Received: 4 July 2022 | Accepted: 27 August 2023

DOI: 10.1111/apps.12501

ORIGINAL ARTICLE

APPLIED PSYCHOLOGY IAAP

What is work engagement? A text mining approach using employees' self-narratives

Henrico van Roekel[1] | Enno F. J. Wigger[2] | Bernard P. Veldkamp[3] | Arnold B. Bakker[4,5]

[1]Utrecht University, Utrecht, The Netherlands

[2]University of Applied Sciences Windesheim, Zwolle, The Netherlands

[3]University of Twente, Enschede, The Netherlands

[4]Erasmus University Rotterdam, Rotterdam, The Netherlands

[5]University of Johannesburg, Johannesburg, South Africa

Correspondence
Henrico van Roekel, Utrecht University, Utrecht, The Netherlands.
Email: h.vanroekel@uu.nl

Funding information
This research was funded by Stichting IZZ, a healthcare employee collective in the Netherlands. Except for the cooperation with IZZ in collecting the data, the funder had no role in the research.

Abstract

We introduce text mining to study work engagement by using this method to classify employees' survey-based self-narratives into high or low work engagement and analyzing the text features that contribute to the classification. We used two samples, representing the 2020 and 2021 waves of an annual survey among healthcare employees. In the first study, we used exploratory sample 1 ($N = 5591$) to explore which text features explain work engagement (unigrams, bigrams, psychological, or linguistic). In the second study, we confirmed whether features persisted over time between exploratory sample 1 and confirmatory sample 2 ($N = 4470$). We find that psychological features classify employees across two samples with 60% accuracy. These features partly validate the literature: High-engaged employees refer more to affiliation and positive emotions, and low-engaged employees refer more to negative emotions and power. We extend the literature by studying linguistics: High-engaged employees use more first-person plural ("we") than low-engaged employees. Finally, some results question the literature, like the finding that low-engaged employees refer more to their managers. This study shows text mining

Henrico van Roekel and Enno F. J. Wigger shared first authorship.

can contribute by confirming, extending, or questioning the literature on work engagement and explores how future research could build on our findings with survey-based or in vivo applications.

KEYWORDS
machine learning, self-narratives, text features, text mining, work engagement

INTRODUCTION

Whether employees are engaged in their work or not has important consequences for employees themselves, the organizations they work for, and the clients they work with. Engaged employees are full of energy, are dedicated toward work, and are often completely immersed in their work activities (Schaufeli & Bakker, 2010). They also experience more positive emotions, think in novel ways, and show better performance (Bakker et al., 2014; Christian et al., 2011). Since the emergence of the concept of work engagement, organizational scholars have been studying its presence, predictors, and outcomes (Bakker & Demerouti, 2017). However, most of the literature assesses work engagement with structured data, that is, measurement scales, and there have been few attempts to innovate measurements. Although structured data have allowed scholars to understand the phenomenon of work engagement, a drawback is the limited potential for new theoretical or applied discoveries (Balducci & Marinova, 2018).

At the same time, within organizations, a vast pool of data, in the form of unstructured (non-predefined) text data, remains scarcely studied. For example, employees generate and share large amounts of written text with each other. Those qualitative data may potentially offer new insights in work engagement and add to the more traditional structured approaches to data analysis. This is because unstructured data are not limited to predefined categories, present the multidimensionality of a phenomenon, and allow to compare these dimensions simultaneously (e.g., combining linguistic and substantive patterns) (Balducci & Marinova, 2018). Text mining offers a unique approach to unlock these insights as it is a method to analyze large amounts of text in a relatively short timeframe (Jurafsky & Martin, 2017). Its benefit compared with traditional quantitative or qualitative research is that it is able to analyze unstructured text, but on a large scale and replicable across studies. There are quite a few studies that show its potential in a variety of disciplines (e.g., Pang et al., 2020), but, although declared a future research avenue, few attempts have been made regarding organizational research (Kobayashi et al., 2021).

Therefore, the purpose of the current study is to explain work engagement through text mining methods by attempting to classify employees' survey-based self-narratives into high or low work engagement and analyze the text features that contribute to the classification. The research question that guided our study is as follows: To what extent can we explain work engagement by analyzing self-narratives through text mining? Using two samples, representing two waves of an annual survey among Dutch healthcare employees during 2 years of COVID-19, this paper conducts two studies to answer that question. We tested multiple text features: unigrams, bigrams, psychological features, and linguistic features. For the psychological features, we conducted a preselection based on the job demands–resources (JD-R) theory. Next, for the first study, we used exploratory sample 1 to explore which features explain work engagement. We then formulated hypotheses based on the main themes that emerged from the features. For the second study, we used both exploratory sample 1 and confirmatory sample 2 to analyze to what extent text features persist over time, across survey waves.

Our study contributes to the literature by being the first to explain work engagement by text mining self-narratives. Theoretically, we increase the understanding of work engagement as a concept. Some of our results confirm the duality of the JD-R model when we find low-engaged employees tend to mention job demands whereas high-engaged employees tend to mention job resources (e.g., Bakker, 2022; Bakker et al., 2022; Wang, Zhu, et al., 2020). Yet because our analysis is exploratory, we are also able to extend and question extant findings. We find linguistic patterns may be markers of work engagement (e.g., Franklin & Thompson, 2005) and observe features that question the literature, like the finding that low-engaged employees mention their managers more often (Toegel et al., 2013). At the same time, we also discuss how our application of text mining is limited in terms of the accuracy with which we are able to explain work engagement, as well as how particular sample characteristics like age and gender may influence the results. Second, methodologically, our findings open multiple avenues for survey-based and in vivo applications of text mining. We discuss how text mining could support or complement structured forms of data collection (Jurafsky & Martin, 2017; Kobayashi et al., 2021), or be used to analyze existing unstructured data in organizations like emails or intranet posts (He et al., 2012). Finally, we explore how our study may have practical implications in the screening and identification of groups of employees based on work-related well-being challenges (e.g., Day et al., 2007; He et al., 2012).

THEORETICAL BACKGROUND

Defining, modeling, and measuring work engagement

Work engagement is a work-related and positive state of mind, characterized by vigor, dedication, and absorption. Vigor refers to a high level of energy and preparedness to invest effort in activities. Dedication refers to enthusiasm and strong involvement with one's work. Finally, absorption is a state of complete immersion in one's work (Bakker et al., 2014; Schaufeli et al., 2002). Whereas vigor and dedication are considered core dimensions of work engagement, absorption is considered an additional dimension (Schaufeli et al., 2001). Whereas our knowledge of work engagement has increased in the past years, there are several remaining questions, for example, on its social–psychological origins and the effectiveness of work engagement interventions (Bakker, 2022; Knight et al., 2019).

Antecedents of work engagement are often studied within JD-R theory, a theory within organizational psychology that explains how job characteristics affect employees through a dual process (Bakker & Demerouti, 2017; Bakker et al., 2022; Schaufeli et al., 2009). In the health impairment process, demanding job characteristics—"aspects of the job that require sustained physical, emotional, or cognitive effort"—cause job strain (including burnout) and health complaints. Burnout refers to the state when employees experience chronic feelings of exhaustion and a cynical attitude towards work and the people with whom they work (Bakker et al., 2014). In the motivational process, resourceful job characteristics—"aspects of the job that help to either achieve work goals, reduce job demands (…), or stimulate personal growth"—foster motivational outcomes (including work engagement) and job performance (Bakker et al., 2014, p. 392). In addition, job demands and resources are proposed to interact: Resources may weaken the impact of demands on burnout, whereas challenge demands may strengthen the impact of resources on engagement (Bakker et al., 2014; Boyd et al., 2011).

Over the years, many resources that stimulate work engagement have been identified. Generally, they have been classified into one of two categories: situational and individual factors (Bakker & Demerouti, 2008). As explained above, antecedents of work engagement are mainly job resources. These include job characteristics like social support from colleagues, task significance, and autonomy as well as leadership-related factors like having a good relationship with supervisors and experiencing transformational leadership (Christian et al., 2011). In addition, individual factors have been found to explain work engagement. For example, employees with higher emotional stability, extraversion, and conscientiousness are more likely to report higher work engagement. Besides these higher-order personality factors, lower-order factors—factors that are more malleable—have been found to predict work engagement, for example, self-efficacy and optimism (Mäkikangas et al., 2013). And employees who are more proactive tend to be more engaged and even positively influence their co-workers through practices of job crafting (Bakker et al., 2012; Tims et al., 2015).

In turn, studies have shown that work engagement can have far-reaching effects (Christian & Slaughter, 2007; Knight et al., 2017; Lesener et al., 2020). Research on work engagement shows positive relationships with more active positive emotions and more novel thinking (Bakker et al., 2014). What is more, there is abundant research that shows work engagement increases task performance (e.g., Christian et al., 2011), although studies also indicate that we know relatively little about the boundary conditions of these effects (Kane-Frieder et al., 2014).

The studies described above commonly measure work engagement with multidimensional scales. The most used scale that defines work engagement as the combination of vigor, dedication, and absorption is the Utrecht Work Engagement Scale (UWES; Schaufeli et al., 2006). Other measures of engagement are very similar to the UWES (see Schaufeli & Bakker, 2010, 2022), or measure concepts that are fundamentally different from engagement. For example, May et al. (2004) and Rich et al. (2010) developed the Job Engagement Scale, which includes cognitive, emotional, and physical engagement. According to Bakker et al. (2022, p. 285), “the wording of the items shows a striking resemblance with those included in the absorption, dedication, and vigor subscales of the UWES, respectively.” The latter authors also discuss other instruments to assess engagement, including the instrument by Soane et al. (2012) and Shuck et al. (2017). Bakker et al. (2022, p. 286) conclude that the items show considerable overlap with the vigor and absorption subscales of the UWES, whereas some of the alternative instruments that aim to assess engagement in fact assess affective organizational commitment and extra-role behaviors.

Although new scales have been developed since (Schaufeli et al., 2017), there have been few attempts to innovate measurement. For example, Bakker et al. (2014) point out that most of the research on work engagement has not attempted to link the concepts to observable outcomes. At the same time, there is some criticism on the UWES, including the fact that factor analyses have not always been able to distinguish between the three components of work engagement (Schaufeli & Bakker, 2022). Here, a new method like text mining may help optimize the measurement of work engagement by approaching it in a completely different way. Similarly, a particular bias of maintaining the same measurement methods is that these structured data limit the potential for new theoretical or applied discoveries (Balducci & Marinova, 2018). Text mining may allow new insights into what observable behaviors of employees are affected by work engagement. Below we address this issue further.

Considering unstructured data to measure work engagement

The vast majority of data in an organization are unstructured. Unstructured data refer to "a single data unit in which the information offers a relatively concurrent representation of its multifaceted nature without predefined organization or numeric values" (Balducci & Marinova, 2018, p. 558). For example, employees continuously exchange spoken or written text via conversations, email, or texting. These text data are seldom used in studies but may present new insights for the study of work engagement through three advantages over structured data. First, structured data (like survey scales) are always limited to the way they are defined and operationalized. In contrast, unstructured text data are neither predefined nor categorized, and this may lead to new insights. Second, unstructured data are multifaceted. There are multiple potential facets to unstructured data to be studied (e.g., there are linguistic and substantive properties to text). Third, unstructured data offer concurrent representation: Through analyzing facets simultaneously (e.g., the combination of linguistic and substantive patterns), we can learn about different phenomena at the same time (Balducci & Marinova, 2018).

