



Unpacking the Quantifying and Qualifying Potential of Semi-Open Job Satisfaction Questions through Computer-Aided Sentiment Analysis

Indy Wijngaards^{1,2} · Martijn Burger^{1,3} · Job van Exel²

Accepted: 10 March 2021 / Published online: 23 March 2021

© The Author(s) 2021

Abstract

Despite their suitability for mitigating survey biases and their potential for enhancing information richness, open and semi-open job satisfaction questions are rarely used in surveys. This is mostly due to the high costs associated with manual coding and difficulties that arise when validating text measures. Recently, advances in computer-aided text analysis have enabled researchers to rely less on manual coding to construct text measures. Yet, little is known about the validity of text measures generated by computer-aided text analysis software and only a handful of studies have attempted to demonstrate their added value. In light of this gap, drawing on a sample of 395 employees, we showed that the responses to a semi-open job satisfaction question can reliably and conveniently be converted into a text measure using two types of computer-aided sentiment analysis: SentimentR, and Linguistic Inquiry and Word Count (LIWC) 2015. Furthermore, the substantial convergence between the LIWC2015 and, in particular, SentimentR measure with a closed question measure of job satisfaction and logical associations with closed question measures of constructs that fall within and outside job satisfaction's nomological network, suggest that a semi-open question has adequate convergent and discriminant validity. Finally, we illustrated that the responses to our semi-open question can be used to fine-tune the computer-aided sentiment analysis dictionaries and unravel antecedents of job satisfaction.

Keywords Job satisfaction · Semi-open question · Validity · Sentiment analysis · Methodology · Measurement

✉ Indy Wijngaards
wijngaards@ese.eur.nl

Martijn Burger
mburger@ese.eur.nl

Job van Exel
vanexel@eshpm.eur.nl

Extended author information available on the last page of the article

1 Introduction

Job satisfaction, “a positive (or negative) evaluative judgment one makes about one’s job or job situation” (Weiss 2002, p. 175), continues to be closely monitored in corporate surveys (Macey and Schneider 2008) and has a long history of scientific study (Judge et al. 2017). There are good reasons for its popularity; job satisfaction has proven to be a robust correlate of subjective well-being (Bowling et al. 2010), health (Faragher et al. 2005) and job performance outcomes (Judge et al. 2001). Over the past few decades of scholarly and practitioner attention, a plethora of survey question instruments have been developed to cover conceptual nuances, e.g., affect-oriented vs. cognition-oriented scales (Kaplan et al. 2009; Organ and Near 1985) and job facet satisfaction vs. general job satisfaction (Scarpello and Campbell 1983; Spector 1997; Weiss 2002). Others have aimed at reducing response burden by shortening multiple-item scales (e.g., Russell et al. 2004) or developing single item measure of job satisfaction (e.g., Fisher et al. 2016; Gardner et al. 1998; Wanous et al. 1997).

The majority of survey instruments share one commonality: they are typically comprised of closed questions or items (e.g., “Overall, I am satisfied with my job.”, Fisher et al. 2016, p. 8) rather than open questions (e.g., “Please give us feedback or comments about your job.” Gilles et al. 2017, p. 4; “How do you think about your job as a whole?”, Wijngaards et al. 2019, p. 5) or semi-open questions (e.g., “What three to five adjectives come to mind when you think of your job as a whole?”, Wijngaards et al. 2019, p. 5). Closed questions have several advantages over open questions, as they typically pose less burden for respondents than open questions (Krosnick 1999; Vinten 1995; Zehner et al. 2016) and they prove to be more straight-forward to code and validate than open questions are (Maxwell and Delaney 1985; Tausczik and Pennebaker 2010).

The unpopularity of semi-open or open job satisfaction questions is unfortunate, because semi-open or open questions hold great potential as a complement to closed questions in surveys. A semi-open or open question can, for example, be used to better *quantify* job satisfaction, as weaknesses of individual methods are likely off-set by the use of multiple methods (Bryman 2006; Jick 1979; Turner et al. 2017). Measuring job satisfaction with both closed and open-ended response formats in a single questionnaire could help mitigate common method error, as respondents are forced into different forms of cognitive processing (Podsakoff et al. 2003). Semi-open or open questions can also be used to *qualify* job satisfaction and thereby obtain a more complete and deeper understanding of a construct (Fielding 2012; Jick 1979; Mauzeri 2016; Turner et al. 2017). Textual responses can be leveraged to contextualize responses to closed questions and obtain insights into the causes and sources of job (dis)satisfaction (Spector and Pindek 2016; Taber 1991). Moreover, they can illustrate and clarify the results from quantitative data analyses to nonexpert audiences (Borg and Zuell 2012; Zhang et al. 2019). Furthermore, the practical disadvantages of constructing textual job satisfaction measures are becoming increasingly obsolete, as computer-aided text analysis, which is a form of content analysis that facilitates the measurement of constructs by converting text into quantitative data based on word frequencies, make the creation of text measures more convenient than ever (McKenny et al. 2018; Short et al. 2010, 2018).

In this article, we aim to unpack the quantifying and qualifying potential of a semi-open job satisfaction question. We focus on a semi-open rather than a completely open job satisfaction question, because semi-open questions impose answering constraints on responses and therefore produce more structured, e.g., fewer meaningless words and less semantic nuance, than completely open questions (Glerum et al. 2014; Wijngaards et al. 2019). This makes text

measures based semi-open questions produced by computer-aided sentiment analysis methods probably more suitable for quantifying the level of job satisfaction than text measures based on completely open questions produced by computer-aided sentiment analysis methods (Wijngaards et al. 2019). We investigate the semi-open questions' quantifying potential by creating text measures using computer-aided sentiment analysis, the practice of automatically detecting opinions, sentiments, attitudes and emotions about certain objects in human-generated texts (Feldman 2013; Liu 2015), and validating this measure using well-established survey scales. We investigate the semi-open questions' qualifying potential by examining which sentiment ratings in sentiment-dictionaries are context-dependent and what unfavorable job characteristics are taken for granted if favorable job characteristics are present.

This study contributes to the literature by building on and extending existing methodological work on validation of textual job satisfaction measures. Previous studies on textual job satisfaction measures lack a systematic validation approach, neglecting content validity and using ad-hoc, single-item job satisfaction measures to test convergent validity (Borg and Zuell 2012; Poncheri et al. 2008; Wijngaards et al. 2019), as well as discriminant validity (Borg and Zuell 2012; Gilles et al. 2017; Poncheri et al. 2008). It is essential to systematically examine the validity of a text measure, as researchers with a preference for traditional survey measures are unlikely to accept text measures as fruitful complement to a closed question if there is no convincing evidence for their validity available. Therefore, we discuss the content validity of the semi-open job satisfaction question. Using correlational analyses and confirmatory factor analyses (CFAs), we also test its fit with a closed question job satisfaction scale. In addition, we assess the semi-open question' correlations with variables falling within and outside job satisfaction's nomological network. As computer-aided text analysis techniques produce measures with attenuated reliability (McKenny et al. 2018), we benchmark the text measures generated by computer-aided sentiment analysis techniques with a text measure with the least possible measurement error, a text measure based on respondents' own sentiment annotations (henceforth: benchmark measure).¹

The qualitative analysis on context-dependency of sentiment ratings help advance the field of computer-aided sentiment analysis, while the exploration on the weight of different job characteristics contributes to scientists and practitioners' understanding of job satisfaction and its causes.

1.1 Using Computer-Aided Sentiment Analysis to Create a Textual Job Satisfaction Measure

Much of the research on constructing text measures from responses to semi-open or open job satisfaction questions have made use of computer-aided sentiment analysis techniques. The techniques' popularity is not surprising, as it is much faster than manual sentiment annotation (Wijngaards et al. 2019) and an individual's choice of words is a plausible manifestation of thoughts and opinions (Pennebaker et al. 2003; Short et al. 2010). As mentioned earlier, computer-aided sentiment analysis is particularly suitable for constructing a job satisfaction measure from text, as job satisfaction classifies as a job attitude and comprises cognitive

¹ This measure is likely a more reliable measure of sentiment than measures produced by independent coders. Independent coders introduce unique measurement errors through individual differences in evaluation strategies (Mikhaylov et al. 2012) and in annotation experience and education (Snow et al. 2008) and, in the end, are mere forecasters of the sentiment rating a respondent would have decided upon.

appraisals, emotions, and beliefs (Weiss 2002). Sentiment analysis software typically classifies respondents that mainly use positive words in their written responses as satisfied individuals, and classifies respondents that mainly use a negative tone as dissatisfied respondents (Liu 2015; Poncheri et al. 2008).

To construct the text measure in this study, we use lexicon-based computer-aided sentiment analysis software, a type of sentiment analysis that annotates texts using dictionaries of words with pre-labelled sentiment orientation. Sentiment orientation concerns words' sentiment polarity (e.g., "good" vs. "bad") and sentiment strength (e.g., "good" vs. "great"). The lexicon-based sentiment analysis approach is characterized by two stages. In the first stage, software is used to pre-process raw textual data. Steps involved in pre-processing are removing stop words, such as "the", "from" and "as", correcting language mistakes, converting words to lowercase and removing punctuation and white spaces (Meyer et al. 2008; Pandey and Pandey 2019). Once this has been done, software can be employed to classify words into a sentiment classes, e.g., very negative (-2), negative (-1), neutral (0), positive (+1) and very positive (+2). Contemporary sentiment analysis software does not only consider an individual words' sentiment rating, but also considers words in the context that they are used in. For example, contemporary software considers negators, such as "not" and "never", as they reverse the semantic polarity and amplifiers, such as "very", and de-amplifiers, such as "reasonably", as they offer an indication of the sentiment strength. After this, the software automatically adds up (and weighs) all the individual ratings and computes an overall sentiment rating. The rating can, in turn, be used as a measure of job satisfaction.

