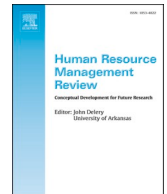




ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Human Resource Management Review

journal homepage: [www.elsevier.com/locate/hrmr](http://www.elsevier.com/locate/hrmr)

# A review of text analysis in human resource management research: Methodological diversity, constructs identified, and validation best practices

Emily D. Campion<sup>a,\*</sup>, Michael A. Campion<sup>b</sup>

<sup>a</sup> Department of Management and Entrepreneurship, Tippie College of Business, University of Iowa, United States

<sup>b</sup> Department of Organizational Behavior and Human Resources, Daniels School of Management, Purdue University, United States

## ARTICLE INFO

## Keywords:

Text analysis  
Construct validity  
Machine learning  
Natural language processing  
Systematic review

## ABSTRACT

Discovering and producing reliable and valid measures of psychological constructs are central aims for human resource management (HRM) researchers and practitioners. While HRM researchers have historically relied on traditional *quantitative* methods, increased accessibility of text analysis techniques enabled by advancements in machine learning make *qualitative* data more convenient to analyze and include in decision-making processes. In this review, we systematically analyze research in HRM, organizational behavior, strategy, and entrepreneurship that has used text analysis to uncover and/or measure constructs. Our goals are to 1) delineate types of text analyses (categorization, dictionaries, supervised machine learning, and unsupervised machine learning), 2) review what constructs can be derived from text data, 3) describe how those constructs have contributed to the core HRM functions, 4) provide guidance on validation efforts that are needed to trust inferences made, and 5) and identify future research opportunities to use text analysis by HRM function. We support these points by conducting two text analyses on the papers in our review: a hand-coded content analysis using an existing framework and building a topic model of the abstracts. We find that while there is convergence (triangulation), there is notable divergence such that the topic model revealed more nuanced and useful clustering in significantly less time, thus illustrating the value of different types of text analysis. We encourage HRM researchers and practitioners to use machine learning to increase efficiency, reduce subjectivity, increase replicability, and facilitate methodological diversity. We close with a brief discussion on the promise of large language models.

A primary goal of organizational scholars is to generate usable theory to understand and improve the experiences of people and processes at work (Sutton & Staw, 1995). Central to this aim is discovering and producing reliable and valid measures of psychological constructs. While human resource management (HRM) researchers have historically relied on quantitative (numeric) data, a powerful

\* Corresponding author at: W224 Pappajohn Business Building, University of Iowa, Iowa City, IA 52242, United States.  
E-mail address: [campion@uiowa.edu](mailto:campion@uiowa.edu) (E.D. Campion).

<https://doi.org/10.1016/j.hrmr.2025.101078>

Received 8 June 2023; Received in revised form 20 December 2024; Accepted 31 January 2025

Available online 14 February 2025

1053-4822/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

and less commonly used type of data are qualitative (text) data. These data require text analysis to derive meaning. Broadly defined, *text analysis* refers to a family of techniques for identifying, summarizing, or measuring topics, themes, or other dimensions of interest in a corpus using a systematic method. Techniques range from hand-coded content analyses and grounded theory to natural language processing (NLP) enabled by machine learning. Text analysis has been crucial to construct identification and theory building. In their foundational article on methodological fit, [Edmondson and McManus \(2007\)](#) assert that theory development typically begins with the observation of new phenomena and that the eventual quantification of such concepts is the result of qualitative groundwork. By extension, we can use text analysis as a way of creating alternative measures of existing constructs, which engenders confidence that the construct exists through triangulation (deductively). We can also use text analysis to promote different theoretical lenses by using a different method on the same text data to uncover new ideas (inductively). We refer to this triangulation and analytic diversification as “methodological diversity” ([Turner et al., 2017](#); [Welch & Piekkari, 2017](#)).

Management and organizational psychologists are not strangers to text data, yet adoption of text analysis has been slow for several reasons. One is the field’s historic prioritization of deductive methods of discovery focusing on quantitative measurement and empirical validity, which might create a perceived incompatibility between inductive text analyses and traditional psychometrics. Another explanation is that text analysis has traditionally required qualitative methods (e.g., grounded theory, ethnographies, case studies; we refer to these approaches in aggregate as “purely qualitative methods” hereafter) in which few graduates receive formal training ([Glaser, 1999](#)). Further, purely qualitative methods have been charged as being less rigorous than quantitative methods creating tension around text data ([Pratt et al., 2020](#)). Developments in advanced text analysis such as NLP may be viewed more favorably due to the perceived objectivity, reproducibility, and potential for quantification, but few researchers have been trained in NLP because it is relatively new to the field of HRM and many view it as being a “black box”<sup>1</sup> ([Campion & Campion, 2023](#)). Such perspectives create misunderstanding as to the value-add of text data, distrust of text analysis, and unnecessary obstacles between HRM researchers and practitioners and a rich data source ([Cheng & Hackett, 2021](#)).

We contend that research in HRM could benefit by expanding its use of text analysis, and particularly of NLP methods, for at least two reasons. First, while we have noted that text-based data have historically been under the province of researchers who use purely qualitative methods, scholarship has shown that qualitative data can be analyzed using NLP to measure job-related knowledge, skills, abilities, and other characteristics (e.g., values, motivation) (KSAOs) that HRM practitioners and researchers can use to improve talent management functions (e.g., [Campion et al., 2016, 2024](#); [Koenig et al., 2023](#); [Speer, 2018](#)). This is valuable because HRM practitioners and researchers have relied almost entirely on employment tests and interviews where text data are immediately quantified through response choices or human judgment to measure KSAOs because text-based data has been resource-intensive to score especially with the exponentially increasing amount of data (the “big data” movement). This means, 1) automated text analysis methods may enable HRM practitioners and researchers to measure additional job-related constructs from applicant and employee language in a systematic, objective, and efficient manner, and 2) we can approach the text data inductively or deductively, which enables flexibility to meet the researcher’s needs such as assessing whether a target construct is present in the data (deductive) or uncovering new or previously unmeasured constructs (inductive). Second, the increasing availability of NLP techniques makes it more accessible and perhaps more acceptable to those who prefer quantified measures ([Campion & Campion, 2020](#)). Although text analysis has traditionally been used in inductive endeavors by purely qualitative researchers, NLP methods permit quicker and reproducible induction for quantitative scholars and it can also be used for deduction, which broadens the opportunities for HRM researchers.

Given these potential benefits of using NLP, the current research contributes to the HRM literature in three ways. First, we describe the methods of analyzing text data and, with a focus on NLP techniques, we discuss how HRM researchers can use these to improve efficiencies and the acceptance of text analysis by reducing subjectivity, increasing replicability, and, where relevant, quantifying constructs so they can be statistically analyzed. We contend that by encouraging scholars to collect or use the text data they already have or have access to in the research context, they may be able to improve measurement of existing constructs, triangulate findings, and diversify analyses to uncover new constructs. Moreover, as noted, text data have often been perceived as cumbersome to score and include in predictive models ([Campion et al., 2024](#)). Thus, it is consequential to equip HRM researchers and practitioners with information on how they can automate analyses to leverage these underutilized but rich data sources.

Our second contribution is to identify the constructs discovered or measured via text analysis in management research and their relevance to HRM. We show how text analysis has already benefitted HRM and other areas of management research (organizational behavior, strategy, and entrepreneurship) through improved or new measurement of constructs to build, elaborate, and refine theory. We organize the constructs according to their ability to inform research on the primary HRM functions (job analysis, selection, performance management, etc.). We then conduct two text analyses. In the first, we use a deductive approach to manually categorize the constructs measured via text analysis in our review based on [Bosco and Uggerslev’s](#) metaBUS taxonomy (“meta-analysis” and “omnibus”; a cloud-based platform that extracts and synthesizes findings from social science research [<http://metabus.org/>]). In the second, we adopt an inductive approach and use NLP to build a topic model based on the abstracts of the papers to evaluate the domains these methods have advanced. In comparing findings, we illustrate how methodological diversity enriches our scholarship via triangulation (replication of results with a new method) and analytic diversification (to uncover new ideas).

Finally, to be prescriptive, we provide guidance as to what types of analyses should be pursued given one’s research goals. Key to the decision is distinguishing between summarization and prediction, which could be described as “qualitative” and “quantitative,” respectively. In light of our goal to encourage use of these methods, we also advance validation best practices. Validation evaluates the

<sup>1</sup> “Black box” refers to vaguely understood statistical models. This perception is common in many applications of artificial intelligence not only due to the complexity of the algorithms, but also the perceived sole focus on prediction to the neglect of interpretability to the non-researcher.

quality of the applications of text analysis using well-known standards. Where text data are quantified to predict or score data, we critically assess authors' efforts to demonstrate construct validity<sup>2</sup> based on established principles ([Standards for Educational and Psychological Testing](#) ["Standards"], 2014); [Uniform Guidelines on Employee Selection Procedures](#) ["Uniform Guidelines"], 1978; [Principles for the Validation and Use of Personnel Selection Procedures](#), 2018). Where text data are retained qualitatively using inductive NLP methods (e.g., topic modeling), we also discuss steps to assure validity based on the principles cited above. In doing so, we hope to correct misperceptions that text analysis and foundational psychometrics are incompatible, and that text analysis may be less rigorous.

## 1. Methods of text analysis

*Text analysis* refers to a family of techniques used to uncover the meaning of text data. Although approaches vary in complexity, all techniques reduce corpora (collection of words) to coherent categories (variables) representative of the ideas in the text. This can be done at any level (word, phrase, line, sentence, paragraph, document) and each response can be grouped into one or more categories. Researchers can use these methods to develop theory from narrative and other text-based data from candidate applications and resumes, interviews, observations, employee surveys, performance reviews, employee or managerial communications, archival records, and other sources. Because the goal of this review is to encourage the use of advanced methods, we use "level of automation" as the primary way to distinguish the four types of methods. While this is a simple to think about their methodological options, in simplifying, we may inadvertently neglect other relevant considerations. To address this in part, we discuss how the purpose of the analysis—to predict or to summarize—is crucial when choosing a method. This will be discussed in the section titled, "Selecting Text Analysis Methods."

Low or no automation approaches refer to purely qualitative hand-coding techniques; and those using high levels of automation refer to NLP approaches. Low automation approaches aim to categorize the data via the application of codes and may or may not use software to help manage the coding scheme (e.g., NVivo, Atlas.ti). Codes (e.g., themes, categories) can either be known a priori (deductively) or emerge from the data (inductively). Identification of patterns and themes among the codes is driven by the human researcher. Included in this type of analysis are content analyses where the researcher identifies categories and codes the data with those labels, as well as more complex grounded theory techniques where the researcher iterates between the data and emerging theory ([Gioia et al., 2013](#); [Murphy et al., 2017](#)). Low automation approaches are typically used to summarize data, but in rare instances may be used to quantify it for use in statistical tests. HRM examples from our review include technology interdependence in work ([Bailey et al., 2010](#)), how knowledge is transferred in teams ([Ben-Menaheem et al., 2016](#)), and the influence of power asymmetries in online work environments ([Curchod et al., 2020](#)).

High automation approaches, on the other hand, refer to techniques where an algorithm is applied to score or summarize data. This is generally executed using *NLP* (sometimes referred to as computational linguistics) is a common machine learning technique used to derive meaning from text by analyzing words and relationships among words and sentences ("Natural Language Processing," 2019). While there are several ways to conceptualize machine learning, we adopt the following definition: "Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention [once developed]" ([Machine Learning: What It Is and Why It Matters, 2019](#), brackets added). A goal of machine learning is to automate resource-intensive procedures to replicate human-like judgments to measure or discover constructs and relationships among them.

