Journal Pre-proofs

Full Length Article

Inter-individual Differences in Intra-individual Variability in Personality Within and Across Contexts

Nadin Beckmann, Damian P. Birney, Jens F. Beckmann, Robert E. Wood, Victor Sojo, David Bowman

PII: S0092-6566(19)30129-1
DOI: https://doi.org/10.1016/j.jrp.2019.103909
Reference: YJRPE 103909

To appear in: *Journal of Research in Personality*

Received Date: 22 January 2019
Revised Date: 12 December 2019
Accepted Date: 16 December 2019

Please cite this article as: Beckmann, N., Birney, D.P., Beckmann, J.F., Wood, R.E., Sojo, V., Bowman, D., Inter-individual Differences in Intra-individual Variability in Personality Within and Across Contexts, *Journal of Research in Personality* (2019), doi: https://doi.org/10.1016/j.jrp.2019.103909

Inter-individual Differences in Intra-individual Variability in Personality Within and Across Contexts

Nadin Beckmann (Durham University, UK)

Damian P. Birney (University of Sydney, Australia)

Jens F. Beckmann (Durham University, UK)

Robert E. Wood (UTS, Australia; UNSW, Australia)

Victor Sojo (Melbourne University, Australia)

David Bowman (YSC Consulting, Australia)

Corresponding author:

Dr Nadin Beckmann

Durham University

Leazes Road

Durham DH1 1TA

United Kingdom

nadin.beckmann@durham.ac.uk

Declarations of Interest: none

1.1 Ethics

All procedures for the recruitment and treatment of participants in the current study were approved by the Ethics Committee of the UNSW.

1.2 Acknowledgement

This research was supported by the Australian Research Council (project LP0669552). The views expressed herein are those of the authors and are not necessarily those of the Australian Research Council.

1.3 Author Contributions

NB, DB, JFB, DPB, and REW have contributed to the study conceptualization and data collection. DPB, NB, JFB and VS have prepared and analysed the data. The report was written by NB, DPB, JFB and REW.

Abstract

This study focuses on intra-individual variability in personality at work, and how it relates to job performance. 288 professionals completed contextualised adjective-based personality assessments in work and non-work contexts, and a non-contextualised personality measure. Ratings of their personality were also obtained from colleagues, family members and friends. Supervisors provided performance ratings for 130 participants. Results indicate that personality is context- and source-dependent, and varies systematically within contexts intra-individually regardless of source. Whilst this variability was predictive of some performance criteria when based on other-ratings, overall predictive effects were small in number and size. This study adds to the relatively small body of research on personality variability and performance and contributes to the conceptualisation of personality as a dynamic construct.

Keywords: intra-individual variability, within-person variability, contingent personality, Big Five, other-report, frame of reference (FOR), job performance

Inter-individual Differences in Intra-individual Variability in Personality Within and Across Contexts

2 Introduction

The past decade has seen renewed interest in integrating trait and social-cognitive perspectives in the study of personality (e.g. Fleeson & Jayawickcreme, 2015; see also recent Special Issues on Dynamic Personality, e.g. Beckmann & Wood, 2017; Rauthmann, Beckmann, Noftle, & Sherman, 2019). Differences in personality are manifest as both differences between individuals (*inter*-individual variability) and differences within individuals across repeated observations (*intra*-individual variability). There is now mounting evidence demonstrating that intra-individual variability in personality is meaningful, and co-exists with inter-individual rank-order stability. Furthermore, the amount of variability in personality at the intra-individual level is often similar to or exceeds the amount of variability for the same factors at the inter-individual level (e.g. Fleeson & Gallagher, 2009; Huang & Ryan, 2011; Judge, Simon et al., 2014; Minbashian, Wood, & Beckmann, 2010; Beckmann, Wood, & Minbashian, 2010).

The majority of empirical research on personality relies on self-ratings. Whilst less is known about observer ratings, they promise to offer important additional insights into personality (e.g. Beer & Vazire, 2017; McAbee & Connelly, 2016; Rauthmann, 2017; Vazire, 2010). For instance, information about how we are perceived by others at work matters for performance and how it is evaluated (Connelly & Ones, 2010). Observer perceptions of *intra-individual variability* in particular are not well understood. The current paper contributes to knowledge and understanding of variability in personality, as experienced by the self and observed by others in work and non-work contexts. We consider the validity of variability

indices and implications for predicting performance at work under conditions that are meaningful and of medium stakes for participants and their observers.

In the following sections we first present evidence for why variability due to context (work/non-work) and rating source (self/other) is important. We then focus on individual differences in intra-individual variability, the operationalisation of intra-individual variability, and the relationship of intra-individual variability with outcome variables of interest.

2.1 Cross-context Variability

Whilst a work context comprises diverse sets of situations to which individuals might respond adaptatively; situations in non-work contexts are arguably even more diverse than work situations (Beer & Vazire, 2017). These differences in context-related situational diversity impact on how individuals experience their own personality as it is manifest in their thoughts, feelings and behaviours, as well as how others perceive and interpret nuances in their behaviours. Importantly however, individuals also differ in their responsiveness to situational cues and consequently in the amount of response variability they exhibit. This has been shown to be linked to their psychological functioning (e.g. Lievens, Lang, De Fruyt, Corstjens, & Van de Vijver, 2018; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007; Hardy & Segerstrom, 2017). In addition, a number of studies have provided evidence for within-person situation contingencies as individual differences variables in their own right (e.g. Berenson, Downey, Rafaeli, Coifman, & Paquin, 2011; Huang & Ryan, 2011; Huang & Bramble, 2016; Judge et al., 2014; Sherman, Rauthmann, Brown, Serfass, & Jones, 2015; Minbashian et al., 2010; Wood, Beckmann, Birney, Beckmann, Minbashian, & Chau, 2019). Clearly, situation-specific response variability captures an important part of personality.

Two principles have been suggested to explain cross-situational variability in personality-related responses; the *situation strength* principle (e.g. Mischel, 1973) and *trait*

activation (Tett & Burnett, 2003). The situation strength principle refers to variability at the *group-level* explained by the situation. Inter-individual differences are under-represented in this perspective. For example, in a strong situation most individuals are expected to show the desired behaviour, e.g. behave conscientiously given the appropriate situational cues (incentives or punishments). Whilst the situation strength principle acknowledges the impact of the situation, it does not explicitly speak to individual differences in the responsiveness to situations, which is what the current study is focussed on. Similarly, trait activation does not assume traits to vary within-person but merely states that trait-relevant responses will only manifest if the situation allows for this to happen. Individual differences in responsiveness to situations are acknowledged only in the sense that those individuals higher on the stable trait are more likely to show trait-relevant responses compared to those lower on the trait. This constitutes a between-person perspective.

Both principles have been used to explain in what situations traits predict best, which is when people can (in weak situations) and are encouraged to (in situations that present trait-relevant cues) respond in ways that correspond with their (stable) traits. From this perspective, it is the situation that changes, not the person. Individuals may, however, also differ in the extent to which they show variability, i.e. how much they adapt and adjust to different situational requirements within a context (e.g. at work). In other words, whilst we accept that situations may differ in their characteristics, in the current study we are specifically interested in individual differences in the responsiveness to the same range of situations (here operationalised using a work frame of reference). As we outline in more detail below, variability is hence operationalised at the individual level (rather than the group-level) in the form of person-specific variability indices.

Taking a person perspective, we are interested in a person's variability in trait-manifestations as indicated by a person-specific variability index. We are interested to

establish whether such individual differences do exist and, if so, whether such person-specific variability is systematic and meaningful. In that sense, we aim to go beyond the more fixed approaches that capture situation-related variability at the group-level.

2.1.1 Contextualised Personality Measures

Studying personality in different contexts (e.g. work, school, home) requires contextualised measurement. Contextualised personality measures can take various forms. One approach is to modify the inventory instruction so as to encourage participants to apply a specific frame of reference. For instance, Hunthausen, Truxillo, Bauer, and Hammer (2003) instructed participants to think about how they are at work when responding to the personality measure. Another, commonly used approach is to simply add a context tag ('at-work', 'at-school') to each personality item (e.g. Schmit, Ryan, Stierwald, & Powell, 1995; Holtz, Ployhart, & Dominguez, 2005).

A number of studies have shown that contextualised personality measures are more valid predictors of various outcome variables of interest when compared to their non-contextualised counterparts. For instance, contextualised personality measures outperformed non-contextualised personality measures in the prediction of job performance (meta-analysis, Shaffer & Postlethwaite, 2013; Bing, Davison, & Smothers, 2014), work-related creative problem solving (Pace & Brannick, 2010), job satisfaction (Heller, Ferris, Brown, & Watson, 2009; Bowling & Burns, 2010), work frustration, turnover intention and absenteeism (Bowling & Burns, 2010), and organizational citizenship behaviour (Wang & Bowling, 2016). Similar findings have been observed in academic contexts (e.g. Reddock, Biderman, & Nguyen, 2011; Bing, Whanger, Davison, & VanHook, 2004; Schmit et al., 1995; Lievens, De Corte, & Schollaert, 2008; Holtrop, Born, de Vries, & de Vries, 2014; Robie, Risavy, Holtrop, & Born, 2017). A likely explanation is that the context of measurement is more closely related to the criterion context. Rather than capturing the full scope of a given personality

trait, contextualised measures may only capture the range that is relevant to the context, i.e. their content validity may be different; although this will depend on the specific operationalisation, or contextualisation, of the measure. Whilst conceptually plausible and demonstrated in the majority of empirical research, a few studies report absences of such contextualisation effects (Robie & Risavy, 2016; Kurtz & Palfrey, 2016; Holtrop, Born, & de Vries, 2014).

2.1.2 Intra-individual Differences Across Contexts

The idea of contextualising personality measures rests on the assumption that people think, feel, and behave differently, at least to some extent, in different situations or contexts. In line with this assumption, systematic mean differences have been observed between contextualised and non-contextualised personality measures for some traits (Schmit et al., 1995; Robie, Schmit, Ryan, & Zickar, 2000; Lievens, 2008; Fisher, Cunningham, Kerr, & Allscheid, 2017). Alternative explanations for such contextualisation effects have been proposed, including social desirability and improved self-presentation (Bing et al., 2004; Schmit et al., 1995), and increased internal consistency of contextualised measures leading to better predictive validities (Lievens et al., 2008; Fisher et al., 2017).

Only a few studies have investigated *intra-individual* differences across specific frames of references or contexts. For instance, Heller et al. (2009) explicitly compared measures for an 'at work' with an 'at home' context. Such comparisons matter because contrasting work with home comes closest to the diversity of typical situations encountered in an adult's life. In the Heller et al. (2009) Study 1, a community sample of 147 recently married couples were asked to describe themselves at work and at home using an adjective list (Goldberg et al., 1992). Systematic context related differences were observed, such that participants reported to be more conscientious, more open to experience, and less extraverted at work than at home. However, in this study (Heller et al., 2009) and another by Lievens et al. (2008), significant

and moderate cross-context correlations between contextualised scales have also been reported suggesting a considerable level of rank order stability. Nonetheless, some shifts in rank order across contexts must have occurred given there were differences in the predictive validity indices between the two contextualised measures.

In sum, there is evidence to suggest that individuals vary systematically in their personality from context to context (mean-level differences), but at the same time also show rank order stability in their personality responses, as evidenced in the cross-context scale correlations (e.g. Lievens et al. 2008; Heller et al., 2009). In other words, and as has been pointed out repeatedly (e.g. Clifton, 2014; Fleeson, 2001; Fleeson & Jayawickreme, 2015; see also Japyassu & Malange, 2014 on variability concepts in evolutionary biology), intra-individual, cross-situational variability coexists with inter-individual stability in personality responses. Consequently, it is important to differentiate context factors from other determinants of intra-individual variability. One such determinant, to which we now turn, is the source of the rating (self/other).

2.2 Cross-source Variability

When asked to provide summary personality evaluations, raters are likely to draw on situation-specific experiences and perceptions. This is true regardless of whether self-ratings or observer ratings are obtained, but to different extents. The experience of situation-contingency and its role in the judgment of personality as perceived by others, can be thought of as being filtered through two processes. The first is a closeness effect, and the second is an observability filter.

2.2.1 Closeness Effect

There is both convergence and divergence in personality ratings across different raters (self vs other, and other vs other raters, e.g. Connelly & Ones, 2010). A considerable amount

of variance in other-ratings is due to the individual observer (Rauthmann, 2017). This might be meaningful variance where target individuals interact with different observers in different contexts. Observers have the opportunity to evaluate the target individual's response to cues relevant in their shared context, and hence such inter-rater differences may present meaningful, context-related variability in the targets' behaviour. Alternatively, differences associated with a specific observer might be seen as an observer's idiosyncratic "bias". Such idiosyncratic "distortions" can be estimated and addressed by aggregating across multiple observers in a given context. Regardless of context, a more robust and theoretically substantive finding is that observer ratings tend to be more positive (i.e. favourable) when compared to self-ratings, particularly when ratings are elicited from well acquainted observers or observers that like the target individual (Leising, Erbs, & Fritz, 2010; Leising, Gallrein, & Dufner, 2014)[1]. For instance, observers tend to evaluate the target as being higher on desirable traits such as emotional stability and conscientiousness, than the target evaluates themselves. If one were to interpret the difference between self and other ratings as indication of a bias of some sort, one would need to consider that such difference could be indicative of a "positivity bias" on the side of others or, alternatively, a "self-deprecation bias" of the self. In any case, this seems to be a uniform effect such that the between-person rank order for the assessed attribute remains relatively stable.

