

The measurement of well-being at work

Kevin Daniels¹, Emma Russell², George Michaelides¹, Emike Nasamu³, Sara Connolly¹.

¹ Employment Systems and Institutions Group, Norwich Business School, University of East Anglia, NR7 4TA, United Kingdom

² Business School, University of Sussex, Brighton, BN1 9SL, United Kingdom

³ Department of Social and Political Sciences, University of Chester, Parkgate Road
Chester, CH1 4BJ, United Kingdom

Contact: Kevin Daniels, Employment Systems and Institutions Group, Norwich Business School, University of East Anglia, NR7 4TA, United Kingdom, kevin.daniels@uea.ac.uk.

Acknowledgement: Sara Connolly, Kevin Daniels and Emike Nasamu were supported by Economic and Social Research Council grants no.s ES/N003586/1 and ES/S012648/1.

Introduction

Well-being has a prominent profile in many academic disciplines. For example, in philosophy, there is Aristotle's conception of well-being as associated with human flourishing or 'eudaimonia'. In political theory, Utilitarianism defines the main goals of policy as maximising pleasure and minimising pain (Bache & Reardon, 2016). Much more recently, some national political leaders, as well as political theorists/scientists, have become interested in well-being as an alternative marker of national progress to economic measures such as gross domestic product. Although there are many debates on the nature of well-being, the dominant view is that well-being is inherently a psychological construct (O'Donnell et al., 2014). Thus, psychological approaches to measuring well-being provide a platform for indexing the effectiveness of policy decisions made at all levels, from workplaces through to nation states (Layard, 2006). In relation to the workplace, such policies may include those directed at reducing absence rates or securing sustainable productivity gains without threatening worker health, wherein psychological well-being may serve as a leading indicator. At regional, national or even supra-national level, relevant policies pertain to labour market regulation and workplace health and safety.

As such, the purpose of this chapter is to outline some of the main and emerging issues in the measurement of workplace well-being. We consider both positive markers of well-being (e.g., job satisfaction) and markers developed from research focused on indexing psychologically harmful effects of working practices. As we shall see, research on positive markers pre-dates by some decades the emergence of positive psychology (Seligman, 2002), that sought to direct researchers away from a primary focus on negative states and psychopathology. Moreover, measures of the major well-being concepts developed reflect not just psychological constructs *per se* but overlap considerably with lay/public views on what constitutes workplace well-being (Daniels et al., 2016), namely around job satisfaction,

happiness, absence of psychological ill-health and a sense of meaning and purpose in life. This overlap is important for two reasons. First, measures of the major concepts have a starting point to establish face validity. Second, the reasons for policies and practices developed from research on workplace well-being can be conveyed with relative ease by reference to scientific concepts that are easily translated into everyday language.

At a theoretical level, the overlap is also important. There are concerns that well-being is a social construction that needs to be understood from the point of view of research participants and their specific contexts (White et al., 2014). This contrasts with the dominant approach in the psychological (and economic) sciences, wherein well-being comprises a series of more specific constructs that can be measured using quantitative rating scales that apply across all contexts. The overlap between the theoretical constructs and corresponding measures developed in the psychological literature and lay/public conceptions of well-being considerably lessens concerns over the ontological/ epistemological status of well-being.

There are concepts we do not cover in the chapter. We do not consider measures of potential workplace causes of well-being (e.g., job demands, resources, person-environment fit) or consequences (in-role performance, organisational citizenship, absence, presenteeism). Neither do we consider indicators of well-being that relate to physiology (e.g., heart rate) or expression of felt emotions (e.g., facial expressions of feelings of happiness that an individual may or may not choose to suppress). Rather we concentrate on the psychological aspects of well-being. Although there is overlap, physiological, expressive and psychological aspects of well-being are only loosely coupled (Lang, 1988).

Psychological well-being has two major components (Waterman, 1993). The first, subjective well-being, consists of summative assessments of one's life (e.g., life satisfaction) or life domain (e.g., job satisfaction) and affective well-being, which is the experience of positive affective states (e.g., joy, enthusiasm) and the relative absence of negative affective

states (e.g., lack of anxiety, feeling calm) (Diener, 1984). The second component is eudaimonic well-being which includes feelings of autonomy, mastery, personal growth, positive relations with others, purpose in life and self-acceptance (Ryff & Keyes, 1995).

Following a brief history of major developments in the measurement of well-being, the chapter will then consider current issues and complexities in well-being assessment. The first of these is one not traditionally of much concern to researchers into workplace well-being, namely, how to establish accepted monetary thresholds for changes in well-being to inform those who take decisions about well-being. For organisational decision makers and policy makers, this is a critical practical issue as it informs investment decisions: If one option returns more well-being gains than another for a lower price, then the former option should be chosen. Monetisation is also important for researchers to draw out the practical implications of their research more fully. We then consider the dynamics of well-being. As noted above, a central element of well-being relates to affective states, which are themselves highly volatile, and so capturing this element of well-being in particular has raised many issues relating to the design of measurement instruments. Looking forward, we then consider emerging issues in the dynamic assessment of well-being. One key element here is indexing variability in well-being in the same person over time. This leads to the final substantive section on considering variability in well-being between people and considering how to index well-being inequalities and why well-being inequalities might matter.

Some highlights in the measurement of well-being

The purpose of this brief section is not to provide a comprehensive listing of every measure of worker well-being, or indeed every concept. Rather, it is to give the reader an overview of the major concepts that have emerged, and some of the measures of those concepts. The choices are subjective and based on the authors' personal favourites from their combined years of researching in this field.

One of the earliest formal, quantitative measures, is that of job satisfaction. In 1951, Brayfield and Rothe published an 18-item measure of job satisfaction that included items pertaining to how interesting workers found their jobs, boredom at work, enthusiasm for their job, liking for the job and how satisfied they are at work. Job satisfaction, most generally defined as the extent to which people like or derive pleasure from their jobs (Locke, 1976), has remained as one of the key indicators of workplace well-being used in work psychology and employment relations research. Multiple measures have been developed, tapping into generalised assessments of how much people like their work or satisfaction with specific facets of their work (e.g. job security, pay, supervision, development opportunities) which are then summed into an overall score. Hackman and Oldham's Job Diagnostic Survey (1974) includes examples of both kinds of job satisfaction scale, although many researchers appear to use generalised assessments and summations of satisfaction with specific facets interchangeably. Generalised assessments of job satisfaction appear to confer two key benefits for the assessment of well-being. First, they provide a summative assessment of well-being in relation to work. Second, they can be assessed with just one, or a small number, of items (e.g., Eurofound, 2015).

Given the growth of models of occupational stress in the 1960s and 1970s (e.g., French et al., 1974; Kahn et al., 1964), typologies of occupational stressors (e.g., Cooper & Marshall, 1976) and measures of those stressors (e.g., House & Rizzo, 1972), researchers needed to incorporate measures of strain as well as satisfaction in their studies. In the UK, the 12-item measure of mental health, the GHQ12 (Goldberg, 1972), became championed as a short, unidimensional measure for workplace studies (Banks et al., 1980) that could potentially capture the influence of workplaces on clinical and subclinical mental health outcomes. The Maslach Burnout Inventory (MBI, Maslach et al., 1986) has been a popular measure to gauge the impact of workplace stressors on well-being, especially in human

service work. The MBI assesses burnout across three dimensions of emotional exhaustion, depersonalisation and reduced personal accomplishment.

To capture both negative and positive well-being reactions to work, measurement developed further in two complementary ways. One line of research sought to augment measurement in studies of burnout, the concept and measurement of work engagement was developed (Schaufeli et al., 2002), which positions positive work-related well-being to consist of three elements of vigour, dedication and absorption in work activities. Although burnout and engagement are considered to be distinct concepts, they are highly correlated (Schaufeli & De Witte, 2017).

Another line of research on negative and positive well-being took as its starting point debates concerning the dimensional structure of affect. On the one hand, Russell (1980) argued for a two-dimensional structure of affect, with dimensions of pleasantness-unpleasantness and arousal. On the other, Watson and Tellegen (1985) argued for two alternative dimensions of negative and positive affect, representing the degree to which highly activated pleasant (e.g., enthusiasm) and unpleasant (e.g., anxiety) affective states are experienced. Larsen and Diener (1992) argued that the difference between Russell's and Watson and Tellegen's models reflected the choice of rotation in factor analytic models.

Research building on models of the dimensional structure of affect has produced a range of measures of workplace affective well-being. Some of these measures (Van Katwyk et al., 2000; Warr et al, 2014) have assessed affective well-being as a composite for four unipolar 'facets' on well-being: i) High pleasure/high arousal (e.g., enthusiastic); ii) high pleasure/low arousal (e.g., relaxed); iii) low pleasure/high arousal (e.g., anxious); iv) low pleasure/low arousal (e.g., depressed). Others have argued differential tendencies to response to positively or negatively items obscures the true bipolarity of dimensions of affective states in conventional factor analytic models. Correspondingly, these researchers sought to assess

affective well-being through measures assessing bipolar dimensions and developed through using more sophisticated factor analytic methods (Daniels, 2000; Warr, 1990). Warr's measures capture two bipolar elements of well-being (depressed to enthusiastic, corresponding most closely to positive affect and anxious to contented, corresponding to {low} negative affect). Daniels' measures capture five bipolar dimensions that have two bipolar second order factors corresponding to negative and positive affect, with high and low arousal states.

