all nine surveys. As expected, response rates at each subsequent time period were high: 97% at Time 2, 100% at Time 3, 98% at Time 4, 100% at Time 5, 98% at Time 6, 100% at Time 7, 97% at Time 8, and 100% at Time 9. At Time 10, we received instructor ratings for all cadets.

To be included in the final sample, cadets needed to have completed all nine surveys, and ROTC leadership needed to have supplied performance data. After matching the data for cadets with the data received from ROTC leadership, the final sample consisted of 160 cadets (a final response rate of 90%).[1] Sixty-nine percent of the final sample was male, and 54% identified as White, 32% identified as Hispanic, 25% identified as Asian American, and 8% identified as Black. Twenty-nine out of the 160 (18%) cadets identified with more than one ethnicity. Their average age was 21.1 years ($SD = 3.4$), and their average ROTC tenure was 1.9 years ($SD = 1.0$).

Measures

Respondents (cadets and ROTC leadership) all used a 7-point Likert scale ranging from 1 = *strongly disagree* to 7 = *strongly agree* when responding to the items below, unless otherwise noted.

Independent Variable: Engagement (Variability)

We measured engagement in the ROTC role using nine items from the Rich et al. (2010) engagement scale: three for cognitive, three for emotional, and three for physical engagement. For cognitive engagement, each week, cadets indicated the extent to which they (a) had paid attention to, (b) were absorbed in, and (c) had concentrated in their ROTC role. For emotional engagement, each week, cadets indicated the extent to which they (a) had been enthusiastic about, (b) were interested in, and (c) were excited about their ROTC role. Finally, for physical engagement, each week, cadets indicated the extent to which they had (a) exerted effort, (b) demonstrated physical striving, and (c) exerted physical energy in their ROTC role. These engagement items formed reliable engagement scales at each time period ($\alpha = .91$–$.98$). In line with prior research on variability constructs (e.g., Fleeson, 2001; Matta et al., 2017; Scott et al., 2012), we calculated the standard deviation in engagement across time periods, Time 2 through Time 9, to create a measure of engagement variability.

Dependent Variable: Performance

To measure performance in ROTC, ROTC instructors at the branch level provided ratings for each cadet using four items developed by Pearce and Porter (1986) that involved rating cadets on a quantile 10-point scale that ranged from *bottom 10%* to *top 10%*. The four items captured cadets' overall performance, ability to get tasks done on time, the quality of their performance, and the achievement of work goals ($\alpha = .99$).

Moderating Variable: Emotional Stability

To assess emotional stability, we asked cadets to indicate their level of agreement with twelve items adapted from Costa and McCrae (1992). Because these items reflected neuroticism, the low end of emotional stability, we reverse-coded these items. A sample item is "I often feel tense and jittery" ($\alpha = .92$).

Control Variables

When determining which control variables to include in our analyses, we statistically accounted for factors that would enable us to eliminate potential alternative explanations for our effects. First, we controlled for underlying factors that may confound the effects of our engagement variability measure such as average engagement in the role. Scholars have found that high levels of engagement positively influence task performance (Christian et al., 2011; Rich et al., 2010). Given this empirical relationship and the fact that any standard deviation measure, such as our engagement variability measure, takes into account the mean of that measure, it was necessary to parse out the average level of engagement in the role from the standard deviation (i.e., engagement variability). In this way, we could determine whether engagement variability had an effect on job performance above and beyond the average level of engagement in the role. To create this measure, we calculated the average engagement across time periods, Time 2 through Time 9. We also controlled for engagement squared in order to rule out whether effects emanating from engagement variability were actually a reflection of possible curvilinear effects of average engagement.

Second, we controlled for potentially relevant situational factors such as average workload as well as the distribution, or standard deviation, in workload across Time 2 through Time 9 (i.e., workload variability). To account for average workload as well as workload variability, we asked cadets to estimate the amount of time, in hours, each week, that they spent "attending ROTC class," "completing assignments for ROTC class," "attending ROTC physical training," "attending as well as preparing for ROTC lab," and "completing other ROTC activities." These categories were developed in conjunction with ROTC instructors to ensure that the categories were collectively exhaustive but also mutually exclusive. There were conceptual reasons to include the average workload and its standard deviation because workload level or changes in workload could influence the extent that one engages in a role and also how one subsequently performs (Demerouti et al., 2001). This may especially be the case because cadets can indicate the same level of engagement, given that it is perceived, but have different actual workloads, rendering average workload and workload variability as potentially important, but distinct, predictors of performance.

Furthermore, we controlled for potentially relevant individual factors such as self-esteem, as well as self-rated performance, because they were likely to be associated with both the independent

[1] We did not find that the pattern of results changed meaningfully when we relied on multiple imputations such that all 177 cadets were included. As expected, average engagement was positively associated with supervisor-rated performance (estimate = 0.58, $SE = 0.20$, $p = .004$, $f^2 = .045$), supporting previous findings in the literature. Furthermore, in support of Hypothesis 1, those higher in engagement variability exhibited lower supervisor-rated performance, even when considered alongside average engagement (estimate = −1.21, $SE = 0.57$, $p = .034$, $f^2 = .026$). Moreover, as predicted in Hypothesis 2, we found a significant negative interaction between engagement variability and emotional stability ($b = -0.75$, $SE = 0.33$, $p = .023$, $f^2 = .032$).

variable, engagement variability, as well as the dependent variable, cadet performance. Self-esteem is defined as one's beliefs about one's own self-worth (Rosenberg, 1965). High self-esteem suggests high psychological availability, defined as a readiness to invest one's resources into a role—a key antecedent illuminated in Kahn's (1990) original theorizing on engagement. Furthermore, self-esteem has long been associated with higher performance (see Judge & Bono, 2001). To measure self-esteem, we asked cadets at Time 1 to indicate their level of agreement with ten items developed by Rosenberg (1965). A sample item is "On the whole, I am satisfied with myself" ($\alpha = .91$). Similarly, self-rated performance likely influences engagement variability to the extent that those high in self-rated performance are also those who vary in engagement less because they believe themselves capable of consistently engaging in their roles based off successful past performance. Self-rated performance was measured using the same scale as supervisor-rated performance above ($\alpha = .92$).

Results

Substantiating the Engagement Variability Construct

Following previous organizational scholars who have studied variability constructs (e.g., Fleeson, 2001; Matta et al., 2017; Scott et al., 2012), we sought to substantiate that engagement variability (Row 1 in Table 1) was indeed a meaningful third parameter that complements between-person and within-person conceptualizations of engagement. To provide such evidence, we looked to ensure that our data supported the between-person and within-person perspectives of engagement as evidenced by variance partitioning. If the between-person and within-person perspectives were not supported, it would make little sense to advance engagement variability as a construct, as its existence is theoretically predicated on the presence of the other two parameters (Fleeson, 2001). Therefore, we determined the between- and within-variance (the person and week variance, respectively) by running a null model in multilevel analyses from which we calculated the intraclass correlation coefficient. Variance decomposition analyses revealed that 59.7% of the variance was accounted for by the person; 40.3% of the variance was accounted for by the week. Accordingly, we found support for both between-person and within-person perspectives: In general, individuals differed from each other (a between-person perspective) and differed from themselves (a within-person perspective), although to a lesser extent, in how much they exhibited engagement in their role.

Confirmatory Factor Analyses

We looked to empirically establish discriminant validity among those measures obtained from the same source at the same time point: self-esteem, self-rated performance, and emotional stability. Specifically, we loaded each item of our measures onto the respective higher order factors of self-esteem, self-rated performance, and emotional stability. Due to the number of self-esteem and emotional stability items relative to the sample size, we created three parcels of items for each of these two measures (Little et al., 2002). A model in which the three factors were distinct fit the data well ($\chi^2(32) = 45.46$, CFI = .99, TLI = .99, SRMSR = .03) and had a superior fit in comparison to more parsimonious two-factor models (self-rated performance and self-esteem combined into one factor: $\chi^2(34) = 454.18$, CFI = .68, TLI = .58, SRMSR = .22; emotional stability and self-esteem combined into one factor: $\chi^2(34) = 156.45$, CFI = .91, TLI = .88, SRMSR = .06; self-rated performance and emotional stability combined into one factor: $\chi^2(34) = 403.95$, CFI = .72, TLI = .63, SRMSR = .23). The three-factor model also had a superior fit to the one-factor model ($\chi^2(35) = 546.96$, CFI = .61, TLI = .50, SRMSR = .19). As such, we concluded that the three measures were distinct.

Multilevel Confirmatory Factor Analysis

We also examined the factor structure of engagement given the multiple time points during which it was measured in order to ensure that it followed the standard structure specified in prior work on engagement (e.g., Rich et al., 2010). To do so, we conducted a multilevel confirmatory factor analysis (engagement at the week level nested within individuals). We modeled the items for cognitive engagement, physical engagement, and emotional engagement at Level 1 (week level) and clustered by the cadet. We found that this model fit the data well ($\chi^2(24) = 33.26$, CFI = .99, TLI = .99, $SRMSR_{within} = .02$).

Table 1
Study 1: Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Engagement variability	0.58	0.36	—										
2. Supervisor-rated performance	7.22	2.35	−0.30***	—									
3. Emotional stability	5.25	1.18	−0.24**	0.14†	—								
4. Self-esteem	5.71	1.01	−0.28***	0.16*	0.75***	—							
5. Self-rated performance	8.37	1.72	−0.24**	0.38***	0.37***	0.47***	—						
6. Average engagement	5.63	0.87	−0.58***	0.29***	0.39***	0.42***	0.43***	—					
7. Average workload (hours)	15.06	8.29	−0.05	0.12	0.01	0.09	0.19*	0.22**	—				
8. Workload variability	7.89	9.49	0.02	0.05	0.09	0.11	0.16*	0.13	0.83***	—			
9. Average engagement squared	32.44	9.30	−0.58***	0.29***	0.39***	0.43***	0.43***	0.99***	0.23**	0.14†	—		
10. Gender	0.69	0.47	−0.03	−0.03	0.19*	0.05	−0.04	−0.01	0.07	0.06	0.004	—	
11. Race	0.54	0.50	0.01	0.01	−0.14†	−0.03	0.08	0.04	0.03	0.04	0.05	0.11	—

Note. $N = 160$. Gender: 1 = male, 0 = not male. Race: 1 = White, 0 = non-White.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed). † $p < .10$.

Measurement Invariance Tests

We also conducted measurement invariance tests, which assess whether each dimension of engagement holds the same meaning to participants across weeks. To do so, first, we looked for support for configural invariance—or whether the same items measured each engagement dimension across weeks—as supported by acceptable goodness-of-fit statistics. We found this to be the case (cognitive: $\chi^2(140) = 201.75$, CFI = .99, TLI = .97, RMSEA = .05; emotional: $\chi^2(140) = 209.50$, CFI = .99, TLI = .97, RMSEA = .06; physical: $\chi^2(140) = 234.39$, CFI = .98, TLI = .95, RMSEA = .07). We then conducted tests for metric invariance, examining goodness-of-fit statistics after setting the factor loadings of the items as equivalent across weeks (cognitive: $\chi^2(154) = 229.19$, CFI = .98, TLI = .97, RMSEA = .06; emotional: $\chi^2(154) = 233.58$, CFI = .98, TLI = .97, RMSEA = .06; physical: $\chi^2(154) = 266.73$, CFI = .97, TLI = .95, RMSEA = .07). Finally, using chi-square difference tests, we tested whether there was a significant difference between the metric and configural invariance models for all three dimensions of engagement. We did not find significant differences ($p > .05$). As such, the dimensions of engagement seemed to be viewed equivalently across the weeks of the study.

Descriptive Statistics and Correlations

Table 1 presents the means, standard deviations, and correlations among the study variables. Of note, not all of our hypothesized control variables correlated significantly with the independent variable as well as the dependent variable. For example, although self-esteem and self-rated performance correlated significantly with the independent variable (self-esteem: $r = -.28$, $p < .001$; self-rated performance: $r = -.24$, $p = .002$) as well as the dependent variable of interest (self-esteem: $r = 0.16$, $p = .045$; self-rated performance: $r = .38$, $p < .001$), neither average workload ($r = -.05$, $p = .502$) nor the standard deviation in workload, that is, workload variability ($r = .02$, $p = .787$) correlated significantly with the independent variable of engagement variability. As such, the latter two variables were potentially "impotent" controls (Becker et al., 2016). Because the lack of significant correlations conflicted with the aforementioned conceptual rationales for their inclusion (see Becker et al., 2016 for a full discussion of statistical control), we report analyses with and without controls.

Hypothesis Testing

Because of the multilevel structure of the data in which cadet performance ratings were nested within ROTC branch leadership, we used hierarchical linear modeling with maximum likelihood estimation. Hypothesis 1 predicted that engagement variability would be negatively related to job performance, above and beyond average engagement. As can be seen in Table 2, Model 1, average engagement was positively associated with performance (estimate = 0.73, $SE = 0.20$, $p < .001$, $f^2 = .074$), supporting previous findings in the literature. Furthermore, in support of Hypothesis 1, as seen in Table 2, Model 2, those higher in engagement variability exhibited lower performance, controlling for average engagement (estimate = −1.25, $SE = 0.58$, $p = .033$, $f^2 = .029$).[2] We found a similar pattern of results when we included average engagement squared, self-esteem, self-rated performance, average workload, and workload variability as covariates (estimate = −1.31, $SE = 0.56$, $p = .022$, $f^2 = .033$; see Table 3, Model 2).