Currently, the most popular NLP technique in our field are dictionaries developed a priori to capture a specific construct and represent a simpler form of NLP. Dictionaries include single (e.g., "career") or multi-word terms (e.g., "career achievement"). A dictionary applied to a corpus yields counts such that the "occurrence of specific words from an existing list indicates the salience of a construct" ([Short et al., 2018](#), p. 420). Scores are then used in predictive models. Existing dictionaries that measure a range of constructs are available (e.g., LIWC; [Pennebaker et al., 2007](#)), but many researchers have developed their own for specific constructs. HRM examples from our review include measuring competencies ([Speer et al., 2019](#)), teamwork processes ([Mathieu et al., 2021](#)), and adaptability within a work culture ([Chatman et al., 2014](#)).

More advanced NLP methods often refer to supervised or unsupervised machine learning. Both methods derive semantically related groupings of words from the corpus to uncover patterns in the data ([Eichstaedt et al., 2021](#)). The key difference is that supervised machine learning includes a label such that classification ("ground truth") of the data is known a priori and the algorithm determines the optimal pattern to predict the correct classification (e.g., criterion, outcome, known status). The ground truth can also be a continuous variable, such as a rater score. Thus, supervised machine learning usually refers to building a model to predict an outcome that can be applied to new, non-labeled data. HRM examples from our review are predicting job performance (e.g., [Koenig et al., 2023](#)), turnover (e.g., [Sajjadijani et al., 2019](#)), and supervisor communication style (e.g., [Meinecke & Kauffeld, 2019](#)).

Unsupervised machine learning is used when there are no labels to develop a model to predict. Instead, the algorithm identifies

<sup>2</sup> While we include purely qualitative methods in our review because it is a form of text analysis and constructs uncovered are relevant to HRM, we do not review the validity efforts (referred to as "trustworthiness" in that scholarship; [Pratt et al., 2020](#); presented in our online supplement) of those papers because the focus of the paper and likely applicability of text analysis to the domain of HRM research and practice, which adheres to the Standards and Principles and is primarily focused on using NLP for prediction. Nevertheless, the advent of tools powered by Large Language Models has enabled automated summarization for HRM researchers. Those validation efforts would need to adhere to the first stages of the Standards and Principles to demonstrate evidence of validity, which we discuss in the "Validation of Text Analysis" section below.

patterns in the data from which the researcher imposes meaning. This is typically accomplished using topic modeling (Hannigan et al., 2019), which automates categorization through analysis of  $n$ -grams and relationships among. Some argue that unsupervised machine learning is less prone to human subjectivity than manual procedures. For example, in their study of more than 52,000 business descriptions over a 10-year period, Menon et al. (2018) wrote that this method “eliminates bias introduced by human coders and enables researchers to discover unknown thematic structures latent in a large textual data set” (p. 15). Nevertheless, unsupervised machine learning still requires substantial human participation to select the optimal topic model (number of topics) and derive meaning. HRM examples from our review include identifying recruitment signals from websites (Banks et al., 2019), understanding public sentiment to enforced remote work during COVID-19 (Zhang et al., 2021), and uncovering social coping mechanisms in response to work stressors (Sajjadiani et al., 2024).

While a more technical description of each method is outside the scope of the current review, we encourage interested readers to locate additional information in our online supplement<sup>3</sup> and refer to exemplar work in our field. This area advances very quickly, but the following papers should be a valuable place to begin for novices, though this list is far from exhaustive: Banks et al. (2018), Mathieu et al. (2021), Campion and Campion (2023), Campion et al. (2024), Eichstaedt et al. (2021), Hannigan et al. (2019), Hickman, Bosch, et al. (2022), Short et al. (2018), Speer et al. (2023), and Valtonen et al. (2022).

## 2. HRM constructs and text analysis

In this section, we describe our literature review method. We follow this with an object lesson on the value of text analysis using two types of text analysis: deductive categorization of constructs by human judgment and unsupervised machine learning on the abstracts. This demonstrates how the categories and themes identified by the distinct methods converge, but also how the methods differ in what they offer theoretically. We then describe the relevance of the results to core HRM functions, drawing out implications for practice and future research.

### 2.1. Literature review method

To generate a comprehensive database, we searched terms related to text analysis, particularly focusing on research that uses computer-assisted text analysis (e.g., computer-aided text analysis, computerized content analysis, sentiment analysis, machine learning, natural language processing), as well as types of software (e.g., Linguistic Inquiry Word Count [LIWC]), in 17 high-quality management and organizational psychology journals.<sup>4</sup> We selected the journals based on the following criteria: (a) those containing original empirical text analysis research in management, (b) represent the range of management journals, (c) include all or most of the top-cited journals, and (d) yield a large sample of studies that represent the domain of relevant studies (Hiebl, 2023). This resulted in 963 articles from 1990 to 2023 across human resources, organizational behavior, strategy, and entrepreneurship. In early 2024, we updated the search specifically focusing on the use of NLP in HRM and added 27 articles, bringing the total to 990. We read all titles, abstracts, methods, appendices, and online supplements to ensure articles applied text analysis and retained 583 articles. Of these, 560 (96.05 %) used text analysis in primary studies and 23 (3.95 %) provided illustrations on how to conduct text analysis. We did not retain the remaining 407 because they were commentaries; cited a text analysis search term, possibly as a future direction, but did not conduct text analysis; or focused on the interface between humans and algorithms but did not conduct text analysis.

We coded the following variables for each article: author(s), year, title, journal, abstract, theoretical perspective, whether they retained the data qualitatively or if they quantified, type of text data, sample description, intended constructs measured, text analysis software, type of text analysis used, whether subject matter experts (SMEs) were used to code or interpret, whether psychometrics on text variables were calculated (e.g., agreement among coders), whether correlations were presented with similar measures to yield information on the nature of text constructs, whether text variables were used to predict other variables or were being predicted to bear on the interpretation of constructs measured, and a brief description of the findings. In the initial round of coding (1990–2018), the first author coded all papers, and the second author cross-checked a 10 % sample to ensure the quality of the coding (Bluhm et al., 2011; Wilhelmy et al., 2016). We discussed differences to consensus based on our emerging understanding of the literature. In the second round (2019–2022), we trained two independent coders to code the papers and the first author cross-checked the coding of all new papers. In the third round (2022–2024), the first author coded the papers with the second author cross-checking a 10 % sample.

Of the papers in our review, 258 (44.25 %) scored their text data for use in a predictive model, 330 (56.60 %) summarized their data, and 11 (1.89 %) used both approaches. This finding offers initial evidence of two groups of studies and a key distinction in the use of text analysis: one summarizing the data emerging from purely qualitative methods and the other scoring the data to make predictions from psychology’s statistical tradition. The two groups of studies differ largely by “should we quantify all data?” The two groups of studies follow different approaches to ensure quality. Whereas quantitative analyses conform to tenets of validity, qualitative analyses aim for “trustworthiness” (Pratt et al., 2020), which they support through procedural considerations (e.g., coding, member

<sup>3</sup> Online supplement link: [https://osf.io/b3hmz/?view\\_only=251e1e5bc55d434fb6fa3a30e5fb6169](https://osf.io/b3hmz/?view_only=251e1e5bc55d434fb6fa3a30e5fb6169)

<sup>4</sup> We searched the terms in the following journals: *Academy of Management Journal*, *Academy of Management Proceedings*, *Administrative Science Quarterly*, *Entrepreneurship Theory and Practice*, *Human Relations*, *Human Resource Management Review*, *Journal of Applied Psychology*, *Journal of Business Venturing*, *Journal of Business and Psychology*, *Journal of International Business Studies*, *Journal of Management*, *Journal of Organizational Behavior*, *Organizational Behavior and Human Decision Processes*, *Organizational Research Methods*, *Organization Science*, *Personnel Psychology*, and *Strategic Management Journal*.

checks). In reviewing the differences in recommendations for validity in this review, it is obvious that 1) the core tenets of validity exist (how can you trust the conclusion?) for both types of studies, but the approaches vary; and 2) without quantification, several of the traditional psychometric validation steps are impossible to complete. Therefore, we argue that the two groups differ in principle (different assumptions about how to observe, measure, and interpret human behavior, affect, and cognitions) and in some validation requirements, but we encourage HRM researchers to view the methods as complementary with the common interest in text data as an opportunity to engage in methodological diversity.

## 2.2. Identifying categories of constructs

To summarize the constructs and show their relevance to HRM research, we conducted two text analyses and aligned the results with HRM functions. Technical details are in the online supplement. In the first analysis, we used a deductive, hand-coding categorization technique (low automation) by grouping constructs according to an existing taxonomy. We categorized the attributes using Bosco and Uggerslev's metaBUS taxonomy because it is backed by extensive research, well-accepted, and allows for logical

**Table 1**  
Categorization of Attributes Measured by Text Analysis Based on Existing Taxonomy of Constructs in the Literature.

Category	Count; Percent	Examples from the Literature Review
1. Behaviors	182; 31.54 %	<b>HR-Specific Examples:</b> Performance evaluations; work experience; job tasks; skills  <b>Examples from Related Disciplines:</b> Boundary management practices, tactics, and work; language and rhetoric (rational, normative, creativity-relevant, political, tone of, cues); performance (ritual, team); job search
2. Attitudes and Evaluations	171; 29.64 %	<b>HR-Specific Examples:</b> Engagement; job satisfaction; embeddedness; person-environment fit  <b>Examples from Related Disciplines:</b> Identity (organizational, authenticity of, marginalized, split, dual) and identity work; media optimism, favorability, or tenor; emotions; sentiment; just and unjust treatment from supervisors; sensemaking; tone; legitimacy
3. Processes	141; 24.44 %	<b>HR-Specific Examples:</b> Team processes; identity creation and maintenance; organizational formation and change  <b>Examples from Related Disciplines:</b> Network creation; dynamics of organizational routines; organizational legitimation; preference shifts
4. Organizational Characteristics	130; 22.53 %	<b>HR-Specific Examples:</b> Organizational recruiting brand; organizational fit; organizational culture and values  <b>Examples from Related Disciplines:</b> Organizational strategies; tone of crowdfunding campaigns; business ecosystems; innovation patterns; organizational competition and cooperation; organizational orientations (exploration, clock-time, social value, economic value)
5. Person Characteristics	99; 17.61 %	<b>HR-Specific Examples:</b> Knowledge; skills (cognitive ability, communication, critical thinking, leadership); abilities; personality traits  <b>Examples from Related Disciplines:</b> CEO personality (narcissism, extraversion); psychological capital; entrepreneurial orientation; adaptability; clout
6. Cognitions	60; 10.40 %	<b>HR-Specific Examples:</b> Cognitive processing and flexibility  <b>Examples from Related Disciplines:</b> Schemas, framings, and rules; team cognitive maps; strategic knowledge structures; CEO mental models, attention, and temporal focus
7. Dyad or Group Characteristics	22; 3.81 %	<b>HR-Specific Examples:</b> Teamwork characteristics; team composition; team faultlines  <b>Examples from Related Disciplines:</b> Top management teams (modesty, cognitive maps, attention patterns); equity distribution among team members; strategic leadership constellations; power asymmetries
8. Miscellaneous	22; 3.81 %	<b>Examples from Related Disciplines:</b> Descriptions of manuscript decision letters; structure and themes of academic articles; lending criteria of loan officers
9. HR Practices	23; 3.99 %	<b>HR-Specific Examples:</b> Recruitment (organizational signals, language-sensitive); sentiment of performance review narratives; selection and assessments; organizational engagement policies; interviewer impression management; candidate materials (accomplishment records, resumes, cover letters)
10. Contextual Characteristics	17; 2.95 %	<b>HR-Specific Examples:</b> Virtual/remote work; institutional context; organizational reputation  <b>Examples from Related Disciplines:</b> Disruptive innovation; technology interdependence; workgroup context; culture-specific problems
11. Intentions	15; 2.60 %	<b>HR-Specific Examples:</b> Turnover intentions; job search intentions  <b>Examples from Related Disciplines:</b> Qualified immigrants' motivation to migrate; intrinsic and extrinsic language cues; motivators and de-motivators
12. Occupational Characteristics	8; 1.39 %	<b>HR-Specific Examples:</b> Job analysis and design  <b>Examples from Related Disciplines:</b> Occupational value maintenance, value of occupational artifacts

*N* = 577 articles that used text analysis.