In concluding this section on some, but not all major developments, in the measurement of well-being, there is one omission. Measures of eudaimonic well-being have historically attracted less attention in much work psychology research, possibly because of the focus on stress and/or health outcomes, including mental health outcomes. One measure developed for use in general populations, rather than working populations, is that developed by Ryff and Keyes (1995). This measure assessed six dimensions of eudaimonic well-being, namely autonomy, mastery, personal growth, positive relations with others, purpose in life and self-acceptance. In an analysis of indicators of psychological and eudaimonic well-being used in the European Social Survey, Huppert and So (2013) found support for two separate dimensions, with one reflecting items with a greater affective content (labelled 'positive characteristics') and another reflecting items with a greater eudaimonic content (learning new things, sense of meaning, sense of accomplishment, positive social relationships). Nevertheless, to date, there is no widely accepted and comprehensive measure of eudaimonic well-being in relation to work.

Common metrics and conversion rates across measures

As discussed in the preceding sections, there are many conceptions of well-being, as well as different measures of well-being. The choice of well-being measure will depend on the use of relevant stakeholders. Within the domain of policy making and at a macro-

economic level, Stiglitz et al. (2009) and Coyle (2014) highlighted the importance of using statistical metrics which can capture aspects of social progress and quality of life, which are absent in traditional economic indicators such as GDP (see Wallace et al (2020) for how Gross Domestic Well-being might be monitored across domains). In particular policy domains, such as health, there was also a dissatisfaction with reliance upon cost benefit analysis which required monetising all elements when justifying policy choices or evaluating policy outcomes. Evaluation techniques have now been developed which compared costs in monetary terms and benefits in quality adjusted life years and these are now embedded in health decision making (NICE, 2013). A similar approach is now being applied for evaluating workplace well-being initiatives where benefits are captured in terms of well-being (Bryce et al, 2020). These techniques of well-being cost-effectiveness analysis can be applied at a policy level (e.g., employment legislation, health and safety regulation), but have been developed primarily for use by employers faced with making choices between different workplace health and well-being initiatives.

As discussed above, well-being is multi-dimensional and may correlate with key indicators such as health, education, material living standard, social connections and so on. Layard (2016) argues that having a singular well-being metric which serves as a common currency is necessary for ease of comparison across policy domains or types of intervention. Measures of well-being need to be meaningful to individuals, in terms of providing a summary or an overview measure of their quality of life. Similarly, there needs to be a clear metric for decision makers who monitor well-being and thereby compare the well-being outcomes of various interventions or activities associated with workforce development for transformation.

National statistical agencies (e.g., UK ONS, 2011) and international organisations (e.g., Eurostat, 2010) responded to the recommendations of Stiglitz et al (2009)¹, by undertaking research on what meaningful and reliable data on well-being could be collected, and which of these could measure people's quality of life and inform decision making. In the UK, the national statistical agency (Office of National Statistics, ONS) identified four questions, reflecting psychological well-being (summative, affective and eudaimonic components): 'Overall, how satisfied are you with your life nowadays?' (summative); 'Overall, to what extent do you think the things you do in your life are worthwhile?' (eudaimonic); 'Overall, how happy did you feel yesterday?' (affective); and 'Overall, how anxious did you feel yesterday?' (affective). The latter two are clearly sensitive to changing events and the second reflects the respondent's value judgement on what is/not worthwhile. Layard (2016) recommended the first question 'Overall, how satisfied are you with your life nowadays?' be used as a common currency in measuring well-being. Responses to the question are made on a scale of 0 - 10, where 0 is not at all and 10 is completely. In relation to the workplace, the choice of a life satisfaction measure may seem unusual, when an index of job satisfaction may appear more relevant to workplace initiatives and policies. However, a metric that captures the entire life experience has the advantage of reflecting the effects of workplace initiatives and policies that reach beyond the workplace, such as flexible working practices that enhance family life or make caring responsibilities less demanding.

In the field of health, a medical intervention which yields an additional quality of life adjusted year is deemed to be cost effective if it costs less than £20-30,000 (or equivalent). Building upon this, Layard (2016) proposes that an extra unit of life satisfaction over a year converts to a threshold benefit of between £2,000-£3,000. The use of a common well-being

¹ Their report resulted from The Commission on the Measurement of Economic Performance and Social Progress

metric – a additional unit of life satisfaction over a year – alongside a monetary benchmark of acceptable costs not only enables comparisons between different workplace well-being interventions but also well-being interventions in other domains of life.

In practice, however, individuals may be interested in other aspects of well-being and researchers or organisations may monitor other variables of interest such as job satisfaction, engagement, mental health, self-esteem and social support. In such cases, Layard (2016) proposes ‘converting’ values of other metrics into a corresponding value for the ONS life satisfaction by making use of conversion rates as in Table 1. The conversion rates in Table 1 are based on empirical estimates from the analyses of Mukuria et al. (2016) and Powdthavee (2012), where panel data has been used to examine the impact of changes in each of the well-being measures upon life satisfaction.

TABLE 1 HERE

The use of life satisfaction as a single item metric measuring well-being is not uncontroversial. Huppert (2017) returns to the argument that well-being is a multidimensional construct which requires measurement of both the internal and external factors which influence it. Recent research (Marsh et al, 2020; Ruggeri et al, 2020) has sought ways to bridge theory, evidence and practice by developing a composite score based on more complex multi-item psychological measures. For decision makers – whether they be HR managers, policy makers – this approach may still be unwieldy and the simplicity of a single item metric such as life satisfaction remains a more pragmatic route to embedding well-being into organisational practice, and does not preclude using other measures in surveys of well-being.

The Measurement of Affective Well-being at Work

Affective well-being (AWB) involves a person’s evaluation of the valence and activation of their feelings, to constitute an emotional expression that has value or meaning

within a specific context. AWB will be higher if a person considers their emotional experience to be positive, meaningful and valuable (Frijda, 2008). A person's AWB may or may not be tied to a particular event or stimulus, is differently structured depending on the duration and intensity of emotions experienced, and reflects an individual's interaction with their environment (Bliese et al., 2017; Frijda, 2008; Wright & Cropanzano, 2000). As such, measures of AWB involve more than just reports of sensations, responses or feelings; AWB is imbued with meaning according to the individual's evaluation of their affective experience. It is therefore important that both the structure of affect, and the context within which it is being considered (time, place, etc.), be appropriately captured and represented, in AWB measurement².

In work contexts, AWB is increasingly recognised as a salient predictor *and* outcome of job-relevant metrics and initiatives. For example, AWB has been found to predict work outcomes such as satisfaction (Hoffmann et al., 2014; Ilies & Judge, 2004) and success (Lyubomirsky et al., 2005). It is also a significant outcome of work-based predictors such as goal conflict (Hoffmann et al., 2014) and provision of job resources (e.g. coaching and autonomy) (Xanthopoulou et al., 2012a). Because of these relationships, measuring AWB at work has become a necessary feature of much organisational-based research, and is likely to influence the extent to which work policies and initiatives are sustained in the long term (Diener et al., 2015). Moreover, AWB is a key element of well-being and, being highly volatile (Xanthopoulou et al., 2012b), measures of AWB are well placed to capture short-term and dynamic influences on changes in a person's well-being whilst still retaining the ability to capture longer-term and more stable differences between people (Xanthopoulou et al., 2020).

² This is not to say that AWB is contained enough as a construct to only represent affect experienced within that context. We recognise that there are clear spillover effects of AWB, whereby how a person feels in relation to work will also impact their feelings at home, and vice versus (Ilies et al., 2014).

Levels of AWB

Much organisational research that is concerned with measuring AWB, as either a predictor or outcome of work-related variables and stimuli, is also concerned with how AWB is constructed. AWB can be hierarchically arranged as representing three broad levels relating to the duration and stability of the construct (Frijda, 1993; Russell & Daniels, 2018). At the most stable level, trait-based characteristics of a person and their propensity to appraise emotions in a particular way, or express a particular emotional style over time, are represented (Beal & Ghandour, 2011; deNeve & Cooper, 1998; Steel et al., 2008). The level down from this involves relatively changeable aspects of emotional experience, usually framing generalised ‘mood’, or a sum of a person’s affective response over a briefer, aggregated period of time (e.g. last week, last month, yesterday) (Brief & Weiss, 2002; Weiss & Cropanzano, 1996). At the lowest, most transitory level, AWB is represented as a momentary construct that fluctuates in terms of discrete emotional expression, often in response to a specific event or stimulus (Frijda, 1993).