Although unstructured data offers new opportunities for research, structured data are important too. Unstructured data provide, besides its general format being either text-based or image-based, little certainty. This type of data does not allow for easy sorting, searching, analyzing, summarizing, or visualizing. Structured data, on the other hand, provide certainty in measurement and analysis as the data are set in predetermined categories or values. This type of data is easily stored, searched, analyzed, summarized, and visualized. The data contain exactly what could be expected and allow for accessible, unbiased analysis. It has been an important part of theory-based research as theory is operationalized into a specific measurable form. In comparison, unstructured data require a different, more thorough approach to hold value (Hänig et al., 2010).

Although unstructured data have been underused with regard to work engagement, previous studies have shown that free-form text can be a rich source of data that contains important insights about mental well-being and allows identification and screening for mental diseases. For example, research has shown that the content of the speech of schizophrenics differs substantively from non-schizophrenics (Franklin & Thompson, 2005; Rosenberg & Tucker, 1979). There are some specific examples of text mining in psychology and organizational research. Pang et al. (2020) succeeded in predicting 24 character strengths, like gratitude, zest, and leadership, based on Twitter language. This study indicated that one can use text mining to measure the character strengths of large populations. Similarly, La Bella et al. (2018) used text mining to track perceived organizational leadership styles almost real-time with Twitter messages. Examples in clinical settings include the screening of posttraumatic stress disorder in self-narratives (He et al., 2012) and the identification of trauma patients (Day et al., 2007). When employing text mining for work engagement, we hope to explore whether and how employees high in work engagement may display different features from employees low in work engagement.

One reason for the limited attention to analyzing textual data may be that traditionally analyzing text was a time-consuming endeavor as manual coding was the only option. However, new techniques derived from machine learning and statistics may enable to study work engagement and other concepts with unstructured data (Kobayashi et al., 2021). One particularly promising avenue to innovate is by text mining as it is a general methodological framework to analyze large corpora of text (Jurafsky & Martin, 2017). Hence, text mining offers large-scale text analysis in short timeframes, only bottlenecked by computing power and the fact that often large amounts of texts are required to generate insights. Its benefit, therefore, compared with traditional quantitative or qualitative research is that it is able to analyze unstructured text, but on a large scale and replicable across studies. Only recently organizational scholars have suggested that this approach could be used to assess organizational concepts like burnout (Kobayashi et al., 2021).

Text mining refers to the analytical process that aims to generate insights or test hypotheses using unstructured text data (Kao & Poteet, 2007). The data are systematically collected, cleaned, and transformed (a process referred to as preprocessing), after which one of multiple text mining operations can be applied to generate insights from the text data. Texts can be analyzed based on textual patterns and linguistic features, as well as dictionary approaches, considering the words used in the texts. Finally, postprocessing requires the interpretation and evaluation of the results by applying specific domain knowledge to them and validating the data (Kobayashi et al., 2018).

There are many text features that can be analyzed through text mining. First, textual patterns such as bag-of-words approaches look at the occurrence of words in texts and try to understand the corpus based on word counts (Aggarwal & Zhai, 2012). Similar approaches include the use of n-grams, which refers to word combinations of two (e.g., "working day"), three (e.g., "busy working day"), or more words. The advantage of using bag-of-words and n-grams lies in their simplicity. The only information lost is the position of the words or n-grams in the text, which means it is a true-to-source feature to analyze. The downside, however, is that respondents with different background characteristics like educational background or social status may use different words to convey identical information. Patterns might emerge based on characteristics that are unrelated to the research at hand.

Second, there are dictionary approaches, which tag words in the texts with categories the words belong to. In our analysis, we can use these categories to understand the texts. An example of such a dictionary approach is Linguistic Inquiry and Word Count (LIWC). LIWC counts words in linguistic and psychologically meaningful categories (Tausczik & Pennebaker, 2010). Dictionary approaches resolve the issue of textual pattern features as the underlying meanings and categories of the words are analyzed, rather than the words themselves. However, the downside is a loss of information as the words themselves are not analyzed further. In choosing the features, one could either test all available features or apply some sort of a priori selection. A priori selection is often applied to prevent overfitting. Overfitting is a common problem in machine learning where a model performs well on the training data but fails to generalize to new, unseen data. This happens because the model is trained on a specific set of data that might not contain a fully balanced representation of all words that do or do not contribute to the classification of the variable. The model will fit to the training data as specifically as possible, even though there might be false patterns that do not hold over multiple samples. Overfitting can be avoided by using a larger and more diverse dataset for training, as well as using regularization techniques to prevent the model from learning overly complex patterns in the data. Our text mining approach was both theory and data driven. Specifically, we used the JD-R theory to select psychological features based upon their resemblance to any aspect of the definition, dimension or items of work engagement. In sum, in the present study, we hope to gain new insights into the concept of work engagement using text mining. Across two studies, we will compare bag-of-words, bigrams, and LIWC dictionary approaches (psychological process and linguistic features) to explore the possibilities offered by text mining.

METHODS

Procedure, samples, and data

We used data from two samples, representing two waves of an annual survey among Dutch healthcare employees who are members of *Stichting IZZ*, a collective of healthcare employees and employers in the Netherlands. This foundation has over 400,000 members of which around 210,000 are healthcare employees, who make up a notable share of the population of around 1.7 million healthcare employees in the country (CBS, 2022; Van der Fels, 2020). The annual survey, executed since 2018, is used to monitor how healthcare employees perceive their work and well-being. It presents an opportunity for employees to share their experiences, which are then shared (at the group level) with healthcare organizations, governments, societal partners, and media. The survey has, among else, been helpful in informing these parties about the challenges COVID-19 posed for healthcare employees. Besides, to add an extrinsic motivation to finish the survey, participants could choose to participate in a separately organized giveaway (with products that stimulate well-being).

For the first sample, data were collected in May and June 2020. For the second sample, data were collected in May and June 2021. Table 1 shows how we arrived at our final sample. All members of the collective that provided an email address were sent an invitation to participate in the survey via email. For a response to be valid, respondents had to provide informed consent and indicate they were currently working in healthcare. At the informed consent page, respondents were informed about the goal of the survey, procedures of participation and opting out, data storage and usage, and the possibility to get in touch with the researchers. The survey itself consisted out of multiple open and closed questions on employee well-being in healthcare, including the questions presented in this study. We did not employ attention checks. All questions used in this study, except the text mining question, used forced response. The text mining question was placed at the end of the survey as it would take considerable time. Therefore, as Table 1 indicates, respondents with partial responses corresponded with respondents who did not fill in the text mining question—these were removed from the dataset. Besides, to be included in our final sample, we set the minimum number of written words at 20. Finally, as in

TABLE 1 Sample justification.

	Sample 1 (2020)	Sample 2 (2021)
No. of survey invitations	138,382	133,322
No. of responses recorded	19,772	8955
~ Response rate (recorded responses/invitations)	**14.29%**	**6.72%**
No. of valid responses[a]	12,630	8132
No. of responses to text mining question	5976	5016
~ % of partial responses (no answer text mining/valid responses)	**52.68%**	**38.32%**
No. of responses with 20 words or more (final sample)	5591	4470
~ Response rate for final sample (final sample/invitations)	**4.04%**	**3.35%**
No. of responses in top/bottom 10%[b] (subsample)	1119	894

[a]Respondents who gave consent to participate and were currently working in healthcare.
[b]Number of responses in the top or bottom 10% of work engagement scores.

our study, we will compare employees in the top 10% with those in the bottom 10% of work engagement scores. Table 1 also presents this subsample. The methodological choices made above will be elaborated on below.

Table 2 presents the characteristics of the respondents in the final sample. The respondents are representative for the population of Dutch healthcare employees in terms of gender (84.3% of employees are female) but somewhat less representative in terms of age (employees in the population are younger: 34% are younger than 35 and 24.2% are older than 55) (CBS, 2020). Especially the gender composition, with a vast majority of female employees, is a typical (but not unique: e.g., in Dutch primary education, 87% of teachers are female; OCW, 2021) characteristic of healthcare sectors. Although the majority in the Netherlands is very large, a WHO report shows across the world women constitute around 70% of the healthcare workforce (World Health Organization, 2019). What is more, the same report indicates that the small minority of men in healthcare is more likely to hold leadership positions. This leadership gap has systemic roots in gender roles (see, e.g., Ryan et al., 2016): Men and women in healthcare (are expected to) work in different jobs. We should take into account that such factors could affect work engagement. Besides, although our sample is older than the population of healthcare employees, the population of healthcare employees is aging rapidly (the group of healthcare employees aged 55 and oversaw a 9% increase in just 10 years; Van Wijk, 2020). Finally, nursing/home care, hospitals, and disabled care constitute the biggest healthcare branches within the Netherlands and are also the largest in our sample. However, although hospitals are the biggest group in our sample, within the population, nursing/home care is bigger

TABLE 2 Sample characteristics ($N_1 = 5591$ and $N_2 = 4470$).

	Sample 1	Sample 2
Gender		
Female	4731 (84.6%)	3816 (85.4%)
Male	845 (15.1%)	636 (14.2%)
Rather not say	15 (.3%)	18 (.4%)
Age		
−25	48 (.9%)	36 (.8%)
26–35	447 (8%)	240 (5.4%)
36–45	991 (17.7%)	709 (15.9%)
46–55	1711 (30.6%)	1366 (30.6%)
56–65	2327 (41.6%)	2081 (46.6%)
66–	63 (1.1%)	36 (.8%)
Unknown	4 (.1%)	2 (< .1%)
Healthcare branch		
Hospitals	1983 (35.5%)	1582 (35.4%)
Nursing/home care	1343 (24%)	1175 (26.3%)
Mental healthcare	911 (16.3%)	681 (15.2%)
Disabled care	981 (17.5%)	743 (16.6%)
Other	373 (6.7%)	289 (6.5%)

(with a total of 28% of healthcare employees in the population; Van Wijk, 2020). In sum, our samples of healthcare employees are fairly representative for healthcare.

It is important to note that there may be systemic reasons to expect differences in work engagement based on demographic characteristics. Table 3 shows that, to some extent, work engagement varies across age, gender, and healthcare branch. Most notably, work engagement is generally higher among women, among 46–55-year-old employees (not taking into account the youngest and oldest categories, which both have low *N*), and among employees in nursing/home care (Additional analysis in the Supporting Information elaborates on these differences). This may have consequences for our study's external generalizability. This study analyzes text-based features among a fairly representative sample of healthcare employees, but the particular sample characteristics (e.g., distribution of age and gender) and how work engagement relates to these characteristics may limit generalization to different sectors.

Finally, to protect the privacy-sensitive information that participants provided in their self-narratives, data are stored on secure servers in compliance with privacy regulations and not made publicly available. We do present multiple Supporting Information that provide extra information on the research process: an overview of the *included features*, the *R script* for our analyses, our approach to deciding the *cutoff*, an overview of all *significant features*, an overview of the relative *feature importance* to the models, *additional analysis* on the role of demographics, and an additional analysis using a different classifier (*Naive Bayes*) (accessible via https://osf.io/jzdx5/?view_only=f981153538214470b2bd1a9e9a538009).