Hypothesis 2 stated that emotional stability would moderate the relationship between engagement variability and performance. Specifically, we expected that although engagement variability would be negatively related to performance when employees were higher in emotional stability, this effect would be attenuated when employees were lower in emotional stability. To test this hypothesis, we began by mean-centering our predictor variables before calculating our interaction effect, in line with recommendations (Dawson, 2014).[3] As can be seen in Table 2, Model 4, we found a significant negative interaction effect ($b = -0.80$, $SE = 0.33$, $p = .018$, $f^2 = .038$). Further, when we controlled for average engagement squared, self-esteem, self-rated performance, average workload, and workload variability, the pattern of results held, as seen in Table 3, Model 4 ($b = -0.72$, $SE = 0.33$, $p = .031$, $f^2 = .032$).

Although the magnitude of the coefficients differed slightly, the results suggested that the shape of the interaction would be similar across models. Thus, we plotted the interaction at higher (+1 *SD*) and lower levels of emotional stability (−1 *SD*) for the more parsimonious model, which featured only average engagement as a control. As can be seen in Figure 2, for individuals higher in emotional stability (+1 *SD*), engagement variability was negatively related to performance (simple effect = −2.33, $t = -2.93$, $p = .004$, $f^2 = .054$). For individuals lower in emotional stability (−1 *SD*), engagement variability was not significantly related to performance (simple effect = −0.44, $t = -1.40$, $p = .166$, $f^2 = .012$; the difference in slopes: $t = -2.38$, $p = .018$, $f^2 = .036$). Thus, Hypothesis 2 was supported.[4,5]

[2] Although, in our theory, we do not distinguish among the resources involved in engagement (i.e., cognitive, emotional, and physical) in terms of how variability relates to performance, we explored whether variability in different engagement resources affected performance in distinct ways. We found that those higher in cognitive engagement variability exhibited significantly lower supervisor-rated performance, controlling for average cognitive engagement (estimate = −1.38, $SE = 0.49$, $p = .006$, $f^2 = .048$). Moreover, at $p < .10$, those higher in emotional engagement variability exhibited lower supervisor-rated performance, controlling for average emotional engagement (estimate = −1.03, $SE = 0.53$, $p = .056$, $f^2 = .024$). Finally, likewise, at $p < .10$, those higher in physical engagement variability exhibited lower supervisor-rated performance, controlling for average physical engagement (estimate = −0.93, $SE = 0.54$, $p = .088$, $f^2 = .018$). Thus, cognitive engagement variability appeared to have the largest statistically significant negative effect on performance.

[3] The pattern and significance of results remain unchanged if one chooses not to mean-center predictors.

[4] Given that emotional stability contains myriad sub-dimensions (e.g., depression, self-consciousness, vulnerability; Judge et al., 2013), we also explored the role that these sub-dimensions play in our hypothesized interactive effect by creating subscales (see Appendix A for exact items). Regardless of the sub-dimension examined as a moderator, our pattern of results remained the same. That is, as hypothesized, we found a significant interactive effect (depression sub-facet: $b = 0.65$, $SE = 0.30$, $p = .033$, $f^2 = .031$; self-consciousness sub-facet: $b = 0.77$, $SE = 0.28$, $p = .006$, $f^2 = .044$; vulnerability subfacet: $b = 0.85$, $SE = 0.31$, $p = .008$, $f^2 = .049$). (*Note*. These interactions are positive given that we did not reverse code the sub-dimensions as we did for our neuroticism measure following our emotional stability theorizing).

[5] To delve deeper into the engagement variability–performance relationship, we also empirically explored whether there was an interactive effect between engagement variability and average engagement on performance. We did not find a significant interactive effect ($b = 0.36$, $SE = 0.52$, $p = .496$, $f^2 = .022$).

Table 2

Study 1: Effect of Engagement Variability on Performance (Only Average Engagement as a Control)

Variable	Model 1				Model 2				Model 3				Model 4			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
Intercept	3.17	1.19	2.67**	.009	5.57	1.62	3.44***	<.001	7.25	0.27	26.73***	<.001	7.18	0.30	23.70***	<.001
Average engagement	0.73	0.20	3.55***	<.001	0.43	0.25	1.74†	.083	0.39	0.26	1.50	.135	0.38	0.25	1.49	.137
Engagement variability	—	—	—	—	**−1.25**	**0.58**	**−2.15***	**.033**	−1.25	0.58	−2.15*	.033	−1.38	0.57	−2.41*	.017
Emotional stability	—	—	—	—	—	—	—	—	0.07	0.16	0.44	.659	0.14	0.16	0.89	.376
Engagement Variability × Emotional Stability	—	—	—	—	—	—	—	—	—	—	—	—	**−0.80**	**0.33**	**−2.38***	**.018**
−2 log likelihood	711.79				707.22				707.03				701.53			
Δ −2 log likelihood	—				4.57*				0.19				5.50*			

Note. $N = 160$. Bolded statistics show test of hypothesis. Because of the multilevel structure of the data in which cadet performance ratings were nested within ROTC branch leadership, we used HLM. In Models 3 and 4, predictors were mean-centered before calculating interaction effect. Pattern of results remains unchanged when not mean-centering. ROTC = Reserve Officer Training Corps; HLM = hierarchical linear modeling; *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed). † $p < .10$.

Discussion

The findings of Study 1, a multisource, ten-wave field study, offered initial support for our hypotheses. As predicted, engagement variability was negatively related to performance and this relationship was moderated by emotional stability such that the relationship attenuated in magnitude and significance as emotional stability decreased. Whereas a strength of Study 1 was that it enhanced external validity through the use of lagged, multisource field data, it did not permit us to infer causality. Moreover, Study 1 did not explore the role of flow as a mechanism of the interactive relationship (i.e., Hypothesis 3). To address these concerns, we ran Study 2, an experiment, which tested the full model (i.e., Hypotheses 1–3) by manipulating both engagement variability and emotional stability.

Study 2

Sample and Procedure

Study 2, an online experiment with a behavioral measure of performance, used a 2 (low engagement variability vs. high engagement variability) × 2 (low emotional stability vs. high emotional stability) between-subjects design to test the full model. We recruited 600 U.S. full-time employees (42% female; $M_{age} = 35.46$ years, $SD = 10.30$) via Prolific, following a sample size calculation of a small to medium effect with 80% power.

In this study, we told participants they would complete a slider task (Gill & Prowse, 2012; Yip et al., 2018) in which they would have 2.5 minutes to move sliders to a designated position. Specifically, participants saw 96 sliders that each ranged from 0 to 30 as well as a target number (e.g., 16) that indicated the position to which participants were expected to move the relevant slider (see Appendix B for an example). To manipulate engagement variability, participants in the low engagement variability conditions received the following instructions:

> Recent research suggests that performance can be improved if one works at a consistent steady pace throughout all of the rounds. That is, the best performance happens when one invests a steady amount of effort, enthusiasm, and mental energy throughout.

In contrast, participants in the high engagement variability conditions received the following instructions:

> Recent research suggests that performance can be improved if one varies their pace throughout all of the rounds. That is, the best performance happens when one varies the amount of effort, enthusiasm, and mental energy throughout, exerting their highest amount of effort, enthusiasm, and mental energy some of the time and their lowest effort, enthusiasm, and mental energy some of the time.[6]

[6] We importantly did *not* instruct those in the high engagement variability conditions to move from their highest effort to no effort in order to minimize the chance that those in these conditions were taking a break, thereby explaining drops in performance. We further assuage this concern methodologically and empirically. Methodologically, we chose a simple rather than complex performance task because some research suggests that downtime during simple tasks may actually increase performance by breaking up the repetitive nature of the task (e.g., Speier et al., 1999, 2003). Empirically, we recruited participants ($N = 600$; 42% female) from Prolific to participate in the same slider task with the same manipulation and found that the average number of clicks across rounds did not differ based on condition (high engagement variability condition: $M = 23.62$, $SD = 16.83$ vs. low engagement variability condition: $M = 22.67$, $SD = 7.43$; $F(1, 598) = 0.82$, $p = .367$, $\eta_p^2 = .001$). This suggests that those in the high engagement variability conditions may not have taken breaks.

Table 3

Study 1: Effect of Engagement Variability on Performance (All Controls)

Variable	Model 1				Model 2				Model 3				Model 4			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
Intercept	3.18	4.73	0.67	.503	4.67	4.70	0.99	.322	7.23	0.21	35.18***	<.001	7.17	0.24	30.28***	<.001
Average workload (hours)	0.04	0.04	0.99	.325	0.04	0.04	0.94	.349	0.04	0.04	0.99	.326	0.04	0.04	1.04	.300
Workload variability	−0.03	0.03	−0.93	.352	−0.02	0.03	−0.73	.468	−0.03	0.03	−0.77	.441	−0.02	0.03	−0.76	.450
Self-esteem	−0.14	0.20	−0.72	.470	−0.17	0.20	−0.87	.386	−0.23	0.26	−0.87	.388	−0.20	0.26	−0.79	.429
Self-rated performance	0.45	0.12	3.87***	<.001	0.46	0.11	3.98***	<.001	0.46	0.11	3.98***	<.001	0.43	0.11	3.78***	<.001
Average engagement	−0.19	1.76	−0.11	.913	−0.08	1.74	−0.05	.962	−0.13	1.74	−0.08	.941	0.65	1.75	0.37	.712
Average engagement squared	0.06	0.16	0.34	.732	0.02	0.16	0.11	.916	0.02	0.16	0.13	.900	−0.05	0.16	−0.32	.746
Engagement variability	—	—	—	—	**−1.31**	**0.56**	**−2.32***	**.022**	−1.31	0.56	−2.33*	.021	−1.45	0.56	−2.60*	.010
Emotional stability	—	—	—	—	—	—	—	—	0.07	0.22	0.33	.745	0.13	0.21	0.62	.539
Engagement Variability × Emotional Stability	—	—	—	—	—	—	—	—	—	—	—	—	**−0.72**	**0.33**	**−2.18***	**.031**
−2 log likelihood	695.87				690.58				690.47				685.87			
Δ −2 log likelihood	—				5.29*				0.09				4.60*			

Note. $N = 160$. Bolded statistics show test of hypothesis. Because of the multilevel structure of the data in which cadet performance ratings were nested within ROTC branch leadership, we used HLM with maximum likelihood estimation. In Models 3 and 4, predictors were mean-centered before calculating interaction effect. Pattern of results remains unchanged when not mean-centering. ROTC = Reserve Officer Training Corps; HLM = hierarchical linear modeling; *SE* = standard error.
* $p < .05$. *** $p < .001$ (two-tailed).

To reinforce this manipulation, participants received reminders of these instructions related to their corresponding condition every 50 seconds (i.e., after a third of the time had passed).

To manipulate emotional stability, participants listened to one of two songs as they completed the task. The songs were not significantly different in beats per minute but did seem to differ in the extent to which they manipulated emotional stability (a notion we validate with our manipulation check). Specifically, in the high emotional stability conditions, participants listened to an instrumental version of Katy Perry's "Roar." In the low emotional stability conditions, participants listened to "Threnody to the Victims of Hiroshima" composed by Krzysztof Penderecki.

Measures

Participants used a 7-point Likert scale ranging from 1 = *strongly disagree* to 7 = *strongly agree* when responding to the items below, unless otherwise noted.

Dependent Variable: Performance

We measured performance by assessing the number of sliders moved to the correct position within the 2.5 minutes allotted.

Mediating Variable: Flow

We measured flow with three items: "I was not able to get in the groove during the slider task," "I was not able to get in the rhythm during the slider task," and "I was not able to get in the flow during the slider task" ($\alpha = .97$; Jackson & Csikszentmihalyi, 1999). Items were reverse-scored such that higher scores indicated greater flow.[7]

Control Variable: Engagement

Given that the hypotheses posed sought to establish the effect of engagement variability on performance above and beyond the average level of engagement, we measured engagement using six items from Rich et al. (2010). Specifically, participants indicated the extent to which they had (a) paid attention to, (b) concentrated on, (c) been enthusiastic about, (d) been excited about, (e) exerted full effort, and (f) exerted a lot of energy when completing the slider task ($\alpha = .78$).

Results

Table 4 presents the means, standard deviations, and correlations among variables, and Table 5 displays the means and standard deviations of the key focal variables by condition.

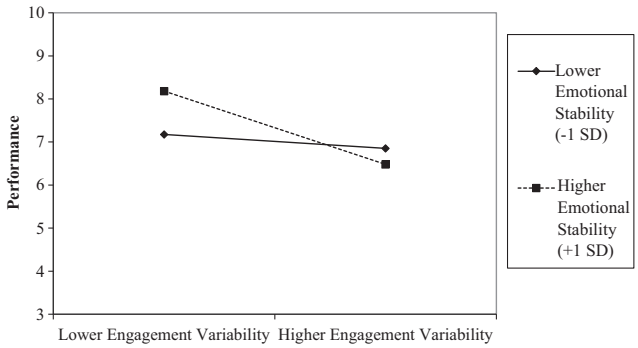
Confirmatory Factor Analysis

We looked to first empirically establish discriminant validity between those measures obtained from the same source—flow and

[7] Given the presence of other measures of flow present in the literature, we also recruited U.S.-based employees working at various organizations via Prolific ($N = 99$; 50% female) and explored the correlation between the three-item scale used (reverse-scored) and the 10-item Rheinberg et al. (2003) measure of flow. We found that the three-item scale used was very highly correlated with the Rheinberg et al. (2003) measure of flow ($r = .81$, $p < .001$).