Measuring AWB at any of these levels requires an adaptation in approach. For example, the focal instruction used with the respondent must make clear which time period is of interest. Asking how someone feels ‘right now’ will not capture stable, trait-based affect, although it is likely to be influenced by this. Further, the terms used to capture and rate the affective experience need to be carefully considered and scored to reflect the level of interest. For example, asking if a person feels ‘good’ or ‘bad’ is more likely to capture a mood construct. Rating discrete emotions, such as how ‘angry’ or ‘calm’ a person feels is likely to be momentary-based, or would need to be aggregated with other items in a composite score if it is to represent an overall trait-based characteristic such as ‘Hostility’ or ‘Neuroticism’. To provide a reliable measure of AWB, researchers also need to weigh up how long an AWB scale needs to be, to capture the construct effectively and reliably without causing survey

fatigue, which can invalidate outcomes (Gabel et al., 2000; Stanton et al., 2002). For example, in measuring momentary AWB, respondents may need to complete an AWB scale several times a day over a period (Xanthopoulou et al., 2012b). Such scales need to be shorter, so that response protocols are not undermined (Cranford et al., 2006; Ouweneel et al., 2012). However, if respondents only need to complete a one-off measure (e.g. to measure trait-based AWB) longer scales can be justified. Table 2 summarises some of these issues.

TABLE 2 HERE

Considerations in the measurement of AWB at work

Balancing the issues outlined above requires careful evaluation for the organisational researcher. In measuring AWB at work, we suggest that there are seven key elements to consider.

Context. The *context* of the affective experience needs to be captured via the use of appropriate focal instructions and scoring bands. When the temporality of the experience is of interest then the scoring band needs to include options of frequency (e.g. Always to Never) and focal instructions need to refer to the time boundedness. For example, if the researcher is interested in momentary AWB then the focal instruction needs to ask the respondent to reflect on how they feel ‘right now’ or ‘at the present moment’. In considering mood, the focal instruction will ask the respondent to sum their experience over ‘today’ or ‘the past week’, for example. For stable traits, the focal instruction needs to look at ‘how you generally/typically feel’, or what one would ‘usually’ do or feel. Apart from temporality, context may involve understanding the intensity of the emotion, so focal instructions may ask about the ‘extent to which’ the affect was experienced, using scoring bands that refer to ‘not at all’ to ‘very much’. Further, the focal instruction can draw out whether AWB in relation to an event, experience or domain is relevant: e.g. ‘in relation to your last customer interaction’, or ‘when at work’. The Job Affective Well-being Scale (JAWS: van Katwyk et al., 2000)

provides a focal instruction that asks participants to think about their job environment, e.g., ‘my job made me feel...’. Choosing a focal instruction that represents the context is therefore significant, and researchers would be well-placed to utilise AWB scales that allow for the focal instructions to be adapted for context (without undermining the reliability and validity of the measure).

Length. The *length* of the AWB measure needs to be appropriate for the frequency with which the respondent is expected to rate their affective experiences. This is especially relevant when the respondent is completing measures alongside normal, day-to-day work/life tasks (Russell & Daniels, 2018). Too much cognitive load on the participant is likely to negatively impact response protocols and create invalid responses or even dropouts from studies (Gable et al., 2000; Scollon et al., 2003). If the respondent is frequently rating their momentary AWB in relation to a specific experience (e.g. over several times a day), the length of the AWB scale needs to be short and convenient with few items (Cranford et al., 2006). A one-off measure, specifically if stable AWB is being captured, can afford to be longer and include more items without creating survey fatigue or invalidating response styles.

Affective Structure. The nature of the ‘*affective*’ structure of well-being needs to be balanced across the scale. Although, other areas of psychology have been concerned with understanding the broad, universal categories of emotional expression (e.g. Ekman, 1992; Izard, 1977), psychologists concerned with measuring AWB have generally overlooked this (Frijda, 2008). As such, AWB measures can include an array of terms that vary in terms of categorisation, representation, activation, specificity and valence.

For example, the 20-item Positive And Negative Affect Schedule (PANAS) (Watson et al., 1988) includes states that are not feelings (e.g. strong and alert), motivational terms (e.g. determined, inspired) and emotional feelings (e.g. afraid, nervous) (Diener et al., 2009). Items are considered to be of negative or positive valence, but there are more negatively

valenced terms relating to categories of anxiety (e.g. jittery, nervous), none relating to sadness, and few representing hostility (Diener et al., 2009). Low activation emotions are also omitted; however, when these are added to the newer 60-item version of PANAS (Watson & Clark, 1999) the factor structure of the PANAS is compromised, no longer cleanly representing two factors of AWB (a positive and a negative factor). Rather, a positive affect factor emerges (with only highly activated positive affect items loading strongly onto this) and a range of negative affect factors representing hostility, fear, low activation, low self-esteem and other categories. Relatedly, a bias towards negative items is seen in both the long and short versions of the Profile of Mood States (POMS: Cranford et al., 2006; McNair et al., 1992). The POMS captures emotional terms that represent negatively valenced items across categories of anxiety, depression, anger/hostility and fatigue. Positive affect is captured with just a fifth of its items, primarily representing vigour.

In other models of affect, the valence and activation of the emotional term has been balanced equally in representations. Russell (1980), Feldman-Barrett and Russell (1998), and Larsen and Diener (1992) present circumplex models where there are two orthogonal factors representing high to low activation, and positive to negative valence. So, 'calm' represents positive valence and low activation, whereas 'angry' represents negative valence and high activation. Any term can be plotted along both factors on a continuum, and scales attempt to provide a balance of items accordingly (Daniels, 2000; Diener et al., 2009). The circumplex model does not necessarily specify from which emotional categories of affect each term should herald (e.g. representation of 'fear', 'joy', 'disgust', 'regret', etc.). Some circumplex scales, such as the Scale of Positive and Negative Experience (SPANE: Diener et al., 2009; Diener et al., 2010) include terms that represent both discrete emotional terms, but also broad categories of 'mood' (e.g. good, negative, pleasant) to overcome the context-dependency of items which, they argue, bias existing AWB measures towards certain groups.

Researchers choosing which AWB measures to use should therefore consider whether the terms used to represent affect in the scale also represent their position as to how affect is structured. If a circumplex structure is favoured, then scales need to represent hedonic tone (valence) and activation (arousal). Whether terms should reflect discrete emotions or broader mood items, probably depends on the level of affect being considered (e.g. using general or 'mood' terms at level 3 is possibly not advisable). Further, although there is no existing measure of AWB that claims to have captured the breadth of relevant emotions that emerge in the workplace, researchers should consider whether the emotions that are referenced are appropriate for predicting (or being predicted by) their contingent variables in the focal research design. For example, using the PANAS to index 'frustration' may not be helpful, as anger items are poorly represented. Scales, such as JAWS (van Katwyk et al., 2000), include terms such as 'satisfied' and 'inspired', which other authors suggest should be considered separately to AWB (Diener et al., 2009; Ilies et al., 2007; Wright & Cropanzano, 2000) and potentially could create a conceptual contamination or tautology if used to predict, for example, job satisfaction.

Scope. Relatedly, the terms used in an AWB scale need to represent the broad *scope* of emotional experiences felt in relation to human activity, without being biased towards particular cultures, age-groups (Diener et al., 2009) or other demographically relevant groups. For example, if a measure includes more 'energy' related terms, Diener et al. (2009) argue that these will be biased towards younger people who are more likely to agree that they are feeling 'active', 'alert', etc. This would result in younger people potentially being misconstrued as having higher levels of well-being in a context, compared with older responders. Across organisational research, many AWB scales have been validated and trialled in other national cultures (Schimmack et al., 2002). This is to be encouraged, but scale developers need to make clear the scope of their scale for use in different organisational

and national settings, by clarifying from which dictionaries original scale terms have been derived (and their embeddedness in the culture in question), and with which sample groups scales have been validated (as per Van Katwyk et al., 2000). Further, there is evidence that there are gender differences in the rating of emotions and tendency towards positive response bias which is not necessarily reflected in the actuality of experience (Fujita et al., 1991). More research is needed to ascertain whether ratings of AWB are biased towards people from certain groups or categories, and the extent to which this impacts the significance of findings. Researchers are encouraged to return to the original papers that detail scale development to understand whether validity and reliability data can applicably relate to the scope of the studies they wish to undertake, and the people with whom AWB will be sampled.