To illustrate this two-stage procedure, we use a response from a semi-open question and a response to an open job satisfaction question, both obtained from the open-access data from Wijngaards et al. (2019). The sample answer to the semi-open question was a list of three adjectives: "Interesting", "Stressful" and "Helpful". The sample answer to the open question was a statement: "I am extremely proud of the work I do and think I do it very well, but I don't enjoy how hard and tiring it is and the people I work with are difficult to work with".

The pre-processing of the response to the semi-open question only involves converting all words into lowercase (i.e., "interesting", "stressful" and "helpful"). Then, the software searches in the pre-processed text for words that either contain non-zero sentiment ratings or sentiment strength and draws from its dictionary to assign sentiment ratings. As the semi-open question asks respondents to come up with individual adjectives, it is important to treat the individual words as separate, de-contextualized textual instances. In this example, all words carry a non-zero sentiment loading: "interesting" (+1) "stressful" (-1) and "helpful" (+1). The three sentiment scores can be summed into a single score (i.e., $1 + -1 + 1 = 1$) and the response would be classified as positive. As an illustration, SentimentR (Rinker 2019), a sentiment analysis software program that by default uses a dictionary of words with 20 sentiment classes (e.g., -0.25, 0, 0.75, scale ranges from -2.0 to 2.0) and considers negators, amplifiers and deamplifiers in its algorithm, assigns the following sentiment ratings to each word: "interesting" (+0.75), "stressful" (-0.50) and "interesting" (+0.75). The final sentiment rating would be 1.0.

The pre-processing and sentiment calculation of the example response to the open question is more complex. To pre-process the sentence, the software has to convert all words into lowercase, omit punctuation and remove stop words. The pre-processed text would then read "extremely proud do work think do very well but don't enjoy hard tiring people work difficult work". As the response is a sentence, it is important to consider the context in which words are used (e.g., valence shifters and amplifiers). In this text, the following terms do not have a non-

zero sentiment score or sentiment strength score: “extremely proud” (+2), “very well” (+2), “don’t enjoy” (-1), “hard” (-1), “tiring” (-1) and “difficult” (-1). Finally, the software solves the equation (i.e., $2 + 2 + -1 + -1 + -1 + -1 = 0$) and classifies the response as neutral. In accordance with this example, SentimentR would classify the response as slightly positive (0.3).

1.2 Quantifying Job Satisfaction

Now we have explained how a textual response to a semi-open or open question can be converted into a text measure, we move to a justification of the semi-open question’s suitability as a job satisfaction measure. We discuss the measure’s theoretical validity and provide a description of our empirical validation procedure.

Theoretical Validity We use the following semi-open question in our study: “What three to five adjectives come to mind when you think of your job as a whole?” (Wijngaards et al. 2019, p. 5). Being based on the fundamentals of the adjective generation technique, a psychological method initially used for personality assessment (Potkay and Allen 1973), the question was designed to tap into both to the affective and cognitive component of job satisfaction and thus measure the construct as a whole (Judge et al. 2012; Weiss 2002). The verb “think” likely spurs both affective and cognitive thoughts and is presumably more suitable than more specific verbs like “appraise”, “evaluate” and “feel”. The first two words would primarily elicit cognitive evaluations of the job, while the third word would likely tap more into the affective component of job satisfaction. The “as a whole” at the end of the question was included, as the semi-open question is purported to measure general job satisfaction. The clause likely triggers respondents into thinking more inclusively (Scarpello and Campbell 1983; Weiss 2002). We restricted the number of words to (i) reduce required effort for respondents and (ii) get an idea of the most salient thoughts.

Convergent Validity Turning to empirical validity, the text measure based on the responses to the semi-open job satisfaction question has to converge with an existing measure of job satisfaction (Edwards 2003; Hinkin 1998). However, we do not expect perfect convergence between the two types of job satisfaction measures, because closed and open questions have divergent epistemological foundations and introduce different sources of measurement error (Fielding 2012; Mauceri 2016; McKenny et al. 2018).

In line with this expectation, previous research on open comment boxes in surveys has demonstrated moderate correlations between sentiment ratings and overall job satisfaction. For example, in a study among military personnel, Poncheri et al. (2008) documented a correlation of .41 between comments’ affective tone and a closed question measure of general job satisfaction. Drawing upon data from a large corporate survey from an information technology organization, Borg and Zuell (2012) documented a correlation between affective comment tone and closed question job satisfaction measures of .38, on average. In a study by Wijngaards et al. (2019), similar correlations were found between the text measures based on a completely open question and a closed question measure of general job satisfaction: $r_{\text{average}} = .40$. The average correlation between the text measures based on the responses to the semi-open question and the closed question measure was significantly higher, $r_{\text{average}} = .56$. As we use a semi-open question in this study, we expect that the text measures correlate positively with a closed question measure of job

satisfaction. Even though neither of these studies provided factor analytic evidence for the convergent validity of the text measures, we expect that the text measures of job satisfaction fit well with closed question job satisfaction measures.

Discriminant Validity To have satisfactory discriminant validity, a text measure must fit within a constructs' nomological network, the abstract representation of constructs, their measures and the interrelationships among them (Cronbach and Meehl 1955). In our case, discriminant validity could be demonstrated by testing the bivariate correlations between the text measures of job satisfaction and theoretically related antecedents, correlates and outcomes of job satisfaction, and looking into its relationship with constructs that fall outside the job satisfaction's broader theoretical context (Edwards 2003; Shaffer et al. 2016). Drawing on the many nomological networks that have been developed in the many decades of job satisfaction research (e.g., Bowling and Hammond 2008; Brief 1998; Crede et al. 2007), we identify various antecedents, correlates, outcomes and unrelated constructs. Research suggests that skill variety, task autonomy and person-environment fit are pertinent examples of antecedents of job satisfaction. Theorists have argued that job variety, "the degree to which a job requires employees to perform a wide range of tasks on the job" (Morgeson and Humphrey 2006, p. 1324), and job autonomy, "the extent to which a job allows freedom, independence, and discretion to schedule work, make decisions, and choose the methods used to perform tasks" (Morgeson and Humphrey 2006, p. 1324), contribute to an employee's experienced meaningfulness and responsibility, respectively (Hackman and Oldham 1976). They reasoned that the sense of meaningfulness and responsibility have the potential to boost intrinsic motivation, which in turn positively correlates with job satisfaction (Fried and Ferris 1987; Hackman and Oldham 1976). Person-organization fit refers to the fit between employees and organisations that occurs when one offers what the other wants, they share similar important characteristics, or both (Kristof 1996). Therefore, person-organization fit is likely associated with job satisfaction (Kristof 1996). Life satisfaction, "a global assessment of a person's quality of life according to his own criteria" (Shin and Johnson 1978, p. 478), is among the most fundamental correlates of job satisfaction, because job satisfaction contributes to a person's overall satisfaction with life, and vice versa (Judge et al. 2012; Judge and Watanabe 1994; for empirical evidence, see Bowling et al. 2010). Organizational citizenship behaviour and turnover intention are two relevant performance indicators for organizations (G. Cohen et al. 2016; Koys 2001). Organizational citizenship behaviour concerns supportive gestures from employees that are valued by organizations, but are not linked directly to individual productivity or their contractual role expectations (Organ 1988). Theory suggests that satisfied employees exhibit organizational citizenship behaviour to reciprocate the favourable job conditions organizations offer (Organ 1988). Meta-analytical evidence supports this theoretical contention (Dalal 2005; LePine et al. 2002). Turnover intention, the intention to willingly change jobs or companies (Schyns et al. 2007), is a likely outcome from job dissatisfaction, as employees tend to avoid unpleasant work situations by displaying withdrawal behaviours. When job dissatisfaction persists, individuals tend to withdraw for good and leave (Hanisch and Hulin 1991). Indeed, previous research suggests that the relationship between job satisfaction and turnover intention is negative (Bowling and Hammond 2008; Tett and Meyer 1993). Personality traits are also related to job satisfaction (Judge et al. 2002), but empirical studies indicate that this is not the case for all traits (Bowling et al. 2018; Bui 2017; Harvey and Martinko 2009; Judge et al. 2002). Two examples are need for cognition, "the need to structure relevant situations in meaningful, integrated ways" (A. R. Cohen et al. 1955, p. 291), and

openness, “the breadth, depth, and permeability of consciousness, and in the recurrent need to enlarge and examine experience” (McCrae and Costa Jr 1997, p. 826). A plausible explanation for this nonsignificant correlation is that openness to experience and need for cognition are positively related to job satisfaction for some jobs (e.g., entrepreneurs and jobs where one can learn), while it may be negatively correlated to job satisfaction in other jobs (e.g., boring and uncreative jobs), rendering the overall correlation between the two personality traits and job satisfaction nonsignificant (Bui 2017).

Taking theoretical justifications and empirical evidence into consideration, we expect that text measures of job satisfaction will significantly correlate with measures of task variety, job autonomy, person-organization fit, life satisfaction, organizational citizenship behaviour and turnover intention, while we do not expect it to correlate significantly with measures of need for cognition and openness.

1.3 Qualifying Job Satisfaction

In this part of the study, we want to show that semi-open questions can be used to obtain insights that closed questions may not provide. As respondents themselves know best what the sentiment in their responses is, we use their ratings for our analyses on qualifying job satisfaction.