2.2.2 Observability Filter

Self and other ratings each capture important components of the person. Some personality components might be observable by others and not necessarily be perceived by the self (e.g. reputation effects, McAbee & Connelly, 2016); other personality components

[1] Similar biases have been observed in other areas also, such as in evaluation of the creative products of others (e.g. Birney, Beckmann, & Seah, 2016).

might only be known to the self but not to others (e.g. identity/façade effects; McAbee & Connelly, 2016). The validity of self as compared to other ratings varies by trait. Both observability and evaluativeness (social desirability) of a given trait likely impact which source (self or other) is more valid for trait assessment (Vazire, 2010). For instance, extraversion is generally better observable (than, e.g. neuroticism) and hence other-ratings tend to be more accurate predictors of extraversion (Beer & Vazire, 2017), or at least correspond well with self-ratings. At the same time other ratings tend to be better predictors of outcome variables such as job performance than self-ratings (Connelly & Ones, 2010). Of course, job performance is also typically rated by others (i.e. supervisors) and hence observability matters for both predictor and criterion. Note that whilst observer ratings might involve a positive "bias", as long as the between-person rank order for the given trait is not unduly impacted, predictive validity coefficients of observer ratings are likely not affected.

In sum, given the closeness effect, it seems sensible to employ a number of well acquainted and knowledgeable observers from different contexts (work, non-work). Using context-specific aggregates of observer ratings reduces the impact of potential idiosyncratic rater-specific biases, whilst still acknowledging meaningful inter-rater differences due to context (work vs non-work). For a work context outcome variable, such as job performance, other-ratings are particularly relevant as they might show stronger predictive validity than self-ratings. The impact of the observability filter needs to be examined experimentally.

2.3 Intra-individual Variability as an Individual Difference Variable

Mean-level differences notwithstanding, individuals also differ in how much variability they show in their thoughts, feelings and behaviour (differential effects). The notion of intra-individual variability as an individual difference construct has a long tradition in personality (Fiske & Rice, 1955; for an overview see e.g. Lievens et al., 2018; Dalal, Meyer, Bradshaw, Green, Kelly, & Zhu, 2014), ability (e.g. Guthke, 1992; Guthke & Beckmann, 2003; Paulus &

Martin, 1988; Martin, & Rubin, 1995; Martin, Anderson, & Thweatt, 1998), and social-cognitive research (Gangestad & Snyder, 2000; Rosenberg, 1965; Webster, Smith, Brunell, Paddock, & Nezlek, 2017). Different operationalisations of intra-individual variability or (in)consistency have been suggested. As described next, these include one-off measures asking participants directly how variable they are, and more indirect measures of response variability as observed within and across measurement occasions (e.g. cross-role, cross-social partner, cross-item, cross-context, cross-time).

2.3.1 Direct Approaches

Direct approaches underlying operationalisations of intra-individual variability often utilize "one off" measures in which data are collected in one sitting. For example, participants might be asked to describe their personality as experienced in various social roles they inhabit (e.g. as a friend, as a co-worker; Donahue, Robins, Roberts, & John, 1993; Baird, Le, & Lucas, 2006), or in relation to specific individuals in their social network (Clifton, 2014). As described above, context effects have been assessed in this way. Of course, direct approaches require a considerable level of self-awareness from respondents, the ability to reflect on one's thoughts, feelings and behaviour across one or more social contexts, and a sufficient willingness to share these reflections. An alternative, though more time-consuming approach is to ask participants directly how frequently and to what extent personality items (e.g. 'I am the life of the party') were descriptive of them over a defined period of time (e.g. the past 6 months), and to use such frequency-based estimates to derive indicators of intra-individual variability (Fleisher, Woehr, Edwards, & Cullen, 2011). More recently, situational judgement tasks have been added as a promising one-off non-questionnaire-type method to capture intra-individual variability, as they allow presenting participants with selected trait-relevant situation descriptions to actively elicit personality-related responses cross-situationally

(Lievens et al., 2018). However, these approaches are challenged by the need for valid and reliable respondent-relevant situational scenarios, and still require sufficient self-awareness.

2.3.2 Indirect Approaches

We consider two indirect approaches, inter-item variability and experience sampling. In terms of the former, the consistency with which a participant responds to different items of the same personality scale can be conceived as a measure of intra-individual variability (or "traitedness") under the assumption that different items represent different situations in which one might show trait-relevant responses (e.g. Reddock et al., 2011; LaHuis, Barnes, Hakoyama, Blackmore, & Harman, 2017). This is a useful approach because it enables researchers to derive variability estimates from any personality assessment implicitly without requiring differential self-reflection on the part of the respondent with regard to their variability.

All the approaches to operationalising intra-individual variability we have so far discussed have in common that data is being captured within one measurement occasion. They do however vary in the extent to which self-awareness and reflection is required of the participant, i.e. how much the measurement process is based on a self-assessment of variability vs elicited response behaviour. Rather than asking participants how variable they think they are, or how they perceive themselves across roles or contexts, an arguably more convincing approach is to use designs that actually assess fluctuations in personality responses across time, such as experience sampling, and hence capture variability in situ (Fleeson, 2001; Wood et al., 2019). A disadvantage of experience sampling is the time and resources required to collect the data, and validity-related concerns that come with asking participants to complete a measure repeatedly within a relatively short period of time (e.g. participant fatigue, reactance, non-random attrition). It has also been suggested that in experience sampling designs, variability that is due to response styles might be misinterpreted

as intra-individual variability (Baird, Lucas, & Donnellan, 2017), though arguably this would apply to all forms of repeated measurement approaches (within and across measurement occasions). However, encouragingly, some evidence is available to suggest that one-off intra-individual variability indices are related to intra-individual variability estimates derived from experience sampling (Baird et al., 2006; Lang, Lievens, De Fruyt, Zettler, & Tackett, 2019; Lievens et al., 2018; Webster et al., 2017), and therefore might be a tool of choice in field studies when an experience sampling design is not feasible.

Of the different approaches discussed above, the most common method used to describe individual differences in intra-individual variability is to calculate person-specific intra-individual standard deviation scores (computed across items within a scale, or across social roles, contexts or measurement occasions). Whilst an intuitive approach, it is not without problems. For instance, the observed degree of across-measure variability (SD scores) is often related to the mean of the scale (e.g. a high mean level on the neuroticism scale indicates consistency in endorsing neuroticism items). In an early study, Baird et al. (2006) demonstrated that negative associations between intra-individual personality variability indices and measures of well-being were attributable to the effect of the mean; correlations were substantially reduced or disappeared when mean-corrected variability estimates were used in the analyses. This was later replicated (Magee, Buchtel, Human, Murray, & Biesanz, 2018; Baird et al., 2017). Thus, it seems to be advisable to correct for means.

More recently, item-response theory (IRT) approaches have been introduced that might offer some additional flexibility and sophistication in modelling intra-individual variability when conceptualised as within-scale item response consistency (e.g. LaHuis et al., 2017; Lang et al., 2019). In the current study, we use the relative variability index (Mestdagh, Pe, Pestman, Verdonck, Kuppens, & Tuerlinckx, 2018) to operationalise intra-individual

variability as an individual differences variable both from the self as well as the observer perspective, as outlined in more detail below.

2.3.3 Intra-individual Variability as a Predictor

Given that intra-individual variability can be framed as an individual difference variable, its relationship with outcome variables such as well-being and performance has been of interest. Three types of association are conceivable, a negative, positive, or no relationship. *A negative relationship* would suggest intra-individual variability to be indicative of vulnerability, instability, or a weak sense of self, and hence a risk factor or maladaptive strategy. A number of studies on variability in personality and affect seem to be in line with this perspective (Donahue et al., 1993; Suh, 2002; Kuppens et al., 2007; Reddock et al., 2011; Hardy & Segerstron, 2017; for a meta-analysis see Bleidorn & Koedding, 2013). For instance, Hardy and Segerstron (2017) reported intra-individual variability in negative affect (collected from the daily experiences of a community sample over eight consecutive days) to be associated with psychological distress and physiological ill health indicators (after controlling for the mean level of negative affect and number of stressor days). Interestingly, when conceptualising intra-individual variability as contingent on contextual cues, the ability to match one's experience of negative and positive affect with the occurrence of negative events, positive associations with well-being outcomes were found. Other research by Reddock et al. (2011) has similarly shown vulnerability links in a student sample, where intra-individual variability (as cross-item variability) in personality was negatively related to cognitive ability and academic performance.

Positive relationships would be expected if intra-individual variability were to indicate flexibility in responding to environmental cues. If this is an adaptive strategy, it should be reflected in positive correlations with well-being and performance measures. For instance, in a recent experience sampling study with students, Magee et al. (2018, Study 1) found that

cross-occasion intra-individual variability in personality states, when corrected for the effect of the mean, was positively associated with two out of three measures of psychological adjustment. When analysing cross-occasion intra-individual variability for each Big Five dimension separately, positive associations with psychological adjustment measures were found for openness, extraversion and neuroticism (and none of the tested associations were negative). However, the reported effects were typically small in size. Lievens et al. (2018) observed a small positive correlation (r = .19) between intra-individual variability in sociability and dutifulness and peer-rated performance in a student sample (Study 1), and a small to moderate correlation (r = .37) between intra-individual variability in personal initiative and supervisor ratings of job performance in an employee sample (Study 3). In both studies intra-individual variability (here operationalised using a situational judgement task paradigm) incrementally added to the prediction of performance after controlling for mean level scores of the relevant traits.

Finally, no relationship would indicate that outcomes such as job performance or well-being vary independently of intra-individual variability in thoughts, feelings and behaviour. Given publication bias against null findings, it is not surprising there is not much research available that suggests a zero relationship. However, recall that negative correlations observed between intra-individual variability and outcomes of interest have been shown to be reduced or even nullified when means were partialled out (Baird et al., 2006; Baird et al., 2017; Magee et al., 2018). This speaks to the importance of considering moderation of personality levels when exploring intra-individual variability as an individual differences construct.

In sum, the majority of studies have focused on implications for health and well-being, fewer studies have tested relationships with performance. Most studies use student samples. Initial findings implied intra-individual variability to represent a liability, particularly with regard to well-being outcomes. More recent work suggests a more nuanced picture with

possible benefits in terms of well-being and performance. Differences in findings might partially be due to differences in operationalisations used to measure intra-individual variability, and possibly the content or substantive area of interest (e.g. safety behaviour vs sociability, see Lievens et al., 2018). Overall however, it remains unresolved whether the more variable or the more consistent person is a happier and more productive one.

2.4 The Current Study

Our conceptual focus is on intra-individual variability in the manifestation of personality at work. In a work context, consistency/rigidity vs. variability/flexibility of one's behaviour forms an important component of the judgment of one's work performance by others. The "traditional" approach has been to centre on how, on average, levels of openness, conscientiousness, agreeableness, extraversion, and neuroticism vary across contexts. Given supervisors tend not to know how their employees think, feel, and behave outside work, changes between work and non-work contexts seem rather irrelevant. Rather than focus on between-context variability, we focus on the variability within a specific context; at work. If one considers the possibility that item responses are manifestations of personality primed by situations in which *the target (self or other) of the rating has been observed*, then variability becomes an outcome of interest in and of itself (Reddock et al., 2011). Our approach is premised on the view that when items, rather than total scores are taken as fundamental, a more nuanced account of personality may be possible. In our study this is achieved by treating responses to items per personality factor as measurement occasions of various manifestations of the respective personality trait. Such operationalisation creates person-specific intra-individual variability indices (IVI) for each personality factor across the experimental conditions realised in our study design (self vs other ratings in work vs non-work contexts).

We assess variability indirectly, by deriving an indicator of intra-individual variability from self- and other-descriptions of trait manifestations within a given context (e.g. work).

Our rationale behind this operationalisation of variability is that different adjectives (across the 60 items of the personality scale used here) trigger recall of experiences in situations that are relevant to the thoughts, feelings and behaviours associated with this adjective. Items vary with regard to their situational prototypicality. For instance, "anxious" makes implicit reference to a slightly different set of situations than "self-conscious", yet both adjectives are subsumed under Neuroticism. In other words, items *implicitly* refer to situations in which these trait manifestations are relevant. Adopting this perspective allows to sample responses across different situations. We are interested in whether variability in thoughts, feelings and behaviours operationalised by using items as triggers of situation-specific experiences is (a) substantive, (b) systematic, and (c) meaningful.

As discussed, a number of operationalisations are available to measure within-person variability. They come with strengths and weaknesses. A disadvantage of experience sampling field studies is that typically there is not much control over the situations that participants encounter at work. This matters as indicators of within-person variability (responsiveness to situations or contexts) reflect situation variability, such that when an individual shows low responsiveness (based on ESM indicators), this may be because they tend to be relatively stable in their responses (i.e. do not adjust to changing situations at work), or because the situations they encountered happen to have been very stable during the period of study (often one or two weeks). An advantage of the current approach, in addition to its efficiency and lower participant burden, is that it reflects better the typical experience at work given that it relies on summary statements which cover the broad range of experienced / observed trait manifestations and, indirectly, the underpinning situations encountered at work. Whilst it is difficult to disentangle person and situation effects in field studies when situations vary between people within the sample, this is an issue that affects the current study somewhat less.

Aiming for ecological validity of our study findings, we employ a sample of high-performing working professionals who completed assessments of their personality in work, non-work and generic contexts. To ensure data collection was meaningful, relevant to participants, and of medium (rather than low or high) stakes, observer personality assessments were submitted anonymously by multiple colleagues, friends and family and only shared with participants in aggregated form.