Scoring. The *scoring* approach applied to the AWB measure needs to represent the hierarchical level of the affective experience of interest. For example, assuming AWB can be expressed as a fluctuating, transitory state at the most volatile level, then momentary AWB is likely to best be captured by scoring affective terms as discrete items or small clusters of items (e.g. anxious, sad, etc.). At the next level, AWB may be experienced as a more stable, but still fluctuating state, such as a mood state. Scoring may therefore focus on summing or averaging discrete affective term items into broader-mood based categories (e.g. happy plus enthusiastic plus joyful may result in a generalised ‘positive’ mood). It is also possible that at this level, terms could themselves provide an affective summary by directly asking respondents if they feel in a ‘pleasant’ mood or a ‘negative’ mood. Thus, single scores for such aggregated terms may be sufficient for capturing mood-based AWB. It is unclear whether summing discrete terms or using single-item aggregate terms are synonyms for capturing ‘mood’-based AWB. At the highest, most stable level, AWB may be expressed as a trait – a representation of how one usually feels. Optimism or Neuroticism are often considered to be personality-based reflections of durable AWB (Brief & Weiss, 2002). These

constructs would usually be ‘scored’ by utilising scales of multiple items reflective of the stable construct, which are then summed or averaged.

Along with the level of affect under consideration, scoring needs to appropriately represent factor structures. For example, when SPANE (Diener et al., 2009) is scored as an overall measure of well-being (taking negative items away from positive items) the two-factor positive and negative factor structure is disrupted. In using the 20-item PANAS, only independent positive and negative valence is captured with the scoring approach, no unique activation factor is scored. Daniels’ (2000) scale can be used to represent different levels and factor structures, depending on how it is scored. Using Daniels’ 10-item measure, momentary AWB is best scored using five 2-item factors, whereas longer term, mood-based AWB (past week) is best scored across 2-3 factors (one PA and two NA factors) (Russell & Daniels, 2018). It is also apparent that if a longer-form scale is used to measure AWB, but only items relating to specific scales are scored, then the factor structure breaks down, as the original contextualisation of terms has not been accounted for (Russell & Daniels, 2018). Researchers are therefore encouraged either to use standalone short scales when brevity is needed, *or*, if using items extracted from long-form scales, to undertake reliability, validity, and factor analysis checks of the reduced range of items before applying them (Boyle, 1991; Kline, 1986; Stanton et al., 2002). Attending to the scoring of scales reveals that it is not just the upfront scale design and validation that matters when using AWB measures, the end-user scoring of scales is equally vital, to ensure that measures retain their worth in application.

Inter and intra-individual measurement. At levels 2 and 3, AWB can be measured as both an inter-individual or intra-individual construct. Dynamic, within-person measures of AWB have been enabled by the increased use of experience sampling methods (ESM), and the analytical tools (such as hierarchical linear modelling) that can examine how variables impact, or are impacted by, repeated measures of AWB (Brief & Weiss, 2002; Ilies et al.,

2007; Schimmack, 2003; Xanthopoulou et al., 2012b, 2020). This has been advantageous as work-related AWB involves understanding both the transient, fleeting feelings associated with work events and outcomes, alongside more enduring (between-person) affective tendencies. By positioning within-person ratings at a moment in time, research into the conceptualisation of AWB has developed substantially, not least because prior ratings can be used as lagged measures or predictors of subsequent ratings, enabling researchers to better understand cycles and fluctuations in affect in relation to other variables and stimuli (e.g. Park et al., 2011). In particular, using within-person measures means that changes in AWB can be directly related to events, especially when measures are captured directly before and directly after the event in question (Zhu et al., 2019).

Further, by moving beyond between-person measurement of affect, discrete item measurement and specificity in terms can be enabled, which allows for greater conceptual concordance with contextual stimuli (e.g. job events) (Brief & Weiss, 2002). Finally, including intra-individual measures in analyses can overcome some of the measurement biases that beset any form of self-report construct. By centring repeated measures data to the individual's mean, variations in AWB can be more accurately related to each participant's specific experience, rather than confounded by personal biases in comparison to the overall group or sample (Schimmack, 2003). In designing studies of AWB, researchers are therefore advised to integrate both inter and intra- individual measurement, in order to better understand the dynamic relationship between the construct and other variables, and in terms of individual participant experiences across a period (Ilies et al., 2007).

Ethics and the participant experience. In addition to the above considerations, relating to the structural and environmental issues involved in scale measurement, perhaps the most significant consideration in rating any psychological construct is the respondent. Much has been written about respondent biases (Schimmack et al., 2002), when it comes to response

styles brought about by insufficient motivation, honesty, or self-awareness (Scollon et al., 2003) or a general tendency to more positive or negative responding (Gotlib & Meyer, 1986; Schimmack et al., 2002). These all need to be attended to in the design of any scale.

Further, researchers can adhere to certain principles in designing their studies to ensure that AWB ratings consider the participant experience. First, researchers need to consider the time of day and week when respondents are asked to complete measures, as AWB tends to show different patterns according to when it is recorded. For example, PA scores appear to rise throughout the day before dropping in the evening (Clarke et al., 1989), suggesting that afternoon ratings of AWB will be more positive than morning ratings, which is important to acknowledge when using daily ratings to capture whole day effects. There are also day-of-week effects. For example, AWB (particularly ‘mood’) is rated lowest at the beginning of the working week, with positive valence more likely to be reported on weekends – specifically Saturdays (Kennedy-Moore et al., 1992; Ryan et al., 2010). Despite this general trend, asking people to rate AWB in relation to work, in their own time (e.g. after work or on the weekend) could feasibly produce more negative ratings from those who do not wish to be disturbed by thoughts of work in their time off (Derks et al., 2016; Park et al., 2011).

Thinking about affect, and rating one’s affect, could also feasibly alter the affective experience. This can mean that a measure designed to capture AWB in relation to a work event may actually have low fidelity as the feelings being captured are – in reality – related to the process of rating them. Although this interference effect has not been examined in empirical research (to the authors’ knowledge), it would be useful to understand the extent to which ratings of AWB can – in and of themselves – create an ‘intervention’ that impacts ecological validity. With other variables, using objective measures can be useful for validating self-reports. However, in rating AWB, which involves a value-laden evaluation or emotion, objective measures may not be helpful. Physiological measures could provide some

construct validity for arousal ratings but cannot suggest how meaning and valence was imbued in the level of arousal.

There are also issues in terms of participant memory recall (Fisher et al., 2016). Because momentary or mood-based AWB (level 2 and 3) can be transient constructs, asking people to recall in the evening how they felt earlier on that morning can be problematic. People may end up ‘summing’ their emotions (Reis & Gable, 2000), or being unduly affected by the memory of a particular strong emotion, depending on their own trait-based AWB (e.g. pessimists may focus more on negative events and show bias towards recalling these) (Taylor, 1991). Finally, we have the issue of measurement fatigue for participants (Stanton et al, 2002, Xanthopoulou et al., 2012b). Repeated or lengthy administrations of AWB scales can be demotivating and tiring (Scollon et al., 2003; Gable et al., 2000). This can result in response styles emerging (e.g. careless or random responding; central tendency bias, etc.) which then invalidates the study design. The message of this section is to consider the psychology of participant responding, and show due ethical concern to respondents in designing any study that requires use of AWB measures.

The CLASSIE framework for measuring AWB at work

For ease of reference, we have organised the above elements into a CLASSIE framework. This provides a summary of the issues that researchers would do well to consider, prior to designing their studies of AWB and engaging in AWB measurement at work. It could also serve as a tool for researchers interested in advancing understanding of the structure and conceptualisation of the AWB construct at work, as the discussion above highlights areas that are still unclear and require further elucidation. Table 3 provides an overview of the framework.

TABLE 3 HERE

Emotion dynamics

Further to the assessment of emotional experiences, patterns of emotion fluctuations over time can be important markers of psychological functioning and AWB (Davidson, 2015; Hollenstein, 2015; Houben et al., 2015; Koval et al., 2016). As we react emotionally to different affective events, our AWB shifts away from our normal baseline levels and at the same time our emotional system tries to regulate our emotions and return them back to baseline levels (Kuppens et al., 2010b). To explore individual differences in the degree to which we react to external stimuli and how quickly or efficiently we return to baseline, Kuppens and Verduyn (2015) propose a 2x2 taxonomy of dynamic features. This taxonomy differentiates between those features that focus on variability of affective states within a specific time period, and other features concerned with time dependency or whether affective states carry over or are sustained over time. Each of these can also be applied to individual discrete emotions or dimensions of AWB, or can be applied to combinations of emotions or multiple AWB dimensions simultaneously. This suggests a combination of four potential dynamic signatures that relate to AWB and emotional regulation in different ways, each of which is discussed below.

Emotional inertia

Emotional inertia was first introduced by Suls et al. (1998) and refers to the persistence of affective states and whether these are sustained for a long time once they are experienced. Such lingering affective states are considered to be a failure in regulating emotions and their homeostatic return back to baseline levels (Houben et al., 2015; Kuppens et al., 2010a). As such, high emotional inertia implies that the emotional system is subject to a self-perpetuating process of affective states that is less open to external influences. High levels of inertia can be evidence of psychological maladjustment and different studies have shown that it is related to other measures of AWB and eudaimonic well-being (Houben et al.,

2015), depression (Koval et al., 2016; Kuppens et al., 2010a), onset of depression in adolescence (Kuppens et al., 2012) and rumination (Koval et al., 2012). It should be noted that inertia is not indicative of lower well-being only for negative affective states. Although, it is more intuitive to think of inertia of negative emotions as an indicator of low AWB, persistence of positive emotions is also considered as indication of maladaptive emotional regulation and has been associated with lower AWB (Houben et al., 2015; Trull et al., 2015). Nevertheless, this association tends to be weaker for inertia of positive emotions and stronger for negative emotions (Koval et al., 2016).