Context-Dependency of Sentiment Ratings Can all words be assigned a single sentiment rating or are sentiment ratings generally context-dependent? Previous research indicates that the answer to this question is probably somewhere in the middle (McKenny et al. 2018; Short et al. 2010). Some words are likely to have a less ambiguous meaning across contexts, such as “love” and “enjoy”. Other words evoke divergent sentiments depending on the sentiment of the words that they are used along with. As an illustration, the word “challenging” might evoke a positive meaning when it is used with adjectives, such as “engaging” and/or “satisfying”, rather than when it is used with adjectives, such as “stressful” and/or “overwhelming”. Following this reasoning, we expect that words will vary substantially depending on the extent to which they can reliably be assigned a single sentiment score.

Balancing Job Characteristics Even if our semi-open question does not produce detailed information about the cognitive appraisal of job facets, it may offer some ideas about the job characteristics that influence it. One avenue of research we found particularly interesting is: examining the unfavourable job characteristics that respondents are willing to accept in light of unfavourable job characteristics. For instance, respondents may accept the boring nature of their job if it gives them job security.

2 Methods

2.1 Participants and Data Collection

We collected our data through Prolific, a virtual crowdsourcing platform where people get compensated to complete tasks. Prolific workers tend to provide reliable data and turned out to

be more honest and more diverse in terms of geographical location and ethnicity than respondents from other crowdsourcing platforms, such as Amazon's Mechanical Turk (Peer et al. 2017). Using Prolific's filtering system, we selected people from the United States of America, who worked at least 20 h a week and had an approval rate of 80% or higher. We followed recommendations from Prolific for the height of the respondent compensation and paid respondents an amount of \$1.31 for 10 min of work. The data collection procedure resulted in 395 responses. Most respondents were male (56.0%). The large majority of respondents had at least some college experience (94.2%). Most respondents had a permanent employment contract (76.5%). The average age was 35.1 ($SD = 10.2$). The average number of work hours and number of years of experience within their organisation were 37.7 ($SD = 7.8$) and 5.2 ($SD = 5.0$), respectively. Of all the respondents, 32.6% had a managerial position in an organisation.

The research context classifies as low-stake environment, which may have introduced the issue of careless responding in our data (Curran 2016; Fleischer et al. 2015). To address this problem, we flagged careless respondents based on three criteria: average response time per item, item consistency on a semantic antonym and Mahalanobis distance (Curran 2016; Meade and Craig 2012). We adopted the cut scores set for 95% specificity from Goldammer et al. (2020) for average response time per item and the Mahalanobis distance. We considered responses to be inconsistent if the absolute difference between the two reverse-item scored items was equal or larger than 2. The criteria pointed towards three different samples of careless respondents. We constructed three reduced samples based on the omission of the three samples of careless respondents.

The first sample contained individuals who excluded respondents who took less than 5.56 s to complete a survey item, on average ($N = 232$). The second sample excluded respondents where the absolute difference in scores to the following two items measuring openness (i.e., "I tend to vote for liberal political candidates." and "I tend to vote for conservative political candidates."), each rated on a 5-point Likert scale, was higher or equal to 2 ($N = 290$). We deemed this item suitable for careless responding analyses, as the two items are antonyms and the bivariate correlation between the items in the whole sample was high ($r = -.79$). The third sample excluded respondents with a Mahalanobis distance higher than 94.81, computed over 52 items ($N = 381$).

2.2 Measures

Internal consistency of the instruments was tested using McDonald's (1999) omega (ω) (Dunn et al. 2014). The closed question survey scales' internal consistency statistics were considered sufficient, as all the values were equal to or above .8 (Nunnally and Bernstein 1994), see Table 3. We constructed measures by computing unweighted averages of all items, unless specified differently.

Job Satisfaction Job satisfaction was measured using the 3-item Michigan Organisational Assessment Questionnaire Job Satisfaction Subscale (MOAQ-JSS, Cammann et al. 1979). The MOAQ-JSS has been validated (Bowling and Hammond 2008). The answer categories from the MOAQ-JSS ranged a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). An example item is "All in all, I am satisfied with my job." In addition, we measured job satisfaction using a semi-open question, which reads "Which three to five adjectives come to mind when you think of your job as a whole?"

Task Variety Task variety was measured using a 4-item scale from the Work Design Questionnaire (Morgeson and Humphrey 2006), with response categories ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). An example item is “The job involves a great deal of task variety.”

Job Autonomy Job autonomy was measured using a 3-item scale from the Work Design Questionnaire with response categories ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). An example item is “The job allows me to make a lot of decisions on my own.”

Person-Organization Fit Person-organization fit was measured using a 3-item scale developed by Cable and Judge (1996), with response categories ranging from 1 (*not at all*) to 5 (*completely*). An example item is “My values match those of current employees in my organisation.”

Life Satisfaction Life satisfaction was measured using the 5-item Satisfaction With Life Scale (SWLS, Diener et al. 1985), with response categories ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The item scores were summed up into one aggregate measure. An example item is “In most ways my life is close to ideal.”

Organizational Citizenship Behaviour Organizational citizenship behaviour was measured using the 10-item short version of the Organizational Citizenship Behaviour Checklist (Spector et al. 2010). The scale had response categories ranging from 1 (*never*) to 5 (*every day*). An example item is “I worked weekends or other days off to complete a project or task”.

Turnover Intention Turnover intention was measured using the 3-item turnover intention subscale in the MOAQ (Cammann et al. 1979), where the questions had to be answered using categories ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is “How likely is it that you will actively look for a new job in the next year?”

Need for Cognition Need for cognition was measured using the 10-item Need For Cognition Scale (Cacioppo and Petty 1982). Answer categories ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is: “I like to solve complex problems”.

Openness Openness was measured using a 10-item scale from the NEO Personality Inventory (Costa and McCrae 1985), and had response categories ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is “I have a vivid imagination”.

2.3 Data Processing and Statistical Analysis

Data pre-processing and validity testing was done in the software program R (R Core Team 2018). All scripts and the data are made available as supplementary material.

Analyses for Text Measure Construction In our study, we used SentimentR (Rinker 2019) and Linguistic Inquiry and Word Count (LIWC) 2015² (Pennebaker et al. 2015) to compute

² The Emotional Tone summary variable from LIWC 2015 was used (Cohn et al. 2004).

the sentiment ratings of the semi-open questions. We selected these software packages, as Wijngaards et al. (2019) found that sentiment scores from these software packages most closely resembled the sentiment scores produced by independent human coders ($r_{\text{SentimentR}} = .772$ and $r_{\text{LIWC2015}} = .775$).

SentimentR is a freely available sentiment analysis package written in R (R Core Team 2018). It uses, by default, the English sentiment dictionary of Jockers (2017), which contains 10,739 annotated words. This software package has been successfully deployed in several studies outside the survey methodology domain (e.g., Ikoro et al. 2018; Naldi 2019; Rinker 2019; Weissman et al. 2019). As SentimentR software incorporates the context of words in its sentiment ratings and the semi-open questions ask respondents about three to five disconnected adjectives, we constructed the SentimentR measure in three steps: (1) classifying the individual adjectives in terms of sentiment, (2) recoding these individual sentiment scores onto a scale from 1 (*very negative*), 2 (*negative*), 3 (*neutral*), 4 (*positive*) to 5 (*very positive*), and (3) averaging the scores into a final sentiment score.

The LIWC software is one of the most widely used computer-aided text analysis techniques in the organizational sciences (Short et al. 2018), and has been validated across a large number of studies (Pennebaker et al. 2015). LIWC2015 draws from an English dictionary of 620 positive words and 744 negative words. LIWC2015 does not explicitly contextualize the valence loadings of individual words. As we wanted to maximize the comparability of the LIWC2015 measure and SentimentR measure, we adopted the aforementioned three-step sentiment calculation procedure to create the LIWC2015 measure.

We tested the reliability of the SentimentR measure and the LIWC2015 measure in the current data by examining their convergence with text measures produced by humans (i.e., parallel-forms reliability, McKenny et al. 2018). We used the benchmark measure for this procedure. At the end of the survey, we asked respondents to separately annotate all adjectives they previously used on a scale from 1 (*very negative*), 2 (*negative*), 3 (*neutral*), 4 (*positive*) to 5 (*very positive*). The question read: "How would you rate your previous answer in terms of sentiment/emotion?"

Analyses for Empirical Validation Correlational analyses and single-factor CFAs were used to test the convergent validity of the text measures. The fit of a single-item CFA was considered adequate if the χ^2 -test is significant, the comparative fit index (CFI) value was above .95 and the standardized root mean square residuals (SRMR) and root mean square error of approximation (RMSEA) values were less than .06 (Brown 2014; Hu and Bentler 1999; Kline 2015). Standardized factor loadings of the text measures should exceed .6 to be satisfactory (Matsunaga 2010). Correlational analyses were used to test discriminant validity.

Analyses for Contextualization of Job Satisfaction For this part of the study, we only used the responses to the semi-open question and the benchmark measure. For the analyses concerning the context-dependency of sentiment ratings, we computed the frequency of individual words and calculated the mean and *SD* of the sentiment ratings associated to each individual word. This dictionary allowed us to discover which words had divergent sentiment connotations depending on the sentiment of the other words provided by the respondent. We produced three sub-dictionaries by splitting the complete dictionary based on the average sentiment rating by respondents: negative (mean < .2.5), neutral ($2.5 \leq \text{mean} \leq 3.5$) and positive (mean > 3.5).