*Note.* Categories based on manual coding into the metaBUS categories (Bosco et al., 2017). Some constructs were coded in multiple categories. Percent = Percent of papers in review.

comparisons among representative categories (Bosco et al., 2017). Though we found text analysis has been used to measure attributes from all metaBUS categories, we also identified an additional category we refer to as “processes” because many of the purely qualitative papers in our review presented topical content as processes. For example, Caza et al. (2018) built a process model via grounded theory on how individuals who hold more than one job reconciled multiple work identities. Table 1 shows the counts and frequencies of constructs by metaBUS categories with example constructs from the review. In the second analysis, we used an inductive, NLP method called topic modeling (high automation) on the abstracts (see OSF for technical details). We used unsupervised machine learning because it is data-driven where patterns in the data are uncovered by an algorithm to create topics that researchers interpret (Eichstaedt et al., 2021). This contrasts with the approach in our first analysis where we imposed a set of categories developed a priori to generate meaning.

Table 2 includes the topics, names, and descriptions of the topic model, as well as the link between the topic model results and metaBUS categorization. While the metaBUS categorization was based on constructs and the topic model was based on abstract, these linkages nevertheless illustrate convergence between the analyses suggesting that the groupings replicate to some degree. This is a particularly relevant finding for HRM researchers using large data sets where hand coding is not feasible. For most of the topics in the topic model, there is a similar metaBUS category. For example, team performance and culture aligns with metaBUS’s dyad or group characteristics; institutional development and maintenance aligns with organizational characteristics; and identity development and change aligns with cognitions and processes. Yet, the topic model results are often more nuanced, which means metaBUS categories provide more high-level and broader characterizations that may appear in, but are not wholly encompassed by, a topic. For example, signaling by ventures and individuals does not align neatly with any metaBUS categories, but instead likely contains pieces of several, including personal characteristics, organizational characteristics, behaviors, and cognitions.

The theoretical richness of the topic model and the process through which researchers make sense of the topics reflects more closely how researchers who use purely qualitative methods engage with their data. As the authors did with this topic model, researchers must iterate between the data source (abstracts) and results (top terms, groups of abstracts by topic), develop labels and descriptions that unite the top terms and data, and ensure the validity of the results using SMEs. The inductive process of deriving meaning from topic models lends itself to unexpected patterns and conceptual leaps that would otherwise be overlooked in a deductive approach. For example, topic 15 grouped a set of abstracts on ecosystems (e.g., Jacobides et al., 2018) and social movements (e.g., Schifeling & Soderstrom, 2022), which brought these ideas together through the lens of challenging existing institutions within structured, hierarchical, and resistant ecosystems (e.g., Claus & Tracey, 2020). These ideas can be viewed through a strategic HRM lens wherein firms’ participation in social movements and larger ecosystems may influence applicant attraction and retention. In another example, topic 3 cohered at the intersection of papers on customer evaluations and incivility (e.g., Curchod et al., 2020) and how managerial systems exert control over workers (e.g., Cameron & Rahman, 2022) representing ideas around power asymmetries between service workers and their customers, as well as employees and management systems. Ideas from this topic have clear connections to job analysis and design, retention, and performance management.

Considering the results of these analyses simultaneously, the topic model offered intricacies specific to the data set that could not be as richly represented in metaBUS and illustrates the potential for discovery by using NLP. Making sense of the categorization is straightforward given the metaBUS categories are from a framework developed to encompass the domain of management constructs and therefore, by nature, may be broader and more easily accessible. Meanwhile, making sense of the topic model requires researchers to participate in deeper theoretical work to uncover the latent pattern identified by the algorithm.

Of extreme importance is the time cost of each method. Discounting the time to find all the articles and review them to understand the state of the literature, which would be necessary for both analyses, classifying the articles into metaBUS took approximately four months of researcher and research assistant hours to carefully read the papers, code information, and make judgments. In comparison, creating the topic model on the abstracts required about two days of researcher time, plus another day to iterate with SMEs. As such, the efficiencies of automated text analysis in the form of unsupervised machine learning are undeniable, making it an appealing and valuable tool for HRM researchers and practitioners. The initial time investment to learn how to use automated text analysis depends on the type of analysis. For an HRM researcher with typical skills, coding papers and classifying attributes is elementary, but time consuming. Meanwhile, learning how to run an unsupervised model without any prior experience might take many weeks, but is a one-time cost. This could be reduced substantially given the accessibility of Large Language Models (LLMs) that can help researchers produce code to program their own topic models, or even help summarize their text data directly. On this last point, we want to be cautious to not suggest that using LLMs to summarize text data is a replacement for human interpretation. Even with the aid of LLMs, researchers still hold the specialized domain knowledge and methodological training to ensure quality and relevance of LLM output. These tools will facilitate analytic efficiency, just as we have seen in previous historic shifts with tools such as SPSS or Excel. These programs reduce the barriers for researchers to access advanced methods, which propels science forward, but does not replace the crucial role humans play in making sense of the findings in light of the research question, the literature, or the context.

To more acutely draw the connections between the constructs and topics uncovered in our analyses and HRM functions, we review these findings in light of an HRM framework. Table 3 includes an overview of how these categories and topics align with HRM functions. Although some of the constructs are broader organizational-level topics, all of them improve understanding of employees in general and so are also relevant to HRM.

### 2.3. Relevance of constructs to HRM practice and future research

In this section, we describe the findings of the text analyses as they pertain to HRM functions, and identify areas of future research in HRM enabled by text analysis. Interested readers will find an interactive table of the 15 most relevant papers by HRM function in the

**Table 2**  
Categorization of Attributes Identified by Text Analysis Based on a Topic Model of the Abstracts of the Papers in the Review.

Topic No.	Topic (Alphabetized)	Top 15 Topic Terms	Topic Description	Matching Categories of Constructs from metaBUS
1	Chief executive officer influence on organizational outcomes	executive, chief, officers, officer, firms, management, top, orientation, focus, firm, reputation, promotion, regulatory, upper, rate	how CEO attributes influence organizations including media appearances, inferred personality, entrepreneurial orientation, and perceived commitment	<ul style="list-style-type: none"> <li>Organizational Characteristics</li> <li>Behaviors</li> </ul>
2	Creativity and resource generation	creativity, returns, word, forward, believed, arena, followed, looking, guidance, advancement, move, unfortunately, methodological, outlined, kirkman	how creativity is fostered in organizations, how external factors facilitate creativity, and creative ways in which resources are created (otherwise a bit incoherent with several papers being solely about text analysis methods rather than psychological constructs)	<ul style="list-style-type: none"> <li>Personal Characteristics</li> <li>Behaviors</li> <li>Contextual Characteristics</li> </ul>
3	Customer-facing labor and power asymmetry	employees, individuals, control, ability, workplace, people, customer, customers, experience, working, power, participants, experiences, way, health	how workers manage customer interactions, particularly when they are negative (e.g., customer incivility, discriminatory) and power differentials are apparent (such that customers appear to have more power)	<ul style="list-style-type: none"> <li>Personal Characteristics</li> <li>Behaviors</li> <li>Attitudes and Evaluations</li> <li>Cognitions</li> <li>Occupational Characteristics</li> </ul>
4	Ethical challenges and emotional expression	ethical, expressed, competing, respondents, party, productive, morally, generates, mediation, ideals, counterproductive, anger, expressions, aware, vulnerable	on the challenge of maintaining ethical principles and the effects of ethical scandals in organizations; particularly relevant to emotional expression and the language used to express emotion	<ul style="list-style-type: none"> <li>Behaviors</li> <li>Occupational Characteristics</li> </ul>
5	Gender differences	women, gender, female, leaders, differences, gain, tasks, men, misconduct, states, consistent, micro, measures, predictions, shaped	gender differences in management research, especially as it relates to leadership (e.g., women-owned businesses)	<ul style="list-style-type: none"> <li>Personal Characteristics</li> <li>Behaviors</li> <li>Attitudes and Evaluations</li> <li>Cognitions</li> </ul>
6	Identity development and change	identity, job, identities, jobs, general, search, success, national, self, turnover, authenticity, constructed, identification, members, simultaneously	how professional identities form and are influenced by environmental factors and role changes; includes authentic self-expression through work	<ul style="list-style-type: none"> <li>Personal Characteristics</li> <li>Behaviors</li> <li>Attitudes and Evaluations</li> <li>Intentions</li> <li>Cognitions</li> <li>Processes</li> <li>Occupational Characteristics</li> </ul>
7	Influences on firm change	social, find, firms, change, firm, theory, corporate, positive, influence, entrepreneurs, negative, less, examine, effect, support	broad topic on firm change and influence of media coverage	<ul style="list-style-type: none"> <li>Organizational Characteristics</li> <li>Processes</li> <li>Contextual Characteristics</li> </ul>
8	Institutional development and maintenance	institutional, legitimacy, markets, institutions, conversation, know, exploration, maintaining, material, actors, occupations, norms, questions, technical, global	how institutions are created and legitimized through interactions with the environment (with a special focus on multinational enterprises); also discusses legitimacy as it pertains to immigrant entrepreneurs, and legitimacy for organizational newcomers (through the lens of PE fit)	<ul style="list-style-type: none"> <li>Organizational Characteristics</li> <li>Processes</li> </ul>
9	Leadership characteristics and influence	leadership, employee, leaders, emotions, safety, leader, affective, positioning, justice, benefit, bias, violation, supervisors, perceive, perceptions	how leader characteristics (e.g., emotion, humility) influence the environment and outcomes	<ul style="list-style-type: none"> <li>Dyad or Group Characteristics</li> <li>Behaviors</li> <li>Attitudes and Evaluations</li> <li>Intentions</li> <li>Cognitions</li> <li>Contextual Characteristics</li> </ul>
10	Management research process	process, management, knowledge, literature, within, social, language, using, new, practices, content, understanding, human, different, learning	on the process of developing knowledge in management research	–

(continued on next page)

Table 2 (continued)