Emotional inertia is typically captured as the autocorrelation coefficient from successive measurements of emotions (e.g. Koval et al., 2013) and can easily be applied to discrete emotions, or to individual dimensions of AWB such as those discussed earlier. More elaborate approaches use a multilevel model to estimate autoregressive effects as random slopes of lagged consecutive measures of affect (Jongering et al., 2015; Koval et al., 2016). Although more complex, this approach has the benefit of being integrated into a bigger model that allows controlling for other time dependent and time independent covariates. It is also possible to use this approach to model predictors of affect inertia by testing cross-level interactions between lagged effects with person-level covariates (Kuppens et al., 2010b).

A key issue for the measurement of inertia is that the actual estimate may change according to the study design and the interval at which the successive measurements of emotions are collected (Ebner-Priemer & Sawitzki, 2007; Koval et al., 2013). One specific concern is that autocorrelation coefficients tend to wane for longer time intervals so shorter intervals would produce higher and incomparable estimates to longer intervals. This can potentially be addressed through the adoption of continuous time models that take into account the length of each time interval in the estimation of inertia (Oravecz et al., 2011). Such models are in effect the equivalent of a random effect autoregressive slope for

continuous time (Oravecz & Tuerlinckx, 2011) and can be ideal for capturing inertia in experience sampling studies with random or unequally spaced data collection. Nevertheless, the link between lower well-being and emotional inertia has been established for different timescales varying from seconds to minutes or days (Koval et al., 2013; Kuppens et al., 2012; Neumann et al., 2011).

Emotional cross-lags.

Emotional cross-lags encapsulate a similar idea as emotional inertia but apply to different discrete emotions or different dimensions and how they influence and perpetuate each other over time. Thus, cross-lagged effects reflect how different emotions can increase or decrease the experience of other emotions. This is referred to as emotional augmentation or blunting and, similar to affect inertia, at high levels can signify a self-contained system that is less open to external stimuli – which is characteristic of mood disorders. Combining emotional inertia with emotional cross-lags can allow the construction of an emotion-network to represent the dynamic relationships between different discrete emotions over time. The autoregressive and cross-regressive estimates for these emotion networks can be obtained as random slopes from a series of multilevel regressions (one for each emotion) or by estimating a multilevel vector autoregressive model for all the emotions or dimensions simultaneously (Bringmann et al., 2016, 2013). The strength of the autoregressive and cross-regressive relationships is typically referred to as the emotional density of the network and reflects the degree to which the whole emotional system is more resistant to change and has been associated with mood disorders (Pe et al., 2015). More complex metrics can be obtained from the emotional network by applying network analysis to further understand the resulting architecture of the networks.

Emotional variability

Emotional variability captures the degree to which emotions fluctuate over time and it is considered to be indicative of the degree to which individuals are more or less sensitive to external stimuli (Kuppens & Verduyn, 2015). High levels of variability, which imply stronger emotional reactions, are generally considered to be maladaptive and meta-analytic evidence suggest that variability is related to numerous indicators of well-being including negative AWB, eudaimonic well-being as well as a number of disorders such as a depression, bipolar disorder, anxiety and borderline personality disorder (Houben et al., 2015). Similar to inertia, variability is considered to be maladaptive regardless of whether it is variability of positive or negative emotions (Gruber et al., 2013). Moreover, whilst it is high variability that is typically associated with low well-being, very low variability or reactivity can be equally problematic. For example, depression is associated with a decrease in emotional responsiveness to either positive or negative stimuli (Rottenberg et al., 2005).

Variability can be easily estimated by calculating dispersion of within person measures using standard deviation or variance (e.g. Eaton & Funder, 2001; Eid & Diener, 1999). This can be applied to either measures of discrete emotions or affect dimensions based on the circumplex model. More sophisticated approaches are founded on modelling variability as a latent constructs using multilevel innovation variance. Innovation variance simply refers to the model residual from a time series model and can be estimated per person as a random model parameter using multilevel location scale models of repeated measures (Jongerling et al., 2015; Schuurman & Hamaker, 2019; Wang et al., 2012). The advantage of using this more complex approach is that these estimates of variability are based on the residual or what cannot be explained by the rest of the model. Thus, these latent estimates exclude any potential emotional inertia or effects of time dependent and time independent covariates.

A related concept to variability is that of instability which refers to the magnitude of change from one point in time to the next. Thus, in contrast to variability which focuses solely on the amplitude of changes in affective states, instability combines variability with temporal dependency (Trull et al., 2015, 2008). To capture this construct a number of different indices have been proposed including the mean squared successive difference between consecutive measurements of affective states, the proportion of acute changes in affect over total changes (Jahng et al., 2008) and aggregate point by point changes (Santangelo et al., 2014). The mean squared successive difference is the most common approach and although it is treated as a separate construct from measures of either variability or temporal variability, it is closely related to both and it is possible to express instability as a formula of variance and autocorrelation of within person measures of affect (Jahng et al., 2008). It is no surprise then that, similar to emotional variability, instability has also been linked to negative AWB, eudaimonic well-being, as well as numerous psychological disorders (Houben et al., 2015; Trull et al., 2015);

Emotional covariation

Emotional covariation transposes the idea of affect variability to multiple dimensions and examines the degree to which different emotions or different dimensions of within person AWB covary. The substantive meaning of such contemporaneous associations of within person variability is that they reflect an inability to differentiate between different discrete emotions or affect dimensions. This is also referred to as emotional differentiation and is considered to be necessary for emotional regulation (Barrett et al., 2001). Individuals with difficulties in differentiating between emotions tend to experience more negative affect, depression and reduced self-esteem (Erbas et al., 2014).

The simplest way through which emotional differentiation can be estimated is via within-person bivariate correlations of different affect dimensions or emotions. More

sophisticated approaches are based on estimating the intraclass correlation coefficient to represent agreement between different emotions over the duration of a study (Tomko et al., 2015; Tugade et al., 2004). A more elaborate approach is to estimate emotional differentiation as a latent variable from the residual covariance matrix of a multivariate multilevel location-scale model. Similar to estimating variability as a residual that varies per person, in a multivariate model it is possible to do the same for the covariance residual matrix (Jongerling et al., 2015). Although this is a complex approach, it has the benefits of (a) estimating both variability and emotional differentiation as latent parameters simultaneously, and (b) that estimating these constructs from a full model allows for controlling for other momentary, personal or contextual variables or for other dynamic processes such as inertia and cross-lagged effects at the same time.

Two related concepts that also capture variability in multiple dimensions using the core affect model are affect pulse and affect spin (Kuppens et al., 2007; Moskowitz & Zuroff, 2004). These measures are based on representing affect scores on the two-dimensional circumplex. Affect pulse, which captures the intensity of changes in affective states, is estimated as the within person standard deviation of the distance of each affect score on the circumplex and the neutral midpoint position. Affect spin, or affect quality variability, is calculated as the standard deviation of the angular displacement of each emotion experienced on the circumplex model (Beal et al., 2013; Kuppens et al., 2007). Thus, pulse captures variability of the intensity of emotions whilst remaining agnostic to the specific emotions, and spin captures the circular variability of changing emotions regardless of their intensity. Similar to other measures of variability and differentiation, high pulse and spin are also considered to be maladaptive and have been positively associated with borderline personality disorder (Russell et al., 2007), and personality traits such as Neuroticism and Pessimism and negatively with Extroversion and Optimism (Kuppens et al., 2007).

Assessing inequalities in well-being between people

Inequalities matter for health and well-being. One of the most widely studied types of inequality is inequalities in income. Studies have consistently found an inverse correlation between country level income inequality and a range of health and social problem indicators, such that countries or regions with the highest levels of income inequality also tend to have worse health and social outcomes (Pickett & Wilkinson, 2015). Recorded outcomes include educational attainment, teenage birth rates, social mobility, crime rates, mental health problems and a range of other physical health outcomes. These associations are explained through social comparison processes, in which inequality is a social stressor that undermines interpersonal trust and social cohesion.

A smaller research stream has examined well-being inequalities across countries, often using life satisfaction as a summative index of well-being. This research stream also indicates well-being inequalities are inversely associated with well-being (Goff et al., 2016). This association holds after taking into account factors that reflect any artefactual influence on the size of the correlation brought about by range restriction (i.e., the measure of well-being is bounded by the extremes of the rating scale meaning people with very high/low and high/low levels of well-being will tend to bunch at the extremes of the rating scale rather than being differentiated). Moreover, where people have poor levels of well-being, social and health outcomes may be even worse for some groups than others (e.g., differentiation by gender and socio-economic class; Linder et al., 2019).