For the analyses on the antecedents of job satisfaction, we concentrated on individual respondents' word use and associated sentiment ratings. In specific, we averaged the individual sentiment ratings for each respondent and calculated the *SD* for each mean score. To illustrate this, let us consider two hypothetical respondents that responded with two adjectives. Respondent A used the adjectives "Bored" and "Safe" and, at the end of the survey, classified them as 1 (*very negative*), and 4 (*positive*), respectively. Respondent B used "Bored" and "Unhappy" and classified them as 1 (*very negative*) and 2 (*negative*), respectively. The mean sentiment (and *SD*) for rating for Respondent A and B would be 2.5 (2.12) and 1.5 (0.71), respectively. A high *SD*, in this context, thus points towards the use of both positive and negative words in a single answer. In our study, we reported findings based on a threshold for *SD* of 1.5, but checked the sensitivity of our findings for a *SD* threshold of 1.0 and 2.0. We manually qualified the responses to the semi-open question to map the job conditions that respondents with a high *SD* generally had to deal with and discussed its relationship with job satisfaction.

3 Results

3.1 Construction of Textual Job Satisfaction Measures

The semi-open question was generally understood well, as 61.6% of all words provided by respondents were adjectives. Nouns and verbs were the second and third most popular word categories, which is 22.0% and 12.8% of the total words, respectively. We did not omit any textual data, as many verbs and nouns have affective loadings too. The mean number of words per respondent was 4.5 with a *SD* of 1.2 words. The median was 5 words. Table 1 presents the fifteen most common words used by respondents who are dissatisfied with their job (MOAQ-JSS ≤ 3) as well as the respondents who are satisfied with their jobs (MOAQ-JSS ≥ 5).

To test whether SentimentR and LIWC2015 are appropriate software packages for sentiment analysis in our context, we compared the text measures based on the two algorithms and

Table 1 Most frequently used words amongst satisfied and dissatisfied respondents

<i>Dissatisfied with their job</i>		<i>Satisfied with their job</i>	
Word	<i>N</i>	Word	<i>N</i>
Boring	28	Fun	57
Stressful	13	Interesting	46
Tedious	9	Challenging	41
Annoying	8	Rewarding	28
Frustrating	6	Stressful	27
Demanding	5	Easy	21
Dull	5	Flexible	20
Exhausting	5	Fulfilling	17
Pointless	4	Helpful	17
Repetitive	4	Good	15
Underpaid	4	Busy	15
Slow	4	Important	14

Note. *N* = Sample size

the benchmark measure. Parallel-forms reliability for both the SentimentR measure ($r = .80$) and the LIWC2015 measure turned out to be satisfactory ($r = .62$). As apparent in the .18 difference in correlation coefficient and as visualized in Fig. 1, the SentimentR measure more closely resembles the density plot from the benchmark measure than the LIWC2015 measure. Table 2 shows twenty examples of responses with corresponding sentiment scores.

3.2 Empirical Validation

Convergent and Discriminant Validity The results of our correlation analyses, shown in Table 3, indicated convergence between the MOAQ-JSS and the text measures in varying degrees: the SentimentR measure ($r = .70, p < .01$), the LIWC2015 measure ($r = .55, p < .01$) and the benchmark measure ($r = .80, p < .01$). The CFAs further corroborated convergent validity, because the CFA model that included the SentimentR measure ($\chi^2 [2] = 3.697, p = ns, CFI = .999, SRMR = .010, RMSEA = .046$), the LIWC2015 measure ($\chi^2 [2] = 0.535, p = ns, CFI = 1.000, SRMR = .003, RMSEA = .000$) and the benchmark measure ($\chi^2 [2] = 0.446, p = ns, CFI = 1.000, SRMR = .002, RMSEA = .000$) fitted the data very well. The standardized factor loadings of the SentimentR measure ($\lambda = .72$) and the benchmark measure ($\lambda = .83$) exceeded .6, while the standardized factor loading of the LIWC2015 measure did not ($\lambda = .57$). Notably, the factor loadings of the SentimentR measure and, in particular, the benchmark measure were generally in line with the loadings of the MOAQ-JSS items (CFA model including the SentimentR measure: $\lambda_{\text{MOAQ-JSS1}} = .93, \lambda_{\text{MOAQ-JSS2}} = .85, \lambda_{\text{MOAQ-JSS3}} = .92$; CFA model including the benchmark measure: $\lambda_{\text{MOAQ-JSS1}} = .94, \lambda_{\text{MOAQ-JSS2}} = .84, \lambda_{\text{MOAQ-JSS3}} = .92$). The factor loading of the LIWC2015 measure diverged quite substantially ($\lambda_{\text{MOAQ-JSS1}} = .94, \lambda_{\text{MOAQ-JSS2}} = .84, \lambda_{\text{MOAQ-JSS3}} = .92$). Taken together, the convergent validity analyses showed that the SentimentR measure has better properties than the LIWC2015 measure, and that neither measure performed as well as the benchmark measure.

With respect to discriminant validity, the results indicated that the relationships between the text measures and their hypothesized antecedents (i.e., skill variety, autonomy and person-organization fit), correlate (i.e., life satisfaction), and outcomes (i.e., turnover intention and organizational citizenship behaviour) were significant and in the expected direction (e.g., positive association with life satisfaction and negative association with turnover intention).

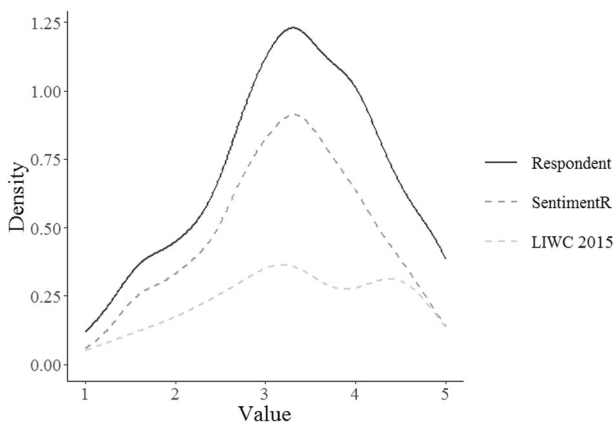


Fig. 1 Density plot of the benchmark measure, SentimentR measure and LIWC2015 measure. *Note.* LIWC = Linguistic Inquiry and Word Count

Table 2 Examples of responses and coding

Response to semi-open question	SentimentR measure ^a	LJWC2015 measure ^a	Benchmark measure ^a	MOAQ-ISS ^b
Exciting, fun, cool, interesting, hard	3.60	4.80	4.00	3.33
Positive, caring, interested, fun, engaging	3.60	5.00	4.80	6.33
Boring, tedious, important, relaxed	2.75	3.00	3.25	5.67
Temporary, part, time, underpaid, remote, flexible	3.00	4.00	3.20	5.67
Interesting, frustrating, focused	3.33	1.80	3.00	6.00
Cold, stringent, red	2.33	2.20	2.67	6.00
Good, decent, well paying, nice, tiring	3.40	4.80	3.80	6.00
Rewarding, interesting, enjoyable, fun, pleasurable	4.20	5.00	5.00	7.00
Beautiful, nice, wonderful, predictable, awesome	3.80	4.80	5.00	7.00
Fun, stressful, tired, happy, strong	3.20	3.80	3.80	6.00
Uninteresting, fast, paced, pointless, obligatory	2.75	2.80	2.50	2.67
Necessity, relaxed, non flexible, boring, annoying	2.80	3.40	2.60	4.00
Hard, flexible, demanding, creative, frustrating	2.80	4.00	3.20	3.00
Creative, hands on, fast, paced, positive, interesting	3.80	4.20	4.20	6.00
Fulfilling, challenging, hectic, rewarding	2.75	3.80	3.50	5.33
Rewarding, flexible, challenging	3.00	3.40	4.33	5.33
Personal, educational, artistic, interesting	4.00	3.40	3.75	6.00
Boring, tedious, dull	1.67	1.00	1.00	2.67
Slow, tedious, surprising, relaxed, calm	3.00	3.80	2.60	3.67
Customer, service, fast pace, stressful, food, serving	3.20	2.60	4.60	7.00

Notes. LJWC = Linguistic Inquiry and Word Count; MOAQ-ISS = Michigan Organisational Assessment Questionnaire - Job Satisfaction Subscale; ^a Scale ranges from 1 (very negative) to 5 (very positive); ^b Scale ranges from 1 (strongly disagree) to 7 (strongly agree)

Table 3 Means, standard deviations, internal consistency statistics and bivariate correlations ($N = 395$)

Variable	Correlations													
	<i>M</i>	<i>SD</i>	ω	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Job satisfaction (SentimentR measure)	3.14	0.73	.54											
2. Job satisfaction (LIWC2015 measure)	3.31	1.01	.66	.60										
3. Job satisfaction (benchmark measure)	3.51	1.07	.84	.80	.62									
4. Job satisfaction (MOAQ-JSS)	5.13	1.64	.93	.70	.55	.80								
5. Task variety	3.79	1.03	.93	.19	.18	.29	.35							
6. Job autonomy	3.80	0.99	.91	.31	.28	.48	.53	.50						
7. Person-organization fit	3.34	1.00	.93	.53	.46	.65	.68	.37	.52					
8. Life satisfaction	21.87	7.72	.92	.39	.34	.47	.50	.21	.33	.42				
9. Turnover intention	2.77	1.34	.92	-.50	-.39	-.60	-.73	-.28	-.39	-.53	-.41			
10. Organizational citizenship behaviour	2.79	0.78	.85	.19	.15	.29	.25	.37	.27	.32	.16	-.15		
11. Need for cognition	3.85	0.57	.82	.02	.10	.10	.12	.11	.11	.09	.14	-.07	.19	
12. Openness	3.93	0.62	.80	-.05	.06	-.03	.02	.06	.01	.06	-.07	.10	.08	.54

Note. All correlations significant at the level of $p < .05$ (2-tailed), except for the ones in italic; LIWC = Linguistic Inquiry and Word Count; MOAQ-JSS = Michigan Organisational Assessment Questionnaire - Job Satisfaction Subscale; *M* = mean; *SD* = standard deviation, ω = McDonald's omega

The data also suggested that the SentimentR measure, the LIWC2015 measure and the benchmark measure only marginally correlated with need for cognition and openness measure. All convergent correlations were higher than the average discriminant correlations.