Topic No.	Topic (Alphabetized)	Top 15 Topic Terms	Topic Description	Matching Categories of Constructs from metaBUS
11	Managerial experiences	narrative, narratives, managers, managerial, network, respond, ratings, characteristics, interpersonal, companies, additionally, activity, feedback, interact, letters	on manager experiences broadly, with some articles focusing on managerial networks and others focusing on narratives (performance management narratives richness, how entrepreneurs (managers in their startups) strategically use narratives to generate funding)	<ul style="list-style-type: none"> <li>• Behaviors</li> <li>• Attitudes and Evaluations</li> <li>• Cognitions</li> <li>• HR Practices</li> </ul>
12	Managing roles and boundaries	value, professional, behaviors, public, values, boundary, competitive, role, roles, collective, level, second, communication, first, action	how workers navigate their professional roles and the boundaries between their roles	<ul style="list-style-type: none"> <li>• Personal Characteristics</li> <li>• Behaviors</li> <li>• Attitudes and Evaluations</li> <li>• Intentions</li> <li>• Cognitions</li> <li>• Processes</li> </ul>
13	Methods	text, topic, modeling, techniques, along, skilled, best, exposure, computational, ended, interested, open, dataset, quantitative, easily	how text analysis is used in management research	–
14	Multinational corporations and managing internationally	activities, multinational, corporations, international, capability, transfer, recruitment, foreign, country, tension, subsidiary, subsidiaries, enterprises, guide, emphasized	how multinational corporations navigate new or foreign country contexts, especially how they recruit and the language used in their recruitment,	<ul style="list-style-type: none"> <li>• Organizational Characteristics</li> <li>• HR Practices</li> </ul>
15	Organizational ecosystems and mobilization	hierarchical, crisis, governance, sets, cannot, activists, complement, enables, ecosystems, coordinate, operate, mobilization, rules, mainstream, ecosystem	how ecosystems are formed, sustained; how new ecosystem entrants (entrepreneurs) galvanize resources within the ecosystems; and how organizations self-mobilize in response to ecosystem disruption (e.g., social movements)	<ul style="list-style-type: none"> <li>• Organizational Characteristics</li> <li>• Processes</li> </ul>
16	Rhetoric and messaging	crowdfunding, logics, visions, funding, rhetoric, tests, racial, sensemaking, minorities, matters, broadly, facilitate, logic, campaigns, inverted	examining the rhetoric (e.g., metaphors) individuals and entrepreneurs use, how that rhetoric may reflect underlying traits (e.g., personality), and how that rhetoric subsequently influences resource generation (particularly for new ventures seeking funding)	<ul style="list-style-type: none"> <li>• Organizational Characteristics</li> <li>• Behaviors</li> <li>• Intentions</li> <li>• Cognitions</li> </ul>
17	Signaling by ventures and individuals	psychological, intervention, intensity, societal, finding, signaling, signal, embodied, norms, causes, surprise, usage, capital, assigned, financial	how entrepreneurs signal possible funders, how organizational cultural norms are signaled, and the psychological experiences (e.g., leader energy) underlying the signal	<ul style="list-style-type: none"> <li>• Personal Characteristics</li> <li>• Organizational Characteristics</li> <li>• Behaviors</li> <li>• Cognitions</li> <li>• Dyad or Group Characteristics</li> </ul>
18	Team performance and culture	team, performance, outcomes, machine, teams, self, information, category, group, high, personality, product, categories, members, others	how team processes unfold and what influence characteristics (e.g., faultlines) have on performance; special focus on entrepreneurship orientation and entrepreneurial teams	<ul style="list-style-type: none"> <li>• Behaviors</li> <li>• Cognitions</li> <li>• Organizational Characteristics</li> <li>• Occupational Characteristics</li> </ul>
19	Technology and innovation	workers, technology, occupational, goal, task, include, efficiently, technologies, interaction, places, encounters, primed, engineering, coordinating, interdependence	largely about technological innovation, how it emerges and its effects, but also on goal orientation and the ways in which technology influences goals and tasks	<ul style="list-style-type: none"> <li>• Organizational Characteristics</li> <li>• Cognitions</li> <li>• Contextual Characteristics</li> </ul>
20	Time and changing frames of organizational strategy	temporal, strategy, career, state, constructs, oriented, experiences, future, time, framing, long, frames, frame, path, dilemma	how organizations change their strategic framing over time with a focus on how the environment and institutional features facilitate that change	<ul style="list-style-type: none"> <li>• Organizational Characteristics</li> <li>• Processes</li> </ul>

Note. Topics generated using Latent Dirichlet Allocation (LDA, Blei et al., 2003). Papers could fall into one or more categories. Topics are alphabetized. More information on this process is in the online supplement.

online supplement, which reflects results from Tables 1–4. We discuss these results by first reporting on papers that use text analysis for the HRM function, describing the results of the text analyses as the findings relate to the HRM function, and identifying opportunities for future research and practice using text analysis to advance each HRM function.



**Table 3**  
Relevance of Constructs Identified in Literature Review to Core HRM Functions.

Analysis	Job Analysis & Design	Recruitment	Selection	Retention & Separation	Performance Management	Training & Development	Employee Relations	Compensation & Benefits	Strategic HRM
<b>metaBUS</b>	Occupational characteristics Contextual characteristics Personal characteristics	Occupational characteristics Organizational characteristics HR practices	Behaviors Attitudes and evaluations Organizational characteristics Personal characteristics Occupational characteristics HR practices	Behaviors Attitudes and evaluations Personal characteristics Cognitions Intentions HR practices	Behaviors Attitudes and evaluations Personal characteristics Cognitions Dyad or group characteristics Intentions HR practices	Behaviors Attitudes and evaluations Personal characteristics Dyad or group characteristics	Attitudes and evaluations Organizational characteristics HR practices	Behaviors Personal characteristics Occupational characteristics Organizational characteristics	Organizational characteristics Personal characteristics Contextual characteristics
	Managing roles and boundaries Technology and innovation Customer-facing labor and power asymmetry	Rhetoric and messaging Signaling by ventures and individuals	Gender differences Managerial experiences Leadership characteristics and influence Customer-facing labor and power asymmetry	Managing roles and boundaries Ethical challenges and emotional expression Customer-facing labor and power asymmetry	Technology and innovation Gender differences Creativity and resource generation Managing roles and boundaries Ethical challenges and emotional expression Leadership characteristics and influence Managerial experiences Customer-facing labor and power asymmetry	Technology and innovation Gender differences Ethical challenges and emotional expression Leadership characteristics and influence Identity development and change Customer-facing labor and power asymmetry	Team performance and culture Leadership characteristics and influence CEO influence	Gender differences Team performance and culture Managerial experiences CEO influence	Time and changing frames of organizational strategy Organizational ecosystems and mobilization Influences on firm change Technology and innovation Institutional development and maintenance Multinational corporations and managing internationally CEO influence Signaling by ventures and individuals Managing roles and boundaries
<b>Topic Model</b>									

### 2.3.1. Job analysis & design

Research has shown that text analysis can be used to make job analysis easier by identifying the KSAOs from job descriptions and other text-based job content (e.g., [Campion & Campion, 2025](#); [Koenig et al., 2023](#)). For example, [Putka et al. \(2023\)](#) built a machine learning model to predict the KSAOs from job descriptions in O\*NET (U.S. Department of Labor's Occupational Information Network) as well as SME ratings. As job analysis and design are the cornerstone of HRM, using text analysis to improve this process and its outcomes should increase the job-relatedness of the subsequent HRM functions, thus enhancing accuracy and making decisions more legally defensible, especially in small organizations or for jobs with few incumbents where formal job analyses may not be feasible.

The metaBUS categories relevant to job analysis include occupational, contextual, and personal characteristics, which capture the KSAOs required by occupation and the contextual characteristics (e.g., workgroup context; [Jehn & Bezrukova, 2004](#)) that influence how a job is completed. For example, [Gibson et al. \(2011\)](#) used word dictionaries to quantify virtual job characteristics (e.g., electronic dependence) to extend job characteristics model. The topics from the model show that research using text analysis has enriched our theoretical understanding of ideas related to job analysis and design including how individuals manage their work roles and boundaries ([Cruz & Meisenbach, 2018](#)), the influence of technology on collaboration ([Bailey et al., 2010](#)), and customer-facing labor and the power asymmetries that can characterize service jobs ([Curchod et al., 2020](#)). These inform the scoping of a role, how technology may facilitate design changes, and how best to re-design labor to reduce burnout.

Future research in this area is promising. Inductive methods may enable discovery of new job design features not included in current theoretical models, such as identifying best design practices around remote work and instant work-related communication, as well as uncovering how workers navigate their professional roles and the boundaries between their roles. In practical applications, these methods could help HRM managers quickly identify what job characteristics facilitate productivity based on employee narratives (e.g., engagement surveys, HR customer reaction surveys, focus groups, blog postings, and discussion boards) that could improve how jobs are designed and the technology required to successfully complete tasks.

### 2.3.2. Recruitment

Research directly on the use of text analysis techniques to understand recruitment has been limited, but nevertheless demonstrates its value. For example, [Banks et al. \(2019\)](#) used topic modeling on website content to assess recruitment signals in multinational firms, and [Peltokorpi and Vaara \(2014\)](#) analyzed the effectiveness of language-sensitive recruitment in multinational firms. Findings from our metaBUS categorization shows that research has used this family of techniques to measure constructs relevant to occupational and organizational characteristics including corporate social responsibility ([Brès & Gond, 2014](#)) and employment quality (e.g., career opportunities, compensation & benefits, [Guo et al., 2024](#)). Similarly, the topic model showed that text analysis can be especially useful when uncovering meaning behind rhetoric and messaging (CEO virtue rhetoric; [Zachary et al., 2023](#)), and organizational signals, as shown by [Banks et al. \(2019\)](#), that may influence applicant attraction and retention.

Future research in this area is ripe to improve recruitment. Theoretically, text analysis might be used to learn how professional identities form and are influenced by organizational recruitment messaging as reflected on the social media. Practically, text analysis could help HRM managers assess organizational reputation and candidate reactions from online sources (e.g., LinkedIn, Indeed, Glassdoor) to refine recruiting messages or understand why reasons candidates turn down job offers from rejection surveys. HRM managers could also summarize comments on competing firms to compare efficacy of recruitment strategies and branding.

**Table 4**  
When to Use Each Type of Text Analysis.

Goal or Context	Type of Text Analysis			
	Categorization	Dictionaries	Supervised Machine Learning	Unsupervised Machine Learning
1. Data are qualitative.	High	High	High	High
2. Measure of construct already exists.	Modest	High	Modest	Modest
3. Goal is discovery of new constructs.	High	Modest	Modest	High
4. Goal is to establish nomological net with existing constructs (convergent and divergent validity)	Low	High	High	Low
5. Goal is theory development, especially in an entirely new area or topic.	High	Modest	Modest	High
6. Goal is theory refinement or testing, especially in well-developed area.	Modest	High	High	Modest
7. Goal is to triangulate by using different methods (methodological diversity).	High	High	High	High
8. Goal is summarization.	High	Low	Low	High
9. Goal is inductive.	High	Low	Modest	High
10. Goal is prediction or scoring.	Modest	High	High	Low
11. External criterion available.	Low	Modest	High	Low
12. Goal is reproducibility.	Low	High	High	Modest
13. Sample size is large.	Low	Modest	High	High
14. Sample size is small.	High	High	Modest	Modest

### 2.3.3. Selection

At present, the use of NLP is most mature in the area of personnel selection. This is likely due to the abundance of text data available in this context paired with the increasing need to score it quickly in the face of resource constraints and competition for talent. For example, [Campion et al. \(2016, 2024\)](#) used text analysis to score work experience in applications to replicate the judgment of human raters and predict job performance; [Sajjadiani et al. \(2019\)](#) used text analysis on applications to interpret work experience relevance; and [Koenig et al. \(2023\)](#) presented several studies quantifying assessment responses, application information, and other narrative data to improve prediction in personnel selection. We only described four illustrations here, but there are many more (e.g., [Fan et al., 2023](#); [Hickman, Thapa, et al., 2022](#); [Liff et al., 2024](#)).

MetaBUS categorization highlights an array of selection-related attributes including behaviors (adaptability, communication; [Liff et al., 2024](#)), attitudes and evaluations (emotional reactions to work challenges; [Sajjadiani et al., 2024](#)), and personal characteristics (skills; [Nadkarni & Narayanan, 2005](#)), yet the topic model did not yield the same robust representation of potential application to selection. This may be because the results from the topic model were induced from the papers and most of the papers did not explicitly examine personnel selection; or because metaBUS is a framework based on all literature and its broader groupings include high-level topics that appear more immediately relevant to selection. Nevertheless, the topic model still yielded useful findings on power asymmetries and impression management during selection interviews ([Waung et al., 2017](#)) and leadership rhetoric ([Bligh et al., 2004](#)).

Despite selection being at the forefront of these methods in HRM, plenty of opportunities remain. The topic model included a topic on manager experiences that identified managerial networks as a new construct potentially relevant to selection that could be revealed by text analysis of past work experience and answers asked in applications and interviews. [Campion et al. \(2024\)](#) used text analysis to discover that there are constructs possessed more equally among racial minority subgroups, thus informing individual differences theory and practically reducing adverse impact in hiring. Another potential application is to identify candidate skills and job-related experiences using work history routinely recorded in personnel records to build skills inventories for internal recruiting and promotions. The viability of skills inventories has long been hampered by the difficulty of archiving employee skills or matching to job openings based on the qualitative nature of the data (e.g., past jobs and tasks in the company). Existing skills inventories mostly consist of a listing of past jobs held by the employee, which may have limited value in identifying internal candidates. Text analysis would allow matching between the skills in the past jobs and the jobs to be staffed. Further, while motivation and effort have historically been difficult to quantify due partly to candidate impression management, text analysis could help recruiters capture candidate motivation indirectly through the passive scoring of motivation-laden terms in interviews and applications ([Campion & Campion, 2024](#)).