TABLE 3 Work engagement across sample characteristics.[a]

Work engagement score	Sample 1	Sample 2
Gender		
Female	3.89 (.62)	3.81 (.65)
Male	3.78 (.67)	3.70 (.72)
Age		
−25	4.01 (.54)	3.68 (.59)
26–35	3.81 (.56)	3.70 (.62)
36–45	3.84 (.59)	3.79 (.61)
46–55	3.90 (.62)	3.83 (.66)
56–65	3.86 (.66)	3.77 (.69)
66-	3.98 (.66)	4.16 (.56)
Healthcare branch		
Hospitals	3.84 (.62)	3.75 (.68)
Nursing/home care	4.00 (.62)	3.91 (.66)
Mental healthcare	3.76 (.62)	3.73 (.63)
Disabled care	3.84 (.64)	3.77 (.66)
Other	3.92 (.62)	3.75 (.67)

[a]This table presents means and standard deviations. Supporting Information Additional analysis presents statistical tests for work engagement scores across sample characteristics and additional descriptives.

Work engagement scale

The UWES-9 work engagement scale includes nine items on three dimensions: vigor, dedication, and absorption (Schaufeli et al., 2006). All dimensions were measured with three items on a 5-point Likert scale ranging from "Never" (1) to "Always (daily)" (5). Example items are "At my work, I feel bursting with energy" (vigor), "I am proud of the work that I do" (dedication), and "I feel happy when I am working intensely" (absorption). The items were summed to create an overall index of work engagement. The reliability of the overall scale was good, Cronbach's alpha was .908 for sample 1 and .910 for sample 2.

Self-narrative question

Below, we present the English translation of the question that was shown to respondents to write their self-narrative:

We have one additional question about how you have experienced your work during this time of COVID-19. We would like to take a closer look at your personal experiences. Could you summarize what you have experienced? How have you experienced the past few months? What impact has this had? How are you feeling now, physically and emotionally? How do you view your work now? And how do you look forward to the coming months?

This multifaceted question functioned as a writing prompt to guide the content of the self-narratives (Kroll & Reid, 1994). Writing prompts make writing easier when they, in our case, promote the structure of the story that respondents are expected to write (Hudson et al., 2005). The question was drafted purposefully and in reiterative discussion between all the authors of this study and contained multiple subquestions that each served their own purpose. The first subquestion was general (Could you summarize what you have experienced?), after which the second subquestion specified the time period (How have you experienced the past few months?). Third, we asked about the consequences of these experiences (What impact has this had?). Fourth, we referred to the energy continuum of exhaustion versus vigor (How are you feeling now, physically and emotionally?). Fifth, we referred to the identification continuum of dedication versus cynicism (How do you view your work now?). Finally, we asked about participants' future perspective (And how do you look forward to the coming months?). After this, a text box provided participants ample opportunity to share their self-narratives.

Analysis

The process of text mining involves four basic steps: (1) data preprocessing; (2) training on a subset of the data; (3) testing on a different subset of the data; and (4) interpreting the results. We will explain these steps below.

Data preparation

In the first data preparation step, the corpus of all texts was cleaned, and features were extracted and selected to prepare for data analysis (Hester, & Bryan, 2022; Silge & Robinson, 2016; Wickham, Miller, & Smith, 2022; Wickham, Wickham, François, et al., 2022).

Packages and code used can be found in the Supporting Information *R script*. We checked whether we needed to apply criteria for minimum or maximum number of words in the self-narratives. An explorative analysis of the data indicated that respondents who replied with fewer than 20 words in their self-narratives commonly responded with variations of "I do not have anything to share." This is not a substantive answer to the question prompt—yet it occurred many times in the initial dataset. We decided to require respondents to have written 20 words at minimum to avoid meaningless self-narratives, but we did not apply a maximum as no self-narrative appeared extremely long. The second part of the data preparation aimed to tokenize the text by removing punctuation, numbers, and capitalization and by splitting the texts word by word (Benoit et al., 2018). The resulting list of words was spell-checked by one of the researchers for all words that occurred at least a total of 10 times to find spelling errors or gibberish that would be included in the model. One returning issue concerned the occurrence of abbreviations alongside the same abbreviations written out in full. This included both general abbreviations as well as job-specific abbreviations. As abbreviations were often unclear and seemed to vary on a text-by-text basis, these abbreviations were not manipulated to full terms. No other issues were found based on this quality check of the data, and after this quality check, we proceeded with the data. Further cleaning steps were the removal of frequently occurring stop words (e.g., "the" and "a") that do not discriminate texts or add little to no meaning to texts. The list of stopwords filtered is based on the Dutch stopwords list as compiled by the Snowball stemming project; this list is included in the corpus package (Perry, 2017; Porter, 2001). Finally, stemming the tokens, reducing all words to their stem, was done to ensure various inflections of the same word are counted together (e.g., "working," "worked," and "work" become "work"). For sample 1, this step reduced the total number of words by 48.44% (from 632,174 to 325,963) and the unique number of words by 29.41% (from 22,524 to 15,899). For sample 2, this step reduced the total number of words by 48.01% (from 443,668 to 230,653) and the unique number of words by 17.77% (from 15,573 to 12,806).

The other steps in data preparation were feature extraction and feature selection. This study used features based on bag-of-words approaches and dictionary approaches to explain work engagement. Importantly, below, we describe how the features were selected for sample 1. For sample 2, we only used the features that contributed to the classification into high and low work engagement in sample 1.

First, the bag-of-words approach assumes no relationship between the order of the words in a text and the meaning of the text. The words are as they are independent from other words in the text. This approach was used for generating unigrams (one word, e.g., "happy") and bigrams (two words, e.g., "not happy"). Second, the dictionary approach used the LIWC dictionary to tag words with categories belonging to Psychological Processes or Linguistic Dimensions (Pennebaker et al., 2015). The translated Dutch version of LIWC 2015 has 67 categories for which the words can be matched (Van Wissen & Boot, 2017). The features selected were both theory driven and data driven. For the Linguistic Dimensions, we explored all features available. For the psychological features, we selected features using the JD-R theory: We checked whether they reflected an aspect of the definition of work engagement, its dimensions, or items. We selected features from the affective, social, perceptual, and biological processes, drives, time orientations, relativity, and personal concerns (Pennebaker et al., 2015). Although this means that there was some a priori selection of features, the selection was necessary to limit the scope of the study. Additionally, because of the theory-related nature of the LIWC psychological process features, a priori selection based on theory could prevent overfitting. Through a priori selection of psychological process features, we narrowed the amount of psychological process features from 54 categories (including all overarching categories and more specific subcategories) to 25 categories. This was done to remove non-work related features, as these would be less relevant for our purposes of explaining work engagement. That is, in this specific analysis, we were looking for factors related to work, not for other factors. That is not to say that work engagement is irrelevant for employees' private lives, as research suggests otherwise (e.g., Wood et al., 2020). It only means that we slightly narrowed the scope of our analysis. Comparing the features to the definition, dimensions or items of work engagement provided a good measure for selection. For example, the category "Time orientations" (including the categories "past focus," "present focus," and "future focus") was included as work engagement is related to time orientations. The dimension absorption includes the phrase "whereby time passes quickly" (Bakker et al., 2014, p. 391). In contrast, some subcategories of "Personal concerns," like the "Leisure" category with words like "home," "chat," and "movie," were deemed less relevant for our endeavors (Pennebaker et al., 2015). Additionally, to prevent overfitting for the Random forest model and to ensure robust features were kept, feature selection was applied for each of the feature representations using chi-squared tests (Forman, 2003, 2004). Features were kept based on the criteria of significant chi-squared outcome for features that occur at least 10 times in the first sample dataset (He et al., 2012). The Supporting Information *Included features* presents an overview of all included features and specifies why features were included.

Training and testing

The training phase consisted of learning from a first subset of the sample to understand how the text is related to the outcome variables. In the exploratory phase, our first study, we tested a number of settings and chose those in which the models performed better for sample 1. First, we used the full UWES-9 scale for the classification of employees (Schaufeli et al., 2006). We did explore whether it would make sense to focus on the energetic component of work engagement by using only one of the dimensions of work engagement (vigor)—or a combination of vigor with exhaustion (a dimension of burnout; Schaufeli et al., 1996)—to classify employees. The argument here was that perhaps this would lead to better classification as studies indicate the energetic component of work engagement is more sensitive than the other components (Bakker & Demerouti, 2017). We decided to focus on the UWES-9 as most studies of work engagement use this full scale. Second, as Forman mentions in his critique on text classification feature selection methods, classifying minority classes is a pitfall for feature selection methods that use scoring methods based on outcome variables (2004). We therefore carefully decided on the criteria for classifying employees into high or low work engagement (the Supporting Information *Cutoff* explains this process in more detail). We explored relative (percentages) and absolute groups (e.g., scores below 2 and above 4 on a 5-point scale). We found the sample is unbalanced: More healthcare employees tend to be relatively high engaged. Therefore, we chose to use a relative, 10% cutoff. For the confirmatory phase, our second study, the same cutoff was set a priori.

Hence, to correctly classify whether an employee is high work engaged or not, we decided to select employees with a self-narrative of at least 20 words, who had a work engagement score in the highest 10% versus a work engagement score in the lowest 10% of each sample. For the first study, sample 1 ($N = 1119$) was split into a training set containing 80% of the self-narratives and a testing set containing the remaining 20% of the self-narratives. For the second study, we wanted to analyze how the features persist over time and across survey waves, so we used sample 1 in its entirety as the training dataset and sample 2 as the testing set. Table 4 reiterates the way the two studies were set up.

The purpose of splitting the sample into a training and testing set is to learn to recognize high versus low work engagement in the training set using the mentioned text features, after which the testing set can be used to assess its performance in recognizing high and low work engagement for previously unseen data. We used Random forest, a machine learning model that generates many decision trees trained and tested using resampling of the sample data. Each tree randomly samples a predefined number of features per split and decides based on the feature that best distinguishes the classes at that point in the tree. After all trees are built, Random forest calculates the best scoring features based on the classification scores per tree with and without the feature. If the trees classify worse when the feature is excluded from the tree, that means the feature has some explanatory power (Breiman, 2001; Kotsiantis et al., 2007).

Random forest allows for hyperparameter optimization, which is the process of modifying the model settings for better model performance. The Random forest model was applied using the randomForest package in R (Liaw & Wiener, 2002). The randomForest package allows for different settings for the nodesize, mtry, and ntree hyperparameters. For study 2, the Random forest hyperparameter optimization was done using the caret package by doing grid search for the minimum node size (nodesize) and the number of variables to sample as candidates for each split of a node (mtry) (Kuhn, 2008). A smaller nodesize hyperparameter value allows for more splits in the tree, resulting in a more complex tree (Breiman, 2001). Additionally, the number of trees (ntree) was optimized by building the Random forest model with 100, 250, 500, and 1000 trees. No noticeable improvement was found after 250 trees for any of the models, test set error rate was lowest with 250 trees, and highest with 500 and 1000 trees, whereas OOB error rate only marginally improved.