Figure 2

Study 1: Emotional Stability as a Moderator of the Relationship Between Engagement Variability and Performance (Only Average Engagement as a Control)



Note. $N = 160$; ROTC = Reserve Officer Training Corps. Performance is rated by ROTC leadership. For individuals (cadets) higher in emotional stability (+1 *SD*), engagement variability was negatively related to performance (simple effect = −2.33, $t = -2.93$, $p = .004$, $f^2 = .054$). For individuals (cadets) lower in emotional stability (−1 *SD*), engagement variability was not significantly related to performance (simple effect = −0.44, $t = -1.40$, $p = .166$, $f^2 = .012$; difference in slopes: $t = -2.38$, $p = .018$, $f^2 = .036$).

engagement—especially given that the two are sometimes thought to be closely related (Bakker, 2011; Schaufeli et al., 2002). Specifically, we loaded each item of our measures onto the respective higher order factors of flow and engagement, allowing the error terms of the pairs of engagement items to correlate. A model in which the two factors were distinct fit the data well ($\chi^2(23) = 102.01$, CFI = .98, TLI = .97, SRMSR = .06) and had a superior fit in comparison to a more parsimonious one-factor model ($\chi^2(24) = 301.39$, CFI = .93, TLI = .90, SRMSR = .15). As such, we concluded that the two measures were distinct.

Manipulation Checks

To check our engagement variability manipulation, we asked participants to rate their agreement with the following three items: "I varied my level of energy throughout the task," "I varied my level of enthusiasm throughout the task," and "I varied my level of focus throughout the task" ($\alpha = .94$). As expected, those in the high engagement variability conditions ($M = 4.33$, $SD = 1.70$) were more likely to agree with these statements than those in the low engagement variability conditions ($M = 3.37$, $SD = 1.85$; $F(1, 597) = 44.49$, $p < .001$, $\eta_p^2 = .069$).[8] To check our emotional stability manipulation, we asked participants to rate their agreement with ten items from Goldberg (1992) that followed the stem "I felt … ." Sample items included "steady" and "moody" (reversed; $\alpha = .91$). As expected, those in the low emotional stability conditions ($M = 3.95$, $SD = 1.47$) were less likely to agree with these items than those in the high emotional stability conditions ($M = 4.90$, $SD = 1.26$; $F(1, 597) = 71.58$, $p < .001$, $\eta_p^2 = .107$).

Hypothesis Testing

Hypothesis 1 posited that engagement variability would be negatively related to performance, above and beyond average engagement. Given the inclusion of average engagement as a covariate, we followed Wellman et al. (2016) and reported estimated marginal means and standard errors where appropriate. As expected, and consistent with what we found in Study 1, a one-way analysis of variance with average engagement as a covariate revealed that those in the high engagement variability condition ($M = 42.11$, $SE = 0.81$) performed worse than those in the low engagement variability condition ($M = 46.44$, $SE = 0.81$; $F(1, 596) = 14.19$, $p < .001$, $\eta_p^2 = .023$).

Hypothesis 2 stated that emotional stability would moderate the relationship between engagement variability and performance. Specifically, we expected that although engagement variability would be negatively related to performance when employees were higher in emotional stability, this effect would weaken when employees were lower in emotional stability. To test this hypothesis, we ran a 2 (low engagement variability vs. high engagement variability) × 2 (low emotional stability vs. high emotional stability) between-subjects analysis of variance on performance, controlling for engagement. There were significant main effects of engagement variability ($F(1, 595) = 13.85$, $p < .001$, $\eta_p^2 = .023$) and emotional stability ($F(1, 595) = 12.15$, $p < .001$, $\eta_p^2 = .020$) which were qualified by a significant interaction effect ($F(1, 595) = 6.66$, $p = .010$, $\eta_p^2 = .011$). Interpreting the simple effects, we found support for Hypothesis 2 (see Figure 3). Specifically, when emotional stability was high, there was a significant negative relationship between engagement variability and performance (high engagement variability condition: $M = 42.66$, $SE = 1.13$ vs. low engagement variability condition: $M = 49.85$, $SE = 1.13$; $F(1, 595) = 18.20$, $p < .001$, $\eta_p^2 = .030$). In contrast, when emotional stability was low, we did not find a significant relationship between engagement variability and performance (high engagement variability condition: $M = 41.61$, $SD = 1.14$ vs. low engagement variability: $M = 42.93$, $SE = 1.16$; $F(1, 595) = 0.42$, $p = .516$, $\eta_p^2 = .001$).

Hypothesis 3 predicted that the interaction between engagement variability and emotional stability on performance would be mediated by the flow. To test this hypothesis, we ran moderated mediation analyses with 10,000 bootstraps following Hayes (2017). First, as seen in Table 6, Model 1, we found a significant negative interaction effect between engagement variability and emotional stability on flow ($b = -0.61$, $SE = 0.29$, $t = -2.08$, $p = .038$, $\eta_p^2 = .007$), which suggested that the effect of engagement variability on flow depended on emotional stability. Moreover, flow was positively associated with performance ($b = 1.46$, $SE = 0.31$, $t = 4.67$, $p < .001$, $\eta_p^2 = .035$). Finally, when evaluating the full moderated mediation model (see Table 6), the index of moderated mediation was significant ($b = -0.89$, $SE = 0.49$; 95% CI [−1.95, −.05]). When emotional stability was high, we found a significant negative indirect effect of engagement variability on performance

[8] As another way to further validate whether those in the high engagement variability conditions varied their engagement more than those in the low engagement variability conditions, we recruited participants ($N = 600$; 42% female) on Prolific (same sample as those reported in Footnote 6) to participate in the same slider task in Study 2 and randomly assigned them to the high or low engagement variability condition. We then calculated the *standard deviation* in the number of clicks across the slider rounds, given that clicks could proxy for how much participants engaged with the sliders and could be less subject to demand effects. As expected, those in the high engagement variability condition ($M = 3.98$, $SD = 4.36$) exhibited higher variability in clicks than those in the low engagement variability condition ($M = 3.07$, $SD = 2.32$; $F(1, 598) = 10.28$, $p = .001$, $\eta_p^2 = .017$).

Table 4
Study 2: Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Engagement variability	0.50	0.50	—						
2. Emotional stability	0.51	0.50	−0.01	—					
3. Performance	44.30	14.66	−0.17***	0.11**	—				
4. Flow	4.28	1.89	−0.04	0.23***	0.16***	—			
5. Average engagement	5.53	0.92	0.10*	0.11**	−0.23***	0.24***	—		
6. Gender	0.42	0.53	0.04	0.04	−0.14***	0.01	0.01	—	
7. Age	35.46	10.30	0.05	0.09*	−0.31***	−0.04	0.19***	−0.06	—

Note. $N = 600$. Engagement variability is a dummy variable where 1 = high and 0 = low. Emotional stability is a dummy variable where 1 = high and 0 = low.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed).

through flow ($b = -0.74$, $SE = 0.35$; 95% CI [−1.51, −.14]), but when emotional stability was low, we did not find a significant indirect effect ($b = 0.15$, $SE = 0.32$; 95% CI [−.48, .80]). Accordingly, we found support for Hypothesis 3.[9,10]

Discussion

Study 2, an experiment featuring a behavioral measure of performance, built upon and extended Study 1 by offering causal support for our full model. Similar to Study 1, we found that engagement variability was negatively related to performance and that this relationship was moderated by emotional stability. When emotional stability was high(er), we found a significant negative relationship between engagement variability and performance. When emotional stability was low(er), we did not find a statistically significant relationship. Completing our model and building on Study 1, we further found that the negative interactive effect between engagement variability and emotional stability on performance was mediated by flow. The primary strengths of Study 2 were that it captured objective performance—given that our theory posited that engagement variability affects actual, not just perceived, performance—and replicated effects along a differing time frame than Study 1. However, the time course in which we observed engagement variability was nonetheless on the shorter side. Thus, there remained an open question as to whether those higher in engagement variability assessed over a mid-range time period, longer than 2.5 minutes, would be truly less productive as we theorized. To address this concern, we ran Study 3, a two-week experience sampling study, which aimed to connect engagement variability (a between-person construct) to both (the between-person components of) daily supervisor-rated and employee-rated performance (as opposed to just performance at a single time point as in Study 1). As a bonus, an experience sampling study provided the opportunity to better connect our work to that on other variability constructs, given that most of this past work has used an experience sampling approach (e.g., Matta et al., 2017; Scott et al., 2012).

Table 5
Study 2: Estimated Conditional Marginal Means and Standard Errors by Condition

Experimental condition	Performance	Flow
High engagement variability; high emotional stability ($n = 152$)	42.66 (1.13)	4.41 (.15)
Low engagement variability; high emotional stability ($n = 153$)	49.85 (1.13)	4.91 (.15)
High engagement variability; low emotional stability ($n = 149$)	41.61 (1.14)	3.93 (.15)
Low engagement variability; low emotional stability ($n = 146$)	42.93 (1.16)	3.83 (.15)

Note. $N = 600$. Standard errors are in parentheses.

Study 3

Sample and Procedure

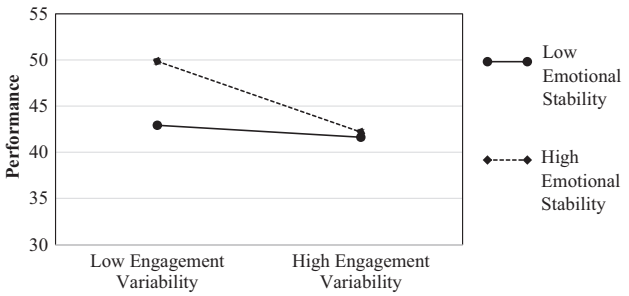
We recruited employees and their supervisors by advertising our study to organizational supervisors via ResearchMatch, an online

[9] Study 2 had an unexpected negative correlation between average engagement (which was measured after our manipulations) and performance (see Table 3). Thus, we sought to examine the robustness of our results excluding average engagement as a covariate (given that this was an experiment and covariates may not be needed). Consistent with our results reported, those in the high engagement variability condition ($M = 41.79$, $SD = 13.73$) performed worse than those in the low engagement variability condition ($M = 46.82$, $SD = 15.15$; $F(1, 597) = 18.22$, $p < .001$, $\eta_p^2 = .030$). Likewise, we found support for the moderating role of emotional stability ($F(1, 596) = 4.35$, $p = .037$, $\eta_p^2 = .007$). When emotional stability was high, there was a significant negative relationship between engagement variability and performance (high engagement variability condition: $M = 42.18$, $SD = 16.00$ vs. low engagement variability condition: $M = 49.59$, $SD = 15.40$; $F(1, 596) = 20.38$, $p < .001$, $\eta_p^2 = .033$). In contrast, when emotional stability was low, we did not find a significant relationship between engagement variability and performance (high engagement variability condition: $M = 41.39$, $SD = 10.99$ vs. low engagement variability: $M = 43.92$, $SD = 14.36$; $F(1, 596) = 2.30$, $p = .130$, $\eta_p^2 = .004$). Finally, flow mediated the interactive effect of engagement variability and emotional stability on performance (the index of moderated mediation was significant: $b = -0.71$, $SE = 0.41$, 95% CI [−1.67, −.07]).

[10] The effects found in this experiment were small. Nonetheless, small effects still matter. According to Prentice and Miller (1992), effect sizes should be interpreted in light of the manipulation, the outcome variable, and the duration of the experiment. It is worth noting that the duration of our experiment was very short. That is, we were able to see a significant negative effect of engagement variability on performance in only 2.5 minutes. Given that small effects accumulate over time to be quite powerful (Abelson, 1985; Funder & Ozer, 2019; Prentice & Miller, 1992), and most workplace performance episodes are longer than 2.5 minutes, the negative effect of engagement variability on performance may be quite meaningful. Moreover, to further shed light on the practical significance of our effects, we note that our effects are comparable to those in other experiments on variability (e.g., Matta et al., 2017).

Figure 3

Study 2: Interaction Between Engagement Variability and Emotional Stability on Performance (Only Average Engagement as a Control)



Note. $N = 600$. Performance is objective performance. When emotional stability was high, there was a significant negative relationship between engagement variability and performance (high engagement variability condition: $M = 42.66$, $SE = 1.13$ vs. low engagement variability condition: $M = 49.85$, $SE = 1.13$; $F(1, 595) = 18.20$, $p < .001$, $\eta_p^2 = .030$). In contrast, when emotional stability was low, we did not find a significant relationship between engagement variability and performance (high engagement variability condition: $M = 41.61$, $SE = 1.14$ vs. low engagement variability: $M = 42.93$, $SE = 1.16$; $F(1, 595) = 0.42$, $p = .516$, $\eta_p^2 = .001$). SE = standard error.

platform for research (for a similar data collection approach, see Gabriel et al., 2021). Supervisors who expressed interest in our study were sent an enrollment survey, which consisted of the consent form and basic demographic information. They were offered up to $26 for their participation: $2 for completion of each daily survey plus bonus amounts to enhance response rates with the goal of receiving at least three complete days. We also asked each supervisor to provide the contact information of up to three employees who would be willing to participate in our study. We then randomly selected one employee for each supervisor and invited them to participate in our study via email. This email included a separate enrollment survey consisting of the consent form, measures of employees' emotional stability and self-esteem, as well as basic demographic information. Employees were offered up to $63 for their participation: $2.50 for completion of each of the two daily surveys plus bonus amounts to enhance response rates with the goal of receiving at least three complete days. To be eligible, participants were required to be at least 18 years of age, work full time in the United States, and have a traditional Monday–Friday work schedule. Furthermore, for eligibility, given that ResearchMatch is a volunteer-based research platform and we wanted to ensure high-quality data, we included a number of quality checks and filtered based on participants' responses.[11] In total, we recruited 166 dyads to participate in our two-week study.