We know of no research on inequalities in well-being in work organisations, although there is a well-developed stream of research on inequalities in how leaders treat subordinates (Martin et al., 2018). There are many reasons why investigating well-being inequalities in organisations is potentially important. If inequalities in well-being are causal in reducing average levels of well-being in an organisation, then there are important implications for

health and social outcomes of workers. If inequalities in well-being in an organisation undermine trust between co-workers and/or managers, then there are important implications for co-operation and conflict. In both cases, there are potential consequences for organisational performance, for example through higher absence rates (health and well-being path), reduced organisational citizenship or industrial disputes and grievances (trust and social cohesion path).

The question then becomes one of deciding on the best means of assessing well-being inequalities. Besides from deciding on the best aspect of well-being to use (life satisfaction, job satisfaction, affective well-being), there are choices concerning the best means of assessing inequality through the distribution of well-being scores in a given unit (country, organisation, team). Quick and Devlin (2018) reviewed a number of ways of assessing well-being inequalities. They divided these into measures of dispersion (e.g., standard deviation, variance, coefficient of variation, Gini coefficient) and measures based on a threshold (e.g., average well-being of the bottom 20% of the well-being distribution compared to average well-being of the top 80% of the distribution, percentage of the distribution falling below a given well-being score).

Quick and Devlin (2018) note that measures based on dispersion each have their own weaknesses. However, one critical weakness shared by most measures of dispersion is that they do not capture the difference in well-being between the best and worst off in well-being terms. Put another way, interventions focused on reducing well-being inequalities through minimising the standard deviation of well-being could work just as easily through reducing the well-being of the best off as increasing the well-being of the worst off. In policy terms at least, this would make very little sense. Measures based on thresholds can capture differences between the best and worst off, but the choice of thresholds would appear to be arbitrary. This could be especially problematic in policy applications, where the choice of threshold

could be manipulated to suit some rather than other policy options. However, the relative advantages and disadvantages of different approaches to assessing well-being inequalities in organisational research has yet to be investigated.

Conclusions

Theoretical approaches to understanding well-being have a long history and research on the assessment of psychological well-being has produced an enormous volume of research that includes the assessment of workplace well-being. This has led to a range of different instruments, some more suited to some theoretical approaches, methodologies and applications than others. Notwithstanding, although we have well developed knowledge of how to assess the level of someone's well-being at a given moment in time, there are four clear conclusions. First, making an assessment of well-being is not straightforward and involves a number of design choices (see Table 3). Second, far less research has examined variability in markers of workplace well-being and the implications of variability, whether variability relates to patterns in the dynamics of well-being over time or variability relates to inequalities in well-being within social groups. Third, although recent developments in well-being economics have provided a way of monetising well-being policy options for (managerial) decision makers, this is a relatively new field of research that has yet to come fully to terms with the multi-dimensional and dynamic nature of well-being and how to incorporate concerns about minimising well-being inequalities between people into the calculations. Fourth, and summarising the first three conclusions, there is much we know but much we still do not know about the measurement of workplace well-being.

References

- Bache, I., & Reardon, L. (2016). *The politics and policy of well-being: Understanding the rise and significance of a new agenda*. Cheltenham: Edward Elgar Publishing.
- Barrett, L.F., Gross, J., Christensen, T.C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion, 15*, 713–724.
- Beal, D.J., & Ghandour, L. (2011). Stability, change, and the stability of change in daily workplace affect. *Journal of Organizational Behavior 32*, 526-546.
- Beal, D.J., Trougakos, J.P., Weiss, H.M., & Dalal, R.S. (2013). Affect spin and the emotion regulation process at work. *Journal of Applied Psychology, 98*, 593-605.
- Bliese, P.D., Edwards, J.R., & Sonnentag, S. (2017). Stress and well-being at work: A century of empirical trends reflecting theoretical and societal influences. *Journal of Applied Psychology, 102*, 389-402
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*, 291-294.
- Brayfield, A.H., & Rothe, H.F. (1951). An index of job satisfaction. *Journal of applied psychology, 35*, 307-311.
- Brief, A.P., & Weiss, H.M. (2002). Organizational behavior: Affect in the workplace. *Annual Review of Psychology, 53*, 279-307.
- Bringmann, L.F., Pe, M.L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., ... Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment, 23*, 425–435.
- Bringmann, L.F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS one, 8*, e60188.

- Bryce, A., Bryan, M., Connolly, S., & Nasamu, E., (2020) *Workplace Cost Effectiveness Analysis (CEA) Calculator: User Manual*. London: What Works Centre for Well-being.
- Clark, L.A., Watson, D., & Leeka, J. (1989). Diurnal variation in the positive affects. *Motivation and Emotion, 13*, 205-234.
- Cooper, C.L., & Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology, 49*, 11-28.
- Coyle, D. (2014). *GDP: a brief but affectionate history*. Princeton: Princeton University Press.
- Cranford, J.A., Shrout, P.E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32*, 917-929.
- Daniels, K. (2000). Measures of five aspects of affective well-being at work. *Human Relations 53*, 275–294.
- Daniels, K., Connolly, S., Ogbonnaya, C., Tregaskis, O., Bryan, M.L., Robinson-Pant, A., & Street, J. (2018). Democratisation of well-being: stakeholder perspectives on policy priorities for improving national well-being through paid employment and adult learning. *British Journal of Guidance & Counselling, 46*, 492-511.
- Davidson, R.J. (2015). Comment: Affective chronometry has come of age. *Emotion Review, 7*, 368–370.
- De Neve K.M., & Cooper, H. (1998). The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197-229.
- Derks, D., Bakker, A. B., Peters, P., & van Wingerden, P. (2016). Work-related smartphone

- use, work–family conflict and family role performance: The role of segmentation preference. *Human Relations*, 69, 1045-1068.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542-575.
- Diener, E., Oishi, S., & Lucas, R.E. (2015) National accounts of subjective well-being. *American Psychologist*, 70, 234-242.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research* 97(2): 143-156.
- Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D. W., & Oishi, S. (2009). New measures of well-being. In Diener E. (Ed.) *Assessing well-Being. Social indicators research series*, vol 39. Dordrecht: Springer (pp. 247-266).
- Eaton, L.G., & Funder, D.C. (2001). Emotional experience in daily life: Valence, variability, and rate of change. *Emotion*, 1, 413.
- Ebner-Priemer, U.W., & Sawitzki, G. (2007). Ambulatory assessment of affective instability in borderline personality disorder: The effect of the sampling frequency. *European Journal of Psychological Assessment*, 23, 238.
- Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, 76, 662.
- Ekman, P. (1992). Facial expressions of emotion: an old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 335, 63-69.
- Erbas, Y., Ceulemans, E., Lee Pe, M., Koval, P., & Kuppens, P. (2014). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition & Emotion*, 28, 1196–1213.

- Eurofound. (2015). *6th European Working Conditions Survey – Technical Report*. Dublin: Eurofound.
- Eurostat, (2010). *Feasibility study for well-being indicators. Task 4: critical review*. Luxembourg: Office for Official Publications of the European Communities.
- Feldman Barrett, L., & Russell, J.A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74, 967-984.
- French, J. R. P., Rodgers, R., & Cobb, S. (1974). Adjustment as person environment fit. In: Coelho, G.V., Hamburg, D.A. & Adams, J.E. (Ed.s) *Coping and adaptation*. New York: Basic Books (316-333).
- Frijda, N.H. (1993). Moods, emotion episodes, and emotions. In: Lewis, M., & Haviland, J.M. (Ed.s) *Handbook of emotions*. New York: Guilford Press (381-403).
- Frijda, N.H. (2008). The Psychologists' point of view. In Lewis, M., Haviland-Jones, J.M., & Feldman Barrett, L. (Ed.s) *Handbook of emotions (3rd edition)*. New York: Guilford Press (68-87).
- Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: the case for emotional intensity. *Journal of Personality and Social Psychology*, 61, 427-434
- Gable, S.L., Reis, H.T., & Elliot, A.J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, 78, 1135–1149.
- Goff, L., Helliwell, J.F., & Mayraz, G. (2016). The welfare costs of well-being inequality. NBER Working Paper Series, 27.
<https://faculty.arts.ubc.ca/jhelliwell/papers/w21900.pdf>
- Goldberg, D.P. (1972). *The detection of psychiatric illness by questionnaire*. London: Oxford University Press.
- Gotlib, I.H., & Meyer, J.P. (1986). Factor analysis of the Multiple Affect Adjective Check