### 2.3.4. Retention & separation

Turnover is another area where we have seen greater interest in using text analysis. For example, [Felps et al. \(2009\)](#) analyzed focus group transcripts to measure embeddedness and job search behaviors to understand turnover contagion, [Sajjadiani et al. \(2019\)](#) used unsupervised and supervised machine learning to uncover the history of reasons public school teachers left their employment to predict turnover in the future., and [Min et al. \(2024\)](#) used a machine learning-based recommender system to predict turnover. Our analysis shows a similar pattern as selection: there are more applicable metaBUS categories than topics from the topic model. The metaBUS categories include cognitions and intentions that may be more likely to predict an employee's intention to leave. For example, [Rothausen et al. \(2017\)](#) analyzed interviews of stayers and leavers to understand how congruent identity and well-being influenced turnover, and [Ng and Sherman \(2022\)](#) used NLP on job descriptions to evaluate intrapreneurship and its role in turnover in high-tech companies. Meanwhile, the topics that aligned with retention and separation includes managing roles and boundaries, ethical challenges and emotional expression, and customer-facing labor and power asymmetry. These make sense as mismanagement of roles and boundaries may drive workers out, as would continually navigating customer incivility ([Walker et al., 2017](#)) or ethical or moral hurdles where employees may disagree with a managerial decision or organizational policy ([Baikovich & Wasserman, 2020](#)).

There are many opportunities to improve our understanding of retention and separation in research and in practice. Theoretically, the process of withdrawal and sources of commitment are likely informed by text analysis of employee communications. Practically, predicting aggregate turnover more accurately from engagement surveys by analyzing the narrative comments is one useful area of inquiry. Another would be to develop turnover risk assessments of individual employees based on comments in career development interviews, employee statements in performance evaluations, and supervisor-documented comments made by the employees informally on the job. This would enable HRM managers to identify interventions to more effectively manage and develop high-risk employees.

### 2.3.5. Performance management

There are a few examples from the literature on the use of text analysis in performance evaluations. [Speer \(2018\)](#), [Speer \(2021\)](#) used NLP to measure attitudinal and behavioral themes in performance appraisal narrative comments, which predicted performance ratings and demonstrated the utility of enhancing performance management using manager narratives. [Meinecke and Kauffoff \(2019\)](#) assessed how supervisor-employee language-style matching from performance appraisal interviews predicted supervisors' empathic communication style and how this related to employee-rated supervisor likeability and their intention to change. Like many other HRM functions, the use of text analysis to improve these functions in research has been less frequent, yet the results from our analyses show potential. Of the HRM functions in [Table 4](#), performance management has the most metaBUS categories and topics. The papers in our review used text analysis to measure many of the metaBUS categories that are relevant to performance management including attitudes and evaluations ("work-from-anywhere" sentiment; [Choudhury et al., 2021](#)), dyad and group characteristics (individual agency and team coordination; [Mattarelli et al., 2022](#)), and intentions (achievement and affiliation language; [Tuncel et al., 2020](#)).

From the topic model, results include creativity and resource generation (how refugees identify and enact entrepreneurial opportunities; [Jiang et al., 2021](#)), ability to navigate tough conversations (“conversational receptiveness”; [Yeomans et al., 2020](#)), and gender differences in behaviors that could influence performance (female leaders less likely to delegate; [Akinola et al., 2018](#)).

Theoretically, text analyses could inform the dimensionality of job performance such as revealing new insights on citizenship behavior and counterproductive behavior. Practically, narrative comments in performance appraisals have been an underutilized source of measurement. Usually, only the quantitative ratings drive HRM outcomes like compensation increases. As such, text analysis may have immense value in enhancing the performance management process and ultimately, promotion decisions. For example, text analysis might be used to study the content of information technology help tickets to provide more actionable customer feedback (e.g., [Baikovich & Wasserman, 2020](#)). Such feedback is often interpreted subjectively and text analysis would more objectively measure tone and content. Further, comments on 360 surveys could be more fully utilized by text analysis. Best practice suggest that collecting and interpreting narrative comments is crucial in the use of 360 for performance evaluation ([Campion et al., 2019](#)). However, their use can be limited to sometimes subjective and informal interpretation, which are challenges text analysis could help address. Text analyses for performance management purposes have two key benefits. First, it makes narrative feedback more objective, which should help the persistent issue with perceived subjective bias in appraisals. Second, it can quantify the feedback, which makes it more useful for personnel decision making (e.g., compensation and promotion).

### 2.3.6. Training & development

Results from our text analysis show that the papers in our review analyzed constructs relevant to training and development. Both approaches uncovered leadership as a domain where text analysis may be particularly helpful—behaviors and dyad or group characteristics in metaBUS, and leadership characteristics and influence in the topic model. Leadership training and development may be facilitated by text analysis, as [Speer et al. \(2019\)](#) demonstrated in their development of dictionaries to measure competencies, one of which was “leading and deciding” and was related to developmental performance ratings. In addition, the topic model showed that text analysis could be helpful in unearthing identity changes through the training process, and training opportunities to manage negative customer interactions.

There are many other opportunities for researchers to use text analysis to advance the training and development literature. For example, text analysis might be used to help leadership development theory and practice such as the topic on how leader characteristics (e.g., emotion, humility) influence effectiveness. As another application, text analysis might be applied to understand how an employee thinks about a problem. [Koenig et al. \(2023\)](#) showed how text analysis could be used to score open-ended responses to assessment questions, which suggests text analysis may be valuable for measuring learning outcomes from training assessments. That is, automated text analysis might make the scoring of narrative responses reflecting cognitive changes due to training easier and more objective, which could inform learning theory as well as be practically useful. Text analysis could also help trainers assess the affective responses to training ([Glerum et al., 2021](#)) to improve participant satisfaction. In academic applications, text analysis on narrative responses could be scored to assess learning outcomes, as opposed to reliance on multiple-choice tests that are often necessary with large classes.

In addition, text analysis has been useful for understanding the content of employee interactions. It has been frequently used to evaluate the development of team processes, which is especially relevant to training and development. For example, [Mathieu et al. \(2021\)](#) used text analysis on team interaction transcripts to measure transition, action, and interpersonal team processes, and [Ben-Menahem et al. \(2016\)](#) text analyzed interviews and observations to evaluate the knowledge coordination in multidisciplinary teams. Further, [Schinoff et al. \(2020\)](#) analyzed interviews and observations to understand the development of team relationships in virtual work settings and whether they related to satisfaction and turnover intentions. Finally, the identification of team cognitive maps using text analysis ([Carley, 1997](#)) might have implications for team training, much like the research on team mental models did ([Stout et al., 1996](#)).

### 2.3.7. Employee relations (engagement, culture, and related topics)

Although the opportunities for using text analysis in employee relations seem obvious (e.g., engagement surveys), only one paper in our review explicitly pursued this area—[Speer et al. \(2023\)](#) analyzed employee engagement survey narratives to develop a set of 28 dictionaries to capture employee work-related attitudes (referred to the Text-Based Attitude and Perception Scoring [TAPS] dictionaries). Yet, our analysis of constructs measured using text-based data in our review offer hopeful signs. From metaBUS, organizational characteristics were analyzed using text data, with clear implications for employee relations. For example, [Jehn and Bezrukova \(2004\)](#) analyzed company documents to score work group contexts and found that members of work groups characterized by high functional diversity that foster people-oriented cultures enjoyed higher composite bonuses. From our topic model, we found that the team performance and culture topic included research relevant to employee relations. For example, [Dixon and Panteli \(2010\)](#) examined the dynamics of team virtuality to uncover processes through which team members attempted to reduce discontinuities between their face-to-face and virtual work with others, and [Ahn and Greve \(2024\)](#) used Glassdoor data to uncover startup cultures.

There are several opportunities for future research to expand the employee relations literature. Engagement surveys, as noted elsewhere, might be the most obvious application. Best practices on how to solicit high-quality responses and text-analyze them might be a useful area for future research, especially as narrative comments in engagement surveys will likely be routinely evaluated. Measuring culture is a prime opportunity because it illustrates how text analysis can be used to capture constructs from existing text data or from inaccessible sources, including organizational documents or executives’ speeches ([Pandey & Pandey, 2019](#)). Scholars should continue to use text analysis on existing text data as it can offer richer theoretical perspectives and nuance than constructs measured through deductively derived close-ended survey items, which are typical in organizations due to perceived ease and

efficiency. Text data could provide more insight into the meaning and causes of engagement, potentially informing theories of identity, career development, commitment, and job satisfaction. Narrative comments explain the reasons behind employee ratings of survey items and are especially attractive to the managers receiving the results for that reason. Because text analysis can also quantify narrative feedback, it can be used for evaluating managers and programs and track trends over time. This is an especially promising area given that high-automation analyses are becoming more accessible through LLMs with easy-to-use interfaces such as ChatGPT that can quicken analyses and provide a feasible tool that can be used by those who are less analytically oriented.

### 2.3.8. Compensation & benefits

The use of text analysis to improve compensation and benefits research and practice is still relatively rare. At the individual level, [Speer \(2018\)](#) showed that performance management evaluations related to future pay increases. Otherwise, it was often used to measure CEO or organizational characteristics to understand the relationship executive compensation. For example, though the CEO characteristics they measured using text analysis was a control (CEO narcissism), [Kang and Han Kim \(2017\)](#) showed that it shared a positive and significant bivariate relationship with CEO compensation. However, our analyses offer several ways text analysis has the potential to inform compensation and benefits. In the metaBUS categorization, some personal characteristics (KSAOs; [Campion et al., 2024](#)) and behaviors (e.g., providing exceptional service, customer focus; [Koenig et al., 2023](#)) reflect compensable factors. In the topic model, managerial experience (gender differences in leader language use; [Dupree, 2024](#)) and team performance effectiveness (absence of dysfunction like group-level faultlines, [Bezrukova et al., 2016](#)) are two areas that influence compensation levels.

Given the limited explicit scholarship on text analysis and compensation, there are several avenues for future research. Theoretically, text analysis may reveal that these other sources of rewards could improve understanding of pay satisfaction and its motivational consequences. Practically, HR managers could use text analysis to evaluate reactions to compensation and benefits in engagement surveys, recruiting messaging or surveys from candidates who rejected job offers to inform their compensation strategy (lag, meet, or lead the market) and the efficacy of that strategy. Like recruitment, HR managers could compare their firm's approach to competitors through open-source data. Further, HR managers can use text analysis to understand the benefits that actually matter to applicants and employees to enhance their packages. Improving the accuracy of job analysis and design using text analysis has direct implications for compensation in at least three ways: 1) identifying compensable factors from job descriptions or job tasks, 2) comparing job tasks to identify benchmark jobs in the labor market for job pricing, and 3) determining the nonpecuniary rewards of jobs (e.g., variety, skill usage, security, etc.) that are very important to satisfaction with job rewards in addition to direct pay.