The daily portion of the study was conducted over a two-week work period, during which employees were surveyed twice a day (midday and end-of-day), and supervisors were surveyed once a day (end-of-day), Monday through Friday. We sent the employee midday survey at 12 p.m., and it included measures of employees' work engagement and flow. We sent the employee end-of-day survey at 4 p.m., and it included measures of employees' self-rated performance as well as their workload that day (as a control variable). Finally, we sent the supervisor end-of-day survey at 4 p.m., and it included a measure of employee performance. On average, the employee midday survey was completed at 1:12 p.m. and the employee end-of-day survey at 6:09 p.m., while the supervisor end-of-day survey was completed at 6:05 p.m. Accordingly, the average time between the employee midday and employee end-of-day surveys was 4 hours and 57 minutes, and the average time between the employee midday and supervisor end-of-day surveys was 4 hours and 53 minutes.

Following prior experience sampling method studies, we reviewed our data and excluded five employee–supervisor dyads for failing to complete at least three days of matched data (e.g., Gabriel et al., 2019) and nine dyads whose reported job descriptions/responsibilities did not align (i.e., supervisors reported having jobs that differed from those reported by their matched employees). Therefore, our final sample was 152 supervisor–employee dyads, with a total of 1,148 day-level observations.

Supervisors in the sample were majority male (60.5%) and Black (48.7%; 31.6% White; 9.2% Hispanic; 5.3% Asian/Pacific Islander; 3.9% American Indian/Alaskan Native; 1.3% multiracial), while their average age was 35.7 years old ($SD = 8.1$). On average, supervisors had an organizational tenure of 7.5 years ($SD = 6.6$), and 81.6% had earned at least a bachelor's degree.

Employees in the sample were majority male (55.9%) and Black (44.1%; 36.2% White; 8.6% Asian/Pacific Islander; 8.6% Hispanic; 1.3% American Indian/Alaskan Native; 1.3% multiracial), while their average age was 32.3 years old ($SD = 6.6$). On average, employees had an organizational tenure of 4.7 years ($SD = 4.7$), and 64.5% held at least a bachelor's degree. They held a variety of positions including lead architect, physician, sales manager, and IT manager.

Measures

Respondents used a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree* when responding to the items below, unless otherwise noted.

Independent Variable: Engagement (Variability)

We measured engagement using six items from Rich et al. (2010). Specifically, participants indicated the extent to which they had (a) paid attention to, (b) concentrated on, (c) been interested in, (d) been excited about, (e) exerted full effort on, and (f) strove hard to complete the job ($\alpha = .80$). As we did in Study 1, and following other variability work (e.g., Fleeson, 2001; Matta et al., 2017; Scott et al., 2012), we calculated the standard deviation in daily engagement across two weeks to operationalize employees' engagement variability.

[11] We removed supervisor–employee dyads from our study based on the following criteria: (a) suspicious and overlapping names and emails, (b) failed attention checks (both enrollment survey as well as daily surveys), (c) insufficient or ineligible responses to open-ended questions, (d) duplicate IP addresses, (e) differing responses to the same demographic questions, (f) careless responding patterns (e.g., reporting that they worked 280 hours per week), and (g) job descriptions that did not align with our study purpose (e.g., teachers "supervising" students).

Table 6
Study 2: Moderated Mediation Results

Variable	Model 1: Flow				Model 2: Performance			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
Constant	1.38	0.45	3.06**	.002	62.09	3.48	17.86***	<.001
Average engagement	0.44	0.08	5.47***	<.001	−4.48	0.63	−7.07***	<.001
Flow	—	—	—	—	1.46	0.31	4.67***	<.001
Engagement variability	0.10	0.21	0.49	.627	−1.46	1.61	−0.91	.363
Emotional stability	1.08	0.21	5.18***	<.001	5.34	1.63	3.27**	.001
Engagement Variability × Emotional Stability	−0.61	0.29	−2.08*	.038	**−4.99**	**2.25**	**−2.22***	**.027**

Note. $N = 600$. Bolded statistics show test of hypothesis. Engagement variability is a dummy variable where 1 = high and 0 = low. Emotional stability is a dummy variable where 1 = high and 0 = low. *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed).

Mediating Variable: Flow

We measured flow using three items developed by Rheinberg et al. (2003).[12] The items were "Today at work, the right thoughts/movements occurred of their own accord," "Today at work, I did NOT notice time passing," and "Today at work, I knew what I had to do each step of the way" (α = .74).

Dependent Variable: Performance

We obtained measures of performance from two sources. First, we asked supervisors to rate employee performance using four items developed by Pearce and Porter (1986): (a) ability to get tasks done on time, (b) quality of their performance, (c) achievement of their work goals, as well as (d) their overall performance (α = .87). All items were rated on a quantile 10-point scale that ranged from *bottom 10%* to *top 10%*. Second, we asked employees to rate themselves using the same items (α = .89).

Moderator: Emotional Stability

We measured emotional stability in the enrollment survey using six items[13] from Costa and McCrae (1992). Again, as in Study 1, because these items assessed neuroticism, the low end of emotional stability, we reverse-coded these items. A sample item is "I often feel tense and jittery" (α = .92).

Control Variables

In alignment with Study 1, we controlled for employees' mean level of engagement over the two-week period as well as engagement squared. In addition, we controlled for employees' average workload and workload variability over the two-week period. Following Study 1, we measured employees' workload in the end-of-day survey by asking employees to report the number of hours they worked that day. We calculated the standard deviation in daily workload across the two weeks to capture employees' workload variability. We also controlled for employees' trait self-esteem using the ten items from Rosenberg (1965). A sample item is "I feel that I'm a person of worth" (α = .83).

Supplementary Analyses: Measuring the Micro-Mechanism of Habit Formation

Our theory posited that engagement variability would be negatively associated with performance via flow because it would impede habituation, given that engagement variability reflects the inefficient allocation of resources.[14] Moreover, we theorized that this negative relationship would be weaker for those lower (vs. higher) in emotional stability because those lower were those who were predisposed to automatically redirect their attention from one stimulus to another. In other words, those lower in emotional stability were more used to—or habituated to—inconsistently allocating personal resources. It follows, then, that habit formation may explain the relationship between the interaction of emotional stability and engagement variability on flow, thereby serving as a micro-mechanism. Accordingly, to capture this micro-mediator, we measured participants' habit formation in the midday survey using four items developed by Verplanken and Orbell (2003). A sample item is "I was able to do my work activities automatically" (α = .72).

[12] We also measured whether depletion served as an alternative mediator of the interactive effect. One could imagine that at higher levels of emotional stability, there may be a positive relationship between engagement variability and depletion because individuals higher in emotional stability are not accustomed to the inconsistent allocation of resources involved with engagement variability, which thereby could impede performance. However, at lower levels of emotional stability, the relationship between engagement variability and depletion may get weaker, given that those lower in emotional stability are more habituated to inconsistently allocating their resources. Thus, we assessed participants' depletion in the end-of-day survey by adapting five items (α = .92) from Twenge et al. (2004).

[13] In a sample of U.S. employees recruited from Prolific ($N = 116$, 43% female), we found that the six items we used correlated very highly with the other items present in the full Costa and McCrae (1992) scale ($r = .92$, $p < .001$). As such, the items we used captured much of the same conceptual space of the items we did not use.

[14] We also measured an alternative micro-mechanism: speed of engagement. Engagement variability could negatively impact performance because it lends itself to multiple (perhaps slower) ramp up periods. Those lower, versus higher, in emotional stability could be better equipped to deal with these ramp-up periods because they are more predisposed to automatically redirecting their attention. To test this possibility, we measured participants' speed of engagement that day in the midday survey by using three items from Vogel et al. (2022). A sample item is "I quickly felt energetic at the start of my work day" (α = .75).

Supplementary Analyses: Clarifying How Flow Manifests (Flow Frequency/Intensity)

The differing time horizons across Studies 1 and 2 and the absence of a flow measure in Study 1 raise interesting questions around how flow may manifest over longer periods of time than that of Study 2. Specifically, over longer time frames, is flow a matter of how *often* the experience occurs for a given individual? Or is it a matter of how *intense* the flow experience is? To answer this question and clarify the manifestation of flow, we measured flow intensity and flow frequency by adapting three items developed by Jackson and Csikszentmihalyi (1999). For flow intensity, we asked participants to report the *extent* to which they experienced flow at work ($\alpha = .83$). Participants responded using a 5-point Likert scale ranging from 1 = *not at all* to 5 = *very much*. For flow frequency, we asked participants to indicate the *frequency* with which they experienced flow at work ($\alpha = .77$). Participants responded using a 5-point Likert scale ranging from 1 = *never* to 5 = *very frequently*. In order to prevent confusion between the two scales, we administered the flow intensity measure only during the first week of data collection and the flow frequency measure only during the second week.

Analysis and Results

Table 7 displays the means, standard deviations, and correlations.

Multilevel Confirmatory Factor Analysis

We conducted a multilevel confirmatory factor analysis in which we modeled the items for emotional stability and self-esteem at Level 2, and the items for engagement, flow, and supervisor-rated performance at Level 1. Drawing from Scott et al. (2010), we person-mean centered Level 1 items and grand-mean centered Level 2 items. Results indicated that this model fit the data well ($\chi^2(165) = 208.84$, CFI = .97, TLI = .96, $SRMSR_{within} = .03$, $SRMSR_{between} = .05$). We also conducted a similar multilevel confirmatory factor analysis with employee-rated performance at Level 1. Results indicated that this second model also fit the data well ($\chi^2(165) = 223.36$, CFI = .96, TLI = .96, $SRMSR_{within} = .03$, $SRMSR_{between} = .05$).

Analytic Approach

Before testing our hypotheses, we took stock of the nested nature of our data (i.e., days nested within employees). Focal Level 1 variables in our model (which are daily observations that may vary within employees across the two-week study period) included flow as well as supervisor- and employee-rated performance. Focal Level 2 variables (which vary between employees) included engagement variability, average engagement, average engagement squared, workload variability, and average workload (all of which were captured by employees' standard deviation or mean values of daily—i.e., Level 1—engagement and workload across the two-week study period, following prior established measurement approaches for other average and variability constructs in organizational scholarship, e.g., Matta et al., 2017, 2020; Scott et al., 2012), as well as employees' emotional stability and self-esteem (which were single measurements obtained at study enrollment). We partitioned the variance in our Level 1 constructs to determine the percentage that resided within-person versus between-person prior to testing our hypotheses. A nontrivial amount of variance existed both at the between- and within-levels for all constructs (all variances > .30; see Table 8).

Next, given the nested nature of our data (i.e., days nested within people), we used multilevel path analysis in Mplus 8.5 to test our model (Muthén & Muthén, 1998–2017). Multilevel path analysis partitions the variance of relevant constructs, thereby producing unbiased estimates at both the between- and within-levels.[15] Because all of our hypotheses were at the between-person level, our analyses were conducted at the between-person level. That is, we estimated the relationships of engagement variability—which is at the between-person level (given that it is an individual's variability across days)—with the between-person components of downstream variables (i.e., flow and performance). Following standard procedures (e.g., Scott et al., 2010), we grand-mean centered our Level 2 predictors (engagement variability, average engagement, average engagement squared, emotional stability, workload average, workload variability, and self-esteem). We adopted a Monte Carlo bootstrap approach with 20,000 simulations to build 95% bias-corrected confidence intervals around the conditional between-person indirect effect estimates (Preacher et al., 2010; Selig & Preacher, 2008). Because unstandardized coefficients are reported with maximum likelihood in Mplus, we manually calculated standardized coefficients following Shockley et al. (2021) to facilitate understanding of effect sizes.

Hypothesis Testing

Tables 9 and 10 present the results of the path model analyses with controls. We report results with controls below. Of note, results also hold without controls (see Tables 11 and 12). Hypothesis 1 predicted that engagement variability would be negatively related to performance, above and beyond average engagement. As predicted, engagement variability was negatively related to supervisor-rated performance ($\gamma = -1.42$, $\gamma' = -0.26$, $p = .001$). Similarly, engagement variability was negatively related to employee-rated performance ($\gamma = -1.15$, $\gamma' = -0.22$, $p = .014$).[16,17]

[15] In other words, we did not aggregate our Level 1 flow and performance variables upward, focusing instead on engagement variability's relationship with the between-person components of flow and performance.

[16] Because each observation of engagement may contain a degree of variability, we also looked to further bolster our findings by seeing if employee-reflected engagement variability over the two-week period would mirror our results that operationalize engagement variability as the standard deviation of daily assessments of engagement (see items in Appendix A). Specifically, we asked participants to reflect on how much they varied in their engagement over the past two weeks as well as their average level of engagement over the past two weeks on Day 10 of the study period. Consistent with our findings, reflected engagement variability was negatively and significantly related to supervisor-rated performance ($\gamma = -0.27$, $\gamma' = -0.05$, $p = .010$) and employee-rated performance ($\gamma = -0.27$, $\gamma' = -0.05$, $p = .010$).