As demonstrated in Appendix Tables 5 and 6, the results remained robust when testing our hypotheses on survey data from respondents who have not been flagged as careless.

3.3 Contextualizing Job Satisfaction

Context-Dependency of Sentiment Ratings The descriptive statistics in Table 4 show that words vary in the extent to which their sentiment rating is context dependent. Eight words, “overwhelming”, “uncertain”, “underpaid”, “complex”, “academic”, “official”, “educational”, and “rewarding”, had a *SD* of 0, and thus an unequivocal meaning. Put differently, we can be quite certain that these words can have a single sentiment rating assigned to them: positive, neutral or negative. Other words had much higher *SD*s, such as “demanding”, “exhausting”, “repetitive”, “different”, “variable” and “easy”, implying that the meaning, and thus the sentiment rating of these words depend on the words they are used in conjunction with.

To illustrate this point, let us consider the word “easy”. With a mean score of 3.72, it was generally rated as positive. When we focused on individual responses and considered the *SD* of 0.96, we noticed that the word was rated as negative when combined with words such as “boring”, “repetitive”, “unchallenging”, “tedious” and “monotonous”. When used in conjunction with words such as “stress-free”, “easy-going”, “fun”, “safe”, “relaxing” and “slow”, it was rated more positively.

Balancing Job Characteristics For this part of the analysis, we were interested in respondents who provided words with varying sentiment meanings, as demonstrated in a *SD* of 1.5 to the

Table 4 Context-dependent sentiment ratings

Negative ($N_{\text{words}}=36$)				Neutral ($N_{\text{words}}=13$)				Positive ($N_{\text{words}}=58$)			
Word	<i>M</i>	<i>SD</i>	N_{obs}	Word	<i>M</i>	<i>SD</i>	N_{obs}	Word	<i>M</i>	<i>SD</i>	N_{obs}
Overwhelming	1.00	0.00	4	Complex	3.00	0.00	3	Educational	5	0.00	5
Uncertain	2.00	0.00	3	Academic	3.00	0.00	3	Rewarding	6	0.00	4
Underpaid	1.30	0.48	10	Official	3.00	0.00	3	Productive	5.8	0.45	5
Frustrating	1.56	0.51	19	Fine	3.25	0.50	4	Freedom	4.8	0.45	5
Long	2.33	0.51	6	Detailed	3.33	0.57	3	Important	4.8	0.45	5
Annoying	1.50	0.52	12	Difficult	2.83	0.81	6	Stimulating	1.2	0.45	5
...				Intense	3.00	0.89	4	...			
Repetitive	2.14	0.66	14	Routine	2.60	0.89	5	Necessary	3.60	0.89	5
Exhausting	1.70	0.67	10	Busy	3.18	0.95	17	Calm	4.14	0.89	7
Demanding	2.20	0.92	10	Hard	2.60	0.67	10	Great	4.25	0.96	4
Temporary	2.00	1.00	3	Work	3.33	1.154	3	Different	3.75	0.96	4
Cheap	2.00	1.00	3	Unpredictable	2.66	1.154	3	Variable	3.75	0.96	4
Stress	1.67	1.15	3	Slow	2.50	1.167	12	Easy	3.72	0.96	29

Note. *M* = mean; *SD* = standard deviation; N_{words} = total number of words that fitted the filtering criteria; N_{obs} = total number of observations per word, ... = words that do fall in the top six of words with the most varying sentiment ratings and bottom six of words with the least varying sentiment rating, i.e., 23 in the “Negative”-category and 45 in the “Positive”-category, please contact the first author for the complete table

average sentiment rating.³ Two examples of responses with high *SDs* were “Boring, Easy, Slow, Secure” and “Flexible, Challenging, Unpredictable, Stressful”. The respondent that provided the first example indicated that the first and third word are negative, the second word is neutral, and the fourth word is positive. The respondent that provided the second example indicated that the first two words are positive and the last two are negative. The mean (and *SD*) sentiment ratings for the two responses were 2.75 (1.48) and 3.00 (1.58), respectively.

Looking at word frequencies, we noticed that certain words were used particularly often. Organized from the most to the least frequently used (count between brackets), respondents used the following positive words: “interesting” (15), “fun” (15), “challenging” (14), “rewarding” (10), “flexible” (9), “easy” (9), “creative” (7), “engaging” (5), “social” (5), “fast” (5), “helpful” (5), “important” (4), “fulfilling” (4), “exciting” (4), “technical” (4) and “satisfying” (4). The following negative words were most often used: “stressful” (23), “boring” (13), “frustrating” (10), “tiring” (9), “underpaid” (6), “exhausting” (5), “demanding” (5), “slow” (4) and “repetitive” (4).

A qualitative analysis of these instances suggests that the common factor behind a substantial part of the most frequently occurring positive words is intrinsic motivation (Hackman and Oldham 1976). People either enjoy doing their work tasks (e.g., “fun”, “interesting”, “challenging” and “engaging”) or believe that their work is important (e.g., “rewarding”, “fulfilling”, “important” and “helpful”).

The negative words can also be categorized into higher-order categories. Except for the word “underpaid”, all words have either a connotation with a job that is too demanding or a job that is not demanding enough. Overall, these findings suggest that a segment of respondents deal with both favourable and unfavourable job conditions at the same time. For example, respondents seem to take boredom or stress for granted when their tasks are sufficiently enjoyable or important. To test whether this combination of favourable and unfavourable job conditions in the responses was also manifested in less extreme (and thus more neutral) scores on the closed job satisfaction measure, we regressed the *SD*-variable against the MOAQ-JSS. As shown in Fig. 2, we found an inverted U-curve. This suggests that respondents with a relatively high *SD* generally had moderate levels of job satisfaction. Respondents with relatively low *SDs* tended to respond more extremely to the job satisfaction question.

4 Discussion

Throughout this study, we investigated the quantifying and qualifying potential of a semi-open job satisfaction question. We showed that computer-aided sentiment analysis is a time-saving method to produce text measures. In less than five seconds, LIWC2015 and particularly SentimentR produced text measures that converged strongly with our benchmark measure ($r_{\text{SentimentR}} = .80$, $r_{\text{LIWC2015}} = .62$). Furthermore, the text measures showed promise as quantitative measures of job satisfaction. The text measures correlated strongly with a closed question measure of job satisfaction ($r_{\text{SentimentR}} = .70$ and $r_{\text{LIWC2015}} = .55$) and CFA models that included the text measure showed adequate fit. Concerning discriminant validity, we found that the text measures had logical associations with closed question measures of constructs that fall within and outside job satisfaction’s nomological network. Finally, we demonstrated that the

³ A sensitivity analysis for other *SD* cut-off values (i.e., 1.0 and 2.0) did not alter the results.

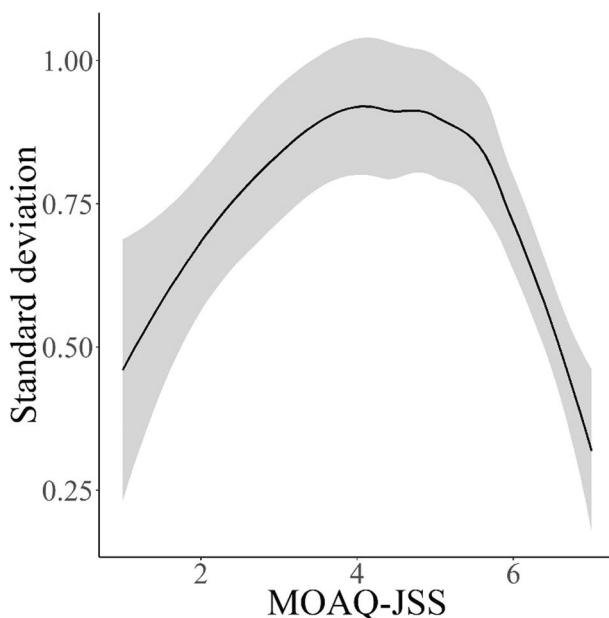


Fig. 2 The relationship between the MOAQ-JSS and *SD* in sentiment rating and pointwise 95% confidence interval on the fitted values. *Note.* MOAQ-JSS = Michigan Organisational Assessment Questionnaire - Job Satisfaction Subscale

responses to a semi-open job satisfaction question can act as a means to fine-tune sentiment analysis dictionaries and unravel antecedents of job satisfaction. Taken together, we conclude that semi-open questions have the potential to quantify and qualify job satisfaction and that computer-aided sentiment analysis is a valuable tool to help researchers to unpack this potential. The theoretical and practical contributions of our study, its limitations and future research directions are discussed below.

4.1 Theoretical Implications

Our study has several theoretical implications. First, we add to the field by illustrating that computer-aided sentiment analysis is not an absolute panacea, as the psychometric qualities of the SentimentR and LIWC2015 measure were inferior to those of the benchmark measure. For instance, the internal consistency score of the benchmark measure ($\omega = .84$) was much higher than the internal consistency scores of LIWC2015 measure ($\omega = .66$) and the SentimentR measure ($\omega = .54$). This limited inconsistency between adjectives, as suggested by the respondents, suggests that the LIWC2015 and SentimentR measure introduce measurement error. This measurement error is likely because the computer-aided sentiment analyses techniques do not explicitly consider the context-dependency of words. The correlation analyses and CFAs showed that the benchmark measure converged more strongly with the MOAQ-JSS than the text measures generated by computer-aided sentiment analysis. These attenuated correlations may be driven by the substantial measurement error in the text measures. Thus, even though the results generally support the appropriateness of text measures produced by computer-aided sentiment analysis as a job satisfaction measure, our findings also suggest that its reliability still has considerable room for improvement.