### 2.3.9. Strategic HRM

While no paper in our review directly analyzed HRM strategy using text analysis, there was no shortage of topics relevant to strategic HRM. This is likely because strategic management was one of the disciplines we reviewed for this research. These topics include how organizations change or communicate their strategic framing (CEO language; [DesJardine & Shi, 2020](#)), how they mobilize and operate in ecosystems (motivation to create a prosocial venture; [Mittermaier et al., 2023](#)), and how multinational corporations are managed ([Elg et al., 2017](#)). However, we can see how papers at lower levels of analysis offer insight into strategic HRM, which were present in topics around managing roles and boundaries, as well as technology and innovation. For example, [Cruz and Meisenbach \(2018\)](#) analyzed interviews to study work-life boundary management, which reflects an opportunity for organizations to use text analysis to inform their development and implementation of revised or new work-life policies. Similarly, changes in job design and employee reactions to those changes can inform policies. For example, [Zhang et al. \(2021\)](#) explored public sentiment on enforced remote work, and [Schinoff et al. \(2020\)](#) analyzed examined how relationships in virtual work settings develop and influence satisfaction and turnover intentions. The results from metaBUS were less robust than the topic model, but nonetheless useful with research on personal (CEO regulatory focus; [Gamache et al., 2015](#)) and contextual characteristics (disruptive industry innovation; [Ansari et al., 2016](#)).

Like other HRM functions we have reviewed, strategic HRM is full of opportunities for future research. Text analysis could be practically useful for summarizing reactions to HR programs, policies, procedures, outcomes, and other HR strategy topics from the range of narrative information available, such as engagement surveys, candidate reaction surveys, executive speeches, annual stockholder reports, customer service feedback, social media, rejection surveys, exit interviews, blogs, e-mail traffic, and the like. HRM practitioners have commonly used close-ended (multiple-choice) surveys of employees to capture reactions to HRM policies, practices, and services. Not only are these deductively derived and not as helpful for discovery, but the summative evaluation they provide ignores formative feedback as to how the practice could be improved according to narrative comments. Text analysis would allow a systematic, objective, and efficient tool for analyzing that data. As a final example, text analysis might be able to identify emerging issues that have yet come to the surface based on the discussions and comments in the various narrative corpuses among executives or managers, employees, candidates, the public, lawmakers, or others with potential future relevance to HRM.

## 2.4. Summary and guidance for HRM researchers and practitioners

From this section, we have generated three takeaways for readers. First, while deductive categorization analyses are intuitive and require basic skills to adequately conduct, they are time-consuming and can result in overly broad groupings. In our review, we used metaBUS as one way to analyze the papers and while it provided high-level insights as to the attributes that were present based on a large nomological network of constructs that covers the gamut of management literature, its lack of specificity reduced our ability to more usefully uncover how text analysis has contributed to our understanding of people at work, theoretically and practically. Meanwhile, the topic model was based solely on the papers in our review and therefore generated more useful or at least incremental

insights. Given the efficiencies, the theoretical richness of the topic model, and its specificity to the data rather than imposing a broad framework, we recommend the high-automation approach for HRM researchers. Though, this recommendation is conditioned on whether the researcher knows what constructs they are aiming to measure. For example, if in a selection or performance management context a job analysis has been conducted or a competency model has been developed, then searching explicitly for those constructs may be more valuable than examining the data for underlying patterns. In these cases, we would likely recommend word dictionaries to those with less machine learning experience and supervised machine learning to those with more machine learning experience (see “Best Practices in Text Analysis” section below). In the case of summarization—as was our goal here—the topic model was more efficient and effective.

Second, as can be observed in Table 4, research in our review could be used to improve all HRM functions. Evidence from our analyses suggest that text analysis could be especially useful in selection, retention and separation, performance management, training and development. There are also significant opportunities for HRM researchers and practitioners to uncover novel, or underappreciated, ways that text analysis can be used in other core HRM functions such as job analysis, recruitment, and employee relations. To achieve initial gains, scholars and practitioners should focus on leveraging existing text-based data rather than collect new data. HR managers have access to text-based data from applicants, employees, and former employees that can be used to 1) shed light on previously unrecognized pain points or challenges for employees (e.g., benefits packages, work-life policies, modified job designs to optimize performance and day-to-day experience), and 2) generate solutions for obstacles employees are facing. Employees can be a great source of potential solutions, so automating feedback and ideas from employees using advanced analyses like unsupervised machine learning can quicken problem-solving and demonstrate to employees that their voices are heard.

Reinforcing our first contribution, a final recommendation is that HRM researchers should not feel compelled to operate within a single disciplinary perspective that limits their ability to achieve methodological diversity. While this is good advice for all scientific domains, and may not immediately appear profound, in our review we have found that such intra-domain separation can be especially troublesome when advancements in one area of management do not quickly spillover over to other areas of management. For example, it is evident from Tables 1–4 that strategy researchers have a head start in using text analysis to advance their discipline. We wonder whether strategy researchers began using text analysis due to the difficulty directly accessing executives and the availability of text data produced, or at least approved, by top executives in the many forms of communications to shareholders and the public. Regardless, text analysis has been a particularly informative method for strategy and entrepreneurship scholars who often rely on archival data. Instead of seeing strategy, entrepreneurship, or even OB as less informative of HRM, HRM researchers might benefit from the knowledge in those disciplines on new ways to improve construct measurement. A similar sentiment can be found in the reluctance of typical quantitative HRM scholars to engage in inductive methods. Historically, this has been the domain of scholars who do purely qualitative research and such methods can be time-consuming and thus impractical for HRM practitioners. However, the increased accessibility of inductive NLP methods, such as topic modeling, means that HRM scholars and practitioners have a new way to engage with their text data that requires less time and is more objective, but necessitates steps closer to what we would see in a purely inductive approach. Cross-pollination of ideas and methods is central to the advancement of HRM research and practice.

### 3. Best practices in text analysis

In this section, we provide guidance to HRM researchers and practitioners in determining the appropriate text analysis method for their research. We then offer best practices for validating text analysis results drawing from the rich history of psychometrics. While HRM researchers and practitioners are familiar with these validation steps, we want to reinforce their application to text analysis, as the process to validate for text data may not be immediately clear to some.

#### 3.1. Selecting text analysis methods

The choice of whether to use text analysis, and which text analysis method to use depends on the research goals and context. At a baseline, text analyses are appropriate when the data are qualitative and the goal is to 1) triangulate with existing measures or discover new ideas through analytic diversification, or 2) refine or develop theory, especially in an entirely new area or topic. In contrast, traditional numeric approaches are typically more appropriate when a valid measure of a construct already exists and 1) the goal is to establish a construct’s placement within a nomological network (convergent and divergent validity), or 2) the goal is theory testing, especially in mature areas (Edmondson & McManus, 2007). The specific type of text analysis approach to use will be determined by the fundamental goals of the researchers: whether they aim to score the data to make predictions or summarize the data, their sample size, availability of a criterion, and need for reproducibility. We detail these decisions according to the types of text analysis techniques used in our review. We discussed these four techniques in the “Methods of Text Analysis” section, but as a reminder, they are: dictionaries, supervised machine learning, unsupervised machine learning, and categorization.

In Table 4, we present recommendations for when to use each type with the conditions listed on the left and the degree to which the condition aligns with the types of text analysis listed in the top row. Alignment is based on research interest, resources, and capabilities, which we describe below. We differentiate among “high alignment” where the goals of a study match the dominant capabilities of the method; “modest alignment” where the method can be used to achieve that goal, but it may depend on other factors; and “low alignment” where the goals of the study are poorly aligned or fundamentally at odds with the method. There are instances where more than one method can be used in a study. We depict these recommendations in the form of a decision tree in the online supplement.

### 3.1.1. Categorization

Categorization, which includes content analysis via hand coding (as we did with the metaBUS taxonomy) and grounded theory, is most useful for summarizing text. It is typically done with no or little automation. However, because the techniques vary, so do the conditions under which they are appropriate. For example, a hand-coded content analysis is best when the data set is small, the constructs to be summarized are known a priori or easily deduced from an initial reviewing of the data (Campion & Csillag, 2022), the goal is to test or refine theory, and reproducibility may be a priority because decisions, categories, and descriptions can be reported and replicated. However, more inductive methods—purely qualitative methods—are best when constructs are not known a priori and the goal is to summarize by building new theory (Pratt et al., 2020). Like others, we consider purely qualitative methods to be an interpretive approach to deriving meaning from text, which involves researcher-based inferences from manifest and latent content (Drisko & Maschi, 2016). Such an approach does not seek reproducibility; therefore, techniques to enhance trustworthiness should be used (see online supplement). Categorization in the form of content analysis has a long history in HRM and will continue to be an important method due to its logical appeal, accessibility, and feasibility for summarizing small datasets. It accounts for 58.15 % of the studies in our review. Our primary recommendation, however, is to move beyond this basic method and to use dictionaries and advanced NLP techniques.

### 3.1.2. Dictionaries

Dictionaries are automated text analyses and are most appropriate when the goal is to score or predict a construct known a priori, an existing dictionary already exists, reproducibility is paramount, or the goal is to demonstrate relationships with existing constructs. Dictionaries are also useful for developing new measures of existing constructs where triangulation is valued. Further, dictionaries can be useful when a researcher has a sample that is too small for machine learning such that models will be unstable, but the sample is too large for hand coding.

For those new to text analysis, we recommend dictionaries as a pathway to more advanced NLP. The content can be directly examined, and dictionaries can be implemented by those with typical statistical skills. There are many validated dictionaries available in our literature such as LIWC (Pennebaker et al., 2007), which contains dictionaries and linguistic measurements (e.g., parts of speech, punctuation), as well as those developed to measure specific constructs (e.g., competencies, Speer et al., 2019; work-related attitudes; Speer et al., 2023; entrepreneurial orientation, Short et al., 2010; verbal behavior in interviews, Hickman, Bosch, et al., 2022). Developing new dictionaries to measure constructs of interest is a well-understood process that involves delineating lists of words that reflect the construct based on hypothesized terms from the theory or topics in the study. This involves generating a seed list of terms and then identifying related terms using a thesaurus or other semantic tools. The list of words are used for measurement by employing one of many available word counters. Dictionaries are often validated against a criterion, if one is available. Without a criterion, researchers can still use SMEs to help support the content and construct validity of the dictionary, based on the development of the word lists (e.g., Madera et al., 2009). Dictionaries have wide applicability in HRM because they can be used on virtually any corpus of text (e.g., applications, interviews, engagement surveys, customer feedback, internet postings). The availability of validated dictionaries and the ease of developing new ones likely accounts for why they are the second most used approach to text analysis in the literature related to HRM. They have been used in 33.45 % of the papers in our review. Of those that used dictionaries, 10.77 % developed their own.

### 3.1.3. Supervised machine learning

Like the dictionary method, supervised machine learning is most useful when the goal is scoring or prediction. Unlike dictionaries, the exact constructs to be measured may not be known a priori. Supervised machine learning requires a criterion to serve as the “ground truth.” In HRM, this is often human ratings of interviews, applications, narrative evaluations, or other qualitative assessments. The algorithm is then trained to optimize the weights of the features (text variables) according to the criterion (Campion et al., 2024; Koenig et al., 2023). This means that a supervised model in HRM may be developed to capture an array of competencies such as leadership, communication, motivation, and the like. If supervised models are built to score job applications where applications are currently scored by humans according to rubrics reflecting competencies, then an algorithm can be trained to score those competencies. However, it will also pick up on other constructs that influence the human raters intentionally or not (e.g., positive tone). Because supervised machine learning includes a “ground truth,” it is used in contexts where goals may be to develop a new measure to test theory, establish a nomological network, and triangulate with existing measures, and do so in a reproducible way.

To achieve stability, supervised machine learning is best used when data sets are sufficiently large, which is in line with data science research that may have sample sizes with thousands to millions of observations. Yet, HRM data sets tend to be much smaller. With smaller data sets, we recommend using unit weights (simple counts) of the construct rather than weights for each  $n$ -gram estimated by the sample to improve cross-validation. Guidance on minimally sufficient sample size depends on a few factors. The first is cross-validation. This requires enough data to have a sizable training and testing set (typically 80 % training, 20 % testing in larger samples, or more equivalent in smaller samples). The second is the number of parameters in relation to sample size. The more parameters, the larger the sample size required (Landers et al., 2023). Third, researchers must think about the heterogeneity of the data content because it relates to model complexity. Should the constructs within the data be more heterogeneous, then a greater number of features may be necessary to extract all the possible ways the construct might be represented. In these instances, a larger sample is also needed to achieve stability, unlike data that are more homogenous. For example, examining text data from a job application may include constructs related to attributes such as leadership, teamwork, and conscientiousness, making the data heterogeneous. However, examining responses to a situational question on a specific leadership behavior in an interview will likely yield more homogenous data.

When developing a text-based measure by training a model against an existing measure or when validating a new measure by

correlating it with an existing or related measure, the important questions are how high must the correlation be to establish equivalence and what if the correlation is low? This is further complicated by the fact that the goal of developing a new measure may be to develop a better measure (e.g., more objective) and a low correlation with that existing measure (e.g., self-report) could be the goal. Obviously, there is no simple answer, but we offer several observations based on the existing literature. First, when training a model against an existing measure, the correlation to be expected will depend on the reliability of the existing measure.<sup>5</sup> For example, if the existing measure is based on human judgment, a common goal is to achieve a correlation as high as the reliability of an individual rater (Campion et al., 2016) because replacing the rater is the intended use of the measure. Second, the expected correlation will similarly be limited by the reliability of the text measure. If it is not highly reliable, that will limit the possible correlation. However, it might be more reliable if it more consistently measures the content or more valid if it captures more relevant content reflecting the construct (Campion & Campion, 2023). Third, in some domains, any positive prediction is a net gain. For example, when predicting an important bottom-line outcome like turnover, any non-zero level of prediction is valuable (e.g., Sajjadi et al., 2019). Fourth, in domains where expected levels of prediction are fairly well known, such as the validity of traditional hiring procedures, matching those levels or showing incremental validity may be the expected level of correlation (see Koenig et al., 2023). Finally, the normal construct validation principles apply, such as expecting the convergent correlations to be larger than the discriminant correlations.

We recommend supervised machine learning where it is applicable and feasible under the conditions above because it is the most advanced approach for prediction. It can be applied to all types of text data in HRM and in any context that has a criterion and a sufficient sample size. It can be more efficient than developing dictionaries because extraction is automatic and requires no manual process of identifying terms. Instead, understanding the constructs measured depends on the criterion. Moreover, and likely why it has been a popular method in selection, it can be used with large samples that would not be possible to analyze manually. To date, supervised machine learning may be the biggest missed opportunity in HRM research, as it has only been used by about 5.66 % of all studies.

#### 3.1.4. Unsupervised machine learning

Unsupervised machine learning is most useful for summarizing text to uncover the content of text data that is unknown a priori, discovering new constructs, developing new theory, and where sample is sufficiently large. It is often used for theory development in new areas but can also be used for theory refinement in established areas. It is more appropriate for inductive than deductive research, though a researcher may know what constructs generally exist in a corpus before using unsupervised machine learning. Like supervised machine learning, the sample size and number of parameters determines the stability of the model. Some researchers using unsupervised machine learning report that results stabilize with a sample size of about 1000 observations (e.g., Schmiedel et al., 2019); however, like supervised machine learning, the minimum sample size depends on content amount and heterogeneity. Unlike supervised machine learning, there is no criterion by which the model is trained (i.e., it is unlabeled). Instead, the algorithm detects an underlying pattern and the researcher makes sense of that pattern post hoc. This inductive sensemaking process resembles that of purely qualitative work, as noted in the development of our topic model of the abstracts but allows for one non-trivial alteration: it can largely be largely reproduced. When we refer to reproducibility here, we mean that equipped with every decision made by the original researcher, a second researcher could follow those steps and yield generally the same findings as the original researcher. Reproducibility is possible with unsupervised machine learning because each decision a researcher makes—the algorithm, the number of topics retained, or the logical labeling and description of the topics—can be specified.

Unsupervised machine learning is the most sophisticated approach to summarizing text. It automates a historically burdensome process of manually coding text and it makes the analysis of large datasets feasible (Min et al., 2021; Zhang et al., 2021). It also elevates text analysis from a process often criticized for its potential subjectivity and susceptibility to bias to a more objective method that can be subject to the same reproducibility standards as other measures in HRM. It can be applied to virtually any text data in HRM with a sufficiently large sample, and it is also another major missed opportunity, accounting for only about 5.83 % of all studies.

### 3.2. Validation of text analysis

The purpose of this section is to evaluate the quality of text analysis and provide recommendations for validation. We use the lens of construct validation because it has a long history with detailed best practices, and it is directly relevant to NLP techniques that quantify text data, which we believe is currently under-recognized. Construct validation has been central to evaluating the quality and meaningfulness of data in psychology and related disciplines for many decades (e.g., Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Lord & Novick, 1968) and such data are often used to make personnel decisions in HRM so must adhere to the Principles (2018), Standards (2014), and Uniform Guidelines (1978), and withstand legal scrutiny. When evaluating the papers in our review, we provide counts of papers that adequately complete each validation step. We interpreted these counts as evidence that not all researchers are aware that text analysis scores should be held to the same standards as traditional numeric scores. As such, we provide best practices for construct validation to guide future research and improve analytic rigor (Table 5). Due to the increased use of unsupervised machine learning to summarize text data we also make note of the validation steps that are relevant for this type of analysis. These are related to, but distinct from, the “trustworthiness” efforts advanced by purely qualitative scholars (Pratt et al., 2020; Table C in the online supplement). For background, validity broadly refers to evidence supporting claims that we are measuring what we think we are

<sup>5</sup> Based on the classic psychometric formula for correcting observed correlations for reliability, the true correlation is equal to the observed divided by the square root of the reliabilities (Ghiselli et al., 1981).



**Table 5**  
Best Practices for Construct Validating Text Analysis Results.<sup>a</sup>

Type of Text Analysis Validity Evidence Based on:	Potentially Applicable Best Practice	Relevant to Text Analysis Method:
1. Clearly defining the content domain	<ul style="list-style-type: none"> <li>Clearly define the domain and link the domain to previous research and theory.</li> <li>Define the domain in terms of observable behaviors.</li> <li>Link the domain to job performance, such as job behaviors or tasks, especially if the data are used for employment decision making.</li> <li>Use job or respondent terminology, when appropriate.</li> <li>Ensure the content domain is adequately represented (e.g., proportion of content covered, domain sampling, test specifications plan, SME judgment, etc.).<sup>4</sup></li> <li>Include hypotheses that test the adequacy of the definition and/or link to the content domain.</li> <li>Quantify the categories in the category-based approach, if possible, so validation methods 2–5 can be conducted.</li> </ul> <p><b>Current Use From Review:</b> Used by virtually all studies that quantified their text data.</p>	<ul style="list-style-type: none"> <li>Categorization</li> <li>Dictionaries</li> <li>Supervised Machine Learning</li> <li>Unsupervised Machine Learning</li> </ul>
2. Using SME judgments to interpret categories	<ul style="list-style-type: none"> <li>Utilize SMEs to help interpret and evaluate the content validity of the variables and categories.</li> <li>Ensure SMEs are actual experts on the content domain (or research process, if applicable) and ensure SMEs are naive to study hypotheses if applicable.</li> <li>Select SMEs to ensure representativeness (e.g., types of SMEs, jobs, experiences, etc.).</li> <li>Use systematic judgment processes such as ratings based on relevant criteria (e.g., measure domain, importance to domain, match job complexity, realism, etc.), independent judgments, consensus discussions, etc.</li> <li>Use an adequate number of SMEs to ensure results are not due to the individual SMEs and can statistically infer to the population of potential SMEs.</li> <li>Train the SMEs on the judgment process, including practice and feedback, and evaluate their performance.</li> <li>Evaluate SME judgments differences (e.g., differences in understanding content domain versus lack of training or carelessness).</li> <li>Use SME ratings as the criterion against which to train machine learning algorithms if applicable.</li> </ul> <p><b>Current Use From Review:</b> Used by 46.83 % of all studies in review.</p>	<ul style="list-style-type: none"> <li>Categorization</li> <li>Dictionaries</li> <li>Supervised Machine Learning</li> <li>Unsupervised Machine Learning</li> </ul>
3. Internal structure of the constructs	<ul style="list-style-type: none"> <li>Calculate internal consistency of text variable composites to evaluate homogeneity and to determine whether differences in the variables would not change the scores and results will replicate.</li> <li>Calculate and report analyses of agreement and reliability (covariation) because they provide different and relevant information about psychometric quality.</li> <li>Evaluate reproducibility of category structure (primary &amp; aggregate) if based on SME/researcher judgment by having independent SMEs recreate structure.</li> <li>Analyze the covariation among the categories and the factor structure, if applicable.</li> <li>Analyze the clustering (distance) between categories, if applicable.</li> <li>Extract variables based on their predictive value and apply other approaches to construct validity to interpret the meaning of the variables.</li> </ul> <p><b>Current Use From Review:</b> Used by 26.59 % of studies in review.</p>	<ul style="list-style-type: none"> <li>Categorization</li> <li>Dictionaries</li> <li>Supervised Machine Learning</li> <li>Unsupervised Machine Learning</li> </ul>
4. Relationships with other constructs	<ul style="list-style-type: none"> <li>Evaluate convergent validity by demonstrating that text analysis measures correlate with other known measures of the construct.</li> <li>Evaluate discriminant validity by demonstrating that text analysis measures do not correlate with different constructs.</li> <li>When other measures of constructs are not available, consider using SME ratings of other constructs to examine convergent and discriminant validity.</li> <li>If the categories are not quantified, such as when only used for theory development, rationally and theoretically justify the convergent and discriminant validity based on similarities and differences with other relevant constructs. This can also be done through the support of SME judgment.</li> </ul> <p><b>Current Use From Review:</b> Used by 22.78 % of studies that scored their data in the review.</p>	<ul style="list-style-type: none"> <li>Dictionaries</li> <li>Supervised Machine Learning</li> </ul>
5. Predicting outcomes or consequences or being predicted by theoretically relevant variables.	<ul style="list-style-type: none"> <li>Analyze the statistical relationship between the text analysis measures and relevant outcomes, if possible.</li> <li>Analyze the statistical relationship between the text analysis measures and their theoretical predictors, if possible.</li> <li>Avoid common method variance by not relying totally on self-reported criteria and predictors collected from respondents providing the text data.</li> <li>Use organizationally relevant criteria to demonstrate the value of text analysis measures, especially objective business outcomes, where possible.</li> </ul>	<ul style="list-style-type: none"> <li>Dictionaries</li> <li>Supervised Machine Learning</li> </ul>

(continued on next page)

Table 5 (continued)

Type of Text Analysis Validity Evidence Based on:	Potentially Applicable Best Practice	Relevant to Text Analysis Method:
	<ul style="list-style-type: none"> <li>• If data are to be used for employment decisions, evaluate subgroup differences (e.g., race, gender, age) to estimate the potential for adverse impact.</li> <li>• Cross-validate or otherwise prevent capitalization on chance of statistical relationships with other variables, especially if text analysis variables are created or selected based on these relationships. In machine learning, it is common to use <i>k</i>-fold cross-validation where data are randomly divided into <i>k</i> sets (typically 5 or 10) and the model is tested against each set.</li> <li>• If the categories are not quantified, such as when only used for theory development, rationally and theoretically justify the expected relationships with outcomes and the causes based on the theory, relevant literature, and field work data (interviews, observations, archival data).</li> </ul> <p><b>Current Use From Review:</b> Used by 45.90 % of studies that scored their data in the review.</p>	

*Note.* Information for this table was derived from a number of highly qualified sources: Uniform Guidelines on Employee Selection Procedures (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978); Principles for the Validation and Use of Personnel Selection Procedures (American Psychological Association, 2018); Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014); Cronbach and Meehl (1955); Campbell and Fiske (1959); Lord and Novick (1968); Nunnally (1978); Messick (1995); and the articles in our review that employed text analysis and yielded statistical results.