[17] We also explored whether engagement variability would help us predict momentary performance ratings by testing, for example, whether one's performance on Wednesday would be a function of one's engagement variability over Monday through Wednesday. As expected, engagement variability across the three days was negatively related to both supervisor-rated performance ($\gamma = -0.98$, $\gamma' = -0.18$, $p = .010$) and employee-rated performance ($\gamma = -2.03$, $\gamma' = -0.38$, $p = .015$) on Wednesday. Next, we explored whether one's performance at the end of the week would be a function of one's engagement variability over the week, taking a similar approach as the prior analysis. Just as we found in the three-day approach, engagement variability was negatively related to supervisor-rated performance ($\gamma = -2.29$, $\gamma' = -0.43$, $p < .001$). Similarly, engagement variability was negatively related to employee-rated performance ($\gamma = -3.08$, $\gamma' = -0.58$, $p < .001$). As such, we found evidence that engagement variability could help us predict momentary performance ratings.

Table 7

Study 3: Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Flow	4.16	0.60	—	0.11***	0.16***	0.35***	0.37***	−0.10***	0.24***	0.35***	0.32***													
2. Supervisor-rated performance	7.75	1.17	−0.004	—	0.33***	0.11***	0.15***	−0.08**	0.13**	0.11*	0.13***													
3. Employee-rated performance	7.73	1.19	0.02	0.90***	—	0.15***	0.28***	−0.10***	0.32***	0.19***	0.26***													
4. Habit formation	4.04	0.59	0.79***	0.10	0.15†	—	0.42***	−0.06*	0.24***	0.30***	0.38***													
5. Speed of engagement	4.23	0.61	0.84***	−0.03	0.05	0.75***	—	−0.11***	0.38***	0.47***	0.46***													
6. Depletion	2.46	1.03	−0.63***	−0.09	−0.15†	−0.52***	−0.65***	—	−0.15***	−0.10*	−0.13***													
7. Flow frequency	4.21	0.59	0.80***	0.04	0.11	0.75***	0.82***	−0.59***	—	n/an/a	0.33***													
8. Flow intensity	4.18	0.69	0.79***	−0.04	0.04	0.70***	0.86***	−0.60***	0.77***	—	0.43***													
9. Daily engagement	4.35	0.52	0.81***	−0.05	0.06	0.73***	0.89***	−0.68***	0.80***	0.79***	—													
10. Engagement variability (*SD*)	0.29	0.22	−0.56***	−0.14†	−0.18*	−0.62***	−0.63***	0.49***	−0.65***	−0.59***	−0.66***	—												
11. Average engagement	4.35	0.52	0.82***	−0.04	0.05	0.73***	0.90***	−0.69***	0.79***	0.80***	0.99***	−0.66***	—											
12. Average engagement squared	19.22	4.35	0.84***	−0.04	0.05	0.75***	0.90***	−0.71***	0.81***	0.81***	0.99***	−0.68***	0.99***	—										
13. Average workload	7.32	1.18	−0.09	0.10	0.11	0.01	0.02	−0.002	−0.05	−0.05	0.02	0.06	0.003	−0.001	—									
14. Workload variability (*SD*)	0.41	0.46	−0.10	−0.12	−0.21*	−0.17*	−0.15†	0.26**	−0.20*	−0.10	−0.30***	0.27***	−0.29***	−0.29***	−0.002	—								
15. Emotional stability	3.60	1.20	0.70***	−0.07	−0.11	0.55***	0.60***	−0.46***	0.53***	0.54***	0.56***	−0.36***	0.58***	0.62***	−0.004	−0.07	—							
16. Self-esteem	3.84	0.75	0.69***	−0.07	−0.13†	0.50***	0.60***	−0.43***	0.46***	0.55***	0.53***	−0.28***	0.55***	0.57***	0.01	0.10	0.83***	—						
17. Employee age	32.28	6.60	0.24**	0.42***	0.42***	0.22**	0.30***	−0.42***	0.25**	0.30***	0.34***	−0.26***	0.35***	0.35***	0.11	−0.30***	0.30***	0.16*	—					
18. Employee gender	0.56	0.50	0.03	−0.23**	−0.23**	0.02	0.02	0.01	0.07	0.03	0.10	−0.12	0.10	0.09	0.03	−0.10	0.03	0.02	−.10	—				
19. Employee race	0.36	0.48	0.25**	0.03	−0.02	0.17*	0.19*	−0.17*	0.12	0.18*	0.16*	−0.05	0.19*	0.20*	−0.03	−0.12	0.38***	0.45***	0.21**	0.03	—			
20. Supervisor age	35.65	8.13	0.26**	0.44***	0.46***	0.30***	0.33***	−0.45***	0.26**	0.31***	0.36***	−0.29***	0.36***	0.38***	0.13	−0.30***	0.28***	0.24**	0.58***	−0.18*	0.31***	—		
21. Supervisor gender	0.61	0.49	0.01	−0.17*	−0.16†	−0.01	0.03	0.03	0.07	0.03	0.09	−0.08	0.10	0.08	−0.14	0.03	−0.05	−0.08	−0.17*	0.29***	−0.01	−0.13	—	
22. Supervisor race	0.32	0.47	0.28***	0.28***	0.22**	0.29***	0.23**	−0.23**	0.19*	0.21**	0.18*	−0.13	0.20*	0.21*	0.04	−0.09	0.35***	0.38***	0.32***	0.01	0.73***	0.40**	−0.06	—

Note. Pairwise Level 1 N = 1,148. Level 2 N = 152. Within-person correlations are presented above the diagonal and between-person correlations are presented below the diagonal. Variables 1–9 are daily within-person correlations. Variables 10–14 are aggregations of daily measurements. Gender: 1 = male, 0 = not male. Race: 1 = White, 0 = non-White. Within-person correlation between flow intensity and flow frequency is not available because flow intensity was measured only in Week 1 and flow frequency was measured only in Week 2.

* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed). † $p < .10$.

Table 8
Study 3: Variance Decomposition

Variable	Within-person variance (σ^2)	Between-person variance (τ_{00})	Percentage of total variance within-person (%)
Employee-rated performance	0.71	1.32	34.9
Supervisor-rated performance	0.68	1.27	35.0
Flow	0.29	0.32	47.3
Engagement	0.13	0.25	34.3
Habit formation	0.26	0.31	45.0

Note. Percentage of total variance within person was calculated as the following: $\sigma^2/(\sigma^2 + \tau_{00})$.

Hypothesis 2 proposed that the negative effect of engagement variability on performance would be weaker for those lower (vs. higher) in emotional stability. As shown in Table 9, the moderating effect of emotional stability on the relationship between engagement variability and supervisor-rated performance was negative ($\gamma = -1.13$, $\gamma' = -0.25$, $p = .037$). Likewise, the moderating effect of emotional stability on the relationship between engagement variability and employee-rated performance was negative ($\gamma = -1.26$, $\gamma' = -0.28$, $p = .034$; see Table 10). Given that the pattern was the same for both supervisor-rated and employee-rated performance, we plotted the interaction effect for the former. As shown in Figure 4, the relationship between engagement variability and supervisor-rated performance was negative for those with higher levels (+1 *SD*) of emotional stability (simple slope = −2.78, $p = .001$), but not significant for those with lower levels (simple slope = −0.06, $p = .932$; difference = −2.71, $p = .037$). In sum, these findings provided support for Hypothesis 2.

Hypothesis 3 posited that the interactive effect of engagement variability and emotional stability on performance would be mediated by flow. Beginning with supervisor-rated performance as the dependent variable, although we found that the interactive effect between engagement variability and emotional stability on flow was significant ($\gamma = -0.62$, $\gamma' = -0.28$, $p < .001$), we did not find that flow was positively related to supervisor-rated performance ($\gamma = 0.30$, $\gamma' = 0.15$, $p = .639$). As such, the index of moderated mediation was not significant (estimate = −0.19, 95% CI [−1.07, .60]). Likewise, when using employee-rated performance as the dependent variable, although the interaction effect between engagement variability and emotional stability on flow was significant ($\gamma = -0.62$, $\gamma' = -0.28$, $p < .001$), flow was not positively related to employee-rated performance ($\gamma = 0.06$, $\gamma' = 0.03$, $p = .930$). As such, the index of moderated mediation was not significant (estimate = −0.04, 95% CI [−.85, .81]). Therefore, we did not find that flow mediated the interactive effect of engagement variability and emotional stability on performance, as specified in Hypothesis 3.[18,19]

Supplementary Analyses

To delve further into our results, we ran a number of supplementary analyses. For all supplementary analyses below, we included the same control variables described above.

Exploring the Lack of Support for Flow as a Mechanism

Given that we found empirical support for Hypothesis 3 in Study 2, but not in Study 3, we probed to try to explain this non-finding. First, we explored whether differences in our results had to do with the differing time horizons across studies. Indeed, maybe flow manifests differently for individuals over a duration of weeks, like in Study 3, as compared to minutes, like in Study 2. Following this logic, flow, as a mechanism in Study 3, may then take on a different form, for example, either as intensity (the strength of flow experienced) or as frequency (the number of flow states experienced). To explore this notion, we specified four distinct multilevel path analytic models that mirrored those in our main analyses, given that we had two performance measures (supervisor-rated and employee-rated) and two mechanisms to test (flow frequency and flow intensity). After running our models, as indicated by nonsignificant indices of moderated mediation, neither flow frequency (supervisor-rated performance index: estimate = 0.08, 95% CI [−.70, .88]; employee-rated performance index: estimate = −0.07, 95% CI [−.69, .54]) nor flow intensity (supervisor-rated performance index: estimate = −0.01, 95% CI [−.53, .24]; employee-rated performance index: estimate = 0.02, 95% CI [−.27, .54]) seemed to mediate the interactive effect of engagement variability and emotional stability on performance.

Given that different manifestations of flow did not explain our interactive effect, we next dug deeper into the data that went into our main analyses. As part of this exercise, we looked to first verify that our data supported flow-related relationships expected given the extant literature. For example, in contrast to our examination of flow at the between-person level in an experience sampling study, prior work has focused on examining flow at the within-person level (Debus et al., 2014; Fullagar & Kelloway, 2009; Nakamura &

[18] Following Footnote 12, we did not find that depletion mediated the interactive effect between engagement variability and emotional stability on supervisor-rated performance, given a non-significant index of moderated mediation (estimate = −0.12, 95% CI [−.46, .01]). Alternatively, it may be the case that significant effects may emerge when one considers the interactive effect between the variability and average parameter on more proximal outcomes (e.g., depletion) that in turn may have implications for performance (e.g., following Matta et al., 2020). Indeed, we found a significant interactive effect between engagement variability and average engagement on depletion ($\gamma = 1.88$, $\gamma' = .21$, $p < .001$). For those with higher levels (+1 *SD*) of average engagement, the relationship between engagement variability and depletion was positive and significant (simple slope = 1.49, $p < .001$). For those with lower levels (−1 *SD*), it was not significant (simple slope = −0.48, $p = .201$; difference = 1.97, $p < .001$).

[19] Like in Study 1, we also empirically explored whether there was an interactive effect between engagement variability and average engagement on supervisor-rated and employee-rated performance. We did not find a significant interactive effect on either supervisor-rated ($\gamma = -1.17$, $\gamma' = -4.28$, $p = .143$) or employee-rated performance ($\gamma = -0.07$, $\gamma' = -0.25$, $p = .932$).

Table 9
Study 3: Moderated Mediation Results—Flow and Supervisor-Rated Performance (All Controls)

	Model 1						Model 2						Model 3					
	Flow			Performance			Flow			Performance			Flow			Performance		
Variable	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*
Intercept	4.16***	0.02	<.001	4.07†	2.12	.054	4.16***	0.02	<.001	4.50*	2.03	.027	4.11***	0.03	<.001	6.38*	2.65	.016
Average workload (hours; L2)	−0.05*	0.02	.027	0.15	0.12	.194	−0.05*	0.02	.024	0.16	0.12	.181	−0.07**	0.02	.002	0.10	0.13	.435
Workload variability (L2)	0.11*	0.05	.019	0.46†	0.23	.051	0.11*	0.05	.017	−0.43†	0.23	.062	0.14**	0.05	.005	−0.37	0.24	.132
Self-esteem (L2)	0.20***	0.04	<.001	−0.22	0.17	.191	0.20***	0.04	<.001	−0.11	0.18	.542	0.13**	0.05	.008	0.10	0.23	.669
Average engagement (L2)	−0.86†	0.45	.056	−0.42	1.03	.682	−0.79†	0.45	.080	0.45	0.98	.645	0.21	0.36	.569	1.26	1.12	.260
Average engagement squared (L2)	0.20***	0.05	<.001	−0.06	0.16	.728	0.19**	0.06	.001	−0.21	0.15	.159	0.05	0.05	.292	−0.29†	0.16	.072
Engagement variability (L2)	—	—	—	—	—	—	−0.10	0.26	.699	−1.60***	0.43	<.001	0.02	0.20	.904	**−1.42****	**0.41**	**.001**
Flow (L2)	—	—	—	0.88†	0.51	.086	—	—	—	0.78	0.49	.112	—	—	—	0.30	0.65	.639
Emotional stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	0.10**	0.03	.001	−0.03	0.15	.830
Engagement Variability × Emotional Stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	−0.62***	0.16	<.001	**−1.13***	**0.54**	**.037**

Note. Level 1 N = 1,148. Level 2 N = 152. Bolded statistics show test of hypothesis. Coefficients are unstandardized. All relationships are at the between-person level (Level 2). L2 = Level 2; *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed). † $p < .10$.

Csikszentmihalyi, 2002). In line with this work, one might expect that within-person flow (the component of flow attributed to the day) would predict within-person performance (the component of performance attributed to the day). Indeed, we found support for this notion (supervisor-rated performance: $\gamma = 0.17$, $\gamma' = 0.09$, $p = .020$; employee-rated performance: $\gamma = 0.25$, $\gamma' = 0.12$, $p = .005$).