Secondly, to the best of our knowledge, we are among the first to test systematically the content, convergent, and discriminant validity of text measures of job satisfaction. Most previous research investigated the convergence between text measures and the responses to ad-hoc job satisfaction measures (e.g., Borg and Zuell 2012; Poncheri et al. 2008; Taber 1991; Wijngaards et al. 2019) or did not consider convergent validity at all (e.g., Jung and Suh 2019; Moniz and Jong 2014; Young and Gavade 2018). In our study, we adopted a traditional instrument validation approach. We tested the text measures validity with well-established multi-item survey instruments and employed techniques that control for same-source variance. Based on our findings, we tentatively argue that semi-open questions can be used to measure the level of job satisfaction.

Finally, we examined the qualitative value of semi-open questions over closed question measures. We found that the responses to our semi-open question could be used to fine-tune the reliability of a computer-aided sentiment analysis dictionary. Analyses showed that words differ in the degree to which they can be labelled with a single sentiment rating. More specifically, certain words can have both positive and negative connotations depending on the words they are used in conjunction with. For instance, words such as “stressful” and “busy” often do not represent stressors that are harmful for well-being if they are combined with more positive words such as “enthusiastic” and “fulfilling”. By contrast, words such as “underpaid” and “overwhelming” always have negative connotations. This finding supports the proposition that stressful job demands such as job complexity and work pressure do not necessarily lead to reduced subjective well-being (Bakker and Demerouti 2017; Van den Broeck et al. 2010). These results highlight the importance of transforming context-free dictionaries into domain-specific dictionaries to guarantee optimal reliability of text measures (e.g., treating “challenging” as slightly positive word, McKenny et al. 2018; Short et al. 2010). Finally, the semi-open question allowed us to discover that respondents who use both positive and negative words are often striking a balance between experiencing boredom or stress, on the one hand, and intrinsic motivation, on the other. This illustrates the intricate ways in which job characteristics contribute to employee well-being, for example, some occupations might be deemed highly meaningful, but at the same time have unfavourable job characteristics (Allan et al. 2018), e.g., nursing (Zangaro and Soeken 2007).

4.2 Practical Implications

In today’s competitive and fast-paced economy, organizations are compelled to maximize employee well-being (Guest 2017). To design high-quality well-being interventions, organizations must first obtain adequate insights about their employees’ work-related well-being and its drivers (Macey and Schneider 2008). To obtain this information, human resource practitioners typically request employees to complete surveys about their work experience and job attitudes (Gerrad and Hyland 2020). In the design of such surveys, they typically face a challenging trade-off between information richness and the minimization of respondent burden (Fisher et al. 2016; Fuchs and Diamantopoulos 2009). For example, practitioners may be reluctant to administer surveys containing multiple, lengthy closed question scales with good measurement quality, because they are associated with considerable opportunity costs and provide no opportunity for contextualization (Krosnick 1999). On the other hand, practitioners may be hesitant to use open questions that may provide valuable context to a survey, because converting raw textual data into reliable text measures is a challenging task and may feel daunting to many. We argue that the semi-open job satisfaction question has the promise to

quantify and qualify job satisfaction, and that it therefore functions as a valuable addition to an employee survey. Practitioners can leverage existing code (e.g., our code included in the supplementary materials) and well-established sentiment analysis software to construct text measures, which can be used both as a valid quantitative measure of job satisfaction, as well as input for the qualification of other quantitative measures. As an illustration of the latter, responses can be used to identify the most pertinent antecedents of job satisfaction in a particular group of employees (e.g., Table 1). Additionally, responses to the semi-open question may help organizational researchers to communicate research findings to individuals with limited experience with quantitative analysis (Borg and Zuell 2012). An anecdote obtained from a semi-open question can, for example, provide an intuitive illustration of improvement areas and make abstract research findings feel more relatable (Glaser et al. 2009; Rynes 2012).

4.3 Limitations and Future Research

This study has several limitations that should be noted. For instance, even though the correspondence of the text measures were substantial, the validation evidence generally corresponded with the evidence of the benchmark measure, as reliability of text measures produced by computer-aided sentiment analysis was not perfect. Considering the importance of the accuracy of a computer-aided text analysis technique in the validity of a text measure (McKenny et al. 2018; Short et al. 2010), we recommend researchers to develop more reliable techniques to construct text measures and come closer to the measurement-error-free sentiment measure. One approach would be to tailor generic dictionaries like the one of Jockers (2017) or the LIWC (Pennebaker et al. 2015), so that text measures better align with the construct they intend or context of study (Taboada et al. 2011; for guidelines, see Short et al. 2010). The results of our study indicate that certain words will inherently differ in their sentiment rating depending on their context. Therefore, it would therefore be valuable to move beyond lexicon-based sentiment analysis and to employ machine learning approaches to sentiment analysis (Zehner et al. 2016). Such approaches leverage pre-labeled texts to train algorithms, with the purpose of predicting (i.e. classifying) unlabeled textual instances, and are particularly useful for recognizing context-specific nuances (Taboada et al. 2011; for guidelines, see Kobayashi et al. 2017). Our data, which is included as supplementary material, can be used as training data for such endeavours. Alternatively, researchers could use publicly available self-rated job reviews from online platforms like Glassdoor (Jung and Suh 2019; Moniz and Jong 2014) or manually annotated texts as training data (Sheehan 2018).

We did not leverage the full potential of the textual data, as we only constructed a measure of job satisfaction from the textual responses. We expect that the information derived from texts could also be used to measure other constructs, such as work engagement, job affect or emotional exhaustion. For example, theory-driven or data-driven dictionaries of particular job affect dimensions can be created (Short et al. 2010). A theory-driven lexicon could be produced by reviewing items taken from the Job-related Affective Well-being Scale (Van Katwyk et al. 2000) and consulting a thesaurus to create emotion-specific dictionaries. The Job-related Affective Well-being Scale's boredom dimension could, for instance, be measured using a dictionary containing words such as "bored", "monotonous", "tedious", "pointless", "dull" and "dreary". Using a more qualitative and data-driven approach, a dictionary could be generated by looking at the most frequently used words, and manually assigning them to word categories or themes. Software programs such as ATLAS.ti, NVivo and CAT Scanner can be used to help researchers annotate texts and create custom dictionaries (Short et al. 2018).

Furthermore, we cannot generalize our conclusions to open questions in general. It is plausible that different kinds of open questions provide different kinds of insights. Our study showed that a semi-open question seems useful for measuring the level of job satisfaction, arguably more suitable than a completely open question (Wijngaards et al. 2019). We expect that the responses to our semi-open question are more suitable for quantifying job satisfaction than responses to an open question, as they are more straightforward to process and analyse. However, the lack of complexity in responses to semi-open questions concurrently functions as its most important limitation compared to responses to entirely open questions: they do not allow respondents to fully contextualize their responses. We, therefore, encourage researchers to investigate the qualifying and quantifying potential of completely open-ended questions and examine the reliability of computer-aided sentiment analysis techniques.

While entirely open questions may be more useful for qualitative research, we expect that semi-open questions could have additional potential for helping researchers quantify and qualify well-being. For example, our semi-open job satisfaction question could easily be reframed into a semi-open life satisfaction question (change ‘job as a whole’ to ‘life as a whole’). Questions, such as “What five aspects of your job contribute the most to your job satisfaction and how do they rank?” and “What words do you associate with work engagement?” could help organizational researchers unravel the constituents of multi-dimensional constructs such as work engagement (Briner 2014; Purcell 2014) and job satisfaction (Hsieh 2012; Mastekaasa 1984). A question like “In a best-case scenario, what job function would have in this organization in five years?” could aid practitioners to map the preferences of employees and design policies aimed at improving well-being.

Our validation procedure is subject to certain limitations. Even though some initial evidence suggests that text measures of job satisfaction have sufficient test-retest reliability (Wijngaards et al. 2019), we did not address it in this study. In addition, our reliance on self-report cross-sectional data is likely to have introduced measurement error, although mixing up different answering formats in theory mitigates the risk for common method bias (Podsakoff et al. 2003) and CFA helps control for same-source variance. Therefore, we recommend future researchers to combine self-report data with other-report data, like supervisor ratings of organizational citizenship behaviour, and adopt a longitudinal research design. This data can be used to more robustly test the hypotheses from this study and explore other components of validity, such as predictive and incremental validity.

Our choice to collect data through Prolific limits the generalizability of our research findings. First and foremost, respondents did not come from the same professional context or company. The findings are therefore not generalizable to a typical employment survey context. Second, respondents from online survey platforms such as Prolific often are very experienced survey takers and thus may have lost their naivety (Peer et al. 2017). This issue may not have biased our results too much, because the purpose of our study was not immediately obvious for Prolific workers and employees in organizations are increasingly often asked to complete employee surveys (Gerrad and Hyland 2020). Third, the motivation of the respondents in our sample to provide high-quality data may be unrepresentative for an average sample in a traditional employee survey. In our study, we did not have any item non-response, because we paid Prolific workers to answer all questions and only considered workers with a high approval rating. In a traditional (corporate) survey context where respondents are not financially compensated, there is likely a higher risk of missing (text) data (Anseel et al. 2010; Scholz and Zuell 2012). For these reasons, it would be interesting to replicate our validation study in other, more natural contexts, such as an employee satisfaction

survey in an organisation, and investigate the generalizability of our results in subpopulations (e.g., blue-collar vs. white-collar workers, different industries) and across nations.