For comparison, in study 2, we also used a different classifier, Naive Bayes (Benoit et al., 2018; Lewis, 1998). Naive Bayes is a probabilistic machine learning algorithm that assumes features are independent of each other. The algorithm is well suited for high-dimensional datasets such as text data because it is efficient due to scaling linearly with the number of predictors and data points. Despite the assumption of independence often being violated, Naive Bayes tends to deliver robust and accurate classification. Naive Bayes, Random forest, and other approaches such as support vector machines and logistic regression are commonly used in text mining classification problems. In data science, there is no consensus on the best method as it depends on the data at hand. This means that in practice researchers use a variety of algorithms based on the conditions for the used case and pick the best performing algorithm.

TABLE 4 Samples and studies.

	Study 1	Study 2
Type of study	Exploratory	Confirmatory (with hypotheses)
Goal	Explore which text features explain work engagement in sample 1	Test the extent to which the text features persist over time (between samples 1 and 2)
Used training set	80% of sample 1 ($N = 895$)	Sample 1 ($N = 1119$)
Used testing set	20% of sample 1 ($N = 224$)	Sample 2 ($N = 894$)

The model results are evaluated primarily using a confusion matrix, the accuracy score, and *P*-value for accuracy score compared with the no-information-rate (NIR), which is an accuracy value that always predicts the most frequently occurring class in the dataset (Kuhn, 2008). The Supporting Information *R script* presents the R code for our main analyses.

Interpreting

In the fourth step, we evaluated the text mining results by comparing them to the domain knowledge on work engagement. For study 1, we used an exploratory approach to interpretation, by comparing the exploratory results to the existing literature on work engagement after conducting the analysis. We developed several hypotheses that explicated the main themes emerging from the analysis of study 1. We use both our observations (our data from study 1) and potential explanations in the existing theories in the literature to inform our hypotheses. For study 2, we used a confirmatory approach to interpretation, by assessing the hypotheses formulated in study 1. These hypotheses guided our discussion in study 2 and enabled to assess whether the same features contribute to explaining work engagement over time (compare our approach to studies on scale development who also distinguish an exploratory and confirmatory phase, e.g., House et al., 2004).

RESULTS STUDY 1

In this section, we present our results in four steps. First, we described the groups of high- and low-engaged employees in the sample. Second, we counted all the features that we analyzed in the self-narratives of these employees, and we assessed whether features are significantly more frequently observed among high- or low-engaged employees. Third, we tested whether the features can be used in a Random forest model that can correctly classify employees into high or low work engagement (using the training and testing approach, as described in our methods). Fourth, we presented the features that contribute most to the accuracy of the model, as these features indicate best how self-narratives of high- and low-engaged employees differ.

First, there are 5591 respondents who answered the text mining question with 20 words or more. The mean number of characters in the self-narratives was 667.24 (SD = 442.01), and the mean number of words was 113.31 (SD = 76.21). The mean work engagement score was 3.87 (SD = .63). Table 5 presents the respondents that are in the highest and lowest 10% of work engagement scores. The variance within the lowest 10% is notably larger than the variance in the highest 10%. This shows that our "lowest 10 percent" group is a varied of group employees ranging from very low on work engagement to moderately engaged.

Second, we counted the features in the groups of employees. The Supporting Information *Significant features* presents all features that are observed significantly more frequently in self-narratives of either high- or low-engaged employees. When we present the most contributing features in Table 7, we use the information from this step to indicate which feature is observed significantly more among high- or low-engaged employees.

Third, we employed Random forest to test whether these features can be used in a model to classify employees in the highest 10% or lowest 10% of work engagement. Table 6 presents the results of four models: unigrams, bigrams, psychological features, and linguistic features. Table 6 first indicates how many features contributed to the models. Next, it shows how the

TABLE 5 Highest versus lowest 10% scores on work engagement.

	No. of self-narratives	Mean score	Median score	Min. score	Max. score
Highest 10%	559	4.87	4.89	4.67	5
Lowest 10%	560	2.60	2.67	1	3

TABLE 6 Results Random forest for unigrams, bigrams, psychological and linguistic features.[a]

		Model 1: unigrams		Model 2: bigrams		Model 3: psychological features		Model 4: linguistic features	
No. of features		156		24		16		6	
Confusion matrix test set (actual/predicted)									
TP	FP	57	35	96	84	64	53	73	44
FN	TN	50	82	11	33	37	70	47	60
Model statistics									
OOB		32.63%		42.01%		42.23%		47.60%	
Accuracy		62.05%		57.59%		59.82%		59.38%	
NIR		52.23%		52.23%		52.23%		52.23%	
P-value (Acc > NIR)		0.002		0.062		0.013		0.019	

[a]The confusion matrix presents the numbers for respondents that score low on work engagement which our model got right (TP), respondents that score low on work engagement which our model got wrong (FN, a type 2 error), respondents that score high on work engagement which our model got wrong (FP, a type 1 error), and respondents that score high on work engagement which our model got right (TN).

models performed. We find that unigrams score best with a 62% accuracy score, whereas the other models have lower accuracy scores (for bigrams, 58%; for psychological features, 60%; and for linguistic features, 59%). Considering that a model based on randomization would have an accuracy score of 50% (an equal chance of true or false classification), we find that all models classify into high or low work engagement better than random. The model with the unigrams appears the most successful.

Fourth, we analyzed the features that best classify into high or low work engagement in the different models. For that, we assessed the discriminatory value of the features based on the mean decrease in accuracy of the models when a specific feature is excluded. The Supporting Information *Feature importance* presents the importance of all the features in the models. Table 7 presents the (most) strongly contributing features. For the unigrams and bigrams, we translated the features from Dutch into English and provided them with a common stem to ease interpretation. For the psychological and linguistic features, the feature categories are presented, and if applicable, the overarching category is presented between parentheses. The Supporting Information *Included features* presents more information on the content of these categories (so does Pennebaker et al., 2015).

Table 7 should be interpreted as follows: The bigram "goes well" contributes most to the accuracy of the bigrams model, and this bigram is significantly more present among employees with high work engagement. In contrast, the bigram "from house," which is the second most contributing feature, is significantly more counted among employees with low work engagement. Likewise, we find that the psychological feature "positive emotion," an LIWC dictionary

TABLE 7 Contribution of features to the models.[a]

	Model 1: unigrams[b]	Model 2: bigrams	Model 3: psychological features	Model 4: linguistic features
No. of contributing features				
	111	19	9	2
Most strongly contributing features				
1	few/little	**goes well**	**Positive emotion**	**3rd person plural**
2	specific	from house	Anger (Negative emotion)	**1st person plural**
3	**nursery home**	**past month**	Power (Drives)	-
4	**family**	high work pressure	**Social processes**	-
5	**good**	unsafe feeling	Present focus (Time orient.)	-
6	workplace	**allowed come**	Negative emotions	-
7	**suspected**	hours per	**Affiliation (Drives)**	-
8	**see**	meter distance	**Reward (Drives)**	-
9	high	**we go**	Work (Personal concerns)	-
10	**allowed**	usual work	-	-
11	**listening**	**colleague s**	-	-
12	face-to-face	**we good**	-	-
13	crisis	come work	-	-
14	**super**	**work we**	-	-
15	**caregiver**	**now then**	-	-
16	burdened	colleague does	-	-
17	**cohort**	**very good**	-	-
18	pension	direct contact	-	-
19	information	**our resident**	-	-
20	**talking**	-	-	-

[a]Features that are present significantly more among high-engaged employees are in bold, features present significantly more among low-engaged employees are in normal font.
[b]The features presented in this column are the 20 most strongly contributing features and have a chi^2-value of 3.7 or higher.

with words like "safe," "trust," and "beloved," contributes most to the accuracy of the psychological features model and is significantly more counted among high-engaged employees. In contrast, the feature "anger," a subdictionary of negative emotions, with words like "aggression," "stupid," and "fight" and the second most contributing to this model, is significantly more present among employees low in work engagement. In Section 5, we interpret what themes are presented in the features and how this relates to the extant literature.

DISCUSSION OF STUDY 1

For study 1, we used an explorative approach to assess whether we can explain work engagement through text mining. We found that models with unigrams, bigrams, psychological features, and linguistic features can correctly classify healthcare employees into high or low work engagement with an accuracy of up to 62% (for the unigrams). Whether we will find similar results in the next study depends on two aspects of the features. First, it will depend on the translatability of the type of feature. Herein, we may expect differences between the types of features we use. A methodological explanation for the success of unigrams in this study is that the high number of features may allow for more discrimination between the groups. However, for study 2, the question is whether unigrams will translate over samples. Potentially, psychological and linguistic features will explain better across samples than unigrams or bigrams as they use dictionary approaches that measure underlying meanings and categories of words rather than the specific words themselves (Tausczik & Pennebaker, 2010).

Second, whether we find the same result in study 2 will depend on the translatability of the content of the feature. If we assume that healthcare employees' work engagement can be explained by factors that are time-insensitive, our models should perform similarly. We will therefore explore whether we observe a few grand themes among the features that contribute to our models. As the process of feature selection was partially data driven, there are many features, and not all are readily interpretable or categorizable. However, across the models, three prominent feature themes emerge from the self-narratives, which we named emotions (24 features), crisis (35 features), and affiliation (19 features) (the Supporting Information *Feature importance* presents the coding). Below, we introduce these themes, compare them with the extant literature, and formulate hypotheses for study 2.

First, the models include strongly contributing features that address the positive or negative emotions in the self-narratives. For the unigrams, we find positive emotion words to be related to high-engaged employees (e.g., good) and negative emotion words to be related to low-engaged employees (e.g., burdened). For the bigrams, we find positive emotion word combinations to be related to high-engaged employees (e.g., goes well and very good) and negative emotion word combinations to be related to low-engaged employees (e.g., unsafe feeling). For the psychological features, we find positive emotions to be related to high-engaged employees, and anger and negative emotions to be related to low-engaged employees. These findings support the conceptualization of work engagement as a "positive motivational state" (Bakker et al., 2014, p. 389) as well as the finding that positive emotions are positively related to work engagement (Ouweneel et al., 2012). Additionally, employees who experience job strain are less able to regulate their emotions (Bakker & Costa, 2014), and scholars have suggested that emotional instability may be a personal demand that affects work engagement (Lorente Prieto et al., 2008). Hence, our first hypothesis is as follows:

> **H1.** Referring to positive emotions contributes to explaining high work engagement, whereas referring to negative emotions contributes to explaining low work engagement.

Second, the models include features that refer to the crisis during which the study was conducted, the COVID-19 pandemic. For the unigrams and bigrams, we find references to the crisis related to low-engaged employees (e.g., high work pressure, meter distance, usual work, direct contact; face-to-face [contact], crisis) and references to the absence of the crisis to be related to high-engaged employees (e.g., family, allowed; goes well, allowed [to] come). A side note here is that we indicated we have reason to expect that unigrams and bigrams translate less well across samples. Nevertheless, the features echo an emerging stream of literature that shows the pandemic may in many cases have deteriorated work engagement (Kniffin et al., 2021) and other aspects of well-being (Van Roekel et al., 2021; Wang et al., 2021). For many healthcare employees, COVID-19 caused higher stress levels and other negative health outcomes (e.g., Shreffler et al., 2020). There is an emerging literature on the effects of a crisis in the context of JD-R theory. Demerouti and Bakker (2023) argue the COVID-19 crisis has increased job demands. Besides, a crisis also tends to make resources scarce. They propose that during a crisis, employees who experience manageable job demands (and high job resources) will maintain higher engagement than employees who experience high job demands (and low job resources). At the same time, changes in engagement are likely not only caused by individual demands and resources but by a more complex interplay of individual and higher-level factors. In sum, our second hypothesis is as follows:

H2. Referring to a crisis contributes to explaining low work engagement, whereas referring to a normal work context contributes to explaining high work engagement.