Buoyed by the fact that our data supported what might be expected from prior work (at least at the within-level), we revisited the non-significant, near-zero between-person correlation between flow and (supervisor-rated and employee-rated) performance that drove our lack of support for Hypothesis 3. Perhaps our near-zero between-person correlation was a reflection of the heterogeneity of the types of jobs participants in Study 3 held (e.g., mail carriers, janitors, and laborers on one end and project managers, marketing specialists, and nurses on the other end), such that in some roles, the relationship between flow and performance could be positive, but, in others, it could be negative. For example, for roles in which an individual's performance is the product of a single individual (rather than a group of individuals), an individual's flow may more clearly predict performance. In contrast, for roles in which an individual's performance is also a product of others' inputs (i.e., the job involves getting others to work together to accomplish tasks), the expected positive relationship between between-person flow and between-person performance may be obscured, or even reversed, given the role that others may have in determining "individual" performance.

To explore this idea, we classified jobs in our sample by the extent to which they required jobholders to coordinate the work and activities of others, a classification present in O*Net that captures the extent to which a job's performance is contingent on others' inputs (see Peterson et al., 2001, for a description of O*Net, a widely used database that classifies jobs along a number of descriptors). Following our intuition, we found that for jobs in which performance was more clearly the product of only the jobholder, between-person flow was positively correlated with between-person performance (supervisor-rated performance: $r = .30$, $p = .009$; employee-rated performance: $r = .28$, $p = .015$). In contrast, for jobs in which performance was less clearly the product of only the jobholder, there was a negative correlation between between-person flow and between-person performance (supervisor-rated performance: $r = -.29$, $p = .010$; employee-rated performance: $r = -.23$, $p = .043$). As such, our lack of support for Hypothesis 3 may have been due to our particular sample—specifically the heterogeneity of jobs held by our participants.

Habit Formation as a Micro-Mediator

Following habit theory, habituation is an implicit micro-mediator in our model.[20] Thus, as a final supplementary analysis, we tested this notion given that we measured habit formation. Specifically, we ran a supplementary analysis in which we explored whether habit formation mediated the interactive effect of engagement variability and emotional stability on flow. As expected, we found that the index of moderated mediation was significant (estimate = −0.15, 95% CI [−.41, −.02]). In detail, the indirect effect of engagement

[20] Following Footnote 14, we did not find evidence for speed of engagement as an alternative micro-mediator. The index of moderated mediation included zero for the interaction on flow via speed of engagement (estimate = −0.04, 95% CI [−.35, .11]).

Table 10
Study 3: Moderated Mediation Results—Flow and Employee-Rated Performance (All Controls)

	Model 1						Model 2						Model 3					
	Flow			Performance			Flow			Performance			Flow			Performance		
Variable	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*
Intercept	4.16***	0.02	<.001	4.97**	1.91	.009	4.16***	0.02	<.001	5.35*	2.19	.015	4.11***	0.03	<.001	7.37**	2.63	.005
Average workload (hours; L2)	−0.05*	0.02	.027	0.15	0.12	.192	−0.05*	0.02	.024	0.16	0.12	.185	−0.07**	0.02	.002	0.09	0.13	.470
Workload variability (L2)	0.11*	0.05	.020	−0.51*	0.23	.026	0.11*	0.05	.017	−0.48*	0.22	.030	0.14**	0.05	.005	−0.42†	0.23	.073
Self-esteem (L2)	0.20***	0.04	<.001	−0.40*	0.17	.021	0.20***	0.04	<.001	−0.30	0.20	.121	0.13**	0.05	.008	−0.06	0.24	.808
Average engagement (L2)	−0.86†	0.45	.057	0.79	1.00	.429	−0.79†	0.45	.081	1.53	1.01	.130	0.21	0.36	.569	2.39*	1.07	.026
Average engagement squared (L2)	0.20***	0.05	<.001	−0.13	0.15	.366	0.19**	0.06	.001	−0.27†	0.16	.087	0.05	0.05	.294	−0.35*	0.16	.029
Engagement variability (L2)	—	—	—	—	—	—	−0.10	0.26	.695	−1.36**	0.50	.007	0.02	0.20	.909	**−1.15***	**0.47**	**.014**
Flow (L2)	—	—	—	0.66	0.46	.154	—	—	—	0.57	0.53	.281	—	—	—	0.06	0.64	.930
Emotional stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	0.10**	0.03	.001	−0.05	0.15	.713
Engagement Variability × Emotional Stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	−0.62***	0.16	<.001	**−1.26***	**0.59**	**.034**

Note. Level 1 N = 1,148. Level 2 N = 152. Bolded statistics show test of hypothesis. Coefficients are unstandardized. All relationships are at the between-person level (Level 2). L2 = Level 2; *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed). † $p < .10$.

variability on flow through habit formation was negative when emotional stability was higher (estimate = −0.30, 95% CI [−.74, −.06]) but was not significant when emotional stability was lower (estimate = 0.05, 95% CI [−.12, .37]; difference = −0.35, 95% CI [−.99, −.04]). Thus, these findings elucidate the role of habit formation in explaining our interactive effect on flow, thereby bolstering the use of habit theory as the foundation for our theorizing.

Discussion

Study 3, an experience sampling study, built upon and extended Studies 1 and 2 by addressing the open question as to whether those higher in engagement variability assessed over a longer time period would be truly less productive as we theorized. We found this to be the case with both supervisor-rated performance and employee-rated performance. Furthermore, like the prior two studies, we found that the negative relationship between engagement variability and both supervisor-rated and employee-rated performance was moderated by emotional stability. When emotional stability was higher, we found a significant negative relationship between engagement variability and both supervisor-rated and employee-rated performance. When emotional stability was lower, we did not find a statistically significant relationship between engagement variability and both supervisor-rated and employee-rated performance. In contrast to Study 2, however, we did not find support for flow as a mediating mechanism. Our supplementary analyses revealed that this non-finding may potentially have been driven by heterogeneity in the between-person associations between flow and both supervisor-rated and employee-rated performance, which resulted in an overall near-zero correlation. We revisit this point in the General Discussion section. Finally, we note that we found support for habit formation as a micro-mediator, lending support to the appropriateness of habit theory as the basis of our theorizing.

General Discussion

In the past two decades, scholars have made considerable headway in outlining the relationships between engagement and critical organizational outcomes. However, there has been little work on how employees' *inconsistency* around their average levels of engagement affects their performance, above and beyond average engagement. Such a lack of consideration is unfortunate because engagement, as originally theorized, is not only the simultaneous investment of one's cognitive, emotional, and physical resources into a work role performance but also the maintenance of such investments over time (Kahn, 1990, 1992). Hence, in this article, we introduced the concept of engagement variability, defined as an individual's tendency to inconsistently engage in their role, and developed a model examining its impact on performance. As noted, variability constructs are crucial for extending theory and empirical work by increasing the power scholars have at their disposal for predicting individual behavior (Fleeson & Leicht, 2006). Further, variability constructs are particularly appropriate in domains like engagement, in which there are unanswered questions about whether a phenomenon of interest is a within-person or between-person difference (Dalal et al., 2008). We conclude by delineating how the study of engagement variability increases clarity and extends theory and empirical work not only in the domain of engagement but also in the domains of emotional stability, flow, and habit theory.

Table 11

Study 3: Moderated Mediation Results—Flow and Supervisor-Rated Performance (Only Average Engagement as a Control)

	Model 1						Model 2						Model 3					
	Flow			Performance			Flow			Performance			Flow			Performance		
Variable	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*
Intercept	4.16***	0.03	<.001	6.84***	1.26	<.001	4.16***	0.03	<.001	7.19***	1.25	<.001	4.10***	0.03	<.001	8.55***	2.00	<.001
Average engagement (L2)	0.94***	0.07	<.001	−0.30	0.29	.295	0.90***	0.10	<.001	−0.66*	0.30	.026	0.58***	0.07	<.001	−0.50	0.37	.185
Engagement variability (L2)	—	—	—	—	—	—	−0.15	0.24	.530	−1.60***	0.43	<.001	0.02	0.19	.932	**−1.20****	**0.37**	**.001**
Flow (L2)	—	—	—	0.22	0.31	.486	—	—	—	0.14	0.31	.663	—	—	—	−0.22	0.49	.651
Emotional stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	0.18***	0.02	<.001	0.02	0.13	.874
Engagement Variability × Emotional Stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	−0.60***	0.14	<.001	**−1.34****	**0.50**	**.007**

Note. Level 1 N = 1,148. Level 2 N = 152. Bolded statistics show test of hypothesis. Coefficients are unstandardized. All relationships are at the between-person level (Level 2). *SE* = standard error; L2 = Level 2.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$ (two-tailed).

Theoretical Implications

By introducing the construct of engagement variability, we broaden the focus that past work has placed on the average level of engagement as a predictor of performance (e.g., Christian et al., 2011; Rich et al., 2010) to show that the extent to which an individual tends to inconsistently engage in their role—their engagement variability—can be an impactful predictor of performance above and beyond the role of average engagement. In doing so, we extend theorizing in at least two ways. First, our focus allows scholars to complete the trilogy of parameters advocated by Fleeson and Leicht (2006) as important for predicting individual behavior: (a) a between-person variability parameter (engagement as measured in cross-sectional studies), (b) a within-person variability parameter (engagement as measured in experience sampling studies), and (c) a between-person parameter summarizing within-person differences (engagement variability as measured in time-lagged studies in which engagement is measured at multiple time periods). Indeed, as Rothbard and Patil (2013, p. 61) observed, "[I]t is important for future research to conceptualize engagement, not as a continuous process filled with constant intensity, but rather as a noncontinuous process," such that individuals may differ from each other in the distribution of engagement levels they display. That is, some may engage more inconsistently in their roles than others (Kahn, 1990, 1992).

Second, and potentially more importantly, our model provides a foundation for those scholars interested in examining other ways in which engagement may boost performance. Indeed, the current work would suggest that encouraging high performance through engagement would best be achieved by increasing individuals' access to cognitive, emotional, and physical resources. However, increasing resources may not always be possible given that resources may be limited (Macey & Schneider, 2008). In this vein, considering the maintenance aspect of engagement–i.e., engagement variability—suggests that manipulating the application of existing resources, rather than simply increasing resources, could be a viable alternative to promoting performance. That is, scholars could begin to consider which contextual conditions may help facilitate low engagement variability. In raising this point, we highlight that, although we conceptualize engagement variability as a relatively stable difference, consistent with other variability constructs, it nonetheless likely has situational antecedents (also consistent with past variability work; Matta et al., 2017; Scott et al., 2012).

Moreover, this research offers insights into the consistency or sustainability of employee engagement levels, a critical open question in the domain of engagement (Knight, Patterson, Dawson, & Brown, 2017; Rothbard & Patil, 2013; Wilson, 2019). Just as consistency reflects an individual's capacity to endure over time, engagement variability represents a between-person difference that captures the extent to which individuals inconsistently engage, that is, are less able to endure. Across three studies, we found that those who are less able to endure—or sustain engagement—in their roles (i.e., are higher in engagement variability) are likely to perform worse than those who are more able to endure (i.e., are low in engagement variability). Importantly, however, we offer evidence to support one moderator that attenuates the negative effect of engagement variability on performance: emotional stability. When employees are higher in emotional stability, engagement variability

Table 12

Study 3: Moderated Mediation Results—Flow and Employee-Rated Performance (Only Average Engagement as a Control)

Variable	Model 1						Model 2						Model 3					
	Flow			Performance			Flow			Performance			Flow			Performance		
	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*	γ	*SE*	*p*
Intercept	4.16***	0.03	<.001	9.02***	1.26	<.001	4.16***	0.03	<.001	9.36***	1.38	<.001	4.10***	0.03	<.001	10.12***	1.99	<.001
Average engagement (L2)	0.94***	0.07	<.001	0.42	0.30	.160	0.90***	0.10	<.001	0.09	0.30	.775	0.58***	0.07	<.001	0.24	0.35	.494
Engagement variability (L2)	—	—	—	—	—	—	−0.15	0.24	.521	−1.48*	0.59	.013	0.02	0.19	.936	**−0.99***	**0.49**	**.042**
Flow (L2)	—	—	—	−0.31	0.31	.320	—	—	—	−0.39	0.34	.251	—	—	—	−0.61	0.49	.209
Emotional stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	0.18***	0.02	<.001	−0.07	0.13	.593
Engagement Variability × Emotional Stability (L2)	—	—	—	—	—	—	—	—	—	—	—	—	−0.60***	0.14	<.001	**−1.48***	**0.57**	**.010**

Note. Level 1 $N = 1{,}148$. Level 2 $N = 152$. Bolded statistics show test of hypothesis. Coefficients are unstandardized. All relationships are at the between-person level (Level 2). L2 = Level 2; *SE* = standard error.
$^{*}p < .05$. $^{***}p < .001$ (two-tailed).

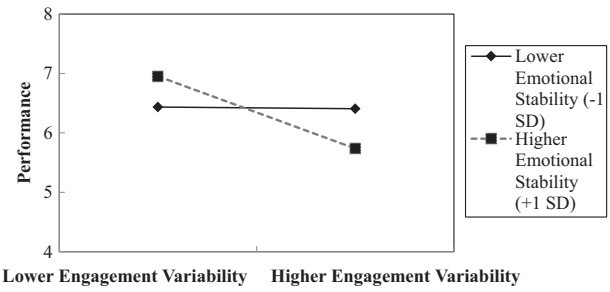
is negatively related to performance, but when employees are lower in emotional stability, this effect weakens. In this way, we put forth a more complete understanding of the relationship between engagement variability and performance that pinpoints for whom the negative effect of engagement variability on performance is likely to be present. Put differently, we identify how individual differences can moderate an individual's capacity to endure and thereby offer an interactive perspective for future scholarship: The effects of engagement variability may be best understood when examined in combination with individual differences like emotional stability.