4.4 Concluding Remarks

We want to emphasize that closed questions are likely still the best strategy for quantifying well-being, because the comparability of closed question measures are high, non-response is relatively low and validation is straightforward. However, we believe that semi-open (and open) survey questions can be asked alongside closed questions to fully realize the closed questions' potential. As complements of closed questions, semi-open questions could serve as a source of qualitative insights and means to cross-validate closed questions. Opportunely, computer-aided text analysis has the promise to mitigate the traditional obstacles, such as labour-intensiveness, that are typically associated with using textual data to study psychological constructs. We expect that the rapid advances in computational linguistics and its applications in psychological science will make computer-aided text analysis more reliable and spur new research avenues on the parallel use of the different types of survey instruments. It is not expected that any particular multiple-item survey scale will soon be labelled the gold standard for measuring all aspects of a psychological construct, such as job satisfaction, in every context. Therefore, we expect that "opening up of standardized surveys" (Singer and Couper 2017, p. 128) and looking at constructs from different epistemological angles (Mauceri 2016) will eventually allow researchers and employers to capture employees' job evaluations and feelings more validly, and to generate better insights into what employees really thinking influences their job attitudes. This could be crucial for the development of more context-specific and allegedly more effective strategies to improve employee well-being.

Appendix

Table 5 Robustness test of confirmatory factor analyses

Sample	Measure	χ^2 df=2	CFI	SRMR	RMSEA	Standardized factor loading of text measure
Full sample (<i>N</i> =395)	SentimentR	3.697	0.999	.010	.046	.72
	LIWC2015	0.535	1.000	.003	.000	.56
	Respondent	0.466	1.000	.002	.000	.83
Sample 1 (<i>N</i> =232)	SentimentR	2.661	0.999	.010	.038	.74
	LIWC2015	1.623	1.000	.009	.000	.58
	Respondent	0.055	1.000	.001	.000	.85
Sample 2 (<i>N</i> =290)	SentimentR	11.543**	0.991	.015	.128	.74
	LIWC2015	5.125	0.997	.011	.073	.60
	Respondent	9.304*	0.994	.010	.112	.86
Sample 3 (<i>N</i> =381)	SentimentR	1.562	1.000	.006	.000	.73
	LIWC2015	1.192	1.000	.005	.000	.56
	Respondent	0.051	1.000	.001	.000	.83

Note. *: $p < .05$; **: $p < .01$; n = sample size; LIWC = Linguistic Inquiry and Word Count; CFI = comparative fit index; SRMR = standardized root mean square residuals; RMSEA = root mean square error of approximation

Table 6 Bivariate correlations in reduced samples

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
Sample 1 (N=232)											
1. Job satisfaction (SentimentR measure)											
2. Job satisfaction (LIWC2015 measure)	.63										
3. Job satisfaction (benchmark measure)	.81	.65									
4. Job satisfaction (MOAQ-JSS)	.71	.55	.82								
5. Task variety	.18	.14	.26	.33							
6. Job autonomy	.34	.26	.48	.56	.47						
7. Person-organization fit	.52	.48	.64	.67	.38	.54					
8. Life satisfaction	.40	.33	.46	.50	.15	.30	.39				
9. Turnover intention	-.55	-.42	-.64	-.77	-.26	-.40	-.53	-.41			
10. Organizational citizenship behaviour	.22	.18	.33	.30	.42	.32	.36	.17	-.21		
11. Need for cognition	.03	.12	.11	.14	.13	.18	.16	.14	-.08	.27	
12. Openness	-.03	.07	-.01	.06	.09	.10	.15	-.05	.06	.14	.56
Sample 2 (N=290)											
1. Job satisfaction (SentimentR measure)											
2. Job satisfaction (LIWC2015 measure)	.63										
3. Job satisfaction (benchmark measure)	.81	.66									
4. Job satisfaction (MOAQ-JSS)	.73	.59	.84								
5. Task variety	.20	.19	.28	.34							
6. Job autonomy	.37	.32	.50	.55	.49						
7. Person-organization fit	.55	.47	.65	.71	.40	.54					
8. Life satisfaction	.40	.31	.45	.52	.22	.33	.39				
9. Turnover intention	-.55	-.42	-.66	-.74	-.32	-.44	-.58	-.44			
10. Organizational citizenship behaviour	.18	.18	.27	.26	.38	.27	.29	.12	-.18		
11. Need for cognition	.01	.04	.09	.10	.14	.14	.05	.13	-.06	.22	
12. Openness	-.04	.04	-.04	-.01	.06	.00	.04	-.09	.14	.11	.49
Sample 3 (N=381)											
1. Job satisfaction (SentimentR measure)											
2. Job satisfaction (LIWC2015 measure)	.60										
3. Job satisfaction (benchmark measure)	.80	.62									
4. Job satisfaction (MOAQ-JSS)	.71	.55	.80								
5. Task variety	.18	.18	.29	.34							
6. Job autonomy	.31	.28	.47	.53	.52						
7. Person-organization fit	.53	.45	.64	.67	.36	.51					
8. Life satisfaction	.39	.34	.46	.50	.21	.31	.40				
9. Turnover intention	-.50	-.40	-.60	-.74	-.28	-.40	-.53	-.41			
10. Organizational citizenship behaviour	.18	.14	.30	.26	.36	.28	.32	.18	-.16		
11. Need for cognition	.02	.09	.09	.12	.11	.10	.08	.15	-.08	.16	
12. Openness	-.04	.07	-.02	.02	.05	.02	.08	-.07	.10	.08	.54

Note. All correlations significant at the level of $p < .05$ (2-tailed), except for the ones in italic; LIWC = Linguistic Inquiry and Word Count; MOAQ-JSS = Michigan Organisational Assessment Questionnaire - Job Satisfaction Subscale

Code Availability The authors confirm that the code supporting the findings of this study are available at <https://osf.io/tkzv4/>.

Data Availability The authors confirm that the data supporting the findings of this study are available at <https://osf.io/tkzv4/>.

Declarations

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allan, B. A., Dexter, C., Kinsey, R., & Parker, S. (2018). Meaningful work and mental health: Job satisfaction as a moderator. *Journal of Mental Health, 27*(1), 38–44.
- Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology, 25*(3), 335–349.
- Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273.
- Borg, I., & Zuell, C. (2012). Write-in comments in employee surveys. *International Journal of Manpower, 33*(2), 206–220.
- Bowling, N. A., Eschleman, K. J., & Wang, Q. (2010). A meta-analytic examination of the relationship between job satisfaction and subjective well-being. *Journal of Occupational and Organizational Psychology, 83*(4), 915–934.
- Bowling, N. A., & Hammond, G. D. (2008). A meta-analytic examination of the construct validity of the Michigan organizational assessment questionnaire job satisfaction subscale. *Journal of Vocational Behavior, 73*(1), 63–77.
- Bowling, N. A., Wagner, S. H., & Behr, T. A. (2018). The facet satisfaction scale: An effective affective measure of job satisfaction facets. *Journal of Business and Psychology, 33*(3), 383–403.
- Brief, A. P. (1998). *Attitudes in and around organizations*. Sage Publications.
- Briner, R. B. (2014). What is employee engagement and does it matter? An evidence-based approach. In D. Robinson & J. Gifford (Eds.), *The future of engagement thought piece collection, engage for success Peer-reviewed thought leadership series*. Engage for success. <http://www.engageforsuccess.org/futureemployee-engagement>.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97–113.
- Bui, H. T. (2017). Big five personality traits and job satisfaction: Evidence from a national sample. *Journal of General Management, 42*(3), 21–30.
- Cable, D. M., & Judge, T. A. (1996). Person–organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes, 67*(3), 294–311.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116.
- Cammann, C., Fichman, M., Jenkins, D., & Klesh, J. (1979). The Michigan organizational assessment questionnaire. Unpublished manuscript. University of Michigan, Ann Arbor.
- Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology, 51*(2), 291.
- Cohen, G., Blake, R. S., & Goodman, D. (2016). Does turnover intention matter? Evaluating the usefulness of turnover intention rate as a predictor of actual turnover rate. *Review of Public Personnel Administration, 36*(3), 240–263.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*(10), 687–693.
- Costa, P. T., & McCrae, R. R. (1985). The NEO personality inventory manual. Psychological Assessment Resources, Odessa, Florida.