Third, the models include features that refer to affiliation and social connection, which appears to explain high work engagement. For the unigrams, we find references to social contact (e.g., listening and talking) to be related to high-engaged employees. For the bigrams, we find multiple plural references to be related to high-engaged employees (e.g., we go, work we, and our resident). For the psychological features, we find social processes (e.g., "talk" and "love") and affiliation (e.g., "friend" and "social") to be related to high-engaged employees. And for the linguistic features, using 1st and 3rd person plural is positively related to high-engaged employees (in addition, the Supporting Information *Significant features* shows that low-engaged employees use significantly more singular forms, but these do not contribute to the models). This supports studies that show the importance of affiliation for work engagement. First, experiencing social support is an important predictor of work engagement, and in turn, employees who are engaged offer more social support (Freeney & Fellenz, 2013). Another study explained how especially new employees' work engagement is highly affected by socialization in the organization (Saks & Gruman, 2018). Additionally, scholars have studied the concept of teamwork engagement, which indicates that work engagement is not merely an individual process but also part of a team process (Costa et al., 2014). Work engagement is contagious, and employees can collectively experience high levels of work engagement (Bakker, 2022; Bakker et al., 2016). Therefore:

H3. Referring to affiliation contributes to explaining high work engagement.

Having defined our hypotheses, we submitted a preregistration at the Open Science Framework that described the hypotheses as well as the plan of analysis. For study 2, our primary aim is to select the features of study 1 and to assess whether these features contribute strongly to the models in study 2. We will evaluate the success of these features both in terms of specific features as well as the themes that they represent (as referred to in the hypotheses). The next section describes the results of study 2.

RESULTS STUDY 2

In study 2, we repeated the analysis with both samples. Again, we present the results in four steps: We described the employees in the sample, we counted all features and tested whether

TABLE 8 Highest versus lowest 10% scores on work engagement.

	No. of self-narratives	Mean score	Median score	Min. score	Max. score
Highest 10%	447	4.84	4.78	4.67	5
Lowest 10%	447	2.46	2.56	1.11	2.89

features are significantly more frequently observed among high- or low-engaged employees, we built the Random forest model, and we presented the features that contribute most.

First, we already introduced the first sample above. In the second sample that we add in this analysis, a total of 4470 respondents answered the text mining question (20 words or more). The mean number of characters in the self-narratives was 583.89 (SD = 397.33), and the mean number of words was 98.03 (SD = 67.99). The mean work engagement score was 3.79 (SD = .66). Again, we selected the respondents that are in the highest and lowest 10% of work engagement scores (Table 8). Like the first sample, the variance within the lowest 10% is notably larger than the variance in the highest 10%. Average scores also appear to be slightly lower for both groups compared with the first sample.

Second, we counted the features. The Supporting Information *Significant features* presents all features that are observed significantly *more* in self-narratives of either high- or low-engaged employees. When we present the most contributing features in Table 10, we use the information from this step to indicate which feature is observed significantly more among high- or low-engaged employees.

Third, we employed Random forest to test if these features can be used in a model to correctly classify employees in the highest 10% or lowest 10% of work engagement. Table 9 presents the models and shows that, compared with study 1, all but one model performed worse. The unigrams, bigrams, and linguistic features performed worse (accuracy scores of 52%, 53%, and 54%) and barely outperformed a random model. However, the model with psychological features still has an accuracy score of 60%. For comparison, we also conducted study 2 using the Naive Bayes classifier rather than Random forest. We find that for unigrams the results improve much (from 52% to 64%), whereas for the other features, the results are only slightly better (bigrams: 56% instead of 53%; psychological features: 61% instead of 60%; linguistic features: 55% instead of 54%). Supporting Information *Naive Bayes* presents all results for Naive Bayes.

Fourth, we analyzed the features that best explain high or low work engagement in the different models. For that, we assessed the discriminatory value of the features based on the mean decrease in accuracy of the models when a specific feature is excluded. Table 10 presents the (most) strongly contributing features that were also significantly more present among either high- or low-engaged employees (the Supporting Information *Feature importance* presents the overview of all features). Notably, the total amount of contributing features decreased drastically for the unigrams and bigrams, indicating that many features that were used in the first sample were not used in the second sample. In the following discussion section, we compare the features to those of study 1.

Additional analysis demographics

We conducted additional analyses to investigate how the sample demographics (gender, age, and healthcare branch, as described in Table 3) may affect the results. First, Supporting Information *Additional analysis* presents some significant differences in work engagement across gender, age, and healthcare branch in both samples. In all cases, effect sizes were small. Second, we explored

TABLE 9 Results Random forest for unigrams, bigrams, psychological and linguistic features.

		Model 1: unigrams		Model 2: bigrams		Model 3: psychological features		Model 4: linguistic features	
No. of features		156		24		16		6	
Confusion matrix test set (actual/predicted)									
TP	FP	313	294	383	356	295	202	313	281
FN	TN	134	153	64	91	152	245	134	166
Model statistics									
Accuracy		0.5213		0.5302		0.604		0.5358	
95% CI		0.4879–0.5544		0.4969–0.5633		0.5711–0.6363		0.5025–0.5689	
NIR		0.5		0.5		0.5		0.5	
P-value (Acc > NIR)		0.108		0.038		<0.001		0.0175	
Kappa		0.0425		0.0604		0.2081		0.0716	
McNemar's test *P*-value		<0.001		<0.001		0.009		<0.001	
Sensitivity		0.7002		0.8568		0.6600		0.7002	
Specificity		0.3423		0.2036		0.5481		0.3714	
Pos Pred value		0.5157		0.5183		0.5936		0.5269	
Neg Pred value		0.5331		0.5871		0.6171		0.5533	
Prevalence		0.5000		0.5000		0.5000		0.5000	
Detection rate		0.3501		0.4284		0.3300		0.3501	
Detection prevalence		0.6790		0.8266		0.5559		0.6644	
Balanced accuracy		0.5213		0.5302		0.6040		0.5358	
Hyperparameter values									
Nodesize		1		5		10		5	
Mtry		4		2		5		1	

whether demographics play a role in explaining work engagement. We used the DALEX package (Biecek, 2018) to analyze how gender and age relate to the text features in explaining work engagement. The results (in Supporting Information *Additional analysis*) show that, next to the text features, gender and age contribute to the models. This suggests that gender and age of healthcare employees contribute to explaining work engagement and that this may partially confound the effects in our main analysis. A next step in future research would therefore be to add these to the analyses. We discuss further implications in the discussion section.

GENERAL DISCUSSION

In this article, we aimed to explain work engagement by analyzing self-narratives through text mining. We compared unigrams, bigrams, psychological features, and linguistic features. After

TABLE 10 Contribution of features to the models.[a]

	Model 1: unigrams[b]	Model 2: bigrams	Model 3: psychological features	Model 4: linguistic features
No. of contributing features				
	39	4	8	2
Most strongly contributing features				
1	pension	**goes well**	**Social processes**	Negations
2	**good**	direct contact	Power (Drives)	**1st person plural**
3	few/little	**we good**	**Positive emotion**	-
4	unsafe	home work	**Reward (Drives)**	-
5	workplace	-	Negative emotions	-
6	**nurtured**	-	Future focus (Time orient.)	-
7	work pressure	-	Work (Personal concerns)	-
8	**resident**	-	**Affiliation (Drives)**	-
9	management	-	-	-
10	**nice**	-	-	-
11	**happily**	-	-	-
12	manager	-	-	-
13	**again**	-	-	-
14	free	-	-	-
15	whereby	-	-	-
16	**our**	-	-	-
17	insufficient	-	-	-
18	leave	-	-	-
19	**sad**	-	-	-
20	unrest	-	-	-

[a]Features that are present significantly more among high-engaged employees are in bold, features present significantly more among low-engaged employees are in normal font.
[b]The features presented in this column are the 20 most strongly contributing features and have a chi^2-value of 3.9 or higher.

the explorative approach in study 1, for study 2, we used a confirmatory approach to assess whether the same text features, dependent on both the type of feature and content of the features, can explain work engagement across two samples. From both studies, we deduce three main findings that we want to highlight. First, psychological features can correctly classify healthcare employees into high or low work engagement with 60% accuracy across samples. Second, the features that contribute to the classification partly confirm the literature on the JD-R theory and work engagement. Third, the features also unlock new insights by extending and questioning work engagement theory.

First, we find that the model with psychological features explained work engagement best in both studies, with 60% accuracy. In the first study, unigrams generated the best model (62% accuracy), but the unigrams performed worse in the second study (52% accuracy). This indicates that dictionary approaches, which measure underlying meanings and categories of words, have more success in explaining work engagement than bag-of-word approaches using specific words

(Tausczik & Pennebaker, 2010). A likely explanation is that although employees may write about similar topics, they may use different words.

Second, some of the features that contribute to the classification partly confirm the extant literature on antecedents and outcomes of work engagement. Based on study 1, we proposed three hypotheses, supported by the literature, regarding prominent features that explained work engagement. In evaluating our hypotheses, we focus on the model with psychological features, as this is the only model that performed consistently. Drawing conclusions from a model that does not outperform a random model would not be appropriate (we will pay some attention to the model with linguistic features as it still performs slightly better than random). First, we expected that referring to positive emotions contributes to explaining high work engagement, whereas referring to negative emotions contributes to explaining low work engagement (H1). This hypothesis is confirmed in study 2 because, again, positive emotions were more present among high-engaged employees and negative emotions were more present among low-engaged employees. Second, we expected that referring to a crisis contributes to explaining low work engagement, whereas referring to a normal work context contributes to explaining high work engagement (H2). This hypothesis was only based on unigrams and bigrams that referred to the COVID-19 crisis. These models were not able to explain work engagement in the second study. Therefore, Hypothesis 2 was not supported. Third, we expected that referring to affiliation contributes to explaining high work engagement (H3). This hypothesis is confirmed too, because again referring to social processes and affiliation explains high work engagement. Besides the hypotheses, three other psychological features contributed across two samples: High-engaged employees referred significantly more to rewards (a dictionary with words like "benefit," "bonus," and "promotion"), and low-engaged employees referred significantly more to power (words like "manager," "attack," and "dependent") and work concerns (words like "job," "burden," and "junior").

Our third finding is that text mining unlocks new insights that extend or question common findings in the literature. First, we are able to uncover that work engagement is related to linguistic patterns. Mainly, across two samples, employees with high work engagement use more first-person plural (e.g., "we" and "our") than employees with low work engagement. Second, some findings are puzzling and allow to question the literature. For example, in the first study, there are multiple unigrams that refer to management, and across two samples, there is a psychological feature that refers to power. What is striking is that these features all contribute to explaining low work engagement, suggesting employees who are low engaged tend to mention their managers more. We also observe features that refer to certain subgroups of employees. For example, in both samples, the unigram "retirement" contributes to explaining low work engagement. Exploratory analyses of self-narratives that include this unigram suggest that these are employees who are close to retirement. And in sample 1, the unigram "caregiver" contributes to explaining high work engagement. Exploratory analyses suggest that these employees are voluntary caregivers besides their regular work. We should be careful to interpret these exploratory findings, and we provide potential explanations below.

Scientific and practical implications

Our findings have multiple implications. Regarding implications for theory, our study furthers the understanding of work engagement as a theoretical concept. First, text mining enables validation of findings in the extant literature and complements these findings with rich context due to a large-scale analysis of self-narratives. As we explained in our theory section, antecedents of work engagement are often studied within the JD-R theory, a theory within organizational psychology that explains how job characteristics affect employees through a dual process. Job resources foster a motivational process leading to positive outcomes like work engagement, whereas hindrance job demands cause a health impairment process and diminish the positive effects of job resources on work engagement (Bakker & Demerouti, 2017; Schaufeli et al., 2009; Van Veldhoven et al., 2020). Our results confirm this duality because the features describe resources and demands. The features that high-engaged employees refer more often to, like affiliation and rewards, are often job resources (Bakker, 2022; Bakker et al., 2014). For example, experiencing social support positively affects work engagement (Freeney & Fellenz, 2013). Even more so, work engagement can be a truly contagious process transferring between employees (Bakker et al., 2016), and even from employees to partners (Bakker et al., 2005) and home life (Culbertson et al., 2012). Likewise, positive emotions, another feature more present among high-engaged employees, can be considered personal resources (Ouweneel et al., 2012). Contrarily, the features that low-engaged employees refer to, like power and work concerns, tend to be job demands. Power refers to words describing hierarchy or dependency, with words like "manager," "attack," and "dependent." This resembles studies that have shown that abusive supervision or bullying is negatively related to work engagement (Einarsen et al., 2018; Wang, Hsieh, & Wang, 2020), and may also point to the absence of autonomy, an important resource and antecedent of work engagement (Christian et al., 2011). Finally, negative emotions, another feature more present among low-engaged employees, may suggest that emotional instability be regarded as a personal demand that affects work engagement (Lorente Prieto et al., 2008).

Second, whereas some findings confirm the duality of JD-R theory, we also found remarkable linguistic patterns that extend it and relatively unexplored antecedents of work engagement that question it. These findings can increase our understanding of work engagement and how it is theorized and measured. First, the finding on linguistic differences between high- and low-engaged employees uncovers a new research area that may focus on work engagement markers within speech or writing. Until now, studies have mostly focused on linguistics in more clinical concepts, like schizophrenia (Franklin & Thompson, 2005). Studying linguistic patterns may increase our understanding of work engagement, especially in the context of diary studies, as these studies allow employees to provide unstructured data on a regular basis (Ouweneel et al., 2012; Zampetakis, 2022). Specifically, the finding that high-engaged employees use more first-person plural is a tangible indication of the social and contagious nature of work engagement (Bakker, 2022). Second, in the self-narratives, low-engaged employees more often referred to their managers. This finding is puzzling. The literature shows that managers can have important, positive influences on employee well-being and often finds positive effects of "good" leadership styles or behaviors on work engagement (Decuypere & Schaufeli, 2020; Tummers & Bakker, 2021). In contrast, managers can also have negative influence, when they bully or execute abusive supervision (e.g., Barnes et al., 2015). Our results suggest that employees are more likely to mention managers if they are a negative influence. A potential explanation is that positive behaviors are more seen as a self-evident part of a managers' role (Toegel et al., 2013). COVID-19 has been a tremendous leadership challenge (Graham & Woodhead, 2021), and especially employees who experienced failing leadership may have wanted to mention this in their self-narratives. In any case, the results suggest a vital role for managers in fostering employee work engagement (Freeney & Fellenz, 2013). Third, some findings beg for further research. For example, a recent study suggests that "mental retirement" among older employees is non-existent (De Wind et al., 2017). At the same time, in our study, employees who were low in engagement more often referred to retirement. One explanation is that working during COVID-19 has been especially burdensome for older employees (Van Roekel et al., 2021). This emphasizes the need for interventions that support older employees in the workplace (Söderbacka et al., 2020). In contrast, the finding that high-engaged employees refer more to being a voluntary caregiver besides their work points to another avenue in which work engagement may affect home life and cause citizenship behavior (Culbertson et al., 2012; Xanthopoulou et al., 2008).

Our main methodological contribution is that, to our knowledge, this is the first study that succeeds in explaining work engagement by text mining self-narratives. The best-scoring model in the first sample uses unigrams (62% accuracy), and the best-scoring model across samples uses psychological features (60% accuracy in the second study). We argue that our study indicates that, for work engagement, classification by text mining cannot easily replace structured forms of data analysis as it is not precise enough yet. Nevertheless, there are multiple avenues in which text mining could support and complement more traditional data analysis. First, it could validate the relative importance of antecedents and outcomes of work engagement. For example, if a relationship between work engagement and another concept, for example, empowering leadership, is analyzed, additional text mining of open questions could indicate differences in the way employees in high or low categories of work engagement discuss their managers (e.g., Tuckey et al., 2012). Text mining could also perform a supporting role by being used in the validation of scales, for example, by analyzing what words employees use to describe being engaged at work. Scales that use this as input for wording may be more ecologically valid (Kobayashi et al., 2021). Text mining could also complement structured forms of data analysis by using it as an exploration of what topics and concepts are associated with work engagement but may have received little attention in the literature. Besides, in situations where lengthy surveys are not preferred, text mining enables efficient analysis of an open question (Jurafsky & Martin, 2017; Kobayashi et al., 2021).

Finally, and this is both a methodological and practical implication, text mining may present a new avenue for in vivo assessment of work engagement. Now that this study has found that survey-based self-narratives explain work engagement to some extent, future research could use existing data to attempt to do the same. Albeit for scientific or managerial purposes, existing texts (like shared diaries or intranet posts) or other forms of unstructured data within organizations may very well allow for the screening and identification of employees whose work engagement is challenged (e.g., Day et al., 2007; He et al., 2012). In addition, studies could employ text mining techniques to present employees with a self-assessment of work engagement. Employees could, after providing a self-narrative, perhaps receive a comparative score and/or a personalized suggestion, like talking to a confidant. By making assessment easier, text mining could perhaps be a preventive HR tool, if employee privacy is maintained and the interest of employees is put first. Besides, the exploration of the features that contribute to explaining work engagement may help employees, (HR) managers, and (healthcare) organizations to more quickly recognize and act upon challenges to work engagement. The features that turned out to be important may indicate resources where organizations should invest in, like guaranteeing adequate social support systems and stimulating social contact between employees. Likewise, organizations should pay attention to employees' emotional state. Gauging healthcare employee work engagement has become increasingly relevant since the COVID-19 crisis, which has been challenging especially among healthcare employees dealing with COVID-19 patients (Van Roekel et al., 2021).

Limitations

There are limitations to this study. First, the data we used present limitations. We compare self-narratives to work engagement scores within the same survey, which may lead to common source bias. Using two survey waves has increased the strength of our design. Still, future research could go beyond survey-based analysis by employing human coders (e.g., psychologists) to assess self-narratives. Likewise, our text mining data were survey-based and created specifically for this study. This somewhat limits the external generalizability of our findings when discussing opportunities for text mining of existing, unstructured data. It is common, however, to begin with manufactured data and then expand to pre-existing data after (e.g., He, 2013). Therefore, future research could use such pre-existing data like email, intranet, or social media messages to address this limitation (e.g., Pang et al., 2020). Finally, the survey did not use attention checks, which may be regarded as a limitation. Nevertheless, there is considerable discussion in the literature about their effectiveness and necessity. Recent findings suggest attention checks do not harm scale validity but removing those who fail attention checks often does not alter substantive analyses either (e.g., Gummer et al., 2021; Kung et al., 2018).

Second, there are limitations related to sample characteristics. The dataset is unbalanced because there are more employees who score high versus low on work engagement (high work engagement: $M = 4.87$ for sample 1 and $M = 4.84$ for sample 2; low work engagement: $M = 2.60$ for sample 1 and $M = 2.46$ for sample 2). This limitation indicates that our analysis strictly explains the differences between very high work engagement and work engagement lower than the midpoint (i.e., 3) of the scale. One explanation is that healthcare employees are generally high in work engagement, so the limited generalizability of our results may be more pronounced in sectors with lower work engagement, like manufacturing (Hakanen et al., 2019). Nevertheless, future research may explicitly include employees with low work engagement by, for example, targeting employees who intend to quit their jobs (e.g., in exit interviews).

Another limitation regarding sample characteristics is that our main analysis focused on text-based features and therefore ignored the role of demographics such as gender and age. However, the literature shows work engagement can vary depending on gender and age (although only to a limited extent; see, e.g., Schaufeli et al., 2006). Yet we also argued that gender and age may affect the results because of the particularities of our sample and, to some extent, the population of healthcare employees. Hence, we can expect gender and age to meaningfully relate to work engagement in our particular samples. Controlling for gender and age in additional analysis confirms that these variables do play a role. Although these findings should be taken into account, our goal was not to develop the best model to explain work engagement but the best fitting model with text features from a representative sample of healthcare employees. Having a fairly representative sample for healthcare is a strength of our study's generalizability within healthcare and comparable sectors but does limit generalization when it comes to sectors with different characteristics. With this restriction in mind, our results contribute to understanding what text-based features contribute to explaining work engagement. Future research may extend our findings by paying more attention to the role of demographics in text mining research and by repeating our methods in different contexts.

A final limitation regarding sample characteristics concerns differences in respondent characteristics between the two samples. In comparison with the first sample, the second sample contains fewer respondents who also wrote shorter texts and had a lower mean work engagement. A potential explanation for the difference in participation rates is respondent fatigue: Respondents may have been more motivated to provide a self-narrative when the request for such a narrative was newly introduced compared with when it was repeated. However, the drop in work engagement may also point to another explanation: As the COVID-19 crisis continued, healthcare employees were exposed to persistent job stress, which may have caused the decrease in general levels of work engagement between the 2020 sample (#1) and the 2021 sample (#2) (Kniffin et al., 2021; Van Roekel et al., 2021; Wang et al., 2021). Sample heterogeneity may have affected the translatability of text features across samples somewhat and may have decreased the reliability of the models. However, we did find that the samples were comparable when it comes to gender, age, and healthcare branch. Still, future research could attempt to collect samples with identical respondent characteristics to counter sample heterogeneity.

Third, limitations apply regarding the methods used. This study aimed to explore the possibility of text mining for work engagement classification using Random forest and Naive Bayes. Our results showed that, with Random forest, we were able to classify, but in study 2, Naive Bayes performed better than Random forest. This may inform future use of classifiers for text-based features. Yet there are more possibilities for future research and further optimization of the methodology. Other approaches, including statistical methods such as LASSO feature elimination and OLS regression, and machine learning methods such as support vector machines (SVMs), could prove better suited to the data at hand. This should be decided on a case-by-case basis depending on the data and project goals. For our project, we primarily used Random forest as it presents feature importance information, which allows us to understand what features contribute most to the models. Besides, compared with regression models, it is able to handle the high dimensionality of text data better. Additionally, compared with regression, Random forest is more robust to outliers. Finally, Random forest is also suited for nonlinear relationships and categorical variables. SVMs share some of the advantages of Random forest but are heavier and harder to interpret. In sum, following up on our study, researchers could employ a variety of methods to provide new insights into the uses of text mining.

Fourth, our results are promising but the models are not nearly 100% reliable. One of the reasons may be that the self-narratives were relatively short (compared with, e.g., He et al., 2012). Longer stories may lead to better explanations. Besides, there is the issue of the "middle 80%": We cannot readily make statements about all respondents in between the highest or lowest 10%. Our approach is a most likely-case scenario, if we do not find differences between these two groups, there most likely will not be any differences found for the 80%. If we do find differences between these two groups, these differences will most likely be more pronounced than the differences for the 80%. Our recommendation for future research is to look beyond binomial categorization. Different feature selection methods, data representations, or neural network approaches to text classification could improve model performance further. Likewise, taking inspiration from our approach to studying work engagement, scholars could expand our study and include different features to test their respective contributions to explaining work engagement. By doing so, scholars can continue to confirm, extend, and/or question the literature. For example, extension could take place by studying unexplored features. We also see opportunities to further question the literature if scholars find features that contribute more to the reliability of the models than the features that we studied, or if features suggest contrary relationships between work engagement and other variables compared with our results or established theories on work engagement. Besides, we explored a bag-of-words and dictionary approach to text classification for work engagement. This means that syntactic and contextual information is not taken into account. Modern approaches that focus on further understanding relationships in the text may help future research do enrich the analysis. Word embeddings such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) allow to keep syntactic information intact instead of considering each word as a standalone feature. Recently, approaches such as BERT (Devlin et al., 2018) and OpenAI GPT (Radford et al., 2018) take this even further by incorporating contextual knowledge in the model using pretraining.

Fifth, we only addressed the concept of work engagement. It remains a question of how other measures, like positive and negative affectivity, compare with our text mining approach. Future research could explore the comparative explanatory accuracy of such measures in the work context.

Finally, the most important limitation regarding the results is inherent to text mining: "text mining procedures in and of themselves cannot support causal inference (i.e. internal validity) unless the study design is such that, next to association, temporal precedence and isolation are also established" (Kobayashi et al., 2021, p. 148). We analyzed associations, not causal relationships.

CONCLUSION

In this paper, we aimed to introduce text mining as a methodological approach to study employee work engagement, and, more generally, text mining as a method in organizational research. Our study attempted to analyze work engagement, and the features that contributed to the models help explain what it means to be (or not to be) engaged in work. Text mining truly allows to assess the multidimensionality of a phenomenon (Balducci & Marinova, 2018), and so, like qualitative research, it offers a richer description of reality, but, like quantitative research, it is able to handle large amounts of data. In sum, text mining is an interesting and innovative approach that may be used to validate but also complement findings from studies with more structured approaches to studying work engagement.

CONFLICT OF INTEREST STATEMENT

The authors report that there is no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

To protect the potentially privacy-sensitive information in the self-narratives, data are not made publicly available. The authors present multiple supplementary materials, including the syntax.

ETHICS APPROVAL STATEMENT

The data collection for this study was reviewed and approved by the Faculty Ethical Review Committee of Utrecht University (nr. 2019-004).

ORCID

Henrico van Roekel https://orcid.org/0000-0002-7179-2741
Bernard P. Veldkamp https://orcid.org/0000-0003-3543-2164
Arnold B. Bakker https://orcid.org/0000-0003-1489-1847

REFERENCES

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-3223-4

Bakker, A. B. (2022). The social psychology of work engagement: State of the field. *Career Development International*, *27*(1), 36–53. https://doi.org/10.1108/CDI-08-2021-0213

Bakker, A. B., & Costa, P. (2014). Chronic job burnout and daily functioning: A theoretical analysis. *Burnout Research*, *1*, 112–119. https://doi.org/10.1016/j.burn.2014.04.003

Bakker, A. B., & Demerouti, E. (2008). Towards a model of work engagement. *Career Development International*, *13*, 209–223. https://doi.org/10.1108/13620430810870476

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and work engagement: The JD-R approach. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 389–411.

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2022). Job demands–resources theory: Ten years later. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*, 25–53. https://doi.org/10.1146/annurev-orgpsych-120920-053933

Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2005). The crossover of burnout and work engagement among working couples. *Human Relations*, *58*(5), 661–689.

Bakker, A. B., Rodríguez-Muñoz, A., & Sanz Vergel, A. I. (2016). Modelling job crafting behaviours: Implications for work engagement. *Human Relations*, *69*(1), 169–189. https://doi.org/10.1177/0018726715581690

Bakker, A. B., Tims, M., & Derks, D. (2012). Proactive personality and job performance: The role of job crafting and work engagement. *Human Relations*, *65*(10), 1359–1378. https://doi.org/10.1177/0018726712453471

Schaufeli, W. B., Bakker, A. B., & Van Rhenen, W. (2009). How changes in job demands and resources predict burnout work engagement and sickness absenteeism. *Journal of Organizational Behavior*, *30*, 893–917. https://doi.org/10.1002/job.595

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*(4), 557–590. https://doi.org/10.1007/s11747-018-0581-x

Barnes, C. M., Lucianetti, L., Bhave, D. P., & Christian, M. S. (2015). "You wouldn't like me when I'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement. *Academy of Management Journal*, *58*(5), 1419–1437.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Biecek, P. (2018). DALEX: Explainers for complex predictive models in R. *Journal of Machine Learning Research*, *19*(84), 1–5. https://jmlr.org/papers/v19/18-416.html

Boyd, C. M., Bakker, A. B., Pignata, S., Winefield, A. H., Gillespie, N., & Stough, C. (2011). A longitudinal test of the job demands-resources model among Australian university academics. *Applied Psychology*, *60*(1), 112–140. https://doi.org/10.1111/j.1464-0597.2010.00429.x

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

CBS. (2020). *StatLine Werknemers met een baan in de zorg en welzijn; persoonskenmerken regio*. Den Haag: Centraal Bureau voor de Statistiek. Retrieved from https://azwstatlinecbsnl/?dl=470D9#/AZW/nl/dataset/24016NED/table

CBS. (2022). *Medewerkers met een baan in de zorg en welzijn*. Den Haag: Centraal Bureau voor de Statistiek. Retrieved from: https://azwstatline.cbs.nl/#/AZW/nl/dataset/24062NED/table?ts=1660929219979

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

Christian, M. S., & Slaughter, J. E. (2007). Work engagement: A meta-analytic review and directions for research in an emerging area. In *Academy of management proceedings* (Vol. 2007, No. 1, pp. 1–6). Academy of Management. https://doi.org/10.5465/ambpp.2007.26536346

Costa, P. L., Passos, A. M., & Bakker, A. B. (2014). Team work engagement: A model of emergence. *Journal of Occupational and Organizational Psychology*, *87*(2), 414–436. https://doi.org/10.1111/joop.12057

Culbertson, S. S., Mills, M. J., & Fullagar, C. J. (2012). Work engagement and work-family facilitation: Making homes happier through positive affective spillover. *Human Relations*, *65*(9), 1155–1177. https://doi.org/10.1177/0018726712440295

Day, S., Christensen, L. M., Dalto, J., & Haug, P. (2007). Identification of trauma patients at a level 1 trauma center utilizing natural language processing. *Journal of Trauma Nursing*, *14*(2), 79–83. https://doi.org/10.1097/01JTN00002787922091382

De Wind, A., Leijten, F. R., Hoekstra, T., Geuskens, G. A., Burdorf, A., & Van Der Beek, A. J. (2017). "Mental retirement?" Trajectories of work engagement preceding retirement among older workers. *Scandinavian Journal of Work, Environment & Health*, *43*, 34–41. https://doi.org/10.5271/sjweh.3604

Decuypere, A., & Schaufeli, W. (2020). Leadership and work engagement: Exploring explanatory mechanisms. *German Journal of Human Resource Management*, *34*(1), 69–95. https://doi.org/10.1177/2397002219892197

Demerouti, E., & Bakker, A. B. (2023). Job demands-resources theory in times of crises: New propositions. *Organizational Psychology Review*, *13*(3), 209–236.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint. arXiv:1810.04805.

Einarsen, S., Skogstad, A., Rørvik, E., Lande, Å. B., & Nielsen, M. B. (2018). Climate for conflict management exposure to workplace bullying and work engagement: A moderated mediation analysis. *The International Journal of Human Resource Management*, *29*(3), 549–570. https://doi.org/10.1080/09585192.2016.1164216

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, *3*, 1289–1305.

Forman, G. (2004). A pitfall and solution in multi-class feature selection for text classification. In C. E. Brodley (Ed.), *Machine learning, proceedings of the twenty-first international conference (ICML 2004)* (p. 38). Banff, Canada: ACM. https://doi.org/10.1145/1015330.1015356

Franklin, C. L., & Thompson, K. E. (2005). Response style and posttraumatic stress disorder (PTSD): A review. *Journal of Trauma & Dissociation*, *6*(3), 105–123. https://doi.org/10.1300/J229v06n03_05

Freeney, Y., & Fellenz, M. R. (2013). Work engagement job design and the role of the social context at work: Exploring antecedents from a relational perspective. *Human Relations*, *66*(11), 1427–1445. https://doi.org/10.1177/0018726713478245

Graham, R. N. J., & Woodhead, T. (2021). Leadership for continuous improvement in healthcare during the time of COVID-19. *Clinical Radiology*, *76*(1), 67–72. https://doi.org/10.1016/j.crad.2020.08.008

Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: Properties and implementation. *Sociological Methods & Research*, *50*(1), 238–264. https://doi.org/10.1177/0049124118769083

Hakanen, J. J., Ropponen, A., Schaufeli, W. B., & De Witte, H. (2019). Who is engaged at work?: A large-scale study in 30 European countries. *Journal of Occupational and Environmental Medicine*, *61*(5), 373–381. https://doi.org/10.1097/JOM.0000000000001528

Hänig, C., Schierle, M., & Trabold, D. (2010). Comparison of structured vs. unstructured data for industrial quality analysis. In *Proceedings of the world congress on engineering and computer science, WCECS 2010*, (Vol. 1, pp. 20–22). San Francisco, USA.

He, Q. (2013). *Text mining and IRT for psychiatric and psychological assessment*. University of Twente.

He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self narratives: A text mining approach. *Psychiatry Research*, *198*(3), 441–447. https://doi.org/10.1016/jpsychres201201032

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.) (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications.

Hudson, R. F., Lane, H. B., & Mercer, C. D. (2005). Writing prompts: The role of various priming conditions on the compositional fluency of developing writers. *Reading and Writing*, *18*(6), 473–495. https://doi.org/10.1007/s11145-004-7042-2

Jurafsky, D., & Martin, J. H. (2017). *Speech and language processing: An introduction to natural language processing 3*. Pearson Prentice Hall.

Kane-Frieder, R. E., Hochwarter, W. A., & Ferris, G. R. (2014). Terms of engagement: Political boundaries of work engagement–work outcomes relationships. *Human Relations*, *67*(3), 357–382.

Kao, A., & Poteet, S. R. (2007). *Natural language processing and text mining*. Springer Science & Business Media. https://doi.org/10.1007/978-1-84628-754-1

Kniffin, K. M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., & Vugt, M. V. (2021). COVID-19 and the workplace: Implications issues and insights for future research and action. *American Psychologist*, *76*(1), 63–77. https://doi.org/10.1037/amp0000716

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

Knight, C., Patterson, M., & Dawson, J. (2019). Work engagement interventions can be effective: A systematic review. *European Journal of Work and Organizational Psychology*, *28*(3), 348–372. https://doi.org/10.1080/1359432X.2019.1588887

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods*, *21*(3), 733–765. https://doi.org/10.1177/1094428117722619

Kobayashi, V. B., Mol, S. T., Vrolijk, J., & Kismihók, G. (2021). Text mining in career studies: generating insights from unstructured textual data. In *Handbook of research methods in careers*. Cheltenham: Edward Elgar Publishing.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3–24.

Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, *3*(3), 231–255. https://doi.org/10.1016/1060-3743(94)90018-3

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Kung, F. Y., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology. An International Review*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

La Bella, A., Fronzetti Colladon, A., Battistoni, E., Castellan, S., & Francucci, M. (2018). Assessing perceived organizational leadership styles through twitter text mining. *Journal of the Association for Information Science and Technology*, *69*(1), 21–31. https://doi.org/10.1002/asi.23918

Lesener, T., Gusy, B., Jochmann, A., & Wolter, C. (2020). The drivers of work engagement: A meta-analytic review of longitudinal evidence. *Work and Stress*, *34*(3), 259–278. https://doi.org/10.1080/02678373.2019.1686440

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In C. Nédellec & C. Rouveirol (Eds.), *Machine learning: ECML-98. ECML 1998. Lecture notes in computer science* (Vol. 1398). Springer. https://doi.org/10.1007/BFb0026666

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. https://CRANR-projectorg/doc/Rnews/

Lorente Prieto, L., Salanova Soria, M., Martínez Martínez, I., & Schaufeli, W. (2008). Extension of the job demands–resources model in the prediction of burnout and engagement among teachers over time. *Psicothema*, *20*, 354–360.

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

Mäkikangas, A., Feldt, T., Kinnunen, U., & Mauno, S. (2013). Does personality matter? Research on individual differences in occupational well-being. In A. B. Bakker (Ed.), *Advances in positive organizational psychology* (Vol. 1, pp. 107–143). Emerald. https://doi.org/10.1108/S2046-410X(2013)0000001008

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *NIPS'13: Proceedings of the 26th international Conference on Neural Information Processing Systems - Volume 2, Lake Tahoe, Nevada* (pp. 3111–3119). Red Hook, NY, USA: Curran Associates Inc.

Ministerie van OCW (2021). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2021*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap. Retrieved from: https://www.voion.nl/media/4089/trendrapportage-arbeidsmarkt-leraren-po-vo-en-mbo-2021.pdf

Ouweneel, E., Le Blanc, P. M., Schaufeli, W. B., & van Wijhe, C. I. (2012). Good morning good day: A diary study on positive emotions hope and work engagement. *Human Relations*, *65*(9), 1129–1154. https://doi.org/10.1177/0018726711429382

Pang, D., Eichstaedt, J. C., Buffone, A., Slaff, B., Ruch, W., & Ungar, L. H. (2020). The language of character strengths: Predicting morally valued traits on social media. *Journal of Personality*, *88*(2), 287–306. https://doi.org/10.1111/jopy.12491

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.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Doha, Qatar.

Perry, P. O. (2017). *Text corpus analysis R package version 0102*. Retrieved from http://corpustext.com

Porter, M. F. (2001). *Snowball: A language for stemming algorithms*. Retrieved from http://snowball.tartarus.org/

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. Preprint retrieved from https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

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

Rosenberg, S. D., & Tucker, G. J. (1979). Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry*, *36*(12), 1331–1337. https://doi.org/10.1001/archpsyc.1979.01780120061008

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly*, *27*(3), 446–455.

Saks, A. M., & Gruman, J. A. (2018). Socialization resources theory and newcomers' work engagement: A new pathway to newcomer socialization. *Career Development International*, *23*(1), 12–32. https://doi.org/10.1108/CDI-12-2016-0214

Schaufeli, W. B., & Bakker, A. B. (2010). Defining and measuring work engagement: Bringing clarity to the concept. *Work Engagement: A Handbook of Essential Theory and Research*, *12*, 10–24.

Schaufeli, W. B., & Bakker, A. B. B. (2022). Work Engagement. A critical assessment of the concept and its measurement. In W. Ruch, A. B. Bakker, L. Tay, & F. Gander (Eds.), *Handbook of positive psychology assessment* (pp. 273–295). Göttingen: Hogrefe Publishing.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, *66*(4), 701–716. https://doi.org/10.1177/0013164405282471

Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). The MBI-general survey. In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *Maslach burnout inventory manual* (3rd ed., pp. 19–26). Consulting Psychologists Press.

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of burnout and engagement: A confirmatory factor analytic approach. *Journal of Happiness Studies*, *3*, 71–92. https://doi.org/10.1023/A:1015630930326

Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2017). An ultra-short measure for work engagement. *European Journal of Psychological Assessment*, *35*(4), 577–591.

Schaufeli, W. B., Taris, T. W., Le Blanc, P., Peeters, M., Bakker, A. B., & De Jonge, J. (2001). Maakt arbeid gezond? Op zoek naar de bevlogen werknemer [Does work make people happy? In search of the engaged worker]. *Psycholoog*, *36*, 422–428.

Shreffler, J., Petrey, J., & Huecker, M. (2020). The impact of COVID-19 on healthcare worker wellness: A scoping review. *The Western Journal of Emergency Medicine*, *21*(5), 1059–1066. https://doi.org/10.5811/westjem.2020.7.48684

Shuck, B., Adelson, J. L., & Reio, T. G. Jr. (2017). The employee engagement scale: Initial evidence for construct validity and implications for theory and practice. *Human Resource Management*, *56*(6), 953–977.

Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *Joss*, *1*(3), 37. https://doi.org/10.21105/joss.00037

Soane, E., Truss, C., Alfes, K., Shantz, A., Rees, C., & Gatenby, M. (2012). Development and application of a new measure of employee engagement: The ISA engagement scale. *Human Resource Development International*, *15*(5), 529–547.

Söderbacka, T., Nyholm, L., & Fagerström, L. (2020). Workplace interventions that support older employees' health and work ability—A scoping review. *BMC Health Services Research*, *20*(1), 1–9.

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. https://doi.org/10.1177/0261927X09351676

Tims, M., Bakker, A. B., & Derks, D. (2015). Examining job crafting from an interpersonal perspective: Is employee job crafting related to the well-being of colleagues? *Applied Psychology*, *64*(4), 727–753. https://doi.org/10.1111/apps.12043

Toegel, G., Kilduff, M., & Anand, N. (2013). Emotion helping by managers: An emergent understanding of discrepant role expectations and outcomes. *Academy of Management Journal*, *56*(2), 334–357. https://doi.org/10.5465/amj.2010.0512

Tuckey, M. R., Bakker, A. B., & Dollard, M. F. (2012). Empowering leaders optimize working conditions for engagement: A multilevel study. *Journal of Occupational Health Psychology*, *17*(1), 15.

Tummers, L. G., & Bakker, A. B. (2021). Leadership and job demands-resources theory: A systematic review. *Frontiers in Psychology*, *12*, 722080. https://doi.org/10.3389/fpsyg.2021.722080

Van der Fels, I. (2020). *Resultaten Monitor Gezond werken 2020 tijdens de COVID-19 uitbraak*. Stichting IZZ. Retrieved from https://izz.nl/sites/default/files/downloads/5f6c8a51c0ef7_0.pdf

Van Roekel, H., van der Fels, I. M. J., Bakker, A. B., & Tummers, L. G. (2021). Healthcare workers who work with COVID-19 patients are more physically exhausted and have more sleep problems. *Frontiers in Psychology*, *11*, 625626. https://doi.org/10.3389/fpsyg.2020.625626

Van Veldhoven, M., Van den Broeck, A., Daniels, K., Bakker, A. B., Tavares, S. M., & Ogbonnaya, C. (2020). Challenging the universality of job resources: Why, when, and for whom are they beneficial? *Applied Psychology*, *69*(1), 5–29. https://doi.org/10.1111/apps.12211

Van Wijk, M. (2020). *Arbeidsmarktprofiel van zorg en welzijn*. CBS. Retrieved from: https://www.cbs.nl/nl-nl/longread/statistische-trends/2020/arbeidsmarktprofiel-van-zorg-en-welzijn?onepage=true

Van Wissen, L., & Boot, P. (2017). An electronic translation of the LIWC Dictionary into Dutch. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 703–715). Lexical Computing.

Wang, C. C., Hsieh, H. H., & Wang, Y. D. (2020). Abusive supervision and employee engagement and satisfaction: The mediating role of employee silence. *Personnel Review*, *49*(9), 1845–1858. https://doi.org/10.1108/PR-04-2019-0147

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied Psychology. An International Review*, *70*(1), 16–59. https://doi.org/10.1111/apps.12290

Wang, N., Zhu, J., Dormann, C., Song, Z., & Bakker, A. B. (2020). The daily motivators: Positive work events, psychological needs satisfaction, and work engagement. *Applied Psychology*, *69*(2), 508–537. https://doi.org/10.1111/apps.12182

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A grammar of data manipulation*. https://dplyrtidyverseorg https://githubcom/tidyverse/dplyr

Wickham, H., Hester, J., & Bryan, J. (2022). *Readr: Read rectangular text data*. https://readrtidyverseorg https://githubcom/tidyverse/readr

Wickham, H., Miller, E., & Smith, D. (2022). *Haven: Import and export 'SPSS' 'Stata' and 'SAS' files*. https://haventidyverseorg, https://githubcom/tidyverse/haven, https://githubcom/WizardMac/ReadStat

Wood, J., Oh, J., Park, J., & Kim, W. (2020). The relationship between work engagement and work–life balance in organizations: A review of the empirical research. *Human Resource Development Review*, *19*(3), 240–262. https://doi.org/10.1177/1534484320917560

World Health Organization. (2019). *Delivered by women, led by men: A gender and equity analysis of the global health and social workforce*. World Health Organization. https://apps.who.int/iris/handle/10665/311322. License: CC BY-NC-SA 3.0 IGO

Xanthopoulou, D., Bakker, A. B., Heuven, E., Demerouti, E., & Schaufeli, W. B. (2008). Working in the sky: A diary study on work engagement among flight attendants. *Journal of Occupational Health Psychology*, *13*, 345–356. https://doi.org/10.1037/1076-8998.13.4.345

Zampetakis, L. A. (2022). Employees' fear at work, job crafting, and work engagement on a daily basis: The case for fear of COVID-19. *Applied Psychology*, *72*(2), 535–558. https://doi.org/10.1111/apps.12388

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: van Roekel, H., Wigger, E. F. J., Veldkamp, B. P., & Bakker, A. B. (2023). What is work engagement? A text mining approach using employees' self-narratives. *Applied Psychology*, 1–32. https://doi.org/10.1111/apps.12501