The identification of the attenuating effect of lower emotional stability on the relationship between engagement variability and performance also contributes to research on neuroticism. Being lower in emotional stability (i.e., higher in neuroticism) has long been associated with a host of detrimental outcomes including intrusive thoughts and distress as well as a predisposition to interpret stimuli and events as negative (Ode & Robinson, 2007; Rusting & Larsen, 1997). However, despite the preponderance of research supporting the "dark" side of being low in emotional stability, a growing stream of work has begun to shed light on the "bright" side of being emotionally unstable (i.e., neurotic). For example, in a study of status attainment among work group members, Bendersky and Shah (2012) found that those low in emotional stability gained in status over time due to their strong desire to avoid social disapproval. In another example, Smillie et al. (2006) found that those low in emotional stability outperformed their more emotionally stable counterparts when the situation merited the use of a great deal of attentional resources such as when a task required tremendous effort. Although we are careful to highlight that this work does not go so far as to suggest there may be gains if one is lower in emotional stability—indeed, those lower in emotional stability do not seem to achieve levels of performance that rival their lower engagement variability/more emotionally stable peers—we show that being lower in emotional stability may *weaken* the negative effect of engagement variability on performance such that we no longer observe a statistically significant effect. In this manner, we offer some balance to the conversation that has highlighted both the "dark" and "bright" sides of being low in emotional stability. Lower emotional stability need not yield direct positive outcomes but rather can serve as a factor that attenuates the negative effect of engagement variability on performance.

The inclusion of emotional stability also allows us to extend habit theory, upon which our model relies. Habit theory highlights the importance of repeated responses in order to utilize one's resources in the most efficient manner possible and thus develop a habit (Ouellette & Wood, 1998). As such, a core premise of this theory is that inconsistency at any level results in the inefficient use of resources. Indeed, we drew upon this logic to theorize that engagement variability would be negatively related to performance. However, much of the literature that has drawn upon habit theory has thus far been largely agnostic about the individual or situational factors that can disrupt the relationships among consistency, resource efficiency, and subsequent positive outcomes. Accordingly, we extend habit theory by qualifying for whom these relationships that underlie habit theory's core premise are likely to hold: those higher in emotional stability. Moreover, in applying habit theory when theorizing the relationship between engagement variability—which involves the inconsistent

Figure 4

Study 3: Interaction Between Engagement Variability and Emotional Stability on Performance (All Controls)



Note. $N = 152$. Performance is supervisor-rated. The relationship between engagement variability and supervisor-rated performance was negative and significant for those with higher levels (+1 *SD*) of emotional stability (simple slope = −2.78, $p = .001$), but not significant for those with lower levels (simple slope = −0.06, $p = .932$; difference = −2.71, $p = .037$).

investment of cognitive, emotional, and physical resources—and performance, we build on an emerging body of work that suggests that habits are not just behavioral in nature but rather may emanate from repeated cognitive or emotional responses as well (e.g., Verplanken, 2006).

Finally, we offer insights for research at the intersection of engagement and flow at work (Csikszentmihalyi, 1990, 1997), a construct attracting burgeoning interest from scholars (Debus et al., 2014; Fullagar & Kelloway, 2009; Quinn, 2005). Although engagement and flow are thought to be related but distinct (Bakker, 2011; Schaufeli et al., 2002), the conceptual bounds of the two constructs and how they relate remain ill-defined, perhaps in part due to early theorizing on flow, which appeared to conflate the two (Csikszentmihalyi, 1997). In this work, we argued that while engagement reflects holistic mobilization of one's cognitive, emotional, and physical efforts, flow reflects a state defined by a perception of almost effortless, automatic action. That is, engagement may be quite effortful, while flow is defined by effortlessness. We encourage future researchers to continue investigating the distinctions between these two constructs.

Limitations and Future Directions

Although this research offers several contributions, it is not without limitations. First, although engagement variability operated similarly across the various time horizons in our studies, it is likely not entirely time invariant. For example, it may not operate similarly over a time horizon of seconds as it does over minutes, days, or weeks (Bakker & Leiter, 2010). Accordingly, future work could benefit from adopting more extreme time spans as a way to further delineate temporal scope conditions. Second, our theory suggests that engagement variability negatively affects actual, not perceived, performance. However, only Study 2 involved objective performance with Studies 1 and 3 relying on supervisor ratings of performance (and employee ratings in Study 3 as well). Indeed, in contexts where employee performance is not easily quantifiable (like in Studies 1 and 3), supervisor ratings may be the best proxies for objective performance (Cho et al., 2022; Murphy et al., 2018). Nonetheless, in future work, we invite researchers to explore our hypotheses in field contexts in which objective performance can be obtained.

Third, future work could explore the antecedents of engagement variability, which can further explicate the etiology of engagement variability. As conceptualized, engagement variability likely emanates from factors internal to the self. Indeed, to date, those studying variability constructs have privileged individual differences as antecedents over situational characteristics (Scott et al., 2012). Although identification of individual differences as antecedents can advance our understanding, it may be fruitful for scholars to begin to identify situational antecedents as well. Research on variability constructs such as affect variability (Kernis et al., 1993), emotional labor variability (Scott et al., 2012), justice variability (Matta et al., 2017), and self-esteem variability (Eid & Diener, 1999) have long suggested that such constructs are products of both individual and situational factors. If we are to comprehensively isolate consistency-related factors that can increase—not just have no effect on—job performance, we need to account for situational factors that are potentially more malleable.

Furthermore, we found mixed evidence for flow as a mediator of the interactive effect of engagement variability and emotional stability on performance, driven by non-significant between-person correlations between flow and supervisor-rated and employee-rated performance in Study 3. Also, in Study 3, we found one instance of our covariate of average engagement significantly negatively predicting supervisor-rated performance (see Table 11). With regard to the former, our supplementary analyses probing the flow mechanism revealed heterogeneity in the jobs participants held that may explain our lack of support for flow as a mechanism. With regard to the significant negative relationship between average engagement and supervisor-rated performance, it may reflect a suppression effect (see Cohen et al., 2003, for a discussion of suppression effects). It is also worth noting that Study 3 was conducted a few short months following the COVID-19 omicron surge during which organizations were at varying return-to-work phases. The stresses of this time may have prevented supervisors from making unobstructed assessments of employee performance. This is why we also included a measure of employee-rated performance. We were heartened to find the same results for the relationship between engagement variability and performance, controlling for average engagement—regardless of the nature of the relationship between average engagement and performance—lending confidence to the focal hypotheses of our model.

Last, although this research illuminated a downside of being higher in engagement variability in the form of decreased performance, there may be unexplored benefits. For example, many great thinkers and scientists such as Isaac Newton and Albert Einstein have claimed that their "eureka" moments occurred when they were immersed in trains of thought unrelated to the creative problem they intended to solve (Baird et al., 2012). Indeed, research on mind-wandering and distraction (Baer et al., 2020; Dijksterhuis & Meurs, 2006) as well as that on incubation effects (Sio & Ormerod, 2009) have all converged to support the notion that "focused deliberation on problems can undermine creativity" (Baird et al., 2012, p. 1117). Following this logic, higher engagement variability may be particularly useful when considered alongside outcomes like creativity because engagement

variability reflects the inconsistent deployment of cognitive, emotional, and physical resources. Additional examination of the relationships engagement variability may have with a broader set of outcomes beyond job performance can further develop the nomological network within which the construct sits.

Practical Implications

By shining a light on engagement variability, we encourage managers to not only consider ways to increase employees' cognitive, emotional, and physical investments into their work roles—an implication laid out in prior work focused on average engagement—but also to attend to managing their employees' inconsistency in engagement. Specifically, it would behoove managers to be on the lookout for engagement variability in the workplace as it can have detrimental effects on performance. However, if engagement variability is deemed inevitable, it may be worthwhile to attend to those who are higher in emotional stability in order to better support them. Indeed, if existing individual or situational factors promote higher engagement variability, those higher in emotional stability may be those who particularly suffer from a performance perspective, even though they may seem to have the most intuitive potential for higher performance.

Finally, we highlight that, although across our studies, the greatest performance seems to come from those who consistently engage and are high in emotional stability, we caution employers against selecting on these attributes because scholars have documented bright sides to low emotional stability (Bendersky & Shah, 2012) and engagement variability may also have upsides not identified here, as we suggested in the prior section (e.g., creativity). Accordingly, practitioners should be mindful of how context may influence how engagement variability operates. For example, perhaps in jobs that demand high levels of emotional labor (Grandey, 2003), engagement variability may be beneficial for employee well-being.

Conclusion

We set out in this article to increase scholarly understanding of the relationship between engagement and performance. Toward this end, we introduced the concept of engagement variability and found that it was negatively related to performance, above and beyond average engagement, which has been the focus of extant work. Furthermore, we also identified for whom this negative relationship was likely to be present: those individuals who are higher, not lower, in emotional stability. In doing so, this model sets the foundation for future scholars to examine the performance consequences of variables besides average level of engagement, such as engagement variability, which can also impact employee performance.

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*(1), 129–133. https://doi.org/10.1037/0033-2909.97.1.129

Baer, M., Dane, E., & Madrid, H. (2020). Zoning out or breaking through? Linking daydreaming to creativity in the workplace. *Academy of Management Journal*, *64*(5), 1553–1577. https://doi.org/10.5465/amj.2017.1283

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*(10), 1117–1122. https://doi.org/10.1177/0956797612446024

Bakker, A. B. (2011). An evidence-based model of work engagement. *Current Directions in Psychological Science*, *20*(4), 265–269. https://doi.org/10.1177/0963721411414534

Bakker, A. B., & Leiter, M. P. (Eds.). (2010). *Work engagement: A handbook of essential theory and research* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203853047

Becker, T. E., Atinc, G., Breaugh, J. A., Carlson, K. D., Edwards, J. R., & Spector, P. E. (2016). Statistical control in correlational studies: 10 essential recommendations for organizational researchers. *Journal of Organizational Behavior*, *37*(2), 157–167. https://doi.org/10.1002/job.2053

Bendersky, C., & Shah, N. P. (2012). The cost of status enhancement: Performance effects of individuals' status mobility in task groups. *Organization Science*, *23*(2), 308–322. https://doi.org/10.1287/orsc.1100.0543

Bendersky, C., & Shah, N. P. (2013). The downfall of extraverts and rise of neurotics: The dynamic process of status allocation in task groups. *Academy of Management Journal*, *56*(2), 387–406. https://doi.org/10.5465/amj.2011.0316

Berdie, R. F. (1969). Consistency and generalizability of intraindividual variability. *Journal of Applied Psychology*, *53*(1), 35–41. https://doi.org/10.1037/h0026720

Britt, T., Dickinson, J., Greene-Shortridge, T., & McKibben, E. (Eds.). (2007). *Self-engagement at work*. SAGE Publications. https://doi.org/10.4135/9781446212752.n11

Cattell, R. B. (1973). *Personality and mood by Questionnaire*. Jossey-Bass.

Cho, I., Berry, C. M., Payne, S. C., & Lee, P. (2022). Too good to be true? Are supervisor-perspective ratings a valid substitute for actual supervisor ratings? *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0001018

Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, *64*(1), 89–136. https://doi.org/10.1111/j.1744-6570.2010.01203.x

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum.

Connor-Smith, J. K., & Flachsbart, C. (2007). Relations between personality and coping: A meta-analysis. *Journal of Personality and Social Psychology*, *93*(6), 1080–1107. https://doi.org/10.1037/0022-3514.93.6.1080

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Crawford, E. R., Lepine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology*, *95*(5), 834–848. https://doi.org/10.1037/a0019364

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. HarperPerennial.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.

Dalal, R. S., Brummel, B. J., Wee, S., & Thomas, L. L. (2008). Defining employee engagement for productive research and practice. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*(1), 52–55. https://doi.org/10.1111/j.1754-9434.2007.00008.x

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1–19. https://doi.org/10.1007/s10869-013-9308-7

Debus, M. E., Sonnentag, S., Deutsch, W., & Nussbeck, F. W. (2014). Making flow happen: The effects of being recovered on work-related flow between and within days. *Journal of Applied Psychology*, *99*(4), 713–722. https://doi.org/10.1037/a0035881

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition*, *15*(1), 135–146. https://doi.org/10.1016/j.concog.2005.04.007

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*(4), 662–676. https://doi.org/10.1037/0022-3514.76.4.662

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4

Fiske, D. W. (1961). The inherent variability of behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience* (pp. 326–354). Dorsey.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Leicht, C. (2006). On delineating and integrating the study of variability and stability in personality psychology: Interpersonal trust as illustration. *Journal of Research in Personality*, *40*(1), 5–20. https://doi.org/10.1016/j.jrp.2005.08.004

Fullagar, C. J., & Kelloway, E. K. (2009). Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology*, *82*(3), 595–615. https://doi.org/10.1348/096317908X357903

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gabriel, A. S., Lanaj, K., & Jennings, R. E. (2021). Is one the loneliest number? A within-person examination of the adaptive and maladaptive consequences of leader loneliness at work. *Journal of Applied Psychology*, *106*(10), 1517–1538. https://doi.org/10.1037/apl0000838

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods*, *22*(4), 969–1006. https://doi.org/10.1177/1094428118802626

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review*, *102*(1), 469–503. https://doi.org/10.1257/aer.102.1.469

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Grandey, A. A. (2003). When "the show must go on": Surface acting and deep acting as determinants of emotional exhaustion and peer-rated service delivery. *Academy of Management Journal*, *46*(1), 86–96. https://doi.org/10.2307/30040678

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology*, *43*(6), 495–513. https://doi.org/10.1016/j.jsp.2005.11.001

Hallberg, U. E., & Schaufeli, W. B. (2006). "Same same" but different? Can work engagement be discriminated from job involvement and organizational commitment? *European Psychologist*, *11*(2), 119–127. https://doi.org/10.1027/1016-9040.11.2.119

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Jackson, S. A., & Csikszentmihalyi, M. (1999). *Flow in sports*. Human Kinetics.