- Crede, M., Chernyshenko, O. S., Stark, S., Dalal, R. S., & Bashshur, M. (2007). Job satisfaction as mediator: An assessment of job satisfaction's position within the nomological network. *Journal of Occupational and Organizational Psychology*, *80*(3), 515–538.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19.
- Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, *90*(6), 12–41.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412.
- Edwards, J. R. (2003). Construct validation in organizational behavior research. In J. Greenberg (Ed.), *Organizational behavior: The state of the science* (2nd ed., pp. 327–371). Lawrence Erlbaum Associates Publishers.
- Faragher, E. B., Cass, M., & Cooper, C. L. (2005). The relationship between job satisfaction and health: A meta-analysis. *Occupational Environmental Medicine*, *62*, 105–112.
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89.
- Fielding, N. G. (2012). Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of Mixed Methods Research*, *6*(2), 124–136.
- Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology*, *21*(1), 3.
- Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*, *8*(2), 196–202.
- Fried, Y., & Ferris, G. R. (1987). The validity of the job characteristics model: A review and meta-analysis. *Personnel Psychology*, *40*(2), 287–322.
- Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195.
- Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*(6), 898–915.
- Gerrad, L., & Hyland, P. (2020). Employee survey research: A critical review of theory and practice. In *Employee Surveys and Sensing* (pp. 374–390).
- Gilles, I., Mayer, M., Courvoisier, N., & Peytremann-Bridevaux, I. (2017). Joint analyses of open comments and quantitative data: Added value in a job satisfaction survey of hospital professionals. *PLoS One*, *12*(3), e0173950.
- Glaser, M., Garsoffky, B., & Schwan, S. (2009). Narrative-based learning: Possible benefits and problems. *Communications*, *34*(4), 429–447.
- Glerum, A., Atasoy, B., & Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models. *Journal of Choice Modelling*, *10*, 11–33.
- Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly*, *101384*.
- Guest, D. E. (2017). Human resource management and employee well-being: Towards a new analytic framework. *Human Resource Management Journal*, *27*(1), 22–38.
- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, *16*(2), 250–279.
- Hanisch, K. A., & Hulin, C. L. (1991). General attitudes and organizational withdrawal: An evaluation of a causal model. *Journal of Vocational Behavior*, *39*(1), 110–128.
- Harvey, P., & Martinko, M. J. (2009). An empirical examination of the role of attributions in psychological entitlement and its outcomes. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *30*(4), 459–476.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121.
- Hsieh, C.-M. (2012). Importance is not unimportant: The role of importance weighting in QoL measures. *Social Indicators Research*, *109*(2), 267–278.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

- Ikoru, V., Sharmina, M., Malik, K., & Batista-Navarro, R. (2018). Analyzing sentiments expressed on twitter by UK energy company consumers. 2018 Fifth international conference on social networks analysis, management and security (SNAMS), 95–98.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602–611.
- Jockers, M. (2017). *syuzhet: Extracts sentiment and sentiment-derived plot arcs from text*. Package ‘Syuzhet.’ <https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf>
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87(3), 530.
- Judge, T. A., Hulin, C. L., & Dalal, R. S. (2012). Job satisfaction and job affect. In S. W. J. Kozlowski (Ed.), *Oxford library of psychology. The Oxford handbook of organizational psychology* (Vol. 1, pp. 496–525). Oxford University Press.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3), 376.
- Judge, T. A., & Watanabe, S. (1994). Individual differences in the nature of the relationship between job and life satisfaction. *Journal of Occupational and Organizational Psychology*, 67(2), 101–107.
- Judge, T. A., Weiss, H. M., Kammeyer-Mueller, J. D., & Hulin, C. L. (2017). Job attitudes, job satisfaction, and job affect: A century of continuity and of change. *Journal of Applied Psychology*, 102(3), 356–374.
- Jung, Y., & Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decision Support Systems*, 113074.
- Kaplan, S. A., Warren, C. R., Barsky, A. P., & Thoresen, C. J. (2009). A note on the relationship between affect(ivity) and differing conceptualizations of job satisfaction: Some unexpected meta-analytic findings. *European Journal of Work and Organizational Psychology*, 18(1), 29–54.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2017). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733–765.
- Koys, D. J. (2001). The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study. *Personnel Psychology*, 54(1), 101–114.
- Kristof, A. L. (1996). Person–organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, 49(1), 1–49.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567.
- LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, 87(1), 52–65.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology*, 1(1), 3–30.
- Mastekaasa, A. (1984). Multiplicative and additive models of job and life satisfaction. *Social Indicators Research*, 14(2), 141–163.
- Matsunaga, M. (2010). How to factor-analyze your data right: Do’s, don’ts, and how-to’s. *International Journal of Psychological Research*, 3(1), 97–110.
- Mauceri, S. (2016). Integrating quality into quantity: Survey research in the era of mixed methods. *Quality & Quantity*, 50(3), 1213–1231.
- Maxwell, S. E., & Delaney, H. D. (1985). Measurement and statistics: An examination of construct validity. *Psychological Bulletin*, 97(1), 85.
- McCrae, R. R., & Costa Jr, P. T. (1997). Conceptions and correlates of openness to experience. In S. R. Briggs, R. Hogan, & W. H. Jones (Eds.), *Handbook of personality psychology* (pp. 825–847). Academic Press.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum
- McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn’t get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, 44(7), 2909–2933.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455.
- Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54.
- Mikhaylov, S., Laver, M., & Benoit, K. R. (2012). Coder reliability and misclassification in the human coding of party manifestos. *Political Analysis*, 20(1), 78–91.
- Moniz, A., & Jong, F. (2014). Sentiment analysis and the impact of employee satisfaction on firm earnings. *Proceedings of European Conference on Information Retrieval*, 519–527.
- Morgeson, F. P., & Humphrey, S. E. (2006). The work design questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology*, 91(6), 1321.

- Naldi, M. (2019). A review of sentiment computation methods with R packages. *ArXiv Preprint ArXiv, 1901, 08319*.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Organ, D. W. (1988). A restatement of the satisfaction-performance hypothesis. *Journal of Management, 14*(4), 547–557.
- Organ, D. W., & Near, J. P. (1985). Cognition vs affect in measures of job satisfaction. *International Journal of Psychology, 20*(2), 241–253.
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods, 22*(3), 765–797.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
- Poncheri, R. M., Lindberg, J. T., Thompson, L. F., & Surface, E. A. (2008). A comment on employee surveys: Negativity bias in open-ended responses. *Organizational Research Methods, 11*(3), 614–630.
- Potkey, C. R., & Allen, B. P. (1973). The adjective generation technique: An alternative to adjective check lists. *Psychological Reports, 32*(2), 457–458.
- Purcell, J. (2014). Disengaging from engagement. *Human Resource Management Journal, 24*(3), 241–254.
- R Core Team. (2018). R: A language and environment for statistical computing . R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rinker, T. W. (2019). *sentimentr: Calculate Text Polarity Sentiment*. University at Buffalo/SUNY, Buffalo, New York. Version 2.7.1.
- Russell, S. S., Spitzmuller, C., Lin, L. F., Stanton, J. M., Smith, P. C., & Ironson, G. H. (2004). Shorter can also be better: The abridged job in general scale. *Educational and Psychological Measurement, 64*, 878–893.
- Rynes, S. L. (2012). The research-practice gap in I/O psychology and related fields: Challenges and potential solutions. *The Oxford Handbook of Organizational Psychology, 1*, 409–452.
- Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts there? *Personnel Psychology, 36*(3), 577–600.
- Scholz, E., & Zuell, C. (2012). Item non-response in open-ended questions: Who does not answer on the meaning of left and right? *Social Science Research, 41*(6), 1415–1428.
- Schyns, B., Torka, N., & Gössling, T. (2007). Turnover intention and preparedness for change: Exploring leader-member exchange and occupational self-efficacy as antecedents of two employability predictors. *Career Development International, 12*(7), 660–679.
- Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*(1), 80–110.
- Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon’s mechanical Turk. *Communication Monographs, 85*(1), 140–156.
- Shin, D. C., & Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social Indicators Research, 5*(1–4), 475–492.
- Short, J. C., Broberg, J. C., Coglisier, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods, 13*(2), 320–347.
- Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior, 5*, 415–435.
- Singer, E., & Couper, M. P. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (Mda), 11*(2), 115–134.
- Snow, R., O’Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and fast—But is it good?: Evaluating non-expert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, 254–263*.
- Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences* (Vol. 3). Sage publications.

- Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology, 95*(4), 781–790.
- Spector, P. E., & Pindek, S. (2016). The future of research methods in work and occupational health psychology. *Applied Psychology, 65*(2), 412–431.
- Taber, T. D. (1991). Triangulating job attitudes with interpretive and positivist measurement methods. *Personnel Psychology, 44*(3), 577–600.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.
- Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology, 46*(2), 259–293.
- Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods, 20*(2), 243–267.
- Van den Broeck, A., De Cuyper, N., De Witte, H., & Vansteenkiste, M. (2010). Not all job demands are equal: Differentiating job hindrances and job challenges in the job demands–resources model. *European Journal of Work and Organizational Psychology, 19*(6), 735–759.
- 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*(2), 219.
- Vinten, G. (1995). Open versus closed questions—an open issue. *Management Decision, 33*(4), 27–31.
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*(2), 247–252.
- Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resource Management Review, 12*(2), 173–194.
- Weissman, G. E., Ungar, L. H., Harhay, M. O., Courtright, K. R., & Halpern, S. D. (2019). Construct validity of six sentiment analysis methods in the text of encounter notes of patients with critical illness. *Journal of Biomedical Informatics, 89*, 114–121.
- Wijngaards, I., Burger, M., & Van Exel, J. (2019). The promise of open survey questions - The validation of text-based job satisfaction measures. *PLoS One, 14*(12). <https://doi.org/10.1371/journal.pone.0226408>.
- Young, L. M., & Gavade, S. R. (2018). Translating emotional insights from hospitality employees' comments: Using sentiment analysis to understand job satisfaction. *International Hospitality Review, 32*(1), 75–92.
- Zangaro, G. A., & Soeken, K. L. (2007). A meta-analysis of studies of nurses' job satisfaction. *Research in Nursing & Health, 30*(4), 445–458.
- Zehner, F., Sälzer, C., & Goldhammer, F. (2016). Automatic coding of short text responses via clustering in educational assessment. *Educational and Psychological Measurement, 76*(2), 280–303.
- Zhang, D. C., Zhu, X., Ritter, K.-J., & Thiele, A. (2019). Telling stories to communicate the value of the pre-employment structured job interview. *International Journal of Selection and Assessment, 27*(4), 299–314.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Indy Wijngaards^{1,2} · Martijn Burger^{1,3} · Job van Exel²

¹ Erasmus Happiness Economics Research Organization, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands

² Erasmus School of Health & Policy Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands

³ Erasmus School of Economics and Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands