DOI: 10.1002/hrm.22085

WILEY

SPECIAL ISSUE ARTICLE

# What you see is what you get? Measuring companies' projected employer image attributes via companies' employment webpages

**Christian P. Theurer**[1†] | **Philipp Schäpers**[2,6†] | **Andranik Tumasjan**[3] | **Isabell Welpe**[4] | **Filip Lievens**[5]

[1]Division Strategy and Organization, Technical University of Munich, Munich, Germany

[2]Department of Psychology, University of Münster, Münster, Germany

[3]Chair of Management and Digital Transformation, Johannes Gutenberg University Mainz, Mainz, Germany

[4]Chair of Strategy and Organization, Technical University of Munich, Munich, Germany

[5]Division of Organisational Behaviour and Human Resources, Singapore Management University, Lee Kong Chian School of Business, Singapore, Singapore

[6]Westfälische Wilhelms-Universität Münster, Germany

**Correspondence**
Christian P. Theurer, Chair of Strategy and Organization, Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany.
Email: christian.theurer@tum.de

Philipp Schäpers, Westfälische Wilhelms-Universität Münster, Germany.
Email: philipp.schaepers@googlemail.com

**Abstract**

Information on a company's employment webpage sends signals about the employer image the company intends to project to applicants. Nonetheless, we know little about the content of recruitment signals sent via company employment webpages. This study develops a method to measure companies' projected employer image attributes based on their employment webpages. Specifically, we analyze companies' projected employer image attributes by applying computer-aided text analysis (CATA) to the employment webpages of 461 Fortune 500 companies (i.e., more than 11,100 individual pages). Our results show that projected employer image attributes remain relatively stable over time. Moreover, we find relatively low levels of employer image differentiation between companies and between industries. Only a small group of companies (<20%) use distinct employer attribute signals to communicate their projected employer image. Finally, there is limited convergence between projected employer image attributes based on employment webpages and ratings on similar attributes on employer review websites. Generally, our results show that CATA is a viable method for assessing companies' projected employer image in the context of employer image management and engineering.

**KEYWORDS**
CATA, content analysis, employer branding, employer image, third party employment branding

## 1 | INTRODUCTION

According to a LinkedIn survey of recruitment trends (Schnidman, Hester, & Pluntke, 2017), 61% of HR officers view a company's employment webpage as one of the best strategies to start and conduct employer branding. Applicants also view company webpages as "one of the first places to look to learn what it is like to be a member of the organization and affiliate with those who currently work there" and "98.2% of respondents used employer websites in their job search" making it "the most frequently used source of information" (Banks et al., 2019, p. 480). Thus, employment webpages represent an excellent starting point for employer

†Christian P. Theurer and Philipp Schäpers contributed equally to this manuscript.

branding because companies can actively control and steer the information provided and consequently their intended image. Furthermore, they provide not only information related to job details but also serve as the first impression of a company's culture and DNA, signaling what it would be like for candidates to work for the organization. Consequently, the information presented on companies' webpages has been conceptualized as a key recruitment signal (Banks et al., 2019).

All of this explains why companies invest substantially in their employment webpage as a way to build and transmit the employer image they want to project to the outside world (Kane, Palmer, Phillips, & Kiron, 2017; Saini, Rai, & Chaudhary, 2014). We refer to this type of image as *projected image* (i.e., communicated [unique] characteristics of a company that reflect an employer's identity; Robertson & Khatibi, 2012). Besides company webpages (Williamson, King, Lepak, & Sarma, 2010), a company can also express its projected image via corporate advertising, brochures, and/or recruitment ads (Collins & Han, 2004; Collins & Stevens, 2002). All of these investments in projected employer image can be understood as stemming from a company's goal to actively control and steer its intended image.

Despite these investments in company webpages, little is known about the content of these websites, their stability over time, or their convergence with what it is like to actually work for the company. Most prior research on companies' webpages has focused on people's reactions to webpages and the effects of aesthetic webpage features (e.g., Cober, Brown, Keeping, & Levy, 2004; Cober, Brown, Levy, Cober, & Keeping, 2003). One reason for this lack of research is that, so far, no method has been developed to quantify employment website information as an important part of companies' projected employer image.

Therefore, the main objective of this article is to develop an efficient and unobtrusive method to measure and quantify companies' projected employer image attributes[1] via computer-aided text analysis (CATA). We apply CATA via automated coding and processing of large amounts of text from the company and employment webpages of US Fortune 500 companies to assess companies' projected employer image attributes. Our study addresses recent calls to provide a validated dictionary and a structured, efficient approach to investigate the content of employer image attributes based on large-scale text data from company webpages (see Banks, Woznyj, Wesslen, & Ross, 2018; Kobayashi, Mol, Berkers, Kismihok, & Den Hartog, 2018). Our approach also fits in with the wider trend of using text mining in organizational research (e.g., Guo, Li, & Shao, 2015; Hickman, Thapa, Tay, Cao, & Srinivasan, 2020; Schmiedel, Müller, & vom Brocke, 2019).

As a secondary objective, we demonstrate and illustrate the usefulness of this innovative approach by examining various questions related to promoting employer image attributes via webpages. That is, we investigate whether the quantified employer image attributes remain relatively stable over time, help companies stand out from their competitors, and converge with how these employer attributes are experienced by company insiders.

At a practical level, this study provides companies with an efficient and unobtrusive tool to assess their projected employer image attributes. The approach allows for comparing a company's projected employer image attributes to its competitors' projected employer image attributes. Similarly, the approach enables monitoring of the key attributes present in a company's projected image (and that of its competitors) over time. This also enables companies to sharpen their projected image and to test, control, and steer the attributes presented on their corporate website. Thus, our approach can also be used when revising or creating new website content. In an increasingly digital world with rapidly changing employee requirements, effective ways for companies to better understand their projected employer image and to position themselves as desirable employers have become a core strategic target (Cascio & Graham, 2016; Kane et al., 2017); this method might be one such novel, automated approach.

## 2 | BACKGROUND AND RESEARCH QUESTIONS

### 2.1 | CATA and projected employer image

Although there exists a rich tradition of content-analyzing traditional company-generated sources and channels such as mass media, annual reports, internal magazines, or mission statements (e.g., Chen & Meindl, 1991; Kabanoff, Waldersee, & Cohen, 1995; see also Arndt & Bigelow, 2000), research on the development of efficient methods to measure companies' projected/intended employer image based on presented webpage information remains scant (Duriau, Reger, & Pfarrer, 2007; Lombard, Snyder-Duch, & Bracken, 2002; Neuendorf, 2016).

So far, there does not exist a method for capturing the intended image attributes that employers want to project. Hence, this study relies on content analysis to capture projected employer image based on company and employment webpages (see also Nolan, Gohlke, Gilmore, & Rosiello, 2013). By applying content analysis, qualitative data can be converted through the systematic evaluation of the text. In particular, we use CATA to process and code large text-based data. Similar to human coding schemes and approaches, CATA builds on word, sentence, or paragraph usage in a text to systematically make inferences about the author's mental models and intentions (Carley, 1997; Morris, 1994). However, in comparison to human coding, which can be error-prone due to insufficient coder training or fatigue, CATA allows for processing large text-based data samples with both high speed and high reliability (i.e., high consistency in measuring the construct; Short, Broberg, Cogliser, & Brigham, 2010; Stevenson, 2001). Specifically, CATA provides adequate coder reliability when applying certain coding rules to a text (Weber, 1990). Another advantage of CATA is that the same text can be analyzed with different category schemes (i.e., dictionaries), which can be modified in case of errors or required changes (Weber, 1990).

## 2.2 | Underlying employer image framework

Employer image attributes can be categorized in numerous ways. Marketing-based brand image theory typically distinguishes between product-related (i.e., attributes directly related to the product or service), and non-product-related image attributes (i.e., external aspects related to the purchase or consumption of a product or service; Aaker, 1991; Keller, 1993). Translated to the employment context, these can be interpreted as job-related and nonjob-related employer image attributes. Lievens and Highhouse (2003) transferred this marketing-based brand image theory to the recruitment context, categorizing employer image into *instrumental* image (i.e., functional, utilitarian job, and organizational attributes) and *symbolic* image attributes (i.e., self-expressive organizational attributes; see also Gardner & Levy, 1955; Keller, 1993, 2011).

Instrumental attributes objectively describe the job and the employing company in terms of tangible, factual, and concrete associations that the employer has or does not have, thus allowing employees to maximize their benefits and minimize their costs (e.g., pay, bonuses, location, working hours; Lievens & Highhouse, 2003). However, employees' attraction to companies cannot be explained solely based on instrumental job and organizational attributes (Lievens & Highhouse, 2003). Therefore, symbolic employer image attributes depict the job and employing company in terms of intangible (non-job/company-related), subjective attributes that (potential) employees attribute to an employer (e.g., specific traits such as prestige, sincerity, or innovativeness; Lievens & Highhouse, 2003). Such traits allow employees to "express themselves, maintain their self-identity, or to increase their self-image" (Aaker, 1997; Lievens & Highhouse, 2003, p. 79; see also Highhouse, Thornbury, & Little, 2007). Many of these symbolic image attributes are based on company brand personality conceptualizations and comprise attributes such as sincerity, competence, or excitement (Slaughter, Zickar, Highhouse, & Mohr, 2004).

As a general thread running through this body of research, both instrumental and symbolic attributes were positively related to both attitudinal and behavioral outcomes, such as employer attractiveness or job pursuit intentions (e.g., Kausel & Slaughter, 2011; Lemmink, Schuijf, & Streukens, 2003; Lievens & Highhouse, 2003; Lievens, van Hoye, & Schreurs, 2005; Slaughter & Greguras, 2009; van Hoye, Bas, Cromheecke, & Lievens, 2013; van Hoye & Saks, 2011). Research revealed that symbolic attributes are almost equally effective across different groups of (potential) employees, whereas instrumental attributes explained the highest variance among actual applicants because they have specific information collection needs in a pre-employment phase (Lievens, 2007).

In sum, the instrumental-symbolic framework has been applied in a variety of contexts to put order to the myriad perceived image attributes. Therefore, it seems to be a good starting point for this study to apply the instrumental-symbolic framework as an overarching conceptual lens for measuring, analyzing, and structuring employer image attributes, although we do not claim that it exhaustively captures employer image in its entirety. Specifically, we acknowledge that company webpages include more information than only employer image attributes. Notably, company webpages include information about corporate social responsibility (CSR) statements and diversity and inclusion (D&I) policies. This information might also relate to people's attraction to the company (see Uggerslev, Fassina, & Kraichy, 2012). Therefore, we ran robustness analyses in which we addressed our first two research questions with a CATA of CSR statements and D&I policies as well.

## 2.3 | Assessing projected employer image attributes via CATA: Advantages

### 2.3.1 | Monitoring projected employer image attributes over time

As previously noted, company employment webpages represent a key company-controlled vehicle for firms to manage their employer brand. One cornerstone of employer branding is "consistency" (Barrow & Mosely, 2005). According to propositions underlying employer branding, companies need to have an enduring and temporally consistent image to increase their credibility and lower the risk of unmet expectations among (prospective) employees (Wilden, Gudergan, & Lings, 2010). Dineen, van Hoye, Lievens, and Rosokha (2019) refer to signal recurrence in this context. That is, there have to be repeated or multiple occurrences of a signal for the signal to be trustworthy. For instance, an annual award given to a company (e.g., Best Place to Work) is only recurrent if the company receives it over several years (Dineen & Allen, 2016). Otherwise, this designation will be interpreted as an exception and not attributed to the employer's image. Applied to the context of employment webpages, this means that companies are generally advised to keep their website information relatively consistent across time to solidify the image they project to the outside world. So far, little is known about the stability of company employment website information and projected employer image attributes. Once a dictionary is developed, CATA can be used for content-analyzing and quantifying companies' projected employer image attributes at multiple points in time. This kind of monitoring makes it possible to test whether projected employer image remains stable over time. Therefore, to illustrate this application domain of CATA, we posit the following research question:

> **Research Question 1.** *Are the projected employer image attributes that emerge based on a CATA of company employment webpages consistent over time?*

### 2.3.2 | Distinctiveness of projected employer image attributes

Apart from stability over time, employer branding also suggests that companies should strive for differentiation. In fact, employer branding is all about promoting a clear view of what makes an organization desirable *and* different from an employer (Backhaus & Tikoo, 2004; Moser, Tumasjan, & Cable, 2020; Moser, Tumasjan, & Welpe, 2017). Several studies have shown that distinctiveness is a core building block in brand management (e.g., Holt, 2004; Rossolatos, 2013; Thellefsen & Sørensen, 2007). Therefore, a key question is whether projected image attributes differ across companies. In addition, it is important to know which projected image attributes allow companies to be differentiated from each other. If such differences are not apparent, there is unused potential for companies to differentiate themselves from other companies and thus opportunities to further sharpen their corporate website profile toward a more distinctive employer brand.

Once a dictionary has been developed, CATA allows for quantifying the projected employer image attributes of a given company and its labor market competitors. Thus, similar to salary surveys, information from a CATA of projected employer image attributes can be used for benchmarking purposes and for building "brand intelligence" through examining how a company's projected employer attributes differ from those of labor market competitors or the industry in general. A CATA of projected employer image attributes can also be used to scrutinize whether different subsidiaries or branches of the same company emphasize different employer attributes on their respective webpages. This might indeed be a relevant application domain of CATA, because Banks et al. (2019) showed that employment information differed between domestic vs. international subsidiaries of multinational enterprises. In sum, to illustrate this application domain of CATA, we posit the following research question:

> **Research Question 2.** *Do projected employer image attributes that emerge based on a CATA of company employment webpages vary across companies and/or industries?*

### 2.3.3 | Convergence with third-party employer ratings

Besides stability across time and across-company differentiation, another capstone of employer branding is that employees (company insiders) should also "live" the brand being projected and promoted to company outsiders (e.g., jobseekers). In other words, projected employer image attributes should not remain a mere promise or a vision. Instead, these projected employer image attributes should exhibit at least moderate convergence with the view of company insiders (i.e., current and former employees; Tumasjan, Kunze, & Bruch, 2020). If this is the case and employees are living and delivering the brand, a company is typically regarded as having a "strong" employer brand (Barrow & Mosely, 2005).

In the last decade, one of the key trends in the digital recruitment arena has been the rise of employer review platforms such as Glassdoor, Indeed, Kanzhun, or Kununu (Etter, Ravasi, & Colleoni, 2019). Research shows that up to 52% of US job seekers read employer reviews before applying (Westfall, 2017). Employer review platforms allow current and former employees to voluntarily and anonymously submit online reviews about their employer via predefined questionnaires. The growing popularity of employer review platforms stems from the fact that current/former employees are company insiders who tend to be perceived as credible. Due to their first-hand experience in the company (e.g., Bone, 1995; van Hoye & Lievens, 2009), their reviews provide information on employer image attributes to the outside world that may potentially contradict their employer's official representation (i.e., projected employer image). That is why employer review websites are considered one form of "third-party employment branding". In contrast to company-controlled employment branding (e.g., via websites), third-party employment branding refers to employer-related communications that occur outside of company control (Dineen et al., 2019). To guarantee quality reviews and reduce idiosyncracies (e.g., overly positive or negative reviews; Dineen et al., 2019), many employer review platforms include safeguards[2] such as averaging out a company's review ratings across a large number of employees. In light of the above, it is not surprising that such employer reviews may affect potential employees' attitudes and intentions toward organizations as a place of work (Evertz, Kollitz, & Süss, 2019; Könsgen, Schaarschmidt, Ivens, & Munzel, 2018; Melián-González & Bulchand-Gidumal, 2016). So, some caveats notwithstanding, it is reasonable to assume that employer image information posted on an employer review platform (Glassdoor) can be used to gauge how current/former employees view and experience their company.

Currently, there is no knowledge about potential convergence between projected employer image and the information presented on employer review websites. Given that a CATA analysis of employment webpages results in quantifiable information, it can be subsequently linked to other information sources and databases. Thus, quantifying projected employer image attributes via CATA makes it possible to conduct a reality check by linking the CATA results to how (current/former) employees experience and view the company. In other words, it enables assessing how a company's projected brand (brand message, brand promise) compares to the brand as lived and experienced by (current/former) employees. Therefore, to illustrate this application domain of CATA, we posit the following research question:

> **Research Question 3.** *Are projected employer image attributes that emerge based on a CATA of company employment webpages related to employer review website ratings posted by former/current employees?*

## 3 | METHOD

### 3.1 | Sample

The sample consisted of the US Fortune 500 list of companies from 2013 (Fortune, 2014). Actual corporate and employment webpage data (text only) were collected in the period from May 3, 2014, to December 29, 2014. The Fortune 500, which comprises a list of the 500 largest

US corporations by total revenue in a fiscal year, is published on an annual basis by Fortune magazine. With a total of around $12.8 trillion in revenue and $1 trillion in profits, Fortune 500 companies represented around two-thirds of US GDP in 2018 (Fortune, 2018).

## 3.2 | Content analysis procedure

Our content analysis procedure was carried out in three steps. We describe these steps below (for an overview, see Figure 1).

### 3.2.1 | Collection of corpus (webpage text)

Several studies have shown that website content, aesthetics, and usability are key factors in website evaluation (e.g., Hartmann, Sutcliffe, & Angeli, 2008; Tarasewich, Daniel, & Griffin, 2001). Our study focuses on how companies' employment webpages can be analyzed via CATA. Thus, we focus on website texts as a key component of website content. The website text data were collected via a stepwise approach. Although websites often follow a similar structure, they usually use slightly different names for similar categories (e.g., about vs. our company) or show minor differences in content and structure. For this reason, we took a top-down approach in our content analysis. That is, in a first step, we identified broad (all-encompassing) webpage categories with employer-relevant information. Second, we assigned all different sub-categories to the broader categories. This approach allowed us to map the content of a heterogeneous medium (i.e., employer webpages) onto one large framework. Therefore, we followed common practice in categorizing website features (see, e.g., Douneva, Jaron, & Thielsch, 2016): That is, an expert with extensive experience with website design, CATA, and employer image attributes screened the content of 30 selected companies' webpages to identify and categorize common webpage sections. This initial categorization leads to the following six areas representing typical names for common webpage sections: (a) “About”, (b) “Careers”, (c) “Diversity and Inclusion”, (d) “Community”, (e) “Values and Responsibility”, and (f) “Culture.”

In a second step, all identified synonymous and sub-categories (e.g., “Our Company,” “Who We Are”) were linked to the broader categories (e.g., “Our Company” is related to the broad category “About”). As a result, a total of 80 subgroups were identified. In a third step, the identified sub-categories were screened in terms of their usefulness and applicability for a CATA. The following criteria had to be met: First, we used only information that was relevant to employer image (see the instrumental-symbolic framework of employer branding, van Hoye & Saks, 2011) and that had the potential to influence (prospective) employees' attraction to the company as an employer. In this context, we drew upon a meta-analytic categorization of job (e.g., compensation, benefits) and organizational characteristics (e.g., organizational image, employee relations) that generate substantial applicant attraction (see Uggerslev et al., 2012). Second, the data had to be directly analyzable through text mining (i.e., only pages with text were analyzed, while graphics, images, PDFs, and similar content were omitted). In the end, 41 out of a total of 80 identified subgroups were used in the sample. For example, within the “About” category, 11 sub-categories were considered not relevant for employer image (e.g., Website Privacy) and five categories non-analyzable through text mining (e.g., “Offices & Locations”). In the “Careers” category, eight categories were retained, while another eight were excluded (e.g., “Hiring Process,” “Recruiting,” “Events,” “FAQ,” and “Requirements”). A second expert verified this categorization (see Douneva et al., 2016 for a similar procedure).[3] The complete list of initially identified and ultimately selected subgroups can be found in Table S1.

The third step focused on actually gathering the text data. Four research assistants collected all hyperlinks and text data for each subgroup and webpage across all included companies. Text was collected in both the original “.html” format and a plain text-only format (i.e., “.txt”). For the subsequent content analysis, only the plain text files were used and aggregated at the company level. Companies with non-accessible webpages (e.g., pages that were offline due to recent mergers or nonloading due to technical restrictions) and companies with a limited online presence not including the aforementioned categories were excluded (e.g., http://www.berkshirehathaway.com). These steps resulted in a text dataset including 486 companies.

In a final step, companies with websites containing <1,000 words were excluded to ensure sufficient input data per company and comparability across companies. Applying this threshold led to a final sample of 461 companies (92%) with a total of more than 11,100 individual webpages and over 4.1 million words. On average, each analyzed company website consisted of around 24 distinct, individual webpages ($SD = 18.42$) and a total of 8,923 words ($SD = 8{,}581$), ranging from 1,000 to 95,689 words per the company website.

### 3.2.2 | Dictionary development

To ensure content validity (i.e., the degree to which a measure captures the full breadth of a specific construct; Nunnally & Bernstein, 1994; Colquitt, Sabey, Rodell, & Hill, 2019), it is recommended that dictionary development be based on both deductively and inductively derived word lists. Whereas deductively derived word lists involve applying theory to develop a coding scheme, inductively derived word lists are explorative and based on relevant texts of interest (Duriau et al., 2007). Therefore, we relied on both of these two approaches.

It was important to start by clearly defining the construct of interest and assessing its dimensionality based on the relevant literature (Short et al., 2010). In this study, employer image is the construct of interest; it is defined as the “set of beliefs that a job seeker holds about the attributes of an organization” (Cable & Turban, 2001, p. 125). Although there exist numerous ways to categorize employer image attribute attributes, we used the well-established categorization by Lievens and Highhouse (2003), which distinguishes between instrumental (i.e., objective, tangible job, and organizational attributes)

| | Step 1: Collection of Corpus (Webpage Text) | Step 2: Dictionary Development | Step 3: Coding via Computer-Aided Text Analysis |
|---|---|---|---|
| *Main Objective* | **Representative Company-Website Sample** | **Development of a Content-valid and Representative Dictionary** | **Scoring** |
| *Details* | 1. Definition of relevant webpage sections and sub-categories with employer-relevant information. Categorization resulted in six areas (e.g., “About”, “Careers”, “Culture”, etc.). *Realized via two authors of this study.*<br>2. Definition and use of 41 website sub-categories (e.g., “Benefits” within the “Career” category). *Realized via two authors of this study.*<br>3. Gathering of website text data in relevant webpage sections : Text was collected in both the original “.html” format and in a plain-text-only format. *Realized via four research assistants.*<br>4. Companies with websites containing less than 1,000 words were excluded to ensure sufficient input data per company and comparability across organizations.<br>➢ **The resulting text dataset is based on 486 companies from the Fortune 500 list.** | 1. Identification of instrumental (e.g., pay, benefits) and symbolic (e.g., sincerity, innovativeness) image dimensions (see also Lievens & Highhouse, 2003).<br>2. Deductive approach: Collection of exhaustive, discrete word lists for each of the theoretical (sub-) dimensions (with the help of the online lexical database WordNet, thesaurus.com, and glassdoor.com). *Realized via two authors of this study and two external experts.*<br>3. Validation of word lists by comparing the identified words with the theoretical definition of each dimension. *Realized via two authors of this study and two external experts.*<br>4. Addition of words based on an inductive approach. That is, on the basis of a randomly generated sub-sample of our Fortune 500 sample (i.e., 369 companies with 8,869 webpages and almost 3.4 million words), a comprehensive list of “insistence” words was generated and assigned to one of the 10 image dimensions. This process resulted in a list of 201 additional words.<br>➢ **The final dictionary consisted of 557 words.** | 1. Calculation of relative occurrence scores using LIWC software.<br>2. The score is build on custom-made dictionaries (i.e., word lists).<br>3. The scores were aggregated at the company level and divided by the overall number of words per company.<br>➢ **The resulting standardization allows to comparable the results across the Fortune 500 companies.** |

**FIGURE 1** Overview of the content analysis procedure

and symbolic (i.e., subjective, intangible, non-job/organization-related attributes) employer image attributes, as the foundation for building the deductively derived word list.

Next, we identified sub-dimensions based on the formal definitions of the instrumental and symbolic employer image constructs in the literature. For example, Lievens (2007) identified nine instrumental factors (i.e., social/team activities, physical activities, structure, advancement, travel opportunities, pay and benefits, job security, educational opportunities, and task diversity) and six symbolic attributes (i.e., sincerity, cheerfulness, excitement, competence, prestige, and ruggedness). Whereas the instrumental attributes were based on semi-structured interviews, the symbolic attributes stemmed from the brand personality scale developed by Aaker (1997).

To identify a comprehensive set of instrumental and symbolic attributes, we analyzed 13 studies in the field of employer image and organizational attractiveness. Overall, we identified 12 instrumental and eight symbolic image attributes that were mentioned in at least two or more studies. The complete list of identified literature and image attributes can be found in Table S2. Our final list consisted of six symbolic image attributes (i.e., sincerity, innovativeness, competence, prestige, ruggedness, and cheerfulness) and four instrumental image attributes (i.e., pay, benefits, advancement, and teamwork).

For the instrumental attributes, we had to omit several sub-dimensions that either represented combinations of terms that could technically not be captured via CATA because we used single words[4] as the unit of analysis (e.g., "challenging work," "task demand," "job security," "working conditions," "working atmosphere," "customer orientation") or that were too diverse to include all potential words (e.g., "location," for which all potential cities and countries would need to be captured). For the symbolic attributes, we treated "excitement" in the same way as "innovativeness," and we regarded "robustness" in the same way as "ruggedness" because they built on similar measures and were often used interchangeably (e.g., Schreurs, Druart, Proost, & De Witte, 2009; van Hoye et al., 2013). An overview of the final list and explanations for the excluded attributes can be found in Table S3.

### 3.2.3 | Creation of discrete word lists

The next step comprised the collection of exhaustive, discrete word lists for each of the theoretical subdimensions. We collected related words from the online lexical database WordNet (Miller, 1995; Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) and the website thesaurus.com to identify meaningfully related words and synonyms. This was done with all terms associated with an attribute in the literature (e.g., competence included adjectives such as "competent," "reliable," "intelligent," and "successful"). The benefits attribute was further complemented by frequently quoted fringe benefits on the employer review website Glassdoor.com[5] (e.g., maternity, childcare, sabbatical). Four expert raters (all male researchers [e.g., postdoctoral researcher] with a minimum 2 years' experience in the field of HRM and OB) then validated the word lists by comparing the identified words with the theoretical definition of each attribute. Inter-rater reliability for these expert ratings across all attributes using Cohen's kappa (Landis & Koch, 1977) ranged from 0.20 to 0.47, demonstrating on average a "fair" agreement between the raters (i.e., 0.30; Landis & Koch, 1977, p. 165). In cases of dissent, agreement on whether or not to include a word in a category was reached via discussion among the raters.

Of the 595 words initially generated from the literature, lexical databases, Glassdoor.com, and words added by the raters, 431 words were retained for further analyses (sincerity = 36 words; innovativeness = 45 words; competence = 37 words; prestige = 32 words; ruggedness = 62 words; cheerfulness = 66 words; pay = 35 words; benefits = 44 words; advancement = 47 words; and teamwork = 27 words).

As noted above, we complemented this deductive approach with an inductive approach. Based on a randomly generated subsample of our Fortune 500 sample, representing 80% of the full sample of companies (i.e., 369 companies with 8,869 webpages and almost 3.4 million words), we generated a comprehensive list of frequently used words (Short et al., 2010). To do this, we used a functionality of the text analysis program DICTION (Short & Palmer, 2008), which can return a list of so-called "insistence words." Insistence words comprise all nouns and noun-derived adjectives that occur three or more times in a standard 500-word section.[6] Ultimately, insistence words capture a text's dependence on often-repeated words. The program returned a list of 5,890 words from our sample texts (including company names and nonsense words resulting from writing mistakes) that were again rated by experts in this field. Each rater was asked to independently determine whether a word was in or out of scope and assign it to one of the 10 attributes. The coding was again guided by the definitions of each attribute that had previously served as the basis for generating the initial deductively derived word lists. We again measured inter-rater reliability for whether a word was considered in or out of scope using Cohen's kappa (Landis & Koch, 1977). Across all attributes, inter-rater reliability was .52 on average, indicating a "moderate" strength of agreement (Landis & Koch, 1977, p. 165). This process resulted in a list of 201 additional words that had not been previously identified by the deductive approach and could potentially be included. Finally, the list was refined through an iterative discussion process between raters in cases where a word was identified by one or two but not all raters, as well as in cases where the raters assigned different attributes. In the end, this inductive approach resulted in a list of an additional 126 words.

Our final dictionary based on both the deductive and inductive approaches consisted of 557 words (sincerity = 43 words; innovativeness = 57 words; competence = 57 words; prestige = 48 words; ruggedness = 66 words; cheerfulness = 69 words; pay = 39 words; benefits = 61 words; advancement = 80 words; and teamwork = 37 words). The combined word lists for each image attribute served as the basis and input for the CATA dictionary. Table 1 presents the detailed word lists for each selected employer image attribute.

### 3.2.4 | Coding via CATA

We used the text analysis software Linguistic Inquiry and Word Count (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015;

**TABLE 1** Word lists for projected instrumental and symbolic employer image attributes

| Employer image attribute/dimension | Content analysis word lists with expert validation |
|---|---|
| Sincerity (symbolic) | Aboveboard, anti-corruption, anti-corruption, artless, authentic, bonafide, dependable, disciplined, down-to-earth, earnest, earnestness, faithful, forthright, frank, genuine, honest, honestly, honesty, natural, no-nonsense, outspoken, plain, pretensionless, real, righteous, serious, serious-mindedness, seriousness, sincere, sincerity, transparency, transparent, true, true-blue, trustworthy, truthful, unaffected, unassumingness, undesigning, unfeigned, unpretentious, up-front, wholehearted |
| Innovativeness (symbolic) | Advanced, all-new, audacious, boldness, breakthrough, conception, contemporary, creation, cutting-edge, dare, daring, enlivened, entrepreneurial, entrepreneurship, excited, excitement, exciting, excogitation, forward-looking, gamy, groundbreaking, hardihood, ingenious, ingenuity, innovate, innovation, innovations, innovator, innovators, inspirit, invention, inventions, inventive, leading-edge, mettlesome, modern, new, newfangled, next-generation, origination, originative, reinventing, spirit, spunky, state-of-the-art, stimulate, transform, transformation, transforming, untested, untried, venturesome, venturous, vernal, young, youthful, zippy |
| Competence (symbolic) | Accurately, achiever, adapted, adequate, analytical, appropriate, capable, certificates, certifications, clever, competence, competencies, competency, competency-based, competent, decent, efficient, endowed, experiences, expertise, functional, high-efficiency, high-performance, high-performing, high-quality, information-driven, intelligence, intelligent, knowing, level-headed, levelheaded, pertinent, polished, practiced, proficient, qualified, quality, reliability, reliable, satisfactory, savvy, schooled, seasoned, secure, skilled, skills-based, skillset, specialists, specialized, studied, succeeder, success, successes, successful, suitable, talented, well-informed |
| Prestige (symbolic) | Ace, award-winning, awarded, awardees, awards, benchmark, celebrated, cool, cultivated, distinguished, doctor, elegance, eminent, esteemed, exalted, famed, great, honorable, honored, illustrious, important, imposing, impressive, invaluable, leading, mundaneness, mundanity, notable, premium, prestige, prestigious, prominent, refinement, renowned, reputable, reputation, respected, respectful, respecting, sophisticated, three-star, top-ranked, valued, winner, winners, world-class, worldliness, worldly |
| Ruggedness (symbolic) | Able-bodied, athletic, athletics, boisterous, brawny, built, bully, full-bodied, goon, hale, hard, hardiness, hardness, hardy, heavy, hefty, hoodlum, hooligan, huskiness, husky, impregnable, inviolable, lustiness, lusty, masculine, muscular, potent, powerful, powerhouse, prosperous, punk, racy, resilient, rigorous, roaring, robust, robustious, robustness, robustuous, rough, roughneck, rowdy, ruffian, ruffianly, rugged, ruggedness, sinewy, snappy, solid, stiff, stout, strong, strong-armer, sturdy, substantial, thug, tough, toughened, toughie, toughness, unattackable, vigorous, vital, yob, yobbo, yobo |
| Cheerfulness (symbolic) | Affable, affectionate, amiable, animated, animation, attentive, beneficial, blitheness, bright, buoyancy, buoyant, cheer, cheerful, cheerfulness, cheery, chipper, chirpy, chummy, comfort, cordial, delight, effervescent, enthusiastic, exuberance, favorable, fraternal, friendly, gaiety, geniality, gladness, glee, good-natured, helpful, hilarity, jauntiness, jaunty, jocundity, jolly, joy, joyful, joyousness, lighthearted, liveliness, lively, loving, loyal, merriment, merry, mirth, neighborly, optimism, optimistic, original, peaceful, peppy, perky, philanthropic, pleasant, rosy, sanguine, smooth, socially, sunniness, sunny, sunshine, sympathetic, upbeat, welcoming, well-disposed |
| Pay (instrumental) | Allowance, bacon, bread, co-pay, commission, compensation, defrayal, defrayment, earnings, emoluments, fee, honorarium, income, indemnity, meed, pay, paycheck, paychecks, payment, perquisite, pittance, proceeds, recompensation, recompense, redress, reimbursement, remuneration, requital, return, reward, salaried, salary, settlement, stipend, stipendium, take-home, takings, wage, wages |
| Benefits (instrumental) | 401, 401k, acupuncture, aerobics, aid, annuities, asset, assistance, beneficiary, benefit, benefit-eligible, benefits, betterment, bonus, book, canteen, childcare, classes, co-insurance, company-paid, courses, daycare, dental, discount, donation, ergonomic, ergonomics, extras, favor, gravy, gym, healthcare, holidays, insurance, loan, massage, massages, maternity, medical, medicare, parental, part-time, paternity, pension, perk, profitsharing, PTO, retirement, rewarding, sabbatical, ski, travel, vacation, volunteer, welfare, wellness, work-life, worklife, worth, yayday, yoga |
| Advancement (instrumental) | Acceleration, achievable, achievers, acquire, advance, advancement, amelioration, betterment, boost, career, careers, careersteps, chance, coaching, coursework, develop, development, educate, education, elevation, empower, empowered, empowering, empowerment, establish, evolution, evolve, expand, expansion, flourish, forward, foster, furtherance, future, gain, grow, growth, guidance, headway, high-potential, improvement, increase, internship, internships, learning, manager-in-training, maturate, maturation, mentor, mentored, mentoring, mentorplace, mentorship, modernize, modernize, ontogenesis, ontogeny, opportunities, opportunity, preferment, prepare, professional-development, progress, progression, promote, promote-from-within, promotes, promoting, promotion, raise, ripen, rise, succession, successors, train, trainee, trainees, training, upgrade, upgrading |
| Teamwork (instrumental) | Alliance, assistance, associate, co-worker, co-workers, coalition, cohort, collaborating, collaborative, communities, community, community-based, companion, confederacy, confederation, cowork, coworker, coworkers, cross-collaborate, federation, harmony, intergenerational, lineup, partisanship, partnered, partnering, partnership, symbiosis, synergism, synergy, team, teams, teamwork, teamworks, union, unit, unity |

*Note*: Deductively derived word lists were based on theoretical definitions of the employer image attributes in the literature, as well as synonym dictionaries (i.e., WordNet, thesaurus.com). Inductively derived word lists were based on commonly used words from suitable text. The final combined word lists were subjected to expert assessment and rating. Of the 790 words initially generated by the deductive and inductive approaches, 557 words were selected by the raters and retained for subsequent analyses.

Pennebaker, Francis, & Booth, 2001; Tausczik & Pennebaker, 2010). LIWC has been frequently used to examine psychological constructs, such as emotional expression or personality traits, across various psychological domains (e.g., Fast & Funder, 2008; Kahn, Tobin, Massey, & Anderson, 2007). LIWC builds on single word counting based on either predefined and validated dictionaries and scales or, as in our case, custom-made dictionaries (i.e., word lists; Boyd & Pennebaker, 2015). In this study, the custom-made dictionary to measure employer image attributes was based on the word lists that were generated in the deductive and inductive steps described above.

Based on the word lists presented above, we calculated relative occurrence scores using LIWC software. The scores were then aggregated at the company level and divided by the overall number of words per company. We standardized the scores by the number of words in each company's webpages to control for discrepancies in text length and thus to make the results comparable across all of the included Fortune 500 companies (Short et al., 2010).

### 3.2.5 | External validity

Finally, to check the generalizability of the created CATA dictionary across multiple settings (Cook, Campbell, & Day, 1979), we assessed employer image using CATA in different samples. Although external validity is typically tested with similar data samples from other sources in which the construct of interest is expected, this approach was not viable in the current study: As the webpage sample already comprised the 500 largest US companies, there was nothing comparable available. Therefore, we decided to split our overall sample and consider the resulting subsample as an equivalent "external" source to test the generalizability of our results.

That is, we used data from 369 companies (i.e., 80% of the companies and around 80% of the webpage content) to develop the inductive word list and the remaining unused sample of $n = 92$ companies to assess external validity. We then compared our subsample with a randomly selected sample from the remaining companies. To ensure that differences between samples were not influenced by the different sample sizes, we selected a sample of the same size in terms of selected companies ($n = 92$).

Tables 2 and 3 present comparisons of the CATA results between the 2014 main sample ($n = 369$) and the 2014 subsample ($n = 92$), as well as the randomly chosen subset of companies from our main sample to control for sample size ($n = 92$). We conducted one-sample *t* tests (compared to a test statistic of zero) for each image attribute as well as for the aggregated scores, to assess the presence of language in line with companies' projected employer image attributes in webpages. Whereas a zero result would have indicated that language in line with the chosen employer image attributes was not present, the results showed that all of the image attributes were present and significant across all samples, indicating that the constructs could also be generally detected and measured in "external", comparable samples.[7]

## 3.3 | Measures

### 3.3.1 | Projected employer image attributes

The section above describes how we used CATA to establish the content validity of the projected employer measure. For an overview, see Figure 1.

**TABLE 2** Evidence of language representing projected employer image attributes in company and employment webpages (2014 samples)

| | 2014 Main sample full ($n = 369$) | | | | 2014 subsample ($n = 92$) | | | |
|---|---|---|---|---|---|---|---|---|
| **Employer image attribute** | *N* | *M* | *SD* | *t* test | *N* | *M* | *SD* | *t* test |
| Sincerity | 369 | 0.14 | 0.13 | 21.58** | 92 | 0.16 | 0.12 | 12.53** |
| Innovativeness | 369 | 0.36 | 0.17 | 40.82** | 92 | 0.34 | 0.18 | 18.79** |
| Competence | 369 | 0.46 | 0.19 | 47.59** | 92 | 0.47 | 0.17 | 27.27** |
| Prestige | 369 | 0.30 | 0.15 | 39.43** | 92 | 0.31 | 0.16 | 18.78** |
| Ruggedness | 369 | 0.15 | 0.08 | 33.33** | 92 | 0.15 | 0.10 | 15.26** |
| Cheerfulness | 369 | 0.07 | 0.05 | 23.51** | 92 | 0.08 | 0.09 | 8.29** |
| Total symbolic | 369 | 1.48 | 0.38 | 75.44** | 92 | 1.52 | 0.37 | 39.66** |
| Pay | 369 | 0.17 | 0.15 | 21.74** | 92 | 0.15 | 0.13 | 10.62** |
| Benefits | 369 | 0.67 | 0.48 | 26.97** | 92 | 0.66 | 0.58 | 10.86** |
| Advancement | 369 | 1.69 | 0.65 | 50.09** | 92 | 1.54 | 0.61 | 24.31** |
| Teamwork | 369 | 0.79 | 0.35 | 42.94** | 92 | 0.75 | 0.44 | 16.55** |
| Total instrumental | 369 | 3.31 | 1.05 | 60.88** | 92 | 3.09 | 1.12 | 26.46** |

*Note*: Results in this table were based on the computer-aided text analysis using the word lists for employer image (i.e., instrumental and symbolic image attributes) presented in Table 1. A one-sample *t*-test was conducted compared to a test statistic of zero. The subsample was used as equivalent to an external data source due to nonavailability of other comparable data of this size. *$p < .05$; **$p < .01$.

**TABLE 3** Evidence of language representing projected employer image attributes in company and employment webpages (2014 samples)

| Employer image attribute | 2014 Main sample reduced ($n$ = 92) | | | | 2014 subsample ($n$ = 92) | | | |
|---|---|---|---|---|---|---|---|---|
| | *N* | *M* | *SD* | *t* test | *N* | *M* | *SD* | *t* test |
| Sincerity | 92 | 0.14 | 0.10 | 12.63** | 92 | 0.16 | 0.12 | 12.53** |
| Innovativeness | 92 | 0.36 | 0.18 | 18.64** | 92 | 0.34 | 0.18 | 18.79** |
| Competence | 92 | 0.50 | 0.23 | 20.69** | 92 | 0.47 | 0.17 | 27.27** |
| Prestige | 92 | 0.31 | 0.20 | 14.97** | 92 | 0.31 | 0.16 | 18.78** |
| Ruggedness | 92 | 0.15 | 0.08 | 17.38** | 92 | 0.15 | 0.10 | 15.26** |
| Cheerfulness | 92 | 0.07 | 0.06 | 10.71** | 92 | 0.08 | 0.09 | 8.29** |
| Total symbolic | 92 | 1.52 | 0.44 | 33.06** | 92 | 1.52 | 0.37 | 39.66** |
| Pay | 92 | 0.18 | 0.16 | 10.72** | 92 | 0.15 | 0.13 | 10.62** |
| Benefits | 92 | 0.77 | 0.62 | 11.87** | 92 | 0.66 | 0.58 | 10.86** |
| Advancement | 92 | 1.71 | 0.66 | 24.99** | 92 | 1.54 | 0.61 | 24.31** |
| Teamwork | 92 | 0.73 | 0.36 | 19.24** | 92 | 0.75 | 0.44 | 16.55** |
| Total instrumental | 92 | 3.39 | 1.22 | 26.63** | 92 | 3.09 | 1.12 | 26.46** |

*Note*: Results in this table were based on the computer-aided text analysis using the word lists for projected employer image (i.e., instrumental and symbolic image attributes) presented in Table 1. A one-sample *t*-test was conducted compared to a test statistic of zero. The subsample was used as equivalent to an external data source due to nonavailability of other comparable data of this size. $^*p < .05$. $^{**}p < .01$.

### 3.3.2 | Employer review website ratings

We collected employer rating data from the website Glassdoor.com.[8] Glassdoor is one of the largest third-party employer review websites and contains a rapidly growing database of millions of company reviews (Glassdoor, 2017a). Current and former employees can anonymously review their (current or former) company, including but not limited to experience reports; ratings of senior leadership, culture and values; and salary and other employee benefits (Glassdoor, 2017a; for an example of a company review, Figure S1). Although companies can flag and respond to the reviews, they cannot manipulate or remove reviews. Glassdoor ensures that reviews are truthful but does not allow disclosure of confidential, non-public internal information (Glassdoor, 2017b).

Comprehensive ratings for the companies included in this study were obtained in July 2015 by downloading employer-based rating data via the Glassdoor API[9] (Application-Programming-Interface). To ensure that the correct companies were selected, we compared the website URLs stated in the reviews to those in our Fortune 500 sample. In cases where subsidiaries were listed as well, we always selected the reviews for the US-based holding company, which was typically the object of a majority of reviews.

Out of the 461 companies included in our 2014 webpage sample, employer ratings for 446 companies (97%) were obtained. Overall, data for the 446 companies comprised 460,117 individual reviews, with an average of 1,032 reviews per company. The ratings in the above-mentioned categories were based on a 5-point Likert-type scale (1 = "very dissatisfied; 5 = "very satisfied"; Glassdoor, 2017c). Across all included companies, the overall average rating was 3.29 ($SD$ = .43), while, for example, the average culture and values rating was 3.25 ($SD$ = .51) and the average senior leadership rating was 2.87 ($SD$ = .44).

## 4 | RESULTS

### 4.1 | Research Question 1: CATA results and temporal stability

Research Question 1 addressed the question of whether the projected employer image attributes that arise from a CATA of company employment webpages are consistent over time. To test the consistency of our results across time, we ran the following analyses. In addition to the 2014 webpage sample, we collected earlier webpage data (i.e., before 2014) from the same companies and categories. To obtain these historical webpage data, we used the "Wayback Machine"[10] which is a freely accessible engine that randomly revisits webpages every few weeks and archives them (Internet Archive, 2017). We collected historical webpage data of 2 years (2010 and 2012) via this tool. The initial search[11] returned available webpage data from 244 of the 461 companies across all three points in time (i.e., 2010, 2012, and 2014). We again excluded companies with less than a total of 1,000 words. This resulted in a final "historical" sample of 163 companies across the years 2010, 2012, and 2014. On average, for these companies and points in time, 13 pages with 4,024 words (2010), and 17 pages with 5,676 words (2012) were available. For 2012, this reflected around 55% of the total words identified in 2014, and for 2010, this reflected around 39% of the total words in 2014.

We started by comparing 2010 and 2014. Table 4 depicts the results of a one-way ANOVA between the two samples of webpage text from these years. Generally, no significant differences in mean values between years were found for most image attributes, with the exception of two instrumental attributes. These were the pay and benefits attributes: The pay image attribute had an average value of 0.22 ($SD$ = .23) in 2010 and an average value of 0.17 ($SD$ = .15) in 2014. The time effect was therefore significant for the pay attribute

**TABLE 4** ANOVA comparisons of samples over time on projected employer image attributes (2014 main sample vs. 2010 Sample)

| Employer image attribute | *Df* | *F* test | *p* |
|---|---|---|---|
| Sincerity | | | |
| Between groups | 1 | 0.12 | .73 |
| Within groups | 324 | | |
| Total | 325 | | |
| Innovativeness | | | |
| Between groups | 1 | 0.47 | .49 |
| Within groups | 324 | | |
| Total | 325 | | |
| Competence | | | |
| Between groups | 1 | 0.87 | .35 |
| Within groups | 324 | | |
| Total | 325 | | |
| Prestige | | | |
| Between groups | 1 | 1.83 | .18 |
| Within groups | 324 | | |
| Total | 325 | | |
| Ruggedness | | | |
| Between groups | 1 | 1.21 | .27 |
| Within groups | 324 | | |
| Total | 325 | | |
| Cheerfulness | | | |
| Between groups | 1 | 0.07 | .79 |
| Within groups | 324 | | |
| Total | 325 | | |
| Total symbolic | | | |
| Between groups | 1 | 2.00 | .16 |
| Within groups | 324 | | |
| Total | 325 | | |
| Pay | | | |
| Between groups | 1 | 5.50* | .02 |
| Within groups | 324 | | |
| Total | 325 | | |
| Benefits | | | |
| Between groups | 1 | 4.42* | .04 |
| Within groups | 324 | | |
| Total | 325 | | |
| Advancement | | | |
| Between groups | 1 | 0.10 | .75 |
| Within groups | 324 | | |
| Total | 325 | | |
| Teamwork | | | |
| Between groups | 1 | 0.02 | .90 |
| Within groups | 324 | | |
| Total | 325 | | |
| Total instrumental | | | |
| Between groups | 1 | 1.60 | .21 |
| Within groups | 324 | | |
| Total | 325 | | |

($F$[1, 324] = 5.50, $p$ = .02). Significant differences in mean values were also found for the benefits attribute between the 2010 sample ($M$ = .82, $SD$ = .64) and 2014 sample ($M$ = .69, $SD$ = .47; $F$[1, 324] = 4.42, $p$ = .04). Taking a broader perspective, neither the aggregated instrumental attribute values nor the aggregated symbolic attribute values exhibited statistically significant differences in mean values over time. Similar results were found when comparing the years 2010 and 2012 as well as the years 2012 and 2014 (details can be obtained from the first author). In these latter two comparisons over time, none of the attributes differed significantly from each other in terms of their mean values.

In sum, the CATA results were relatively consistent across settings. Hence, companies do not seem to substantially change their projected employer webpage attributes over time because few (if any) attributes exhibited changes over time.

## 4.2 | Research Question 2: CATA results and differences among companies/industries

Research Question 2 investigated whether companies/industries can be differentiated from each other in terms of the employer image attributes presented on their employment websites. Given the very large number of companies in our data set, we started by running a two-step cluster analysis (using an agglomerative hierarchical clustering method, Norusis, 2011)[12] to identify $k$ homogeneous groups of companies that are similar with regard to their employer image attributes (e.g., in other domains, see Nielsen & Knardahl, 2014; Rundle-Thiele, Kubacki, Tkaczynski, & Parkinson, 2015). We ran this cluster analysis separately for instrumental and symbolic employer image attributes. If companies differed a lot in terms of employer image attributes, one would expect to find many different clusters of companies. Instead, the fit indices (i.e., average silhouette measure of cohesion and separation; Norusis, 2011) indicated the best fit for a solution with only two homogenous clusters of companies. In fact, this two-cluster solution emerged as the best when companies were clustered in terms of instrumental as well as in terms of symbolic employer image attributes (i.e., average silhouette measures of cohesion and separation of 0.5 and 0.3, respectively). The first, small cluster consisted of 83 (18%) out of 461 companies for the instrumental attribute cluster solution and of 74 (16%) out of 461 companies for the symbolic attribute cluster solution. In each case, the second, much larger cluster consisted of all other companies. Next, we investigated how the clusters differed from each other.

### 4.2.1 | Differentiations across companies

We inspected differences in instrumental and symbolic attributes between the two clusters (using independent sample $t$-tests). The results revealed that the cluster with the small set of companies had higher scores on 3 out of 4 (all $p$ values < .01, Cohen's $d$ ranged from .62–2.78) instrumental attributes and on 5 out of 6 symbolic attributes (all $p$ values <.008,[13] Cohen's $d$ ranged from .79–1.38). No differences occurred for the advancement (instrumental) and innovativeness (symbolic) attributes (all $p$ values > .05, Cohen's $d$ ranged from .10–.17). Thus, overall, a small set of companies (<20%) seemed to more strongly present employer image attributes than other companies as signals on their employment websites to communicate their projected employer image. Moreover, these companies highlighted multiple attributes at the same time. Strikingly, the large majority of companies were in one cluster and did not seem to set themselves apart from each other in terms of employer image attributes.

### 4.2.2 | Differentiations across industries

Apart from the cluster variable, we investigated whether industry (i.e., basic materials, consumer cyclicals, consumer non-cyclicals, energy, financial services, healthcare, industrial, technology, telecommunications services, and utilities; number of companies ranging from 8 to 83 per industry) is a feature that differentiates companies in terms of their employer image attributes. For instance, companies in healthcare might build their projected employer image on different attributes than telecommunications companies. Therefore, we conducted two separate one-way MANOVAs with industry as the independent factor and the six symbolic or four instrumental attributes as dependent variables. There was a significant overall effect for both instrumental, $F$(36, 1,680.601) = 3.249, $p$ < .001, Wilk's $\Lambda = 0.777$, partial $\eta^2 = .061$) and symbolic attributes, $F$(54, 2,278.757) = 2.896, $p$ < .001, Wilk's $\Lambda = 0.713$, partial $\eta^2 = .055$).[14] Specifically, we found differences on two out of four instrumental attributes, namely pay, $F$(9, 451) = 2.580, $p$ < .01, partial $\eta^2 = .049$, and benefits, Welch's $F$(9, 105.815) = 6.590, $p$ < .001, partial $\eta^2 = .138$. Post-hoc tests revealed that only 14 comparisons out of 90 (i.e., ~16%) were significant for instrumental attributes. For instance, financial services companies had higher scores on pay than companies in two other industries; companies in healthcare and finance scored higher on benefits than several other industries.

Differences also emerged for three out of six symbolic attributes, namely sincerity, Welch's $F$(9, 108.324) = 4.391, $p$ < .001, partial $\eta^2 = .123$, innovativeness, $F$(9, 451) = 3.192, $p$ < .01, partial $\eta^2 = .060$, and competence, $F$(9, 451) = 3.825, $p$ < .001, partial $\eta^2 = .071$. Despite these omnibus tests, post-hoc tests (Hochberg's GT2) only seldom revealed differences between specific industries. That is, post-hoc tests indicated only 19 significant comparisons out of 135 (i.e., ~14%). For instance, companies in technology scored higher on innovativeness than other industries (e.g., energy). Thus, overall, we found a similar picture as in our cluster analysis: There is some differentiation across industries, but the extent of differentiation is relatively low (for an overview see Figures S2 and S3). More detailed results can be obtained from the first author.

## 4.3 | Research Question 3: CATA results and third-party employer information

Research Question 3 addressed the convergence between projected employer image attributes and ratings of attributes found on employer review websites. In this context, it was important to compare categories of projected employer image attributes and Glassdoor dimensions that were as similar as possible. However, we acknowledge that not all of the selected instrumental-symbolic attributes from the employer webpage image dictionary were represented in the Glassdoor ratings. Therefore, we could not compare all (projected) instrumental and symbolic employer image attributes to Glassdoor information. Specifically, only the following (mostly instrumental) categories from Glassdoor could be used: "career opportunities rating" (representing the advancement attribute), "compensation and benefits rating" (representing the pay attributes), "work-life balance rating" (representing the benefits attribute as many benefits aim to improve work-life balance), and the "overall rating" (representing all employer instrumental-symbolic image attributes). Thus, we conducted multiple linear regressions to predict third-party employer ratings based on projected employer image attributes. One can assume that large companies usually have a larger budget to create their corporate websites. Thus, we used a number of Glassdoor ratings, industries, and firm size (as measured through total assets at year-end 2014, Hansen & Wernerfelt, 1989) as control variables in our analyses. We also only included companies with at least 30 Glassdoor ratings. In sum, data for 387 companies comprised 430,636 individual reviews with an average of 1,113 reviews per company.

Overall, it seems that the two measures were not related to each other. Regression analyses revealed that the advancement attribute was no substantial predictor for Glassdoor career opportunities ratings ($\beta = .08$, $p > .05$) and pay attributes were not substantially related to Glassdoor compensation and benefits ratings ($\beta = .004$, $p > .05$). We found that the image attribute benefits was a substantial predictor for Glassdoor work-life balance ratings, but indicating a negative effect ($\beta = -.15$, $p < .05$). Notably, using a step-wise regression model including all employer instrumental-symbolic image attributes predicting Glassdoor overall ratings, we found that sincerity ($\beta = .13$, $p < .05$) and advancement ($\beta = .12$, $p < .05$) were positive predictors for the ratings, whereas competence ($\beta = -.21$, $p < .01$) and benefits ($\beta = -.12$, $p < .05$) were negatively related to Glassdoor overall ratings (Table 5). In sum, our results indicated that projected employer webpage image attributes from the CATA and employer review website ratings were only slightly related (for a similar picture when companies were separated via above-mentioned clusters, see Figures S4 and S5).

## 4.4 | Robustness analyses

As mentioned above, we ran robustness analyses in which we also addressed our first two research questions based on a CATA of CSR statements and D&I policies. We could not address our third research question because Glassdoor does not have categories that correspond to CSR and D&I, respectively. In these robustness analyses, we used the same systematic approach for developing dictionaries for CSR statements and D&I policies (see the ESM) as we did for the employer image attributes. Results for both research questions when using the CATA scores for CSR and D&I echoed the ones presented above for the employer image attributes. That is, similar to the results of the instrumental and symbolic attributes in our paper, we found almost no differences across the time points of 2010, 2012, and 2014. The CSR and D&I attributes showed no significant differences in mean values between these years (all $p$ values > .05), with the only exception of CSR for the comparison between 2010 and 2014, ($F[1, 324] = 4.10$, $p = .04$, partial $\eta^2 = .013$). This effect size indicates a small effect, which may be due to the fact that CSR has become even more important for companies across the years (see Carroll, 2015).

To test RQ2, we followed the same procedure as in our paper and thus ran a two-step this cluster analysis for CSR and D&I. The fit indices indicated the best fit for a solution with three homogenous

**TABLE 5** Regression analysis summary for webpage projected employer image predicting image attributes from Glassdoor

| | | Model | | | | | |
|---|---|---|---|---|---|---|---|
| **Dependent variables (Glassdoor ratings)** | **Predictor** | $R^2$ | $B$ | $SE\ B$ | $\beta$ | $t$ | $p$ |
| *Model 1* | | | | | | | |
| Career opportunities | Advancement | .02 | .05 | .03 | .08 | 1.66 | .10 |
| *Model 2* | | | | | | | |
| Compensation and benefits | Pay | .03 | .01 | .15 | .004 | 0.77 | .93 |
| *Model 3* | | | | | | | |
| Work-life balance | Benefits | .04 | −.12 | .04 | −.15 | −2.93 | .004 |
| *Model 4* | | | | | | | |
| Overall rating | Competence | .07 | −.44 | .11 | −.21 | −3.93 | .001 |
| | Sincerity | | .42 | .17 | .13 | 2.51 | .013 |
| | Benefits | | −.10 | .04 | −.12 | −2.42 | .016 |
| | Advancement | | .07 | .03 | 12 | 2.20 | .029 |

*Note*: Number of Glassdoor ratings, industries, and firm size (as measured through total assets at year-end) were included as control variables in these models.
Abbreviation: SE, standard error.

clusters of companies (i.e., average silhouette measure of cohesion and separation of 0.6). The clusters consisted of 72 (16%), 165 (36%), and 224 companies (48%) out of 461 companies. Inspection of differences between the clusters revealed that each group is different from each other on both the CSR and D&I attribute-level (all $p$ values < .01). For CSR, Cohen's $d$ ranged from .68 to 2.80; for D&I, Cohen's $d$ ranged from 0.335 to 2.95. We also found a significant main effect for industry as an independent factor, with $F(9, 451) = 3.616$, $p < .001$, partial $\eta^2 = .067$. Post-hoc tests showed only 10 out of 45 comparisons to be significant for CSR (i.e., ~22%) and 2 out of 45 (i.e., ~4%) for D&I. Thus, overall, we found a similar picture for CSR and D&I as in our main analysis in the paper: There is some differentiation across companies and industries but the extent of differentiation is relatively low. More detailed results can be found in the ESM.

## 5 | DISCUSSION

Extending the emerging research on webpages as recruitment signals (Banks et al., 2019), this article develops an unobtrusive and efficient method to measure projected employer image from company webpages. We relied on CATA (Banks et al., 2018; Kobayashi et al., 2018) and built validated dictionaries of selected instrumental and symbolic employer image attributes, as well as other webpage information that also relates to people's attraction to the company (i.e., CSR statements and D&I policies).

### 5.1 | Main conclusions and contributions

Our study focused on the projected employer image attributes that companies transmit through one of their key communication channels (i.e., webpages). As we were able to analyze which image attributes companies actually communicate and project to job seekers (see also Banks et al., 2019), this study advances our knowledge of projected employer image and webpages as recruitment signals in at least three ways. First, the evidence for the temporal consistency of the CATA results shows that companies do not frequently change their projected webpage-based employer image. Few (if any) projected employer attributes seem subject to changes over time. These findings are in line with propositions in employer image management that companies need to have an enduring and consistent image to increase their credibility and lower the risk of unmet expectations among (prospective) employees (Wilden et al., 2010). That said, our findings show that companies do selectively adjust their image over time on attributes such as pay and benefits.

Second, our study advances the field by providing insights into the distinctiveness of companies' projected employer images, an important prerequisite for establishing an attractive employer brand (Backhaus & Tikoo, 2004; Moser et al., 2020). Thus, we examine to what extent companies actually take advantage of the opportunity to use their website to differentiate their employer image from that of their competitors. Our findings indicate that there is a relatively small subset of companies (<20%) that put more emphasis on building a distinctive employer image. Moreover, it seems that these companies not only differ with regard to single attributes; the result indicates an overall effect that spanned across various attributes. We found a similar picture across industries: the projected attributes differed only marginally between different industries. What are possible explanations? One reason might be that—as compared to job ads that run only for a short term—companies play it safe with their employment websites because these are longer lasting and all the information provided can be used against them. Or perhaps most companies follow best practices, so that they do not stand out and are not regarded as "strange" (Cable, 2007; Cromheecke, van Hoye, & Lievens, 2013). A common theme underlying these findings is that unused potential remains for companies to differentiate themselves and to further sharpen their employer brand on their webpages, especially given that employer websites are still a central element for applicants to get an impression of what it is like to be a member of the organization (Banks et al., 2019).

Third, this article investigates whether the quantified employer image attributes converge with the experiences of third parties (current/former employees). While previous research underlined the importance of corporate websites as an important vehicle to communicate a company's values and employer attributes (Kane et al., 2017), our findings indicate that projected employer image (as expressed through company and employment webpages) and evaluations of current/former employees (as expressed through third-party employer ratings) emerged as different constructs. Thus, it seems that the projected employer image communicated by company employment webpages represents mere "rhetoric" rather than a valuable recruitment signal that provides credible information about a company's culture and DNA. Although we were not able to investigate this convergence for D&I policies, audit research shows that the same effects can be found in that domain because "employers that adopt pro-diversity statements are in fact just as likely to engage in discrimination" (Kang, DeCelles, Tilcsik, & Jun, 2016, p. 496). In sum, "window-dressing" statements by employers on employment webpages might diverge quite a bit from their actual HR practices.

### 5.2 | Implications for practice

Our research has various implications for companies interested in measuring and managing their projected employer image. Given that surveys of employer image attributes are cost- and time-intensive for both companies and respondents, we introduced and developed an approach for efficiently assessing projected/intended employer image "on the fly." Using CATA and the developed and validated employer image dictionaries, researchers and practitioners can efficiently and quickly measure and monitor attributes of projected employer image based on large-scale text data. In our example, this was done for US Fortune 500 companies' employment webpages.

The improved measurability of projected employer image is likely to uncover "image gaps" between projected image and other constructs of interest. As shown in this study's comparison of company-controlled and third-party (i.e., Glassdoor) information, there might be large gaps between intended image projection and perceptions of company insiders and outsiders, which should motivate companies to adjust and fine-tune their intended image. This applies not only to the recruitment but also to the internal context (current employees' perceptions). Although our study does not suggest what to undertake in such cases (see, e.g., Dineen et al., 2019), it provides a useful tool for identifying gaps.

All of this fits into a broader strategy of employer image management and engineering (Schwaiger, 2004). That is, our CATA approach opens up a variety of opportunities for evaluating employer image. For example, our method could serve as an additional component of employer image audits (i.e., as extra input in addition to best employer rankings). Companies could also rely on our method to map the projected employer image of their labor market competitors and uncover how they compare to these competitors on a variety of image attributes.

## 5.3 | Limitations

As a first limitation, our study analyzed only website text, thereby neglecting other media content and webpage features such as pictures, interactive elements, videos, vlogs, PDFs, or website aesthetics, which may also convey valuable image information about an employer (e.g., Dineen, Ling, Ash, & DelVecchio, 2007). Whereas CATA as applied in this paper examined only text data, future research should take into consideration how other media content could be efficiently measured to complement a text-based approach. Specifically, the analysis of PDFs (representing, for instance, customer recruitment brochures, articles, press releases, or annual reports) that were presented on employer webpages might be a promising approach to detect differences between companies. Moreover, recruitment signals on social media channels (e.g., LinkedIn, Twitter) could be scraped (Moser et al., 2020). In contrast to websites, social media channels offer an opportunity for employers to interact directly with customers and applicants. As social media sites have high accessibility and offer numerous possibilities for direct communication (e.g., via videos, postings, and private messages; for further information, see Kissel & Büttgen, 2015), employers can more flexibly adapt their projected employer image and apply additional tools to strengthen their intended image.

Second, our CATA approach examined the use of single words in sample texts to make inferences about the mental models of the texts' authors. Although this method is in line with recommended procedures (e.g., Short et al., 2010), it neglects the broader context found in co-occurrences of specific words as well as negations or negative meanings. Simultaneously, the single-word approach limited the constructs and attributes that could be investigated. For example, some of the instrumental attributes had to be excluded because they could not be expressed in a single word or term (e.g., "challenging work" or "customer orientation"). Moreover, the approach assumed that (prospective) employees seek out and process all available webpages. In real life, however, webpage visitors might only browse a fraction of them.

Third, this study included only large companies because they usually provided sufficient webpage content. Our conclusions might thus not generalize to small and medium-sized enterprises (Tumasjan, Strobel, & Welpe, 2011). In addition, it may be challenging to fully measure the webpage image of companies that do not include sufficient textual data on their webpage. Finally, large companies usually have a larger budget to create their corporate websites. This may make it possible for them to create websites with fancy features that small companies may not be able to include on their websites (Tumasjan et al., 2011).

## 5.4 | Directions for future research

This study opens up several intriguing future research avenues. First, future studies should build upon our approach and include other stakeholders and constituents. For example, current employees are exposed to the company's webpage content, but it can be assumed that they also consider internal media such as the company intranet. As an extension, one might apply a similar approach taking the internal perspective (i.e., current employees and data from the company intranet). For example, one might use CATA to compare webpage attributes to dimensions of internal culture to identify and analyze potential "vision" versus "culture" gaps and examine the impact of such gaps. Other extensions are also possible, such as sentiment analysis/text analysis of Twitter comments about companies.

Second, we found initial evidence that webpage content is relatively stable. Both clarity and consistency are important for enhancing the credibility and perceived quality of an employer brand and to reduce perceived risk and employee information cost (Dineen et al., 2019; Wilden et al., 2010). Our approach can be used to address the need for companies to efficiently measure and monitor the recruitment signals issued via their webpages and those of their competitors (Theurer, Tumasjan, Welpe, & Lievens, 2018). It can also detect how specific events (e.g., bad press, scandals) affect such consistency (Edlinger, 2015). Future research might investigate reasons why companies create webpages that mostly not stand out from among their competitors.

Third, additional research is needed to fully understand the process of website evaluation and the interaction between website text, videos, and other elements. Given that successful companies are able to invest a lot of money in their website communication, it would be helpful to explore key drivers of employer images (besides webpage text). For instance, preferences for a certain website (Van der Heijden, 2003) as well as perceived usability (Tractinsky, Katz, & Ikar, 2000), and content (De Angeli, Sutcliffe, & Hartmann, 2006) are significantly influenced by aesthetics. In addition, aesthetics positively influence the assessment of a website's credibility (Fogg et al., 2003) and trustworthiness (Karvonen, 2000). Although many areas within computer-human interaction research initially focused on usability, nowadays aesthetics is regarded as an important factor in webpages (Cober et al., 2004).

Finally, further research is needed to investigate the reasons behind the gap between employment-related webpage information and third-party information image types. We need to better understand how content on webpages is processed and integrated with information from other sources to determine subsequent outcomes (e.g., site visits, job interviews, etc.). People's perceptions of an employer are created based on multiple impressions from different sources. Future research is thus needed to analyze people's processing of employer information via eye-tracking, verbal protocols, mouse clicks, and so forth.

## 6 | CONCLUSION

This study is among the first to use CATA to analyze and quantify projected employer image attributes as transmitted through company employment websites. Our approach allows for comparing a company's projected employer image to its competitors' projected employer images across time. We found that projected employer image attributes remain relatively stable over time and differentiate companies and industries only to a certain extent. Projected employer image attributes stemming from employment webpages (company-controlled employer brand information) did not converge with ratings on similar attributes on employer review websites (third-party employer brand information). From a practical perspective, we present a useful, unobtrusive, and efficient text-mining tool for analyzing, monitoring, and adjusting a company's external communication as an employer of choice.

### ACKNOWLEDGMENTS

Philipp Schäpers would like to thank the State of North Rhine-Westphalia's Ministry of Economic Affairs, Innovation, Digitalization, and Energy as well as the Exzellenz Start-up Center. NRW program at the REACH-EUREGIO Start-Up Center for their kind support of his work. Open access funding enabled and organized by Projekt DEAL.

### ENDNOTES

[1] Note that our focus is on *employer image attributes* depicted on webpages instead of employment information in general. We acknowledge that this focus excludes other potentially relevant information that might also relate to potential employee's attraction to the company.

[2] Other safeguards are that people must register with a valid email address and agree to comply with review guidelines. Adherence to these guidelines is monitored via technical security measures and a community management team.

[3] Both experts involved in the development of the corpus (webpage text) were also authors of this study. These two authors were also two of the four experts involved in the creation of discrete word lists described in the Section 3.2.3.

[4] In line with recommended approaches in management and psychological research, we decided to use CATA with *single* words as the unit of analysis (see Pennebaker, Mehl, & Niederhoffer, 2003; Short et al., 2010).

[5] URL: https://www.glassdoor.com/blog/top-20-employee-benefits-perks-for-2017/.

[6] As stated in the manual of DICTION version 7.1.3 (http://www.dictionsoftware.com).

[7] We further analyzed the mean differences between samples by conducting one-way analyses of variance (ANOVA). First, we conducted a one-way ANOVA between the reduced 2014 main sample ($n = 92$) and the 2014 subsample ($n = 92$; equivalent to a similar external data source). There were no statistically significant differences in mean values between the samples across image attributes, indicating that language consistent with employer image attributes was consistently communicated and measured in different samples.

[8] URL: https://www.glassdoor.com/index.htm.

[9] URL: https://www.glassdoor.com/developer/index.htm.

[10] https://web.archive.org.

[11] To access the 2010 and 2012 webpages, we pasted the 2014 URLs (uniform resource locator) of our 2014 sample pages into the search field of the Wayback Machine. If this delivered a valid result (i.e., the link was valid at that time), we copied the text from each year as we had done for 2014. If the page was not available (page not archived), we started from the historical home page, navigated through the relevant categories (where available) and recorded the text and alternative links. If the Wayback Machine returned that the page could not be crawled (i.e., "page cannot be displayed due to robots.txt"), the company had to be excluded from the historical sample.

[12] This approach encompasses a three-step procedure: calculating distances between groups, linking clusters, and choosing the right number of clusters based on a Bayesian Information Criterion (BIC).

[13] We corrected for alpha inflation via Bonferroni correction (see Cabin & Mitchell, 2000).

[14] Homogeneity of variance was not violated, with the exception of sincerity and benefits. Therefore, we used Welch's $F$ as the test statistic and Games–Howell as post-hoc tests for these attributes.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

### ORCID

*Philipp Schäpers* https://orcid.org/0000-0002-8270-5105

*Andranik Tumasjan* https://orcid.org/0000-0002-0127-8499

### REFERENCES

Aaker, D. A. (1991). *Managing brand equity*. New York, NY: Free Press.

Aaker, J. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*, 347–356. https://doi.org/10.1177/002224379703400304

Arndt, M., & Bigelow, B. (2000). Presenting structural innovation in an institutional environment: Hospitals' use of impression management. *Administrative Science Quarterly*, *45*, 494–522. https://doi.org/10.2307/2667107

Backhaus, K., & Tikoo, S. (2004). Conceptualizing and researching employer branding. *Career Development International*, *9*, 501–517. https://doi.org/10.1108/13620430410550754

Banks, G. C., Woznyj, H. M., Wesslen, R. S., Frear, K. A., Berka, G., Heggestad, E. D., & Gordon, H. L. (2019). Strategic recruitment across borders: An investigation of multinational enterprises. *Journal of Management*, *45*, 476–509. https://doi.org/10.1177/0149206318764295

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*, 445–459. https://doi.org/10.1007/s10869-017-9528-3

Barrow, S., & Mosely, R. (2005). *The employer brand: Bringing the best of brand management to people at work*. Chichester, UK: John Wiley & Sons Ltd.

Bone, P. F. (1995). Word-of-mouth effects on short-term and long-term product judgments. *Journal of Business Research*, *32*, 213–223. https://doi.org/10.1016/0148-2963(94)00047-i

Boyd, R., & Pennebaker, J. (2015). A way with words: Using language for psychological science in modern era. In C. V. Dimofte, C. P. Haugtredt, & R. F. Yalch (Eds.), *Consumer psychology in a social media world* (pp. 222–236). New York, NY: Routledge. https://doi.org/10.4324/9781315714790

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America*, *81*, 246–248.

Cable, D. M. (2007). *Change to strange: Create a great organization by building a strange workforce*. Upper Saddle River, NJ: Pearson.

Cable, D. M., & Turban, D. B. (2001). Establishing the dimensions, sources and value of job seekers' employer knowledge during recruitment. *Research in Personnel and Human Resources Management*, *20*, 115–163. https://doi.org/10.1016/s0742-7301(01)20002-4

Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior*, *18*, 533–558. https://doi.org/10.1002/(sici)1099-1379(199711)18:1+<533::aid-job906>3.3.co;2-v

Carroll, A. B. (2015). Corporate social responsibility: The center piece of competing and complementary frameworks. *Organizational Dynamics*, *44*, 87–96. https://doi.org/10.1016/j.orgdyn.2015.02.002

Cascio, W. F., & Graham, B. Z. (2016). New strategic role for HR: Leading the employer-branding process. *Organization Management Journal*, *13*, 182–192. https://doi.org/10.1080/15416518.2016.1244640

Chen, C. C., & Meindl, J. R. (1991). The construction of leadership images in the popular press: The case of Donald Burr and People express. *Administrative Science Quarterly*, *36*, 521–551. https://doi.org/10.2307/2393273

Cober, R. T., Brown, D. J., Keeping, L. M., & Levy, P. E. (2004). Recruitment on the net: How do organizational web site characteristics influence applicant attraction. *Journal of Management*, *30*, 623–646. https://doi.org/10.1016/j.jm.2004.03.001

Cober, R. T., Brown, D. J., Levy, P. E., Cober, A. B., & Keeping, L. M. (2003). Organizational web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection and Assessment*, *11*, 158–169. https://doi.org/10.1111/1468-2389.00239

Collins, C. J., & Han, J. (2004). Exploring applicant pool quantity and quality: The effects of early recruitment practice strategies, corporate advertising, and firm reputation. *Personnel Psychology*, *57*, 685–717. https://doi.org/10.1111/j.1744-6570.2004.00004.x

Collins, C. J., & Stevens, C. K. (2002). The relationship between early recruitment-related activities and the application decisions of new labor-market entrants: A brand equity approach to recruitment. *Journal of Applied Psychology*, *87*, 1121–1133. https://doi.org/10.1037/0021-9010.87.6.1121

Colquitt, J. A., Sabey, T. B., Rodell, J. B., & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology*, *104*, 1243–1265. https://doi.org/10.1037/apl0000406

Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Boston, MA: Houghton Mifflin Company.

Cromheecke, S., van Hoye, G., & Lievens, F. (2013). Changing things up in recruitment: Effects of a 'strange' recruitment medium on applicant pool quantity and quality. *Journal of Occupational and Organizational Psychology*, *86*, 410–416. https://doi.org/10.1111/joop.12018

De Angeli, A., Sutcliffe, A., & Hartmann, J. (2006). Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the 6th conference on designing interactive systems* (pp. 271–280). New York, NY: ACM Press.

Dineen, B. R., & Allen, D. G. (2016). Third party employment branding: Human capital inflows and outflows following "best places to work" certifications. *Academy of Management Journal*, *59*, 90–112. https://doi.org/10.5465/amj.2013.1091

Dineen, B. R., Ling, J., Ash, S. R., & DelVecchio, D. (2007). Aesthetic properties and message customization: Navigating the dark side of web recruitment. *Journal of Applied Psychology*, *92*, 356–372. https://doi.org/10.1037/0021-9010.92.2.356

Dineen, B. R., van Hoye, G., Lievens, F., & Rosokha, L. M. (2019). Third party employment branding: What are its signaling dimensions, mechanisms, and sources? *Research in Personnel and Human Resources Management*, *37*, 173–226. https://doi.org/10.1108/s0742-730120190000037006

Douneva, M., Jaron, R., & Thielsch, M. T. (2016). Effects of different website designs on first impressions, aesthetic judgements and memory performance after short presentation. *Interacting with Computers*, *28*, 552–567. https://doi.org/10.1093/iwc/iwv033

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods*, *10*, 5–34. https://doi.org/10.1177/1094428106289252

Edlinger, G. (2015). Employer brand management as boundary-work: A grounded theory analysis of employer brand managers' narrative accounts. *Human Resource Management Journal*, *25*, 443–457. https://doi.org/10.1111/1748-8583.12077

Etter, M., Ravasi, D., & Colleoni, E. (2019). Social media and the formation of organizational reputation. *Academy of Management Review*, *44*, 28–52. https://doi.org/10.5465/amr.2014.0280

Evertz, L., Kollitz, R., & Süss, S. (2019). Electronic word-of-mouth via employer review sites—The effects on organizational attraction. *International Journal of Human Resource Management*, 1–30. https://doi.org/10.1080/09585192.2019.1640268

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*, 334–346. https://doi.org/10.1037/0022-3514.94.2.334

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of websites? In *Proceedings of the 2003 conference on designing for user experiences* (pp. 1–15). San Francisco, CA: ACM Press.

Fortune. (2014). *Fortune 500 2013 List*. Retrieved from http://fortune.com/fortune500/2013/.

Fortune. (2018). *100 Best Companies to Work for—Methodology*. Retrieved from http://fortune.com/best-companies/2018.

Gardner, B. B., & Levy, S. J. (1955). The product and the brand. *Harvard Business Review*, *33*, 33–39.

Glassdoor. (2017a). *About us*. Retrieved from https://www.glassdoor.com/about/index_input.htm.

Glassdoor. (2017b). *Common questions*. Retrieved from https://www.glassdoor.com/employers/start/common-questions.htm.

Glassdoor. (2017c). *Ratings on Glassdoor*. Retrieved from http://help.glassdoor.com/article/Ratings-on-Glassdoor/en_US.

Guo, Y., Li, Y., & Shao, Z. (2015). An ant colony-based text clustering system with cognitive situation dimensions. *International Journal of Computational Intelligence Systems*, *8*, 138–157. https://doi.org/10.1080/18756891.2014.963986

Hansen, G. S., & Wernerfelt, B. (1989). Determinants of firm performance: The relative importance of economic and organizational factors.