2.4.1 Expectations

Given prior research as outlined above, we have two sets of expectations. First, at the group-level, we expect reported personality to vary as a function of context (work vs non-work) and source (self vs other). Second, we expect *intra-individual variability* in personality responses to be systematic and meaningful. We test this using person-specific trait variability estimates based on self- and other-ratings to predict supervisor evaluations of job performance. Our assumption is that individuals differ in their responsiveness to different situations within a work context, and this is expected to have implications for their functioning at work, including their performance as observed by their supervisors. Whilst we expect person-specific variability estimates to be predictive of job performance, particularly variability related to the two main personality predictors, conscientiousness and neuroticism (e.g. Barrick & Mount, 1991; 2000; Barrick, Mount & Judge, 2001), we do not make any specific predictions with regard to the directionality of such effects, given the inconclusiveness of prior findings with regard to the adaptive versus maladaptive nature of intra-individual variability in personality and affect. As one's standing on the trait matters in terms of the amount of observable intra-individual variability (e.g. Baird et al., 2006; Lievens et al., 2018), we also test whether trait-specific intra-individual variability effects on performance are moderated by trait level.

3 Materials and Methods

3.1 Participants

We recruited a sample of 288 professionals who participated in a postgraduate-level management expertise development program. The sample is a subset drawn from the *A.L.L. Flexible Expertise data base*[2] (*N* = 423). Participants were working in middle-level management roles (24 to 57 years, *M* = 34.77, *SD* = 6.49, 37.5% female) at large Australian companies (an internationally operating bank, a major international airline, an insurance company, and a national broadcasting company). On average, participants had 5.6 years (SD = 4.58) of experience in management and had worked 2.09 years (*SD* = 1.94) in their current role within the respective organisation. Of these, 75% had completed a university degree (33% postgraduate; 42% undergraduate) and 13% reported 'high school' as their highest level of education. The remaining 12% of participants reported having completed a different degree ("other") or did not indicate their degree (1%). Participants also recruited 944 observers across work and non-work contexts which we describe in more detail below.

3.2 Measures

3.2.1 Context-specific Measures

In order to assess Big Five personality traits in work and non-work contexts and from the self and observer perspectives, five questionnaires were developed.

Self-report measures. An adjective list based on existing inventories (e.g. Goldberg et al., 1992) was compiled resulting in sixty adjectives that reflected the five broad dimensions

[2] The Accelerated Learning Laboratory (A.L.L.) conducts expertise research and provides a 2-year leadership training program for mid-level managers from large organizations. The assessment and professional development component is a core feature of the program and has a theory based, elaborated assessment-for-learning focus. The overarching objective was to foster the development of flexible expertise in managerial leadership that extends beyond domain-specific routine expertise.

of personality (neuroticism, conscientiousness, agreeableness, openness to experience and extraversion, see Appendix A). Care was taken to ensure that all facets were represented (Costa & McCrae, 1992). The answer format was a visual analogue slider scale with the anchors 'extremely inaccurate' and 'extremely accurate'. Responses were subsequently translated into a numeric scale from 0 to 100. The same adjective list was presented in three different versions: one referring to the respondent's work context ('work'), one referring to the respondent's personal life ('non-work'), and one referring to no specific context ('generic'). The generic version served as a control condition. Participants were instructed to rate how accurately each adjective describes themselves in these different contexts, as follows.

> Self in work context: *"Please use the list of common human traits to describe your behaviours and interactions with others AT WORK. We are interested in how you see yourself within a work context*." Each item was then prefaced with "*Use the sliding scale to indicate how accurately the term below describes you AT WORK*".
>
> Self in non-work context: "*Please use the list of common human traits to describe your behaviours and interactions with others IN YOUR PERSONAL LIFE. We are interested in how you see yourself outside of a work context (i.e., with family, friends, and new acquaintances)".* Each item was prefaced with *"Use the sliding scale to indicate how accurately the term below describes you OUTSIDE WORK*".
>
> Self generally: *"Please use the list of common human traits to describe your behaviours and interactions with others. We are interested in how you see yourself*". Each item was prefaced with *"Use the sliding scale to indicate how accurately the term below describes you."*

Internal consistencies for the five dimensions were acceptable for each context (work context: $N = 282$; Cronbach's $\alpha_N = .83$, $\alpha_C = .87$, $\alpha_O = .73$, $\alpha_A = .74$, $\alpha_E = .83$; non-work

context: $N = 278$; $\alpha_N = .82$, $\alpha_C = .86$, $\alpha_O = .83$, $\alpha_A = .70$, $\alpha_E = .85$; generic context: $N = 284$; Cronbach's $\alpha_N = .80$, $\alpha_C = .84$, $\alpha_O = .74$, $\alpha_A = .70$, $\alpha_E = .86$).

Other-report measures. Two variants of the adjective list were used for the other-report measures. One was completed by the participant's work colleagues, and the other by their family and friends. The instructions were as follows.

Others in work context: *"Please use this list of common human traits to describe the person you are rating's behaviours and interactions with others AT WORK. We are interested in how you see the person you are rating within a work context."* Each item was then prefaced with "*Use the sliding scale to indicate how accurately the term below describes the person you are rating at WORK*".

Others in non-work context: *"Please use this list of common human traits to describe the person you are rating's behaviours and interactions with others IN THEIR PERSONAL LIFE. We are interested in how you see the person you are rating outside of a work context (i.e. with family, friends and new acquaintances)."* The non-work items were prefaced with: *"Use the sliding scale to indicate how accurately the term below describes the person you are rating OUTSIDE WORK*".

We derived the average item response for observers for each participant in each context to calculate the internal consistencies of these personality scales. The internal consistencies for the five dimensions were acceptable for both contexts (work context: $N = 203$; Cronbach's $\alpha_N = .91$, $\alpha_C = .90$, $\alpha_O = .74$, $\alpha_A = .85$, $\alpha_E = .86$; non-work context: $N = 200$; $\alpha_N = .84$, $\alpha_C = .87$, $\alpha_O = .80$, $\alpha_A = .71$, $\alpha_E = .78$).

Personality Indices. The derivation of the personality indices entailed two steps. First, item-level response-aggregates were derived by averaging across ratings, separately for each of the others-work and others-non-work conditions (aggregation was not necessary for the self-work and self-non-work conditions). These aggregates are the basis of the internal

consistency estimates reported above. Second, for each of the four conditions (self-work, self-non-work; others-work, others-non-work), we calculated two person-specific scores across the 12 items for each trait – a level score and a variability score. The level score is simply the mean of the 12 item-responses (item-response aggregates for other-ratings) for a given context. The variability score reflects the variability across those same 12 item-responses (item-response aggregates for other-ratings) around the individuals' mean, and is operationalised in SD units. We refer to this latter score as the Intra-individual Variability Index (IVI). The size of IVI is functionally linked to the mean. That is, the further away the mean is from the scale mid-point the more constrained the intra-individual variability is going to be. As this is primarily a technical necessity but not necessarily a conceptual one, we use a *relative variability index* (i.e. a relative SD) that reflects the proportion of observed variability relative to the maximum possible variability given the observed mean (Mestdagh et al., 2018). Such transformation does not create an artificial independence of the variability from the mean (as a partial correlation would), it merely redresses the effect of the boundedness of the trait scale (in our case 0 to 100) that results in a functional dependency of variability from the mean (Mestdagh et al., 2018, p. 695). It is important to also acknowledge that this functional dependency is not reciprocal, that is, the mean is not a function of variability.

3.2.2 Baseline Big Five Personality Scale

As part of the management expertise development program, participants were assessed on a range of cognitive and personality variables at baseline. The reported internal consistencies were assessed in a larger sample of which the current participants are a sub-sample, see footnote 2. For the purpose of the current study the widely used and validated International Personality Item Pool (IPIP) version of the NEO inventory was employed to gauge the validity of the adjective lists. The IPIP NEO inventory (Goldberg et al., 2006; see http://ipip.ori.org/) is based on the five-factor model of personality (Costa & McCrae, 1992)

and contains 50 items assessing five broad dimensions of personality (neuroticism, conscientiousness, agreeableness, openness to experience, and extraversion). Participants were instructed to describe themselves as they generally are compared to other people of the same sex and roughly the same age. The answer format for all items was a visual analogue scale that required participants to place a marker along a line with the polar ends labelled 'strongly disagree' to 'strongly agree'. The visual analogue scale was later translated into a numeric scale from 0 to 100 (Cronbach's $\alpha_N = .85$, $\alpha_C = .87$, $\alpha_A = .76$ $\alpha_O = .78$, $\alpha_E = .88$).

3.2.3 Job Performance Measure

Job performance was assessed using supervisor ratings. A 50-item questionnaire was created to assess four dimensions of job performance: Leadership (14 items), Problem Solving (14 items), Emotional Competence (11 items), and Impact (12 items) (see Appendix B). Ratings were placed on a visual analogue scale that was anchored by the labels 'not at all accurate' to 'extremely accurate', and later translated into a numeric scale from 0 to 100. Internal consistencies were high for all subscales (Cronbach's alpha: $\alpha_{leadership} = .96$, $\alpha_{problem\text{-}solving} = .97$; $\alpha_{emotional\text{-}competence} = .95$; $\alpha_{impact} = .95$). An exploratory factor analysis with ML extraction of four factors and oblique-rotation indicated the factors accounted for 67.92% of the variance in the data. The four factors mapped on to the four dimensions as expected. Factor-intercorrelations ranged from .42 to .56 suggesting separation in the factors in capturing dominant aspects of job performance (see Appendix B, Table B1-C).

3.3 Design and Procedure

The study was undertaken in the context of a two-year postgraduate-level management expertise development program. A repeated measurement design was employed. Participants took part in 4 sessions. In the first session, and at the beginning of the two-year programme, participants completed a number of baseline measures, including a demographics

questionnaire and the IPIP inventory relevant to the current study. About six to nine months later participants took part in a 3-day training and development module. Over the course of the three days, they completed the generic personality measure (control) in one session and the work-and non-work personality measures in subsequent sessions. The order of presentation for the work and non-work self-report measures was counter-balanced. Participants then were instructed to nominate up to five individuals (three from their work environment and two from their personal life) who knew them well and could therefore comment on the participant's personality in a specific context. Within their work environment they were encouraged to nominate (a) an individual who was in a more senior position than they were, (b) an individual who was in a position of approximately the same level as they were, and (c) an individual who was in a less senior position than they were. Within their personal life context participants were asked to nominate (a) an individual with whom they were in a close personal relationship (e.g. a spouse or family member who lives in the same household), and (b) an individual with whom they share a different type of relationship, yet who knows them well (e.g. a close friend). Nominated individuals (which we refer to as 'observers') were then contacted by the authors and asked to complete an online questionnaire. They responded to a range of demographic questions first (e.g. age, gender, position relative to or relationship with participant and how well they knew the participant), and then completed the questionnaire by indicating how accurately they believed each adjective described the target participant in the relevant context. Taken together, up to eight completed personality reports were available for each participant (3 self-reports completed by the participant, and up to 5 other-reports completed by different observers). Finally, about 12 months after completion of the personality data collection process supervisor ratings of the

participant's job performance were collected. Due to the longitudinal nature of the study we were able to gather supervisor ratings for only a subset of N = 130 participants[3].

3.4 Data Analyses

We first analysed the validity of the adjective list that was developed for the purpose of the current study by analysing its correlation patterns with the IPIP NEO inventory. We then examined descriptive statistics of the sample of observers that our participants recruited in order to judge whether observers were in a position to rate the participant's personality.

The focal analyses are reported in two sections. Our first set of analyses explored the degree of systematicity in the personality indices as a function of the contextual frame and source. In a subset of the data, our second set of analyses investigated the extent to which these personality indices account for incremental validity in the four job performance criteria as rated by the participants' supervisors. Note, the generic context condition was excluded from the second set of analyses as we were interested in the four different source-context constellations (self-work, self-non-work, others-work, others-non-work). Analyses were conducted for participants with available supervisor ratings. As common in applied settings, our sample was a convenience sample. The smallest sub-sample for which analyses were conducted was N = 124. An effect of at least partial η^2 = .061 (a medium effect) is detectable with sufficient statistical power (1 - β = .80; α = .05) in a sample of this size (G*Power, Faul, Erdfelder, Buchner, & Lang, 2009). More detail about the analyses is provided in the results section. All analyses were run in R 3.61 (R Core Team, 2019).

---

[3] The study was not preregistered. In accordance with our ethics obligations (HREC HC06294) at the time of data collection, we are required to store electronic data in a password protected location on the university's internal computer server. Accordingly, we are unable to make the data freely available. However, data and analyses scripts will be made available on request.

4 Results

4.1 Validity of the Big Five Adjective List

Table 1 presents the Pearson correlations between the Big Five dimensions assessed in the current study with the adjective list based self-report measure compiled for this study and the IPIP NEO inventory (Goldberg et al., 2006) to provide an indication of measurement validity. To calculate these correlations, we used the scores from the adjective list that was presented without a specific context ('generic', i.e. the control condition). All five adjective scales correlated moderately to strongly with the corresponding IPIP scales (r = .57 to r = .77, $N_{listwise}$ = 283), and comparably less with the non-corresponding IPIP scales (r = .00 to r = -.41).

[Insert Table 1 about here]

4.2 Observers

Overall 563 work observers (62% male) aged between 23 and 83 years (M = 39.60, SD = 8.76), and 381 non-work observers (55% female) aged between 17 and 70 years (M = 37.73, SD = 9.36) provided personality reports (M = 4.60 observers per participant, SD = .78, ranging from 1 to 5 observers per participant).

Work observers. The vast majority of work colleagues (98%) reported to be in regular contact with the participant, either every day, several days a week, once a week or once a fortnight. Seventy-six per cent of work colleagues reported to have no contact with the participant outside work, and amongst those who were in contact outside work, the majority reported no regular contacts (82%). This supports the notion that the colleagues' judgements were mainly based on their observations of the participant in one context, i.e., at work. On

average, work observers evaluated their relationship with the participant as close (M = 75.96 on a scale from 0 to 100, SD = 16.26), and reported to know the person they were rating well (M = 73.99 on a scale from 0 to 100, SD = 15.73). The average time of working together was reported as being about 3 years (SD = 3.4 years) ranging from 2 months to 26 years.

Non-work observers. Almost half of the group of non-work observers (42%) indicated they lived in the same household as the participant at the time of the study. The vast majority (88%) reported to be in regular contact with the participant; either every day, several times a week, at least once a week, or once a fortnight. On average, non-work observers rated their relationship with the participant as very close (M = 86.57 on a scale from 0 to 100, SD = 14.87), and indicated to know the person they were rating very well (M = 87.29 on a scale from 0 to 100, SD = 13.49). The average duration of the observer-participant relationship was of about 15 years (SD = 10) ranging from 6 months to 58 years.

In sum, the observers selected by the participants were in regular contact with the participant, knew them well and felt close to the participant. This nurtures optimism that the intensity of contacts across a wide range of situations either at work or outside work enabled observers to pass valid judgments about the participant's personality in the respective context. For the majority of work observers, the contact was limited to contact at work suggesting that their ratings reflect the participant's personality at work rather than outside work.

4.3 Personality as a Function of Context and Source

4.3.1 Descriptive Statistics and Preliminary Analyses

The means and SDs for the level and variability indices, broken down by trait, source of the rating, and contextual frame, are reported in Table 2 and Table 3, respectively. To remind the reader, the level scores are the person-specific means across the 12 item-responses (item-response aggregates for other-ratings) for a given context. The variability scores reflect

the person-specific proportion of variability across those same 12 item-responses (item-response aggregates for other-ratings) around the individuals' mean relative to the maximum possible variability given the individual's mean. The distributions for level and variability are presented in Figure 1, showing the full range of individual differences in level and variability scores within- and between contexts (work, non-work) and sources (self, other). Considerable consistency in the distribution patterns across contexts and sources hint at "non-randomness" of this variability.

[Insert Table 2 and Table 3 about here]

[Insert Figure 1 about here]

The next focus is on exploring intra-individual variability indices and their incremental prediction of job performance over and above traditional inter-individual differences conceptualisations (i.e., BFF level). To situate this work within the traditional, level-oriented approach, we explored differences in BFF level as a function of source and contextual frame using linear mixed-effects regression. Results are presented in Table 4A. The source and context variables were weighted effect coded, an approach consistent with te Grotenhuis, Pelzer, Eisinga, Nieuwenhuis, Schmidt-Catran, & Konig, (2017a) and (2017b) to ensure that the mean across conditions is 0, so that an effect can be interpreted as conditional on the other.

Source of the rating was a significant predictor of BFF level for all traits, with self-ratings being on average higher than other-ratings for Neuroticism ($\beta = .17$, $p < .001$), and lower than other-ratings for all other traits ($-.35 \leq \beta \leq -.13$, all $p < .001$), suggesting that observers had a more favourable view of the target participants than participants had of themselves. In terms of *context*, Agreeableness and Openness ratings were on average lower

at work than outside work (β = .18, $p < .001$, and β = .06, $p = .01$, respectively). However, interaction effects were also observed for Agreeableness and Conscientiousness (see Table 4A). In both cases observers provided higher, i.e. more favourable, ratings, and these source-related differences were more pronounced in non-work compared to work contexts. For example, simple-effect analyses indicated that with regard to Conscientiousness, ratings were higher at work compared to outside work when based on self-ratings only (the effect was significant but small: β = -.17, CI = -.22 to -.12, $p < .001$). Observers outside work rated the target participant as being higher in Conscientiousness than did observers at work; an even smaller effect but in the opposite direction (β = .09, CI = .00 to .18, $p = .039$). This may well be because work observers (including peers and superiors) have higher (or at least different) expectations with regard to Conscientiousness than have partners and friends in a non-work context. Higher "benchmarks" regarding Conscientiousness in a work context are due to context-typical demand profiles and are informed by what is observed in the context-specific reference group of colleagues and peers at work.

[Insert Table 4 here]

4.3.2 Systematicity in Intra-individual Variability

To examine the extent to which IVI systematically varied as a function of context and source, a series of analogous linear-mixed effect regression analyses were conducted – one for each BFF trait. Results are presented in Table 4B.

With regard to *context*, collapsing across source, intra-individual variability tended to be significantly smaller at work than outside work for all personality traits ($.17 \leq \beta \leq .28$, all $p < .001$). For *source*, collapsing across context, intra-individual variability in Openness and Agreeableness tended to be smaller for self-ratings than other-ratings (β = -.14, $p < .001$, and

β = -.16, $p < .001$, respectively), whereas intra-individual variability in Conscientiousness and Neuroticism tended to be greater for self-ratings than for other-ratings (β = .17, $p < .001$, and β = .18, $p < .001$, respectively). There were no significant overall differences in intra-individual variability in Extraversion ratings for self and other (β = -.01, $p = .935$), although this was qualified by an interaction with contextual frame. In fact, statistically significant interactions were observed for all traits (see Table 4B). For all traits, the difference in intra-individual variability between work and non-work contexts was more pronounced for other-raters than self-raters. Whilst this suggests there was more context-related variability in other- compared to self-ratings, it is important to note that other-ratings were provided by different observers in each context. An increased amount of variability in observer ratings across contexts might simply reflect observer-specific variability.

To further explore the systematicity of (relative) intra-individual variability indices we analysed their intercorrelation pattern. Table 5B shows that person-specific variability indices were consistently correlated (see Table 5A for analogous correlations for level). In line with other studies (e.g. Reddock et al., 2011), we found that variability indices based on self-report data correlated across traits. That is, those who tended to vary in one trait (as estimated based on their own item responses) also tended to vary in other traits. Effects varied in size from moderate to large ($r = .49$ to $r = .72$, mean $r = .58$), with all correlations being positive and statistically significant.

[Insert Table 5 about here]

Similarly, person-specific variability indices based on observer data (up to 3 work observers per person) tended to correlate across traits. That is, those who were perceived to vary in one trait tended to also be perceived to vary in other traits (mean $r = .39$, with a range

from r = .17 to r = .57; all coefficients were positive and statistically significant). Surprisingly perhaps, trait variability indices were not correlated across sources (self and other; mean r = .02, see diagonal in Table 5B). In other words, we observe positive manifolds within sources, which does not cut across[4].

To our knowledge, this is the first study to report person-specific cross-item variability estimates for observer ratings. One possible explanation for the observed correlation pattern is an underlying variability factor or trait. An alternative explanation is that the correlation pattern is to do with response styles. Although either of these accounts cannot be ruled out entirely, there are a number of observations that support an assumption of non-randomness. First, observer ratings are based on multiple, different raters of each participant (there are 563 different work observers providing ratings for the 288 participants). In deriving the average measure of trait variability, we are aggregating across the (presumably variable) response styles of these many different raters (that is, of up to three raters per target participant). This aggregation would serve to attenuate response style effects and to accentuate randomness. Both would result in correlations that vary (randomly) around zero. This, however, is not what we observed. Effects – taken at face value – were small to moderate, but positive (Table 5B). In contrast, for trait level substantive negative correlations across traits were observed regardless of source, as would be expected (see Neuroticism, Table 5A). Second, the rank order of the self at work variability indices across traits follows the rank-ordering of the variability indices derived from (multiple) observer(s) at work (Table 3, work context). This relative parallelism is depicted in Figure 2A (see 2B for the analogous plot for level). For instance, when the variability scores were higher for one trait relative to another trait for self-

[4] The intercorrelation pattern of variability indices was very similar when using the more conventional partial correlation approach to control for the effects of the mean reflected in *absolute* variability indices (see Mestdagh et al., 2018).

ratings (e.g. Agreeableness relative to Conscientiousness), they also tended to be higher relative to the other for observer ratings. Finally, randomness of effects would not have resulted in variability indices being systematically greater for the self than when derived from observers (partial $\eta^2 = .126$, Figure 2A). Whilst such result pattern is aligned with the assumption of a greater diversity in personally lived experience than would be accessible to observers, it is hardly reconcilable with the notion of mere randomness.

[Insert Figure 2 about here]

4.3.3 Intra-individual Variability and Job Performance

The second set of analyses considered the extent to which IVI from the perspective of *observers at work*, predicted job performance in a sub-sample of our data. For comparison, we also report the same analyses from the perspective of the *self at work*. Separate analyses were conducted for each of the four job performance criteria (Leadership, Problem Solving, Emotional Competence, and Impact) and for each BFF trait (i.e., there were 20 separate analyses[5]). The results for both other- and self-ratings are summarised in Table 6. The columns labelled 'r_{OW}' report the zero-order correlation between the BFF level and IVI indices for others at work. Analogously, the 'r_{SW}' columns report the self-at-work correlations. These reflect the total relationship between each indicator and the performance criteria. The β_{OW} and β_{SW} columns in Table 6 represent the respective standardised regression coefficients in which level, IVI and their cross-product interaction were regressed on to each

[5] Separate analyses were conducted as it was the total BFF variance of interest, rather than the unique effect after controlling for other BFF traits.

of the performance criteria for each BFF trait for others and self (respectively) at work. Whilst the relative variability index used in the current study (Mesdagh et al., 2018) adjusts for the effect of the boundedness of the scale, it does not control for an effect of the mean that goes beyond scale boundedness. Hence, we include trait level and the respective moderation effect in these analyses. In all cases, positive effects for level are indicative of a positive relationship with supervisor-rated performance. Positive effects for IVI indicate greater intra-individual variability is associated with higher performance ratings.

Trait level: In terms of the performance criterion *Problem Solving*, statistically significant incremental validity (over and above intra-individual variability) was observed for Conscientiousness (β = .19) and Neuroticism (β = -.34). For *Emotional Competence*, Openness (β = .25), Agreeableness (β = .23), and Neuroticism (β = -.30), were significant incremental predictors. For *Impact*, Conscientiousness (β = .26) appeared to be the only incremental predictor (see Appendix C1 for forest plots). Conscientiousness (β = .16) was also the strongest incremental predictor of *Leadership* in the regression models, but the effect was not statistically significant at alpha 5% (p = .072). Note that of the 20 standardised regression coefficients (5 BBF by 4 performance criteria), six had values of .19 or above for other-ratings and were statistically significant; only one was statistically significant for self-ratings (the largest observed effect for self-ratings was for Extraversion predicting Leadership, β = .18; p = 0.046).

Intra-individual variability (IVI): In terms of BFF trait IVI, correlational analysis results suggest that greater intra-individual variability in Agreeableness (r = .22) and Openness (r = .21) were associated with higher Problem Solving evaluations by supervisors. Greater intra-individual variability in Agreeableness was also associated with Emotional Competence (r = .20) and Impact (r = .20) evaluations. Less intra-individual variability in Neuroticism was associated with greater Emotional Competence evaluations by supervisors (r

= -.21). Of these, the regression analyses suggested that the association between intra-individual variability in Neuroticism provided incremental validity in predicting supervisor evaluations of Emotional Competence (β = -.25). This effect was more pronounced the higher the rated trait level (interaction β = -.22). Finally, intra-individual variability in Agreeableness provided incremental validity in predicting supervisor evaluations of Impact (β = .22) over and above the differences in job performance ratings explained by the traditional BFF trait level scores (see Appendix C1 for forest plots). Note again that all significant predictors were derived from *other-ratings*.

By ways of a "sensitivity analysis" we tested whether the observed effects could have been inflated (or supressed) by the unlikely, but possible situation where supervisors have provided both personality ratings and job performance for our participants[6]. We therefore have run the same analyses without the personality ratings of all those observers classified as superiors[7]. Note, that because the work personality observer ratings represented an aggregate based on ratings from at least two work observers, the sample size (N = 124) remained the same and statistical power was not affected by excluding personality ratings from superiors. Overall, findings were similar, except for two differences. First, small variability effects (main and / or interaction effects) were now also observed for Openness and Extraversion in addition to Neuroticism (see Appendix C2 for forest plots). However, the incrementally predictive effect of variability in Agreeableness with regard to supervisor evaluations of

[6] We thank an anonymous reviewer for making this suggestion.

[7] As personality ratings were submitted anonymously, we were unable to identify those raters who provided personality ratings as superiors and have also served as a source for participants' job performance ratings. It must be noted that this constellation was rather unlikely given the interval of about 12 months between collecting personality ratings and job performance ratings. In other words, those who provided personality ratings are unlikely those who also provided job performance ratings and vice versa. By excluding personality ratings of all superiors, we have created a rather conservative test of the stability of the reported effect pattern. Possibly more importantly, we may well have removed an important component of the other at work evaluation – that is removing those ratings most likely to have been made from an experienced work-evaluation frame.

Impact at work was reduced and no longer statistically significant (β = .10, *p* = .280). Second, at the same time, main effects for trait level were reduced and only three of the six trait level effects remained statistically significant when personality ratings from superiors at work were excluded.

None of the intra-individual variability coefficients based on *self-ratings* at work were significant incremental predictors (i.e. controlled for level) of the performance criteria (see Appendix C3 for forest plots). The largest effect was however again observed for intra-individual variability in Neuroticism predicting Emotional Competence (β = -.17, *p* = .055).

Type 1 Error: The nature of the question we are investigating dictates a relatively large number of comparisons, and accordingly a potential to compound Type 1 errors. We consider two approaches to gauge this impact. First, as already indicated, we report confidence intervals of the standardised effects in forest plots reported in Appendix C to allow readers to evaluate the observed nature of the effects outside the somewhat arbitrary constraints of predetermined *p*-values. Second, for completeness, a family-wise adjustment[8] of *p*-values using the procedure outlined by Benjamini and Hochberg (1995) was conducted. As reported in Table 6, for level, three of the six incremental effects with $p < .05$ remain statistically significant; for IVI, one of the two incremental effects remained statistically significant (intra-individual variability in Neuroticism predicting Emotional Competence). The one significant interaction (also related to Neuroticism and Emotional Competence) remained significant after Type 1 adjustment.

[Insert Table 6 about here]

[8] The "family" was based on four comparisons for each BFF trait — the four job performance DVs, for each main-effect (level and variability) and the interaction. Analyses were conducted using the p.adjust function in the base package of R.

5 Discussion

The overarching aim of this study was to investigate variability in personality responses in work compared to non-work contexts, not just from the self- but also from an observer perspective. We conceptualised intra-individual variability as an individual differences construct. That is, we expected participants to differ in the amount of variability they show, specifically at work, and that such individual differences reflect systematic variability that is predictive of job performance.

In a first set of analyses we investigated variability at the group-level. We found that personality responses varied as a function of context (work, non-work) and source (self, other). Context-related main effects were generally in line with what would be expected, e.g. on average higher self-ratings of Conscientiousness at work than outside work. Similarly, Lievens et al. (2008, study 2) report significant context-related differences for the three traits included in their study (Conscientiousness and two of its facets, Achievement Striving and Self-discipline) with higher self-ratings reported by students on the work- as compared to the school-related personality measure. Such context-related differences provide further evidence for the context-dependency of personality (as reflected in self- and observer ratings) at the group-level.

In addition, and in line with prior research (Leising et al., 2010), we found observers to have, on average, a more favourable view of participants than participants had of themselves, i.e. they rated the target participants higher on desirable traits (Conscientiousness, Openness, Agreeableness, Extraversion) and lower on undesirable traits (Neuroticism). For two traits (Agreeableness, Conscientiousness) this effect was particularly pronounced for non-work observers, i.e. family and friends. Both non-work and work observers indicated to know

the target participant well, but non-work observers felt closer to the participant, which might explain their somewhat more positive view (see Leising et al., 2010). Such source-related effects may suggest a "bias" in observer ratings as compared to self-ratings. However, the following argument can be made: Self-ratings reflect thoughts, feelings and behaviour. Other ratings are primarily based on observed behaviour. Observers, obviously, do not have direct access to the thoughts and feelings of the target individual as it is the case for self-ratings. Following this argument, other-ratings could be seen as "incomplete", as these ratings seem primarily based on *observed behaviour* (from which inferences might be drawn as to the underlying thoughts and feelings). For selected traits, judgements offered by work observers in particular (as reflected in level and variability indices) were predictive of job performance, whilst in the current study self-ratings were generally not. Work observers were in a good position to evaluate the participant given they knew the participant well without feeling particularly close to them on a personal level. Work observers seem to have focussed in their evaluations on what is relevant for performance at work which might have facilitated the emergence of associative links with performance measures.

In a second set of analyses we investigated *differential* variability effects using a person-specific, cross-item operationalisation of relative intra-individual variability. We found some evidence to suggest intra-individual variability to be systematic and meaningful. First, person-specific trait variability estimates varied systematically with context and source (Table 4B). For instance, intra-individual variability tended to be smaller at work than outside work (context effect), which is to be expected (Beer & Vazire, 2017). Secondly, person-specific trait variability estimates substantially correlated (Table 5B) across traits, particularly when based on self-ratings, suggesting that the cross-item operationalisation of intra-individual variability we used in the current study represents, at least to some extent, non-random variability. Variability indices based on ratings from multiple observers were also

correlated across traits (e.g. variability in Openness with variability in Extraversion and Agreeableness, r = .50, and r = .57, resp.). Taken together, the latter findings seem to suggest that the observed correlation patterns cannot simply be explained by response styles. Even so, self-report based variability indices were not aligned with their respective other-report based variability indices for the five traits (Table 5B, diagonal) suggesting a lack of agreement between self and observers.

Thirdly, for two traits (Neuroticism and Agreeableness) person-specific variability estimates were predictive of job performance criteria above and beyond any effect due to the respective trait level, though the effects were small in size. Specifically, intra-individual variability in Neuroticism as observed by others at work proved to be a liability with regard to job performance (specifically evaluations of Emotional Competence at work), and this effect was more pronounced the higher a participant's reported standing on the trait. In contrast, intra-individual variability in Agreeableness was beneficial with regard to supervisor evaluations of performance (specifically Impact at work). Note, similar effects for Agreeableness were also evident in bi-variate correlation coefficients with Problem Solving (r = .22) and Emotional Competence (r = .20), and for Openness with Problem Solving (r = .21), but these effects were reduced and no longer statistically significant when controlling for trait level and interaction effects.

However, the reported predictive effects of intra-individual variability need to be interpreted with caution. When we excluded personality ratings from work colleagues in superior positions relative to the target individual, the predictive effect of variability in Agreeableness was considerably reduced and no longer statistically significant. Small predictive effects of variability (controlled for level) were still observed for three traits: Neuroticism, Extraversion and Openness. Interestingly, these are the same traits highlighted in the Magee et al. (2018) experience sampling study (Study 1) which used a cross-occasion

measure of intra-individual variability in personality states. They found small positive associations with psychological adjustment measures; whilst in the current study all associations with outcome measures (here assessing performance) were negative when including main and interaction effects of the mean in the analyses. The strongest variability effect was observed for Neuroticism (β = -.25, or β = -.32 when personality ratings from superiors were excluded; and β = -.17, p = .055, for self-ratings) indicating that intra-individual variability in Neuroticism is associated with Emotional Competence. The robustness of this effect is signified by its persistence after applying family-wise error adjustment of p-values. We therefore suggest that variability in Neuroticism is likely to be of importance in performance contexts and worthy of further study.

In sum, we provide evidence for the *substantiveness* and *systematicity* of variability captured by cross-item intra-individual variability estimates based on observer ratings; the *meaningfulness* of these variability components (here tested in terms of predictive validity) is less certain at this stage and warrants replication and further study. Notably, no statistically significant predictive effects were found for intra-individual variability estimates based on *self-ratings*.

Why were variability effects evident for observer ratings but not for self-ratings? There might be a number of reasons. First, contexts differ in their elicitation of variability. The situations in which non-work observers experience the target's behaviours are more diverse, relative to the way observers see the target at work. The work context is primarily focused on performance and includes norms, rewards and sanctions that channel behaviour within a narrower scope. Using situation strength terminology, work situations tend to be strong situations. The fact that the distributions of intra-individual variability in the work context is the narrowest supports this notion. This is why when it comes to predicting performance, the work observer perspective should provide meaningful prediction and in fact,

the perspective of the self, might be secondary. Secondly, whilst observer ratings tended to be somewhat more positive than self-ratings, a person's reputation as captured by other-ratings (Hogan, 1996; Connelly & Ones, 2010) might be all that matters at work. Participants might be able to conceal internal experiences from observers, such as a tendency to feel sad or anxious, at least to some extent (façade/identity effects; McAbee & Connelly, 2016). That is, participants might be able to present themselves at work in a way that is conducive to their goal strivings (Hogan, 1996). Equally observers might be able to perceive parts of a person's personality that stays unnoticed or remains unknown to the person themselves (blind spot/reputation effects, McAbee & Connelly, 2016) but is relevant for their successful functioning at work (e.g. non-verbal behaviour when interacting with peers). In other words, what matters is not an individual's "true standing" on a trait, but the level and variability in manifestations of the trait that is being perceived and inferred by others at work. Stronger validity of other- compared to self-ratings have so far been reported for level estimates (Connelly & Ones, 2010); here we find similar source-related predictive validity effects for variability estimates. To our knowledge, this is the first study to explore cross-item intra-individual variability estimates in observer ratings.

Finally, supervisor ratings are most commonly used as an operationalisation of job performance. When this is the case both the predictor (observer ratings of personality) and the criterion (job performance) are based on other-ratings, and hence they both tap into variability due to what is observable, or 'known or noticeable to others', about a person.

Is displaying intra-individual variability adaptive or maladaptive? In the current study, intra-individual variability in Agreeableness as perceived by others at work was found to be beneficial for performance (at least when ratings of superiors were included in the analyses); whilst intra-individual variability in Neuroticism was not. Indeed, being perceived as variable in Neuroticism (i.e., being emotionally unstable) was detrimental for performance.

This finding is in line with recent work employing a very different operationalisation of intra-individual variability in Neuroticism (Wood et al., 2019). Using a field-based experience sampling design Wood et al. (2019) modelled an individual's self-reported state neuroticism as a function of their evaluation of task demand during work days (specifically task urgency and difficulty). They showed that a stronger positive contingency between task demand and neurotic states was related to lower supervisor ratings of job performance (controlled for trait neuroticism). Note that in the Wood et al. (2019) study variability in neuroticism was operationalised as a patterned response to situational characteristics (i.e. a situation contingency unit); whilst in the current study the total amount of observable response variability in neuroticism was used as an operationalisation. Variability in neuroticism at work might indicate a vulnerability in regard to performance expectations. Similarly, Reddock et al. (2011) reported negative relationships between cross-item variability estimates and cognitive performance and GPA scores in a student sample, though they did not differentiate between different traits in these analyses and used an aggregate measure of variability instead. The performance-facilitative effect of variability in Agreeableness seems in line with Lieven's et al. (2018) findings, who reported a small positive effect for variability in sociability and performance rated by peers in a student sample. The current study adds to the relatively small body of research on the relationship between variability and performance.

Integration: both level and variability matter. Whilst intra-individual variability was systematic and predictive of work performance for several traits, intra-individual variability effects were however few, and small in size. Level effects were also observed and these effects were overall more frequent and somewhat stronger (when ratings from superiors were included in the analyses). Note, all variability effects were observed in addition to level effects. Whilst these findings are interesting and certainly suggest further investigation of intra-individual variability as an individual differences construct, particularly given the

applied context of the current study, we certainly do not argue that level effects should be ignored in favour of studying variability. Instead, variability offers additional insights. An integrated view that considers both level and variability is more informative. Personality measures are developed and fine-tuned to reduce variability with the aim to increase internal consistency of measures (e.g. high Cronbach alphas). In that sense, one could argue the variability effects reported here are conservative estimations.

5.1 Limitations and Future Directions

Whilst this study has a number of strengths (e.g. the operationalisation of intra-individual variability, the real-world nature of the investigation which involved studying activities that were of personal importance to participants, the use of a non-student sample, the fact that both self- and other-ratings in both work and non-work contexts were considered in one design), as for any study there are also limitations to be noted. Some limitations are shared with other related studies, whilst others might be unique to the current one.

One is the recruitment of observers, which was based on participant nomination. Targets tend to select observers they like, and 'liking' has been associated with increased positivity in personality judgements of the target ("letter of recommendation effect", Leising et al., 2010). Leising et al. (2010) therefore suggested to select other-raters independently of the target's own preferences. However, given the nature of activities in the current study this was deemed not feasible and perhaps ethically questionable, as this would have compromised the trust that was an important element in the research and development program that participants were enrolled in. Participants were however encouraged to find individuals who knew them well and therefore were in a position to comment on their personality with the aim to aid the target's personal and professional development. Of course, realistically, and particularly in a non-work context, those who know us best due to spending time with us, e.g. a spouse, a long-term friend, tend to also be those who like us. Clearly a balance needs to be

struck between 'knowing' and 'liking' when choosing observers in studies with multi-rater designs. The fact that both the non-work observers and the work observers were selected by the participants themselves supports comparability of their ratings. The same argument could be made in relation to supervisors. In general, supervisors are expected to be at least reasonably satisfied with their subordinates' job performance given the likely consequences of sustained perceived under-performance.

Another limitation is the type of performance measure used. Supervisor ratings are a common measure of performance at work, however they reflect more or less subjective evaluations of target participants. When personality and work performance are rated in a work context, both are based on observer judgments and tend to focus on a person's perceived attributes leading to some level of overlap in item content between personality and performance measures. Future studies should consider a range of performance outcome operationalisations, including more objective ones. In addition, the criteria and the way they tend to be operationalised deserve further scrutiny too. Strictly speaking, performance ratings are "level" focussed (i.e. aggregation of ratings across items). This might explain why our search for meaningfulness, i.e., correlation with job performance, was of limited success. If we were to operationalise performance in a way that reflects individual differences in variation then results might be more converging. The conceptual question however is whether a good performer is expected to show high performance across all facets of performance as they are addressed in performance rating items. Future research needs to exert more efforts in identifying construct relevant criteria. From that perspective, our findings again reflect rather conservative estimations what the prediction of meaningful performance criteria is concerned, and predictive effects of variability may be stronger when using a criterion that is stronger aligned with the concept of variability.

Third, using different observers in different contexts means that, strictly speaking, observer judgments are not comparable across contexts; an unavoidable effect. This is because observers in each context have different expectations and frames of references. For instance, expectations of conscientiousness are likely higher at work compared to outside work given work task requirements and the likely more conscientious behaviour of the comparison group (i.e. the participants' co-workers). Observers were chosen because of their experience with the target participant in their context-specific role. Rarely will an observer be able to provide valid judgements equally well across both work and non-work contexts, and even if this was possible, expectations regarding, say conscientiousness, will differ between contexts.

Fourth, our findings are limited in terms of generality. We expect our findings to be generalisable to a population of high-performing professionals, a relatively difficult to reach group for this kind of research. The use of contextualised measures is fairly well established, and findings are expected to hold when similar measures (see Appendices) and operationalisations of intra-individual variability (within-context, cross-item) are used.

Fifth, whilst our operationalisation of intra-individual variability had the advantage of being less burdensome, it would be interesting to see whether replication with experience sampling designs based on both self- and other-ratings might yield similar or additional valuable insights to better understand the dynamics of personality.

Finally, our investigations are limited by the plausible power of our sample what the analyses relating to job performance are concerned. As reported, our sensitivity analysis indicated an effect of partial $\eta^2 = .061$ (a medium effect) was detectable in the smallest sub-sample ($N = 124$). The median observed standardized regression coefficients were on the whole somewhat smaller than this.

5.2 Conclusions

The current study adds to the growing body of literature that is supplementing the established effects for levels of traits through the integration of individual level variability into the study of individual differences. Whilst previous studies have examined variability in actual responses (e.g. Fleeson, 2007) and the variability linked to situational contingencies (e.g. Minbashian et al., 2010), the current study examined variability as a function of the frames of references for self and social observers within different contexts. Assessments of personality are influential determinants of other social judgements, such as appraisals and selection decisions. Further research will help develop understanding of the implications of different types of variability across frames of references and whether that variability is adaptive or maladaptive.

5.3 Ethics

Information provided on title page as requested.

5.4 Acknowledgement

Information provided on title page as requested.

5.5 Author Contributions

Information provided on title page as requested.

6 References

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: reliability, validity, and associations with well-being. Journal of Personality and Social Psychology, 90(3), 512-527. doi:10.1037/0022-3514.90.3.512

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. Journal of Research in Personality, 69, 170--179. doi:10.1016/j.jrp.2016.06.015

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1–26.

Barrick, M. R., & Mount, M. K. (2000). Select on conscientiousness and emotional stability. In E. A. Locke (Ed.), Handbook of principles of organizational behaviour (pp. 15–28). Oxford, UK: Blackwell.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? International Journal of Selection and Assessment, 9(1-2), 9–30.

Beckmann, N., & Wood, R. E. (2017). Editorial: Dynamic personality science. Integrating between-person stability and within-person change. Frontiers in Psychology, 8. doi:10.3389/fpsyg.2017.01486

Beckmann, N., Wood, R. E., & Minbashian, A. (2010). It depends how you look at it: On the relationship between neuroticism and conscientiousness at the within- and the between-person levels of analysis. Journal of Research in Personality, 44(5), 593-601. doi:10.1016/j.jrp.2010.07.004

Beer, A., & Vazire, S. (2017). Evaluating the predictive validity of personality trait judgments using a naturalistic behavioral criterion: A preliminary test of the self-other knowledge asymmetry model. Journal of Research in Personality, 70, 107-121. doi:https://doi.org/10.1016/j.jrp.2017.06.004

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57, 289–300. http://www.jstor.org/stable/2346101.

Berenson, K. R., Downey, G., Rafaeli, E., Coifman, K. G., & Paquin, N. L. (2011). The rejection–rage contingency in borderline personality disorder. Journal of Abnormal Psychology, 120(3), 681-690.

Bing, M. N., Davison, H. K., & Smothers, J. (2014). Item-level frame-of-reference effects in personality testing: An investigation of incremental validity in an organizational setting. International Journal of Selection and Assessment, 22(2), 165-178. doi:10.1111/ijsa.12066

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. Journal of Applied Psychology, 89(1), 150-157. doi:10.1037/0021-9010.89.1.150

Birney, D. P., Beckmann, J. F., & Seah, Y. (2016). The eye of the beholder: Creativity ratings depend on task involvement, order and methods of evaluation, and personal characteristics of the evaluator. Learning and Individual Differences, 400-408. doi:http://dx.doi.org/10.1016/j.lindif.2015.07.007(51)

Bleidorn, W., & Ködding, C. (2013). The divided self and psychological (mal) adjustment – A meta-analytic review. Journal of Research in Personality, 47(5), 547-552. doi:https://doi.org/10.1016/j.jrp.2013.04.009

Bowling, N. A., & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. Personality and Individual Differences, 49(2), 95-101. doi:https://doi.org/10.1016/j.paid.2010.03.009

Clifton, A. (2014). Variability in personality expression across contexts: A social network approach. Journal of Personality, 82(2), 103--115. doi:10.1111/jopy.12038

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. Psychological Bulletin, 136(6), 1092-1122.

Costa, P. T., Jr. , & McCrae, R. R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual. . Odessa, FL: Psychological Assessment Resources.

Dalal, R. S., Meyer, R. D., Bradshaw, R. P., Green, J. P., Kelly, E. D., & Zhu, M. (2014). Personality Strength and Situational Influences on Behavior: A Conceptual Review and Research Agenda. Journal of Management, 41(1), 261-287. doi:10.1177/0149206314557524

Debusscher, J., Hofmans, J., & De Fruyt, F. (2016). Do personality states predict momentary task performance? The moderating role of personality variability. Journal of Occupational and Organizational Psychology, 89(2), 330-351. doi:10.1111/joop.12126

Donahue, E. M., Robins, R. W., Roberts, B. W., & John, O. P. (1993). The divided self: concurrent and longitudinal effects of psychological adjustment and social roles on self-concept differentiation. Journal of Personality and Social Psychology, 64(5), 834-846. doi:10.1037//0022-3514.64.5.834

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods & Instrumentation, 41*, 1149-1160. doi:10.3758/BRM.41.4.1149

Fisher, D. M., Cunningham, S., Kerr, A. J., & Allscheid, S. P. (2017). Contextualized personality measures in employee selection: Extending frame-of-reference research with job applicant samples. International Journal of Selection and Assessment, 25(1), 18-35. doi:10.1111/ijsa.12156

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

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. . Journal of Personality and Social Psychology, 97(6), 1097-1114.

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. Journal of Research in Personality, 56, 82-92. doi:https://doi.org/10.1016/j.jrp.2014.10.009

Fleisher, M. S., Woehr, D. J., Edwards, B. D., & Cullen, K. L. (2011). Assessing within-person personality variability via frequency estimation: More evidence for a new measurement approach. Journal of Research in Personality, 45(6), 535--548. doi:10.1016/j.jrp.2011.06.009

Gangestad, S. W., & Snyder, M. (2000). Self-monitoring: appraisal and reappraisal. Psychological Bulletin, 126(4), 530-555. doi:10.1037/0033-2909.126.4.530

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26-42.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality, 40, 84-96.

Guthke, J. (1992). Learning Tests - the concept, main research findings, problems and trends. Learning and Individual Differences, 4, 137-151.

Guthke, J., & Beckmann, J. F. (2003). Dynamic assessment with Diagnostic Programs. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), Models of intelligence. International perspectives (pp. 227-242). Washington, DC: American Psychological Association.

Hardy, J., & Segerstrom, S. C. (2017). Intra-individual variability and psychological flexibility: Affect and health in a national US sample. Journal of Research in Personality, 69, 13--21. doi:10.1016/j.jrp.2016.04.002

Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The Influence of Work Personality on Job Satisfaction: Incremental Validity and Mediation Effects. Journal of Personality, 77(4), 1051-1084. doi:10.1111/j.1467-6494.2009.00574.x

Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), The five-factor model of personality: Theoretical perspectives (pp. 163-179). New York, NY, US: Guilford Press.

Holtrop, D., Born, M. P., de Vries, A., & de Vries, R. E. (2014). A matter of context: A comparison of two types of contextualized personality measures. Personality and Individual Differences, 68, 234-240. doi:10.1016/j.paid.2014.04.029

Holtrop, D., Born, M. P., & de Vries, R. E. (2014). Predicting performance with contextualized inventories, no frame-of-reference effect? International Journal of Selection and Assessment, 22(2), 219-223. doi:10.1111/ijsa.12071

Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the Rules of Justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. International Journal of Selection and Assessment, 13(1), 75-86. doi:10.1111/j.0965-075X.2005.00301.x

Huang, J. L., & Bramble, R. J. (2016). Trait, state, and task-contingent conscientiousness: Influence on learning and transfer. Personality and Individual Differences, 92, 180-185. doi:10.1016/j.paid.2015.12.043

Huang, J. L., & Ryan, A.-m. (2011). Beyond personality traits: A study of personality states and situational contingencies in customer service jobs. Personnel Psychology, 64(2), 451-488. doi:10.1111/j.1744-6570.2011.01216.x

Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. Journal of Applied Psychology, 88(3), 545-551. doi:10.1037/0021-9010.88.3.545

Japyassú, H. F., & Malange, J. (2014). Plasticity, stereotypy, intra-individual variability and personality: Handle with care. Behavioural Processes, 109(Part A), 40-47. doi:10.1016/j.beproc.2014.09.016

Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. Journal of Applied Psychology, 99(2), 199-221. doi:10.1037/a0034485

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion, 7(2), 262-274. doi:10.1037/1528-3542.7.2.262

Kurtz, J. E., & Palfrey, A. P. (2016). On being the same in different places: Evaluating frame-of-reference effects across two social contexts. Journal of Personality, 84(1), 113-120. doi:10.1111/jopy.12144

LaHuis, D. M., Barnes, T., Hakoyama, S., Blackmore, C., & Hartman, M. J. (2017). Measuring traitedness with person reliabilities parameters. Personality and Individual Differences, 109, 111-116. doi:https://doi.org/10.1016/j.paid.2016.12.034

Lang, J. W. B., Lievens, F., De Fruyt, F., Zettler, I., & Tackett, J. L. (2019). Assessing meaningful within-person variability in Likert-scale rated personality descriptions: An IRT tree approach. Psychological Assessment, 31, 474-487.

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. Journal of Personality and Social Psychology, 98(4), 668-682. doi:10.1037/a0018771

Leising, D., Gallrein, A.-M. B., & Dufner, M. (2013). Judging the behavior of people we know: Objective assessment, confirmation of preexisting views, or both? Personality and Social Psychology Bulletin, 40(2), 153-163. doi:10.1177/0146167213507287

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. Journal of Applied Psychology, 93(2), 268-279. doi:10.1037/0021-9010.93.2.268

Lievens, F., Lang, J. W. B., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. Journal of Applied Psychology, 103(7), 753-771. doi:10.1037/apl0000280

Magee, C., Buchtel, E. E., Human, L. J., Murray, D. R., & Biesanz, J. C. (2018). Is personality variability associated with adjustment? Journal of Research in Personality, 72, 22-43. doi:https://doi.org/10.1016/j.jrp.2016.08.005

Martin, M. M., Anderson, C. M., & Thweatt, K. S. (1998). Individuals' perceptions of their communication behaviors: A validity study of the relationship between the Cognitive Flexibility Scale and the Communication Flexibility Scale with aggressive communication traits. Journal of Social Behavior and Personality, 13, 531 – 540.

Martin, M. M., & Rubin, R. B. (1995). A New Measure of Cognitive Flexibility. Psychological Reports, 76(2), 623-626. doi:10.2466/pr0.1995.76.2.623

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods, 23*(4), 690–707.

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. Psychological Review, 123(5), 569-591.

Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. Journal of Applied Psychology, 95(5), 793-806. doi:10.1037/a0020016

Pace, V. L., & Brannick, M. T. (2010). Improving prediction of work performance through frame-of-reference consistency: Empirical evidence using openness to experience. International Journal of Selection and Assessment, 18(2), 230-235. doi:10.1111/j.1468-2389.2010.00506.x

Paulus, D. L., & Martin, C. L. (1988). Functional flexibility: A new conception of interpersonal flexibility. Journal of Personality and Social Psychology, 55, 88-101.

R-Core-Team. (2019). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Rauthmann, J. F. (2017). What are other-rated scales composed of? Sources of measurement error and true trait variance in other-ratings of the Big Five. Journal of Research in Personality, 70, 45-55. doi:https://doi.org/10.1016/j.jrp.2017.05.002

Rauthmann, J. F., Beckmann, N., Noftle, E. E., & Sherman, R. A. (2019). Personality dynamics: A new frontier in personality psychology. Personality and Individual Differences, 136, 1-1. doi:10.1016/j.paid.2018.07.012

Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. International Journal of Selection and Assessment, 19(2), 119-131. doi:10.1111/j.1468-2389.2011.00540.x

Robie, C., & Risavy, S. D. (2016). A comparison of frame-of-reference and frequency-based personality measurement. Personality and Individual Differences, 92, 16-21. doi:10.1016/j.paid.2015.12.005

Robie, C., Risavy, S. D., Holtrop, D., & Ph. Born, M. (2017). Fully contextualized, frequency-based personality measurement: A replication and extension. Journal of Research in Personality, 70, 56-65. doi:10.1016/j.jrp.2017.05.005

Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. Organizational Research Methods, 3(4), 348-365. doi:10.1177/109442810034003

Rosenberg, M. (1965). Society and the adolescent self-image. Princton, NJ: Princton University Press.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. Journal of Applied Psychology, 80(5), 607-620. doi:10.1037/0021-9010.80.5.607

Shaffer, J. A., & Postlethwaite, B. E. (2013). 'A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures': Corrigendum. Personnel Psychology, 66(2), 529-529. doi:10.1111/peps.12040

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. Journal of Personality and Social Psychology, 109(5), 872-888. doi:10.1037/pspp0000036

Suh, E. M. (2002). Culture, identity consistency, and subjective well-being. Journal of Personality and Social Psychology, 83(6), 1378-1391. doi:10.1037//0022-3514.83.6.1378

te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017a). When size matters: advantages of weighted effect coding in observational studies. *International Journal of Public Health, 62*(1), 163-167. doi:10.1007/s00038-016-0901-1

te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017b). A novel method for modelling interaction between categorical variables. *International Journal of Public Health, 62*(3), 427-431. doi:10.1007/s00038-016-0902-0

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. Journal of Personality and Social Psychology, 98(2), 281-300. doi:10.1037/a0017908

Wang, Q., & Bowling, N. A. (2016). A comparison of general and work-specific personality measures as predictors of organizational citizenship behavior. International Journal of Selection and Assessment, 24(2), 172-188. doi:10.1111/ijsa.12139

Webster, G. D., Smith, C. V., Brunell, A. B., Paddock, E. L., & Nezlek, J. B. (2017). Can Rosenberg's (1965) Stability of Self Scale capture within-person self-esteem variability? Meta-analytic validity and test–retest reliability. Journal of Research in Personality, 69, 156-169. doi:10.1016/j.jrp.2016.06.005

Wood, R. E., Beckmann, N., Birney, D. P., Beckmann, J. F., Minbashian, A., & Chau, R. (2019). Situation contingent units of personality at work. Personality and Individual Differences, 136, 113-121. doi:10.1016/j.paid.2018.01.026

7 Tables

Table 1

Correlations between self-ratings of the Big 5 traits using the adjective lists (AL) and the IPIP scales.

		Means	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IPIP O	(1)	67.51	13.21	-								
IPIP C	(2)	70.88	13.61	.01	-							
IPIP E	(3)	61.52	15.78	**.29**	**.13**	-						
IPIP A	(4)	72.15	10.69	**.13**	**.26**	.02	-					
IPIP N	(5)	30.80	14.22	**-.18**	**-.29**	**-.35**	**-.36**	-				
AL O	(6)	59.88	8.93	***.59***	.11	**.33**	**.17**	**-.17**	-			
AL C	(7)	73.34	10.22	.02	***.69***	.11	**.25**	**-.19**	**.22**	-		
AL E	(8)	63.90	12.21	**.28**	.09	***.77***	.00	**-.27**	**.43**	**.24**	-	
AL A	(9)	65.92	8.37	.01	**.14**	**-.17**	***.57***	-.11	**.13**	**.34**	-.07	-
AL N	(10)	36.18	10.60	-.07	**-.18**	-.09	**-.41**	***.61***	**-.21**	**-.34**	**-.21**	**-.36**

Note: AL = adjective list (generic self-report, no specific context was indicated); IPIP = International Personality Item Pool scales (Goldberg et al., 2006); italic: convergent validities; bold: $p < .05$; $N = 283$

Table 2

Descriptive statistics of BFF rating level by source (Other/Self) and context frame (Non-Work/Work)

		Non-Work			Work		
		N	Mean (SD)		N	Mean (SD)	
Openness	Other	200	64.97	7.79	203	63.46	6.54
	Self	279	59.90	9.89	282	59.46	8.45
Conscientiousness	Other	200	82.13	9.38	203	80.46	8.38
	Self	279	71.72	11.57	282	75.38	10.65
Extraversion	Other	200	69.11	9.36	203	68.91	9.06
	Self	279	65.78	12.69	282	65.91	10.91
Agreeableness	Other	200	72.19	8.07	203	68.06	8.68
	Self	279	69.13	8.80	282	67.18	8.73
Neuroticism	Other	200	28.30	10.76	203	29.40	10.40
	Self	279	32.43	11.23	282	32.00	10.77

Notes: *N* represents the number of participants for which there are responses.

Table 3

Descriptive statistics of BFF variability (IVI) by source (Other/Self) and context frame (Non-Work/Work)

		Non-Work			Work		
		N	Mean (SD)		N	Mean (SD)	
Openness	Other	200	0.50	0.12	203	0.41	0.09
	Self	279	0.43	0.14	282	0.41	0.13
Conscientiousness	Other	200	0.36	0.14	203	0.26	0.10
	Self	279	0.37	0.15	282	0.36	0.14
Extraversion	Other	200	0.42	0.11	203	0.35	0.08
	Self	279	0.39	0.14	282	0.38	0.13
Agreeableness	Other	200	0.53	0.13	203	0.43	0.11
	Self	279	0.45	0.15	282	0.43	0.14
Neuroticism	Other	200	0.38	0.11	203	0.28	0.08
	Self	279	0.40	0.15	282	0.36	0.14

Notes: *N* represents the number of participants for which there are responses.

Table 4

Regression analyses of (A) Level and (B) IVI by source and context frame.

A. Level	**Openness**				**Conscientiousness**			
Predictors	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*
context	0.56	.06	.02 – .11	.010	-0.28	-.03	-.08 – .02	.302
source	-2.80	-.27	-.32 – -.22	<.001	-4.52	-.35	-.40 – -.30	<.001
context x source	-0.33	-.03	-.08 – .01	.167	-1.57	-.13	-.18 – -.08	<.001
ICC / R^2	.44	.479			.37	.459		

	Extraversion				**Agreeableness**				**Neuroticism**			
Predictors	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*
context	0.11	.01	-.04 – .06	.663	1.55	.18	.13 – .23	<.001	-0.25	-.02	-.07 – .03	.382
source	-1.92	-.15	-.20 – -.10	<.001	-1.30	-.13	-.18 – -.08	<.001	2.20	.17	.12 – .22	<.001
context x source	-0.11	-.01	-.06 – .04	.692	-0.63	-.06	-.12 – -.01	.013	0.49	.04	-.01 – .09	.122
ICC / R^2	.53	.539			.41	.433			.42	.433		

B. IVI	Openness				Conscientiousness			
Predictors	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*
context	0.03	.26	.20 – .31	<.001	0.03	.23	.18 – .29	<.001
source	-0.02	-.14	-.20 – -.09	<.001	0.03	.17	.11 – .22	<.001
context x source	-0.02	-.15	-.20 – -.09	<.001	-0.03	-.17	-.22 – -.11	<.001
ICC / R^2	.26	.319			.23	.297		

	Extraversion				Agreeableness				Neuroticism			
Predictors	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*	*Estimates*	β	*CI*	*p*
context	0.02	.17	.11 – .23	<.001	0.03	.23	.18 – .29	<.001	0.04	.28	.22 – .33	<.001
source	~0.00	-.01	-.06 – .05	.935	-0.03	-.16	-.22 – -.10	<.001	0.03	.18	.13 – .24	<.001
context x source	-0.02	-.10	-.16 – -.06	<.001	-0.02	-.14	-.20 – -.08	<.001	-0.02	-.13	-.18 – -.07	<.001
ICC / R^2	.25	.275			.18	.246			.31	.385		

Note: Regression coefficients are weighted-effect coded (equivalent to ANOVA comparison); $N = 964$ observations on 288 participants. Positive regression coefficients for Context indicate higher scores for non-work than work; Positive regression coefficients for Source indicates higher ratings for self than other.

Table 5

Correlations within (A) trait levels and (B) trait variability, between BFF self-ratings (upper-triangle), other-ratings (lower-triangle), and self-other intercorrelations (diagonal) at work.

A. Level

	O	C	E	A	N
O	**.22**	**.31**	**.41**	**.15**	**-.34**
C	**.44**	**.20**	**.31**	**.34**	**-.45**
E	**.54**	**.27**	**.45**	-.03	**-.31**
A	**.28**	**.36**	-.04	**.27**	**-.41**
N	**-.44**	**-.51**	**-.20**	**-.64**	**.29**

B. Variability (IVI)

	O	C	E	A	N
O	.04	**.56**	**.63**	**.72**	**.63**
C	**.17**	.03	**.50**	**.50**	**.59**
E	**.50**	**.39**	.01	**.49**	**.56**
A	**.57**	**.33**	**.49**	-.05	**.57**
N	**.30**	**.45**	**.43**	**.29**	.06

Note: Bold, $p < .05$. For self-ratings (upper-triangle), $N = 282$; other-ratings (lower-triangle), $N = 202$; self-other (diagonal), $N = 197$.

Table 6

Incremental prediction of job performance for other-at-work ratings and self-at-work ratings. Level and IVI zero-order correlations (*r*) and standardised regression coefficients (β) for BFF traits.

	Leadership				Problem Solving				Emotional Competence				Impact			
	r_{ow}	r_{sw}	β_{ow}	β_{sw}	r_{ow}	r_{sw}	β_{ow}	β_{sw}	r_{ow}	r_{sw}	β_{ow}	β_{sw}	r_{ow}	r_{sw}	β_{ow}	β_{sw}
Level																
O	.07	-.02	.08	-.07	**.23**	.02	.16	-.02	**.24**	.03	**.25**	~.00	**.19**	.04	.18	.01
C	.16	.12	.16	.12	**.20**	-.04	**.19**	.05	.06	.13	.05	-.12	**.27**	.10	**.26**	.10
E	.08	**.17**	.08	**.18**	.10	.07	.09	.08	.13	.08	.13	.08	.14	.08	.14	.08
A	.06	-.04	.01	-.06	**.22**	-.02	.16	-.03	**.26**	-.01	**.23**	.01	.08	-.03	.01	-.04
N	.16	-.06	.15	-.06	**.34**	-.07	**-.34**	.09	**.35**	-.09	**-.30**	-.11	.16	.01	.18	.01
Variability (IVI)																
O	.07	.07	.04	.09	**.21**	.03	.14	.04	.11	-.06	.00	-.06	.14	.08	.07	.08
C	.07	-.03	.09	-.03	.09	.01	.07	.07	.11	-.11	.09	-.08	.10	.02	.07	-.01
E	.17	-.03	.18	-.04	.16	.03	.15	.02	.02	-.10	.01	-.10	.15	.11	.14	.11
A	.16	-.08	.19	.09	**.22**	.01	.19	.01	**.20**	-.05	.12	-.06	**.20**	.02	**.22**	.03
N	-.09	-.02	-.08	.01	-.14	-.02	-.10	.03	**-.21**	.16	**-.25**	-.17	.05	.06	.09	.07
Interaction (Level x Variability)																
O	-	-	-.08	.03	-	-	.04	-.08	-	-	-.04	~.00	-	-	-.05	-.05
C	-	-	-.08	-.07	-	-	-.01	-.16	-	-	.01	-.04	-	-	-.01	.03
E	-	-	-.07	-.02	-	-	~.00	-.03	-	-	~.00	.02	-	-	-.02	-.02
A	-	-	-.10	-.07	-	-	-.03	-.10	-	-	.04	.09	-	-	-.06	-.03
N	-	-	-.02	.01	-	-	.03	-.14	-	-	**-.22**	-.04	-	-	.09	.10
Model R^2																
O	-	-	.01	.01	-	-	**.07**	.01	-	-	.06	~.00	-	-	.05	.01
C	-	-	.04	.02	-	-	.04	.03	-	-	.01	.03	-	-	**.08**	.01
E	-	-	.04	.03	-	-	.03	.01	-	-	.02	.02	-	-	.04	.02
A	-	-	.04	.01	-	-	**.08**	.01	-	-	**.08**	.01	-	-	.05	~.00
N	-	-	.03	~.00	-	-	**.13**	.03	-	-	**.20**	.04	-	-	.04	.02

Note: ow = other at work; sw = self at work; *r* = zero-order correlation; β = regression coefficient; N_{OW} = 124; N_{SW} = 130; bold: $p < .05$; Shaded cells indicate that β remained significant at $p < .05$ after adjusting for Type 1 error using the family-wise error based approach of Benjamini and Hochberg (1995).

8 Appendix A

Table A1.

Manifestations of Personality

O	C	E	A	N
Unimaginative*	Hard-working	Submissive*	Honest	Temperamental
Intellectual	Ineffective*	Serious*	Considerate	Calm*
Progressive	Efficient	Bold	Modest	Impulsive
Contemplative	Undependable*	Cheerful	Shrewd*	Moody
Unreflective*	Efficacious	Reserved*	Sympathetic	Anxious
Practical*	Disorganized*	Talkative	Selfish*	Stable*
Unsophisticated*	Reliable	Active	Suspicious*	Relaxed*
Conservative*	Thorough	Friendly	Difficult*	Secure*
Adventurous	Careless*	Timid*	Cooperative	Vulnerable
Cultured	Procrastinative*	Quiet*	Trustful	Reasonable*
Conventional*	Unmotivated*	Passive*	Self-promoting*	Self-conscious
Creative	Systematic	Assertive	Tough-minded*	Resilient*

Note: * Reverse coded for scoring

9 Appendix B

Table B1-A.

Problem Solving and Leadership job performance items and Factor Loadings

Problem Solving	F1	Leadership	F2
1. Demonstrates high levels of flexibility in his or her problem solving.	.88	1. Actively facilitates the learning and development of co-workers.	.70
2. Approaches problems and situations from multiple, diverse perspectives.	.85	2. Facilitates the development of others through assisting in goal-setting or identifying objectives.	.81
3. Successfully applies prior experience to new situations.	.69	3. Provides subordinates with regular, specific, and timely feedback in relation to developmental goals or objectives.	.92
4. Identifies innovative approaches or solutions to problems.	.76	4. Provides subordinates with appropriate learning and development opportunities.	.89
5. Effectively adapts to new situations and novel or unpredictable work demands.	.60	5. Provides encouragement to others with respect to developmental efforts and achievements.	.77
6. Anticipates changes that may require action to promote organisational success.	.60	6. Makes effective use of rewards to facilitate employee learning and development.	.80
7. Is creative in his or her approach to problem solving.	.81	7. Is an effective leader.	.61
8. Demonstrates a high level of problem-solving ability.	.80	8. Clearly specifies to subordinates what needs to be done.	.65
9. Identifies situations that may require action to promote organisational success.	.62	9. Actively delegates decision making authority to qualified subordinates.	.61
10. Makes sound decisions based on available information.	.68	10. Effectively motivates subordinates using rewards and/or disciplinary action.	.68
11. Efficiently processes information regardless of the mode or format it is presented in.	.86	11. Inspires others to achieve high standards.	.53
12. Effectively differentiates between relevant and irrelevant information.	.86	12. Effectively maintains authority.	.52
13. Makes decisions in accordance with broader organisational strategies or objectives.	.68	13. Develops systems and processes to achieve and/or maintain high standards.	.39
14. Understands and effectively responds to sources of complexity in performing his or her job.	.75	14. Exerts a calming influence on co-workers during difficult or challenging periods.	.10

Table B1-B

Emotional Competence and Impact job performance items and factor loadings

Emotional Competence	F3	Impact	F4
1. Effectively perceives emotions in his or her co-workers.	.84	1. Displays a high level of job performance.	.74
2. Understands how his or her emotions influence the decisions that he or she makes.	.77	2. Has a high level of productivity.	.65
3. Anticipates the emotional reactions of co-workers.	.62	3. Consistently produces high quality work.	.56
4. Effectively manages his or her emotions in the workplace.	.63	4. Has a reputation as a high performer within the organisation.	.58
5. Effectively responds to the emotional reactions of co-workers.	.75	5. Accomplishes goals set by self and others.	.50
6. Displays a high level of interpersonal skills in his or her interactions with co-workers.	.78	6. Consistently displays a high level of effort.	.62
7. Works effectively with others to achieve organisational goals.	.59	7. Demonstrates initiative and takes action when required without direction.	.34
8. Seeks to maintain or enhance customer satisfaction.	.50	8. Responds quickly to urgent work demands.	.92
9. Initiates and maintains friendly interactions with others to facilitate positive work relations.	.85	9. Persists on tasks despite obstacles, setbacks, or failure.	.92
10. Asserts him or herself in a positive manner.	.56	10. Demonstrates sufficient levels of attention to detail in performing tasks.	.75
11. Utilises social networks to achieve organisational goals.	.42	11. Maintains task focus despite difficulty or ambiguity.	.61
		12. Does not hesitate in making tough decisions.	.81

Table B1-C

Factor Intercorrelations (reliability along diagonal)

Factors		(F1)	(F2)	(F3)	(F4)
Problem Solving	(F1)	.97			
Emotional Competence	(F2)	.56	.95		
Leadership	(F3)	.51	.45	.96	
Impact	(F4)	.56	.42	.48	.95

Note: N = 137. The factor analysis was conducted on all participants in the larger sample who had supervisor ratings.

Figures

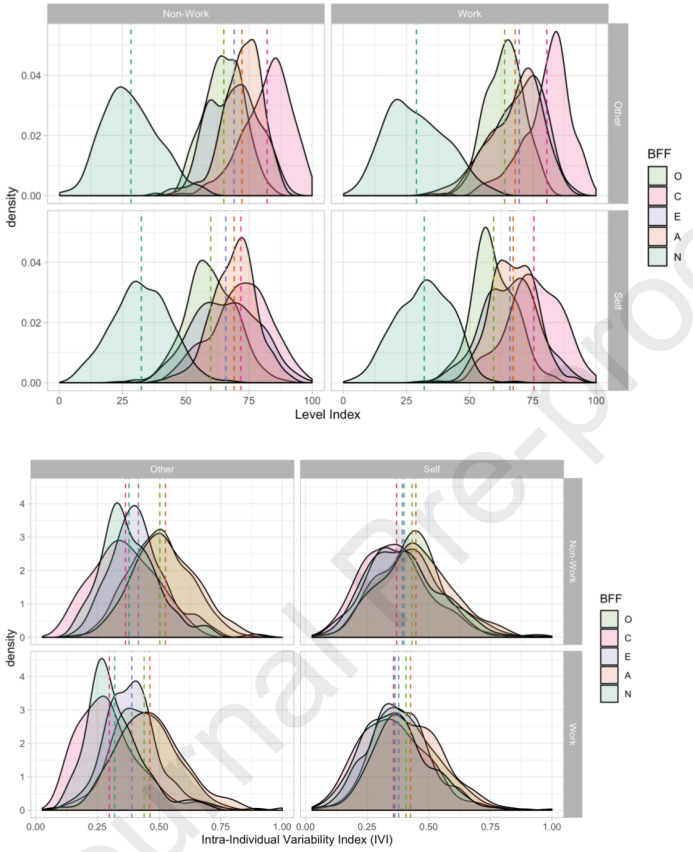

Figure 1. Density distributions of BFF indices by source and context for (A) Level aggregate and (B) intra-individual variability. Note: Dashed lines represent point-estimates of means.

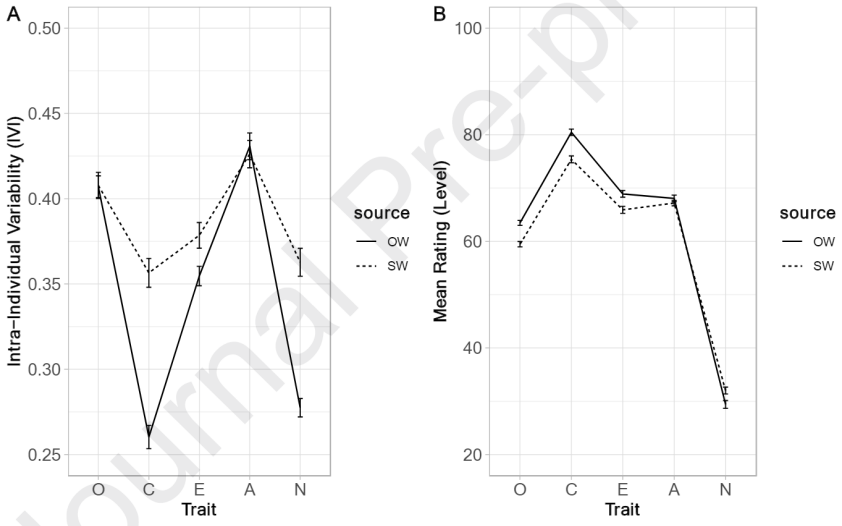

Figure 2. Comparisons across traits of (A) intra-individual variability, and (B) level, for self at work (SW) and others at work (OW).

Appendix C

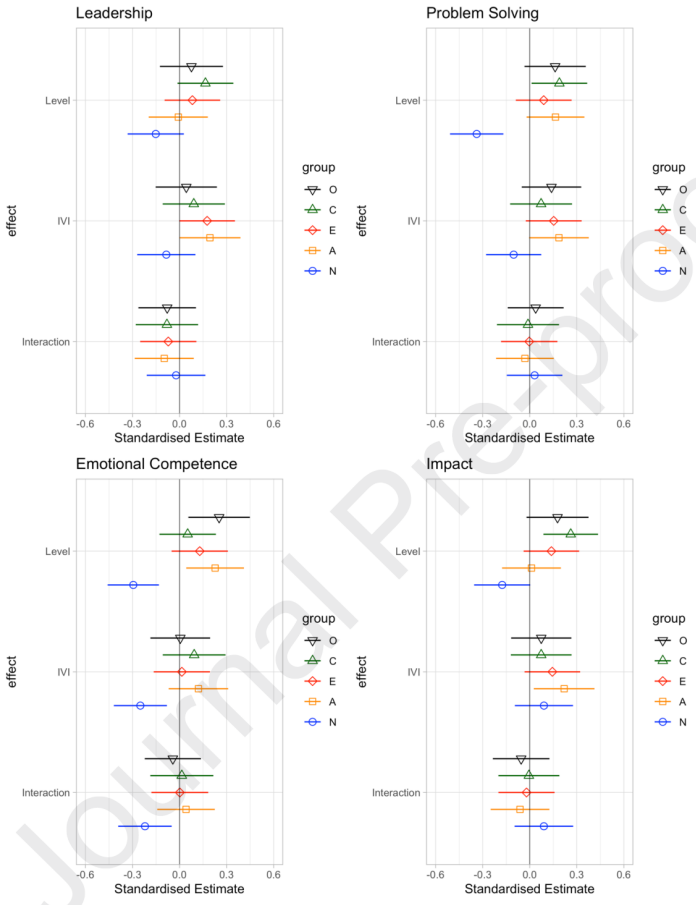

Figure C1. Forest plots of trait mean-level, intra-individual variability index (IVI) and their cross-product interaction effects on four job performance criteria based on personality ratings provided by work observers.

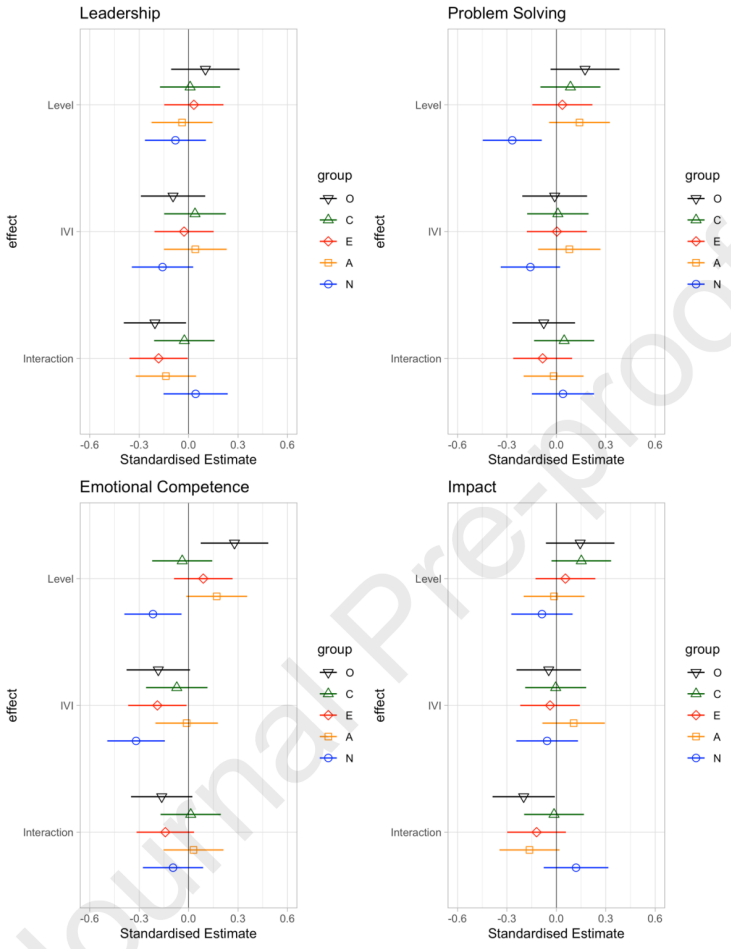

Figure C2. Forest plots of trait mean-level, intra-individual variability index (IVI) and their cross-product interaction effects on four job performance criteria based on personality ratings provided by work observers, but excluding ratings from colleagues in superior positions relative to the target participant (e.g. supervisors).

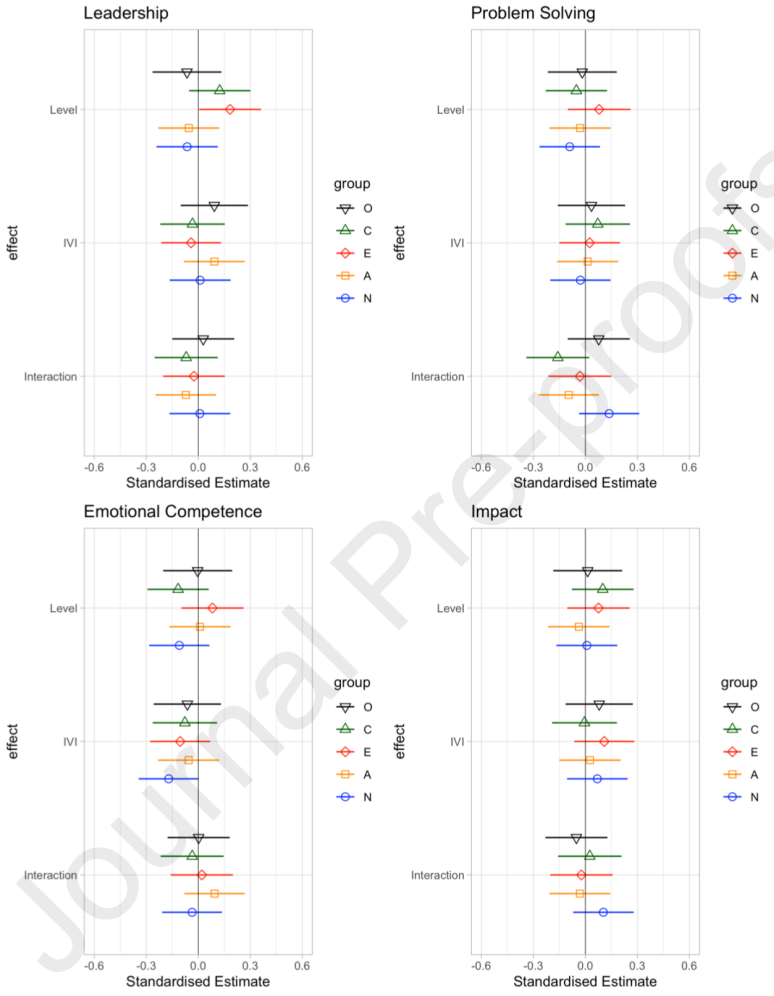

Figure C3. Forest plots of trait mean-level, intra-individual variability index (IVI) and their cross-product interaction effects on four job performance criteria based on self-report personality ratings

Highlights

- studies intra-individual variability from self and observer perspectives
- personality and performance ratings were collected for a sample of professionals
- up to eight self/other-reports were available per person from home/work contexts
- found systematic individual differences in intra-individual variability
- intra-individual variability in Neuroticism matters for job performance