- List: A separation of positive and negative affect. *Journal of Personality and Social Psychology*, 50, 1161-1165.
- Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion*, 13, 1-6.
- Hackman, J.R., & Oldham, G.R. (1974). The Job Diagnostic Survey: An instrument for the diagnosis of jobs and the evaluation of job redesign projects. Yale: Yale University School of Organization and Management.
- Hofmann, W., Luhmann, M., Fisher, R.R., Vohs, K.D., & Baumeister, R.F. (2014). Yes, but are they happy? Effects of trait self-control on affective well-being and life satisfaction. *Journal of Personality*, 82, 265-277.
- Hollenstein, T. (2015). This time, it's real: Affective flexibility, time scales, feedback loops, and the regulation of emotion. *Emotion Review*, 7, 308–315.
- Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological bulletin*, 141, 901-930.
- House, R.J., & Rizzo, J.R. (1972). Toward the measurement of organizational practices: Scale development and validation. *Journal of Applied Psychology*, 56, 388-396.
- Huppert, F.A., (2017). *Measurement really matters*. London: What Works Centre for Well-being.
- Huppert, F. A., & So, T. T. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, 110, 837-861.
- Ilies, R., & Judge, T.A. (2004). An experience-sampling measure of job satisfaction and its relationships with affectivity, mood at work, job beliefs, and general job

- satisfaction. *European Journal of Work and Organizational Psychology*, 13, 367-389.
- Ilie, R., Schwind, K.M., & Heller, D. (2007). Employee well-being: A multilevel model linking work and nonwork domains. *European Journal of Work and Organizational Psychology*, 16, 326-341.
- Izard, C.E., (1977). *Human emotions*. New York: Plenum Press.
- Jahng, S., Wood, P.K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13, 354-375.
- Jongerling, J., Laurenceau, J.-P., & Hamaker, E.L. (2015). A multilevel ar (1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, 50, 334–349.
- Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D., & Rosenthal, R.A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley
- Kammann R, & Flett R. (1983)/ Affectometer 2: a scale to measure current level of general happiness. *Australian Journal of Psychology*, 82, 1007-1022.
- Kennedy-Moore, E., Greenberg, M.A., Newman, M.G., & Stone, A.A. (1992). The relationship between daily events and mood: The mood measure may matter. *Motivation and Emotion*, 16, 143-155.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Koval, P., Kuppens, P., Allen, N.B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & Emotion*, 26, 1412–1427.
- Koval, P., Pe, M.L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13, 1132-1141.

- Koval, P., Sütterlin, S., & Kuppens, P. (2016). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology, 6*, 1997.
- Kuppens, P., & Verduyn, P. (2015). Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry, 26*, 72–79.
- Kuppens, P., Allen, N.B., & Sheeber, L.B. (2010a). Emotional inertia and psychological maladjustment. *Psychological Science, 21*, 984–991.
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010b). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*, 1042-1060.
- Kuppens, P., Sheeber, L.B., Yap, M.B., Whittle, S., Simmons, J.G., & Allen, N.B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion, 12*, 283.
- 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*, 262.
- Lang, P.J. (1988). What are the data of emotion. In Hamilton, V., Bower, G.H. & Frijda N. (Ed.s), *Cognitive perspectives on emotion and motivation* (NATO ASI Series D Vol. XXXIV). Dordrecht: Kluwer, pp 173-191.
- Larsen, R.J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology, 13*, 25-29.
- Layard, R., (2016), *Measuring well-being and cost-effectiveness analysis using subjective well-being*, London: What Works Centre for Well-being.
- Locke, E.A. (1976). The nature and causes of job satisfaction. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally (pp. 1297–1343).

- Linder, A., Gerdtham, U.G., Trygg, N., Fritzell, S., & Saha, S. (2020). Inequalities in the economic consequences of depression and anxiety in Europe: a systematic scoping review. *European Journal of Public Health, 30*, 767-777.
- Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131*, 803-855.
- Marsh H.W., Huppert, F.A., Donald, J.N., Horwood, M.S., & Sahdra, B.K., (2020). The well-being profile (WB-Pro): Creating a theoretically based multidimensional measure of well-being to advance theory, research, and practice. *Psychological Assessment, 32*, 294-313.
- Martin, R., Thomas, G., Legood, A., & Dello Russo, S. (2018). Leader–member exchange (LMX) differentiation and work outcomes: Conceptual clarification and critical review. *Journal of Organizational Behavior, 39*, 151-168.
- Maslach, C., Jackson, S. E., Leiter, M. P., Schaufeli, W. B., & Schwab, R. L. (1986). *Maslach burnout inventory* (Vol. 21). Palo Alto, CA: Consulting Psychologists Press.
- McNair, D.M., Lorr, M., & Droppleman, L.F. (1992). *EdITS manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- Moskowitz, D., & Zuroff, D.C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology, 86*, 880.
- Mukuria, C., Rowen, D., Peasgood, T., & Brazier, J. (2016). *An empirical comparison of well-being measures used in UK*. Research Interim Report RR0027: University of Sheffield and the University of York.
- Neumann, A., Van Lier, P.A., Frijns, T., Meeus, W., & Koot, H.M. (2011). Emotional dynamics in the development of early adolescent psychopathology: A one-year longitudinal study. *Journal of Abnormal Child Psychology, 39*, 657–669.

- NICE (2013). *Guide to the methods of appraisal*. London: National Institute of Health and Care Excellence.
- O'Donnell, G., Deaton, A., Durand, M., Halpern, D. & Layard, R. (2014). *Well-being and Policy*. London: Legatum Institute.
- ONS. (2011). *Initial investigation into subjective well-being from the opinions survey*. Cardiff: Office for National Statistics.
- Oravecz, Z., & Tuerlinckx, F. (2011). The linear mixed model and the hierarchical Ornstein–Uhlenbeck model: Some equivalences and differences. *British Journal of Mathematical and Statistical Psychology*, 64, 134–160.
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, 16, 468–490.
- Ouweneel, E., Le Blanc, P.M., Schaufeli, W.B., & van Wijhe, C. I. (2012). Good morning, good day: A diary study on positive emotions, hope, and work engagement. *Human Relations*, 65, 1129–1154.
- Park, Y., Fritz, C., & Jex, S.M. (2011). Relationships between work-home segmentation and psychological detachment from work: the role of communication technology use at home. *Journal of Occupational Health Psychology*, 16, 457–467.
- Pavot, W., & Diener, E. (2008). The satisfaction with life scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3, 137–152.
- Pe, M.L., Kircanski, K., Thompson, R.J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., ... Gotlib, I.H.. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, 3, 292–300.
- Pickett, K.E., & Wilkinson, R.G. (2015). Income inequality and health: a causal review. *Social Science & Medicine*, 128, 316–326.

- Powdthavee, N. (2012). Jobless, friendless and broke: What happens to different areas of life before and after unemployment? *Economica*, 79, 557-575.
- Quick, A. & Devlin, S. (2018). *Measuring well-being inequality: Working paper on the selection of a headline indicator*. London: New Economics Foundation.
- Reis, H.T., & Gable, S.L. (2000). Event-sampling and other methods for studying everyday experience. In Reis H.T. & Judd, C.M. (Ed.s) *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press (190-222).
- Rottenberg, J., Gross, J.J., & Gotlib, I.H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology*, 114, 627-639.
- Ruggeri, K., Garcia-Garzon, E., Maguire, A., Matz, S., & Huppert, F.A., (2020). Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes*, 18, 1-16.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- Russell, E., & Daniels, K. (2018). Measuring affective well-being at work using short-form scales: Implications for affective structures and participant instructions. *Human Relations*, 71, 1478-1507.
- Russell, J.J., Moskowitz, D., Zuroff, D.C., Sookman, D., & Paris, J. (2007). Stability and variability of affective experience and interpersonal behavior in borderline personality disorder. *Journal of Abnormal Psychology*, 116, 578-588.
- Ryan, R.M., Bernstein, J.H., & Brown, K.W. (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. *Journal of Social and Clinical Psychology*, 29, 95-122.
- Ryff, C.D. & Keyes, C.L.M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.

- Santangelo, P., Reinhard, I., Mussgay, L., Steil, R., Sawitzki, G., Klein, C., ... Ebner-Priemer, U. W. (2014). Specificity of affective instability in patients with borderline personality disorder compared to posttraumatic stress disorder, bulimia nervosa, and healthy controls. *Journal of abnormal psychology, 123*, 258-272.
- Schaufeli, W.B., Salanova, M., Gonzalez-Roma, V., & Bakker, A.B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*, 71–92.
- Schaufeli, W., & De Witte, H. (2017). Outlook work engagement in contrast to burnout: Real and redundant!. *Burnout research, 5*, 58-60.
- Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies, 4*, 79-106.
- Schimmack, U., Radhakrishnan, P., Oishi, S., Dzokoto, V., & Ahadi, S. (2002). Culture, personality, and subjective well-being: integrating process models of life satisfaction. *Journal of Personality and Social Psychology, 82*, 582-593.
- Schuurman, N.K., & Hamaker, E.L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods, 24*, 70-91.
- Scollon, C.N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5-34.
- Seligman, M.E.P. (2002). Positive psychology, positive prevention, and positive therapy. In Snyder, C.R., & Lopez, S.J. (Ed.s.), *Handbook of positive psychology*. New York: Oxford University Press (pp. 3–9).
- Stanton, J.M., Sinar, E.F., Balzer, W.K., & Smith, P.C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167-194.
- Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*, 138-161.