Strategic Management Journal, 10, 399–411. https://doi.org/10.1002/smj.4250100502
Hartmann, J., Sutcliffe, A., & Angeli, A. D. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *15*, 1–30. https://doi.org/10.1145/1460355.1460357
Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2020). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 1–33. https://doi.org/10.1177/1094428120971683
Highhouse, S., Thornbury, E. E., & Little, I. S. (2007). Social-identity functions of attraction to organizations. *Organizational Behavior and Human Decision Processes*, *103*, 134–146. https://doi.org/10.1016/j.obhdp.2006.01.001
Holt, D. B. (2004). *How brands become icons: The principles of cultural branding*. Boston, MA: Harvard Business Press.
Internet Archive. (2017). *About the internet archive*. Retrieved from https://archive.org/about/.
Kabanoff, B., Waldersee, R., & Cohen, M. (1995). Espoused values and organizational change themes. *Academy of Management Journal*, *38*, 1075–1104. https://doi.org/10.5465/256621
Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, *120*, 263–286. https://doi.org/10.2307/20445398
Kane, G. C., Palmer, D., Phillips, A. N., & Kiron, D. (2017). Winning the digital war for talent. *MIT Sloan Management Review*, *58*, 17–19. https://doi.org/10.7551/mitpress/11633.003.0006
Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened résumés: Race and self-presentation in the labor market. *Administrative Science Quarterly*, *61*, 469–502. https://doi.org/10.1177/0001839216639577
Karvonen, K. (2000). The beauty of simplicity. In *ACM Proceedings on the 2000 Conference on Universal Usability* (pp. 85–90). New York, NY: ACM Press.
Kausel, E. E., & Slaughter, J. E. (2011). Narrow personality traits and organizational attraction: Evidence for the complementary hypothesis. *Organizational Behavior and Human Decision Processes*, *114*, 3–14. https://doi.org/10.1016/j.obhdp.2010.08.002
Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *The Journal of Marketing*, *57*, 1–22. https://doi.org/10.1177/002224299305700101
Keller, K. L. (2011). *Strategic brand management: Building, measuring, and managing brand equity*. Harlow: Pearson Education.
Kissel, P., & Büttgen, M. (2015). Using social media to communicate employer brand identity: The impact on corporate image and employer attractiveness. *Journal of Brand Management*, *22*, 755–777. https://doi.org/10.1057/bm.2015.42
Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. (2018). Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, *21*, 766–799. https://doi.org/10.1177/1094428117719322
Könsgen, R., Schaarschmidt, M., Ivens, S., & Munzel, A. (2018). Finding meaning in contradiction on employee review sites—Effects of discrepant online reviews on job application intentions. *Journal of Interactive Marketing*, *43*, 165–177. https://doi.org/10.1016/j.intmar.2018.05.001
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310
Lemmink, J., Schuijf, A., & Streukens, S. (2003). The role of corporate image and company employment image in explaining application intentions. *Journal of Economic Psychology*, *24*, 1–15. https://doi.org/10.1016/s0167-4870(02)00151-4
Lievens, F. (2007). Employer branding in the Belgian army: The importance of instrumental and symbolic beliefs for potential applicants, actual applicants, and military employees. *Human Resource Management*, *46*, 51–69. https://doi.org/10.1002/hrm.20145
Lievens, F., & Highhouse, S. (2003). The relation of instrumental and symbolic attributes to a company's attractiveness as an employer. *Personnel Psychology*, *56*, 75–102. https://doi.org/10.1111/j.1744-6570.2003.tb00144.x
Lievens, F., van Hoye, G., & Schreurs, B. (2005). Examining the relationship between employer knowledge dimensions and organizational attractiveness: An application in a military context. *Journal of Occupational and Organizational Psychology*, *78*, 553–572. https://doi.org/10.1348/09631790x26688
Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, *28*, 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x
Melián-González, S., & Bulchand-Gidumal, J. (2016). Worker word of mouth on the internet: Influence on human resource image, job seekers and employees. *International Journal of Manpower*, *37*, 709–723. https://doi.org/10.1108/ijm-09-2014-0188
Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41. https://doi.org/10.1145/219717.219748
Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–244. https://doi.org/10.1093/ijl/3.4.235
Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management*, *20*, 903–931. https://doi.org/10.1016/0149-2063(94)90035-3
Moser, K., Tumasjan, A., & Welpe, I. M. (2017). Small but attractive: Dimensions of new venture employer attractiveness and the moderating role of applicants' entrepreneurial behaviors. *Journal of Business Venturing*, *32*, 588–610. https://doi.org/10.1016/j.jbusvent.2017.05.001
Moser, K. J., Tumasjan, A., & Cable, D. (2020, June). Don't be so emotional: How social media communication affects potential applicants' engagement. In *Paper presented at the 28th European conference on information systems (ECIS)*. Marrakech: Morocco. https://aisel.aisnet.org/ecis2020_rp/123
Neuendorf, K. A. (2016). *The content analysis guidebook*. Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781071802878
Nielsen, M. B., & Knardahl, S. (2014). Coping strategies: A prospective study of patterns, stability, and relationships with psychological distress. *Scandinavian Journal of Psychology*, *55*, 142–150. https://doi.org/10.1111/sjop.12103
Nolan, K. P., Gohlke, M., Gilmore, J., & Rosiello, R. (2013). Examining how corporations use online job ads to communicate employer brand image information. *Corporate Reputation Review*, *16*, 300–312. https://doi.org/10.1057/crr.2013.19
Norusis, M. J. (2011). *IBM SPSS statistics 19 procedures dompanion*. Austin, TX: Addison Wesley.
Nunnally, J. C., & Bernstein, I. (1994). The assessment of reliability. *Psychometric Theory*, *3*, 248–292.
Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: The University of Texas at Austin.
Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001 manual*. Mahwah, NJ: Erlbaum.
Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041
Robertson, A., & Khatibi, A. (2012). By design or by default: Creating the employer identity. *IUP Journal of Brand Management*, *9*, 31–47.
Rossolatos, G. (2013). An anatomy of the verbo-visual rhetorical landscape of the world's most valuable brands. *International Journal of Marketing Semiotics*, *1*, 73–125. https://doi.org/10.2139/ssrn.2251819

Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., & Parkinson, J. (2015). Using two-step cluster analysis to identify homogeneous physical activity groups. *Marketing Intelligence & Planning*, *33*, 522–537. https://doi.org/10.1108/mip-03-2014-0050

Saini, G. K., Rai, P., & Chaudhary, M. K. (2014). What do best employer surveys reveal about employer branding and intention to apply? *Journal of Brand Management*, *21*, 95–111. https://doi.org/10.1057/bm.2013.10

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, *22*, 941–968. https://doi.org/10.1177/1094428118773858

Schnidman, A., Hester, L., & Pluntke, P. (2017). Global recruiting trends 2017. *LinkedIn Talent Solutions*, 1–32. Retrieved from https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/resources/pdfs/linkedin-global-recruiting-trends-report.pdf.

Schreurs, B., Druart, C., Proost, K., & De Witte, K. (2009). Symbolic attributes and organizational attractiveness: The moderating effects of applicant personality. *International Journal of Selection and Assessment*, *17*, 35–46. https://doi.org/10.1111/j.1468-2389.2009.00449.x

Schwaiger, M. (2004). Components and parameters of corporate reputation-an empirical study. *Schmalenbach Business Review*, *56*, 46–71. https://doi.org/10.1007/bf03396685

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. C. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*, 320–347. https://doi.org/10.1177/1094428109335949

Short, J. C., & Palmer, T. B. (2008). The application of DICTION to content analysis research in strategic management. *Organizational Research Methods*, *11*, 727–752. https://doi.org/10.1177/1094428107304534

Slaughter, J. E., & Greguras, G. J. (2009). Initial attraction to organizations: The influence of trait inferences. *International Journal of Selection and Assessment*, *17*, 1–18. https://doi.org/10.1111/j.1468-2389.2009.00447.x

Slaughter, J. E., Zickar, M. J., Highhouse, S., & Mohr, D. C. (2004). Personality trait inferences about organizations: Development of a measure and assessment of construct validity. *Journal of Applied Psychology*, *89*, 85–103. https://doi.org/10.1037/0021-9010.89.1.85

Stevenson, R. L. (2001). In praise of dumb clerks: Computer-assisted content. In M. West (Ed.), *Theory, method, and practice in computer content analysis* (Vol. 16, pp. 4–12). Westport, CI: Ablex Publishing.

Tarasewich, P., Daniel, H. Z., & Griffin, H. E. (2001). Aesthetics and web site design. *Quarterly Journal of Electronic Commerce*, *2*, 67–82.

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*, 24–54. https://doi.org/10.1177/0261927x09351676

Thellefsen, T., & Sørensen, A. C. (2007). A semiotic note on branding. *Cybernetics & Human Knowing*, *14*, 59–69.

Theurer, C. P., Tumasjan, A., Welpe, I. M., & Lievens, F. (2018). Employer branding: A brand equity based literature review and research agenda. *International Journal of Management Reviews*, *20*, 155–179. https://doi.org/10.1111/ijmr.12121

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*, 127–145. https://doi.org/10.1016/s0953-5438(00)00031-x

Tumasjan, A., Kunze, F., Bruch, H., & Welpe, I. M. (2020). Linking employer branding orientation and firm performance: Testing a dual mediation route of recruitment efficiency and positive affective climate. *Human Resource Management*, *59*, 83–99. https://doi.org/10.1002/hrm.21980

Tumasjan, A., Strobel, M., & Welpe, I. (2011). Employer brand building for start-ups: Which job attributes do employees value most? *Journal of Business Economics/Zeitschrift für Betriebswirtschaft*, *81*, 111–136. https://doi.org/10.1007/s11573-011-0507-2

Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology*, *65*, 597–660. https://doi.org/10.1111/j.1744-6570.2012.01254.x

van der Heijden, H. (2003). Factors influencing the usage of websites: The case of a generic portal in The Netherlands. *Information & Management*, *40*, 541–549. https://doi.org/10.1016/s0378-7206(02)00079-4

van Hoye, G., Bas, T., Cromheecke, S., & Lievens, F. (2013). The instrumental and symbolic dimensions of organisations' image as an employer: A large scale field study on employer branding in Turkey. *Applied Psychology*, *62*, 543–557. https://doi.org/10.1111/j.1464-0597.2012.00495.x

van Hoye, G., & Lievens, F. (2009). Tapping the grapevine: A closer look at word-of-mouth as a recruitment source. *Journal of Applied Psychology*, *94*, 341–352. https://doi.org/10.1037/a0014066

van Hoye, G., & Saks, A. M. (2011). The instrumental-symbolic framework: Organisational image and attractiveness of potential applicants and their companions at a job fair. *Applied Psychology*, *60*, 311–335. https://doi.org/10.1111/j.1464-0597.2010.00437.x

Weber, R. P. (1990). *Basic content analysis*. Newbury Park, CA: Sage. https://doi.org/10.4135/9781412983488

Westfall, B. (2017). *How job seekers use Glassdoor reviews*. Retrieved from https://www.softwareadvice.com/resources/job-seekers-use-glassdoor-reviews/.

Wilden, R., Gudergan, S., & Lings, I. (2010). Employer branding: Strategic implications for staff recruitment. *Journal of Marketing Management*, *26*, 56–73. https://doi.org/10.1080/02672570903577091

Williamson, I. O., King, J. E., Lepak, D., & Sarma, A. (2010). Firm reputation, recruitment web sites, and attracting applicants. *Human Resource Management*, *49*, 669–687. https://doi.org/10.1002/hrm.20379

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Theurer, C. P., Schäpers, P., Tumasjan, A., Welpe, I., & Lievens, F. (2022). What you see is what you get? Measuring companies' projected employer image attributes via companies' employment webpages. *Human Resource Management*, *61*(5), 543–561. https://doi.org/10.1002/hrm.22085