Jackson, S. A., Thomas, P. R., Marsh, H. W., & Smethurst, C. J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *Journal of Applied Sport Psychology*, *13*(2), 129–153. https://doi.org/10.1080/104132001753149865

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—Self-esteem, generalized self-efficacy, locus of control, and emotional stability—With job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*(1), 80–92. https://doi.org/10.1037/0021-9010.86.1.80

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875–925. https://doi.org/10.1037/a0033901

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724. https://doi.org/10.2307/256287

Kahn, W. A. (1992). To be fully there: Psychological presence at work. *Human Relations*, *45*(4), 321–349. https://doi.org/10.1177/001872679204500402

Kernis, M. H., Cornell, D. P., Sun, C.-R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal of Personality and Social Psychology*, *65*(6), 1190–1204. https://doi.org/10.1037/0022-3514.65.6.1190

Knight, C., Patterson, M., & Dawson, J. (2017). Building work engagement: A systematic review and meta-analysis investigating the effectiveness of work engagement interventions. *Journal of Organizational Behavior*, *38*(6), 792–812. https://doi.org/10.1002/job.2167

Knight, C., Patterson, M., Dawson, J., & Brown, J. (2017). Building and sustaining work engagement-a participatory action intervention to increase work engagement in nursing staff. *European Journal of Work and Organizational Psychology*, *26*(5), 634–649. https://doi.org/10.1080/1359432X.2017.1336999

Lanaj, K., Foulk, T. A., & Erez, A. (2019). Energizing leaders via self-reflection: A within-person field experiment. *Journal of Applied Psychology*, *104*(1), 1–18. https://doi.org/10.1037/apl0000350

Leybourne, S. A., & Cook, P. (2015). "Rockin'All Over The World": Organisational improvisation lessons from the music-based practitioner. *International Journal of Management Concepts and Philosophy*, *9*(1), 1–19. https://doi.org/10.1504/IJMCP.2015.075102

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*(3), 151–159. https://doi.org/10.1037/h0026141

Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*(1), 3–30. https://doi.org/10.1111/j.1754-9434.2007.0002.x

Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. *Journal of Applied Psychology*, *93*(3), 498–512. https://doi.org/10.1037/0021-9010.93.3.498

Matta, F. K., Scott, B. A., Colquitt, J. A., Koopman, J., & Passantino, L. G. (2017). Is consistently unfair better than sporadically fair? An investigation of justice variability and stress. *Academy of Management Journal*, *60*(2), 743–770. https://doi.org/10.5465/amj.2014.0455

Matta, F. K., Scott, B. A., Guo, Z. A., & Matusik, J. G. (2020). Exchanging one uncertainty for another: Justice variability negates the benefits of justice. *Journal of Applied Psychology*, *105*(1), 97–110. https://doi.org/10.1037/apl0000425

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology*, *77*(1), 11–37. https://doi.org/10.1348/096317904322915892

McCrae, R. R., & Costa Jr., P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509–516. https://doi.org/10.1037/0003-066X.52.5.509

McGrath, J. E. (1981). Dillematics: The study of research choices and dillemas. *American Behavioral Scientist*, *25*(2), 179–210. https://doi.org/10.1177/000276428102500205

Murphy, K. R., Cleveland, J. N., & Hanscom, M. E. (2018). *Performance appraisal and management*. Sage Publications.

Murray, H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age*. Oxford University Press.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.).

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford University Press.

Newton, D. W., LePine, J. A., Kim, J. K., Wellman, N., & Bush, J. T. (2020). Taking engagement to task: The nature and functioning of task engagement across transitions. *Journal of Applied Psychology*, *105*(1), 1–18. https://doi.org/10.1037/apl0000428

Ode, S., & Robinson, M. D. (2007). Agreeableness and the self-regulation of negative affect: Findings involving the neuroticism/somatic distress relationship. *Personality and Individual Differences*, *43*(8), 2137–2148. https://doi.org/10.1016/j.paid.2007.06.035

Ostroff, C., Kinicki, A. J., & Clark, M. A. (2002). Substantive and operational issues of response bias across levels of analysis: An example of climate-satisfaction relationships. *Journal of Applied Psychology*, *87*(2), 355–368. https://doi.org/10.1037/0021-9010.87.2.355

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*(1), 54–74. https://doi.org/10.1037/0033-2909.124.1.54

Parke, M. R., Weinhardt, J. M., Brodsky, A., Tangirala, S., & DeVoe, S. E. (2018). When daily planning improves employee performance: The importance of planning type, engagement, and interruptions. *Journal of Applied Psychology*, *103*(3), 300–312. https://doi.org/10.1037/apl0000278

Pearce, J. L., & Porter, L. W. (1986). Employee responses to formal performance appraisal feedback. *Journal of Applied Psychology*, *71*(2), 211–218. https://doi.org/10.1037/0021-9010.71.2.211

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., Campion, M. A., Mayfield, M. S., Morgeson, F. P., Pearlman, K., Gowing, M. K., Lancaster, A. R., Silver, M. B., & Dye, D. M. (2001). Understanding work using the Occupational Information Network (O* NET): Implications for practice and research. *Personnel Psychology*, *54*(2), 451–492. https://doi.org/10.1111/j.1744-6570.2001.tb00100.x

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*(1), 160–164. https://doi.org/10.1037/0033-2909.112.1.160

Quinn, R. W. (2005). Flow in knowledge work: High performance experience in the design of national security technology. *Administrative Science Quarterly*, *50*(4), 610–641. https://doi.org/10.2189/asqu.50.4.610

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens [The assessment of flow experience]. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation* [Diagnosis of motivation and self-concept]. (pp. 261–279). Hogrefe.

Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, *53*(3), 617–635. https://doi.org/10.5465/amj.2010.51468988

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press. https://doi.org/10.1515/9781400876136

Rothbard, N. P., & Patil, S. V. (2013). Being there: Work engagement and positive organizational scholarship. In K. C. G. Spreitzer (Ed.), *The oxford handbook of positive organizational scholarship* (pp. 56–68). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734610.013.0005

Rothman, A. J., Baldwin, A. S., Hertel, A. W., & Fuglestad, P. T. (2011). Self-regulation and behavior change: Disentangling behavioral initiation and behavioral maintenance. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 106–122). Guilford Press.

Rusting, C. L., & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and Individual Differences*, *22*(5), 607–612. https://doi.org/10.1016/S0191-8869(96)00246-2

Sarwar, S., & Abugre, J. B. (2013). Does provision of feedback increase work engagement and reduce cynicism among employees? *The Business & Management Review*, *3*(4), 20–31.

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, *25*(3), 293–315. https://doi.org/10.1002/job.248

Schaufeli, W. B., Martinez, I., Marques-Pinto, A., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology*, *33*(5), 464–481. https://doi.org/10.1177/0022022102033005003

Schaufeli, W. B., & Salanova, M. (2007). Efficacy or inefficacy, that's the question: Burnout and work engagement, and their relationships with efficacy beliefs. *Anxiety, Stress, and Coping*, *20*(2), 177–196. https://doi.org/10.1080/10615800701217878

Schüler, J. (2007). Arousal of flow experience in a learning setting and its effects on exam performance and affect. *Zeitschrift für Pädagogische Psychologie*, *21*(3–4), 217–227. https://doi.org/10.1024/1010-0652.21.3.217

Scott, B. A., Barnes, C. M., & Wagner, D. T. (2012). Chameleonic or consistent? A multilevel investigation of emotional labor variability and self-monitoring. *Academy of Management Journal*, *55*(4), 905–926. https://doi.org/10.5465/amj.2010.1050

Scott, B. A., Colquitt, J. A., Paddock, E. L., & Judge, T. A. (2010). A daily investigation of the role of manager empathy on employee well-being. *Organizational Behavior and Human Decision Processes*, *113*(2), 127–140. https://doi.org/10.1016/j.obhdp.2010.08.001

Selig, J. P., & Preacher, K. J. (2008). *Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects* [Computer software]. http://quantpsy.org/medmc/medmc

Shipp, A. J., & Cole, M. S. (2015). Time in individual-level organizational studies: What is it, how is it used, and why isn't it exploited more often. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 237–260. https://doi.org/10.1146/annurev-orgpsych-032414-111245

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L., & Ezerins, M. E. (2021). The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology*, *106*(8), 1137–1155. https://doi.org/10.1037/apl0000948

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, *135*(1), 94–120. https://doi.org/10.1037/a0014212

Smillie, L. D., Yeo, G. B., Furnham, A. F., & Jackson, C. J. (2006). Benefits of all work and no play: The relationship between neuroticism and performance as a function of resource allocation. *Journal of Applied Psychology*, *91*(1), 139–155. https://doi.org/10.1037/0021-9010.91.1.139

Sonnentag, S., Mojza, E. J., Demerouti, E., & Bakker, A. B. (2012). Reciprocal relations between recovery and work engagement: The moderating role of job stressors. *Journal of Applied Psychology*, *97*(4), 842–853. https://doi.org/10.1037/a0028292

Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, *30*(2), 337–360. https://doi.org/10.1111/j.1540-5915.1999.tb01613.x

Speier, C., Vessey, I., & Valacich, J. S. (2003). The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences*, *34*(4), 771–797. https://doi.org/10.1111/j.1540-5414.2003.02292.x

Stanko, T. L., & Beckman, C. M. (2015). Watching you watching me: Boundary control and capturing attention in the context of ubiquitous technology use. *Academy of Management Journal*, *58*(3), 712–738. https://doi.org/10.5465/amj.2012.0911

Suddaby, R. (2010). Editor's comments: Construct clarity in theories of management and organization. *Academy of Management Review*, *35*(3), 346–357. https://doi.org/10.5465/amr.35.3.zok346

Twenge, J. M., Muraven, M., & Tice, D. M. (2004). *Measuring state self-control: Reliability, validity, and correlations with physical and psychological stress* [Unpublished manuscript]. San Diego State University.

van Woerkom, M., Oerlemans, W., & Bakker, A. B. (2016). Strengths use and work engagement: A weekly diary study. *European Journal of Work and Organizational Psychology*, *25*(3), 384–397. https://doi.org/10.1080/1359432X.2015.1089862

Verplanken, B. (2006). Beyond frequency: Habit as mental construct. *British Journal of Social Psychology*, *45*(3), 639–656. https://doi.org/10.1348/014466605X49122

Verplanken, B. (2018). *The psychology of habit: Theory, mechanisms, change, and contexts*. Springer. https://doi.org/10.1007/978-3-319-97529-0

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, *33*(6), 1313–1330. https://doi.org/10.1111/j.1559-1816.2003.tb01951.x

Vogel, R. M., Rodell, J. B., & Agolli, A. (2022). Daily engagement and productivity: The importance of the speed of engagement. *Journal of Applied Psychology*, *107*(9), 1579–1599. https://doi.org/10.1037/apl0000958

Wallace, J. F., & Newman, J. P. (1997). Neuroticism and the attentional mediation of dysregulatory psychopathology. *Cognitive Therapy and Research*, *21*(2), 135–156. https://doi.org/10.1023/A:1021828628571

Wallace, J. F., & Newman, J. P. (1998). Neuroticism and the facilitation of the automatic orienting of attention. *Personality and Individual Differences*, *24*(2), 253–266. https://doi.org/10.1016/S0191-8869(97)00153-0

Wang, L., Law, K. S., Zhang, M. J., Li, Y. N., & Liang, Y. (2019). It's mine! Psychological ownership of one's job explains positive and negative workplace outcomes of job engagement. *Journal of Applied Psychology*, *104*(2), 229–246. https://doi.org/10.1037/apl0000337

Wellman, N., Mayer, D. M., Ong, M., & DeRue, D. S. (2016). When are do-gooders treated badly? Legitimate power, role expectations, and reactions to moral objection in organizations. *Journal of Applied Psychology*, *101*(6), 793–814. https://doi.org/10.1037/apl0000094

Wilson, L. (2019). *Sustaining workforce engagement: How to ensure your employees are healthy, happy, and productive*. CRC Press.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*(6), 1281–1297. https://doi.org/10.1037/0022-3514.83.6.1281

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Yip, J. A., Schweitzer, M. E., & Nurmohamed, S. (2018). Trash-talking: Competitive incivility motivates rivalry, performance, and unethical behavior. *Organizational Behavior and Human Decision Processes*, *144*, 125–144. https://doi.org/10.1016/j.obhdp.2017.06.002

Yohn, D. L. (2019). Why every company needs a chief experience officer. *Harvard Business Review*. https://hbr.org/2019/06/why-every-company-needs-a-chief-experience-officer

Received January 15, 2022
Revision received June 7, 2023
Accepted June 8, 2023 ■