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009).

Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a Rasch analysis using data from the Scottish health education population survey. *Health and Quality of Life Outcomes*, 7, 15-22.

Stiglitz, J.E., Sen, A., & Fitoussi, J.P., (2009). *Report by the commission on the measurement*

of economic performance and social progress. Paris: Commission on the

Measurement of Economic Performance and Social Progress.

Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective

inertia, and neuroticism. *Personality and Social Psychology Bulletin*, 24, 127–136.

Taylor, S.E. (1991). Asymmetrical effects of positive and negative events: the mobilization-

minimization hypothesis. *Psychological Bulletin*, 110, 67-85.

Taylor, M.F., Brice, J., Buck, N., & Prentice-Lane, E. (2018). *British Household Panel*

Survey user manual volume A: Introduction, technical report and appendices.

Colchester: University of Essex.

Tomko, R.L., Lane, S.P., Pronove, L.M., Treloar, H.R., Brown, W.C., Solhan, M.B., ...

Trull, T. J. (2015). Undifferentiated negative affect and impulsivity in borderline

personality and depressive disorders: A momentary perspective. *Journal of Abnormal*

Psychology, 124, 740-753.

Trull, T.J., Lane, S.P., Koval, P., & Ebner-Priemer, U.W. (2015). Affective dynamics in

psychopathology. *Emotion Review*, 7, 355–361.

Trull, T.J., Solhan, M.B., Tragesser, S.L., Jahng, S., Wood, P.K., Piasecki, T.M., & Watson,

D. (2008). Affective instability: Measuring a core feature of borderline personality

disorder with ecological momentary assessment. *Journal of Abnormal Psychology*,

117, 647-661.

- Tugade, M.M., Fredrickson, B.L., & Feldman Barrett, L. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72*, 1161–1190.
- Van Katwyk, P.T., Fox, S., Spector, P.E., & Kelloway, E.K. (2000). Using the Job-Related Affective Well-Being Scale (JAWS) to investigate affective responses to work stressors. *Journal of Occupational Health Psychology, 5*, 219-230.
- Wallace, J., Ormston, H., Thurman, B., Diffley, M., McFarlane, M., & Zubairi, S., (2020). Gross Domestic Well-being (GDWe): an alternative measure of social progress. Dunfermline: Carnegie UK Trust.
- Wang, L.P., Hamaker, E., & Bergeman, C. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods, 17*, 567-581.
- Warr, P. (1990). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology, 63*, 193-210.
- Warr, P., Bindl, U. K., Parker, S. K., & Inceoglu, I. (2014). Four-quadrant investigation of job-related affects and behaviours. *European Journal of Work and Organizational Psychology, 23*, 342-363.
- Waterman, A.S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology, 64*, 678-691.
- Watson, D., & Clark, L.A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*. University of Iowa: Department of Psychology Publications.
- Watson, D, Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology 54*, 1063-1070.

- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.
- Weiss, H.M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes, and consequences of affective experiences at work. *Research in Organizational Behavior* 18, 1–74.
- White, S.C., Gaines Jr, S.O. & Jha, S. (2014). Inner well-being: Concept and validation of a new approach to subjective perceptions of well-being. *Social Indicators Research*, 119, 723-746.
- Wright, T.A., & Cropanzano, R. (2000). Psychological well-being and job satisfaction as predictors of job performance. *Journal of Occupational Health Psychology*, 5(1), 84-94.
- Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2012 a). A diary study on the happy worker: How job resources relate to positive emotions and personal resources. *European Journal of Work and Organizational Psychology*, 21, 489-517.
- Xanthopoulou, D., Bakker, A.B., & Ilies, R. (2012b). Everyday working life: Explaining within-person fluctuations in employee well-being. *Human Relations*, 65, 1051-1069.
- Xanthopoulou, D., Daniels, K., Sanz-Vergel, A.I. (2020). The temporal perspective on well-being at work: Lessons learned and future trends. In Griep, Y., Hansen, S., Vantilborgh, T., Hofmans, J. (Ed.s), *Handbook of dynamic organizational behavior. volume 1: A dynamic look at organizational behavior topics*. Cheltenham: Elgar (290-303).
- Zhu, Z., Kuykendall, L., & Zhang, X. (2019). The impact of within-day work breaks on daily recovery processes: An event-based pre-/post-experience sampling study. *Journal of Occupational and Organizational Psychology*, 92, 191-211.

Table 1: Conversion rates for different measures of well-being into life satisfaction.

<i>Well-being measure</i>	<i>Range</i>	<i>Exchange rate</i>
Life satisfaction (ONS ³)	0-10	1
Satisfaction with Life Scale ⁴	5-35	0.24
Worthwhile (ONS)	0-10	0.75
Happy (ONS)	0-10	0.72
Anxious (ONS)	0-10	0.35
General Health Questionnaire ⁵	0-36	-0.21
Short Warwick Edinburgh Mental Well-being Scale ^{6 7}	7-35	0.25
Satisfaction with job (BHPS ⁸)	1-7	0.49
Satisfaction with income (BHPS)	1-7	0.61
Satisfaction with amount of leisure time (BHPS)	1-7	0.57
Satisfaction with use of leisure time (BHPS)	1-7	0.62
Satisfaction with social life (BHPS)	1-7	0.60
Satisfaction with health (BHPS)	1-7	0.63

Source: Bryce et al. 2020

³ Office of National Statistics (ONS, 2011)

⁴ Pavot and Diener (2008)

⁵ Goldberg and Williams (1988)

⁶ Kammann and Flett (1983)

⁷ Stewart-Brown et al. (2009)

⁸ British Household Panel Survey (Taylor et al., 2018)

Table 2: Measuring AWB at different levels.

<i>Level</i>	<i>Affective structure</i>	<i>Duration of affective experience</i>	<i>Example</i>	<i>Suggested focal instruction</i>	<i>Suggested scoring approach</i>	<i>Recommended scale length</i>
1	Trait-based	Stable	Optimism; Neuroticism	“To what extent are you generally...”	Summing/averaging scale items to provide an overall score representing each trait-based factor (potentially bipolar)	Can be lengthier; usually a one-off administration
2	Summative, aggregate of emotion	Fluctuates somewhat	Positive Affect (PA); Bad mood	“Over the past week, to what extent have you felt...”	Summing/averaging scale items of discrete emotional terms (potentially related to valence and/or arousal) OR using individual scores from aggregated item terms (such as bad/good, positive/negative mood states). Several factors may be represented (potentially bipolar).	Can be one-off or repeated administration (e.g. every day for 10 working days), so needs to be shorter.
3	Discrete emotions or feelings	Momentary fluctuation	Tired; Enthusiastic	“At the present time, to what extent do you feel...”	Individual item scores or summing/averaging item scores to represent discrete affective clusters of items (potentially bipolar).	Repeated administrations likely (e.g. in response to daily events), so needs to be shortest form with the fewest scale items.

Table 3: The CLASSIE framework for measuring AWB at work

	Measurement issue	Consideration	Future research
C	Context	The focal instruction (and scoring bands) should be made relevant to the temporality and intensity of the construct and/or the event, experience or domain (e.g., work) of interest.	Does the reliability and validity of the scale change when the focal instruction is changed to reflect a different context?
L	Length	Inclusion of the number of items must consider the participants' available time and cognitive load in rating (e.g., the frequency with which the scale is administered, concurrent use of other scales, and engagement with other tasks).	Ensure short-form measures are valid and reliable in their own right.
A	Affect Structure	Choose scales that reflect the theoretical affective structure of interest. Ensure concordance of scales with associated variables of interest.	What are the broad, universal categories of AWB in relation to valence and arousal constructs at work?
S	Scope	Check the scope of the sample used in the original scale development, and check that scale items will not adversely bias results from any demographic group or culture.	More research is needed to ascertain how scale items may bias well-being ratings for different groups.
S	Scoring	Ensure hierarchical levels and factor structures of AWB is appropriately represented by the aggregation or summation of item scores.	Do aggregate terms (e.g., positive) capture level 2 mood in the same way as summing discrete terms (e.g., joy, enthusiasm, vigour)?
I	Inter- and intra-individual measurement	Use within- and between- measures and analyses of AWB, wherever possible. Measure affect change by taking pre-event ratings.	Continue to investigate work-based relationships between inter- and intra-individual AWB and work activity (including cyclical or longitudinal fluctuations).
E	Ethics and the participant	Capture variations in participant responding that may be due to response bias, time of day or week effects, memory recall biases, etc.	Examine interference effects in studies and how rating affect can change the affective experience, potentially negating ecological validity.