<sup>a</sup> SME = Subject Matter Expert. Similar authorities from the United Kingdom also recognize the importance of these types of validity evidence, including the International Test Commission Guidelines on Test Use (2013) and Psychological Testing: A Test User's Guide by The British Psychological Society.

measuring, and in the case of HRM, that the measure is job-related. In psychometric research, there are five traditional types of evidence, which inform the meaning of the construct and strength of the research evidence supporting inferences about it.

### 3.2.1. Evidence based on clearly defining the content domain

This is considered content validity by virtually all validity authorities (Principles, p. 28; Standards, no. 1.11; Uniform Guidelines, Section 14C4). It includes defining the domain theoretically, collecting information on the phenomenon (e.g., descriptions, examples, observations), and explicitly generating item content to measure the construct domain. This step is important for those who intend to score and predict, as well as those who aim to simply summarize. All the studies in our review included this type of evidence and we expect that this foundational practice will continue as it is central to all social science research.

### 3.2.2. Using SME judgments to interpret categories

SMEs, or individuals who have expertise in the domain, have been essential to historic validity efforts (e.g., Principles p. 27–28). Of the papers in the review, 46.83 % used this technique. Most of these studies used SMEs to support interpretation of codes or topics. For example, Patton and Johns (2007) recruited SMEs to code 100 popular press articles for constructs related to gender and absenteeism, and Kanze et al. (2018) used SMEs to examine the content validity of word lists to study workplace discrimination. One reason SMEs were used in less than half of the articles may be due to the use of previously validated dictionaries. SME judgments are especially important when other sources of validity evidence are not possible. As with the previous step, this is relevant to all types of text analyses.

### 3.2.3. Internal structure of the constructs

This type of evidence refers to statistical indicators of proper measurement. It typically includes evidence of reliability and factor structures (Principles p. 32–33; Standards, no. 1.13; Uniform Guidelines, Section 14C5). Of the studies in our review, 26.59 % provided this type of evidence. Generally, these were estimates of agreement because the interest was in the reproducibility of the categories, but scholars also used more sophisticated indices such as Cohen's Kappa, Krippendorff's Alpha, and Kendall's W. Traditional measures of reliability (defined as covariation as opposed to agreement) such as intraclass correlations among coders (e.g., Harrison et al., 2019) were less common. Analyses of agreement and covariation provide different information, and both are relevant to the psychometric quality of text analysis measures. Another relevant consideration is that automated text analysis tends to extract a large number of concepts, so reducing dimensionality is desirable. Dimensionality reduction can also be based on covariation or agreement like factor or cluster analyses, respectively. For example, Palmer et al. (1997) used factor analysis to identify themes from 275 downsizing references. The relevant internal structure analysis will depend on the purposes of the study, but indicators of agreement and reliability are probably relevant to all studies because they index data quality and support inferences. Of the list of validation steps, this is the final step that has shared importance for those intending to score and predict as well as for those intending to summarize.

### 3.2.4. Relationships with other constructs

Although historically this has been called "construct validity" (Campbell & Fiske, 1959; Uniform Guidelines on Employee Selection Procedures, 1978, Section 14D), it is described as evidence based on relationships with other constructs because all types of evidence bear on construct validity (Principles for the Validation and Use of Personnel Selection Procedures, 2018, p. 14; Standards for

*Educational and Psychological Testing*, 2014, no. 1.16). Despite its importance, only 22.78 % of the studies that scored their text data reported construct validity (convergent or discriminant). For example, [Campion et al. \(2016\)](#) correlated their NLP measures of past work accomplishments with a battery of employment tests to show similarities and differences with traditional hiring procedures. Further, [Speer \(2018\)](#) used manager ratings of job performance to demonstrate convergent validity with their NLP model of performance evaluation comments. Because construct validity is fundamental to understanding the meaning of new measures, it should be considered when feasible and relevant to the research question with text analyses aimed at quantifying the data (e.g., dictionaries or supervised machine learning), as it is with traditional quantitative measures.

### 3.2.5. Predicting outcomes from theoretically relevant variables

Related to relationships with other constructs is predicting outcomes from the target variable, which is often called “criterion-related” validity (Principles, p. 8; Standards, no. 1.17; Uniform Guidelines, Section 14B). This evidence may only be relevant in studies aimed at predicting as opposed to summarizing. Nearly half of the studies (45.90 %) that scored text data predicted outcomes from their quantified text variable(s). For example, [Campion et al. \(2024\)](#) and [Koenig et al. \(2023\)](#) predicted job performance after hire and [Sajjadiani et al. \(2019\)](#) predicted turnover. In selection, cross-validation in subsequent samples is a crucial step to support the operationalization of a predictive model. In machine learning, they often approach this using a method called *k*-fold cross-validation where the data set is divided into *k* sets (often 5 or 10) and the validities of each set and the aggregate are reported (e.g., [Koenig et al., 2023](#)). This is useful to inform the stability of a model in instances where subsequent data collections may not be feasible. The other half of the studies used their text variable as a control or as a test of group differences, such as [Madera et al.’s \(2009\)](#) study of gender bias in letters of recommendation. When using text analysis to score, predicting important outcomes is probably the ultimate evidence of validity and strongest proof of usefulness, especially if the text-based measures reflect new constructs, are intended to predict beyond existing measures of those constructs, or are used to supplement existing employment tests where those constructs may not be captured. Such evidence should also be bolstered with steps to ensure the avoidance of capitalization chance, such as cross-validation, which is a long-recognized concern in the validation authorities (Principles, p. 26; Uniform Guidelines, Section 14B7).

## 4. Conclusion: making the case for natural language processing approaches

The primary conclusion we draw from this review is that HRM researchers and practitioners should endeavor to learn and use the more advanced NLP methods. This is especially relevant for those who consider themselves exclusively or largely quantitative researchers. In this section, we support this stance through the reinforcement of points made previously: 1) text analysis methods are accessible and they increase efficiencies, 2) leveraging text data facilitates methodological diversity and therefore methodological rigor, and 3) reproducibility is possible (and expected) in text analysis using advanced techniques.

First, with these advanced methods, HRM researchers and practitioners can use a type of data that has, at times, been ignored as a source of discovery and measurement due to cost and inefficiencies. In the past, text data in HRM were only analyzed through time-consuming hand-coding content analyses. Thus, quantitative scholars have overlooked a rich data source and possible insights, as illustrated in the current review. HRM is rife with underutilized text data. In this review, we have seen opportunities to use NLP to score applications ([Campion et al., 2016, 2024](#)), assessment center data ([Koenig et al., 2023](#)), performance appraisal narratives ([Speer, 2018](#)), and job history ([Sajjadiani et al., 2019](#)), as well as identify competencies in job analyses ([Koenig et al., 2023](#); [Putka et al., 2023](#)) and uncover constructs in engagement surveys ([Speer et al., 2023](#)). Yet, additional gold mines abound. For example, large organizations could utilize candidate postings of comments on recruiting websites to efficiently monitor and evaluate organizational reputation using NLP, as [Ahn and Greve \(2024\)](#) did to examine startup culture. Customer comments on the services provided by HRM units could similarly be evaluated. Virtually every HRM service requires a computerized interaction (e.g., benefits, compensation administration, training, promotion opportunities, appraisal, hiring), which usually allow narrative reactions from users. Automating summaries of these comments through unsupervised machine learning has the dual benefit of informing HRM processes while potentially improving comment quality over time as employees learn these data are actually considered. That user experience via these comments is ignored is a known challenge in organizations.

Further, NLP allows for the analysis of large datasets that are cumbersome or simply impossible to hand code. With the explosion in the amount of information being collected (the “Big Data” movement), much of which is qualitative, the ability to automate analyses will be essential to HRM. A prime example from our review is [Min et al.’s \(2021\)](#) analysis of 1.56 million tweets in the early months of the pandemic to assess responses to the shift to working from home. Such an examination on a mass scale would not be feasible without the automation afforded by machine learning. Only the perspectives of workers willing to voice directly to supervisors would likely be heard otherwise. More commonly, organizations of any size receive more employment applications than recruiters can evaluate, more comments in employee surveys than can be systematically analyzed by the survey sponsors, and HRM researchers are often faced with more literature to summarize than is realistic to complete manually. All of these lend themselves to automated assistance by NLP to advance HRM research and practice.

Of importance to HRM specifically, text analysis can be useful for developing measures of constructs discovered through purely qualitative research. HRM researchers and practitioners need to create and validate measures to aid in decision making, which is different than other areas of management that have used text analysis primarily to seek understanding. HRM researchers are often trying to create systems to manage organizations better including selecting new talent, improving employee performance, promoting internal candidates, and laying employees off. Another recommendation from this review, then, is to quantify qualitative ideas so that those constructs can be modeled explicitly in these high-stakes decisions.

Second, HRM researchers should use text analysis to enhance methodological diversity and the rigor of their research.

Methodological diversity refers to the triangulation of results through replication with different methods, and the discovery of new ideas through alternative analysis of the same data (Turner et al., 2017; Welch & Piekkari, 2017). Text data are well-positioned to offer methodological diversity of our traditional quantitative measures (e.g., test scores, rating scales). One could imagine an instance where internal candidates are being considered for promotion. Each have quantified performance scores, narratives from supervisors on their leadership potential, and self-written narratives on their motivation to lead. An HRM practitioner has two opportunities to make the process more objective and efficient with NLP: score the narrative data to triangulate with the quantified performance scores and capture additional leadership-related constructs not captured in the performance scores, or summarize the constructs in the narratives to uncover KSAOs that may be valuable but under-recognized in leadership positions. The former offers triangulation and the latter offers discovery. Multi-operationalism complements reproducibility as each method can compensate for the other's weaknesses. Thinking differently about a phenomenon through diverse analysis is less common but is where much of the potential value of text analysis lies. As such, a clear recommendation is for HRM researchers to normalize the collection of text-based data, where possible, and to use these data to triangulate and complement findings from traditional numeric techniques.

An illustration of methodological diversity from the papers in our review is on leadership, which was the only content domain where multiple text analysis approaches were used. Speer et al. (2019) used dictionaries to examine "leading and deciding" as one of the Great Eight Competencies and showed that it positively related to performance ratings. Mavin and Grandy (2016) used grounded theory to delineate the process through which elite women leaders manage the societal preoccupation with their appearance by shifting focus to their achievements. Campion et al. (2016, 2024) used supervised machine learning to score job candidates' leadership skills and accomplishments to train a model to predict human hiring ratings and post-hire training performance in leadership roles. Finally, Banks et al. (2018) used unsupervised machine learning to evaluate potential dimensions of leader-member exchange and found three themes: individual relationships between leaders and members, task-oriented helping by the leader, and the extent to which a leader directs team performance. While no study captured the exact same construct, comparing these results, we can observe that 1) there are substantial opportunities to triangulate measurement using text analysis, and 2) alternative measurements (though not on the same data in this example) afford new ways of thinking about existing ideas. This is especially apparent in instances where ideas were induced from data (e.g., Banks et al., 2018; Mavin & Grandy, 2016).

Third, reproducibility is a primary principle of HRM research and practice because it supports validity and facilitates fairness in decision making. Though purely qualitative research has not typically been used to inform decision making, it has nevertheless received criticism as to its replicability, validity, and rigor (Gioia et al., 2013). By leveraging automated approaches, researchers can quantify their text data to inspect the model's validity and do so in a reproducible and transparent manner where each step and decision can be documented and replicated by others. While many have written clarifications as to the purpose of text analysis approaches such as grounded theory—including that replicability is *not* a core tenet of such methods (Pratt et al., 2020)—there remains hesitancy toward qualitative data analysis by many scientists for this reason. We see the more advanced text analysis methods reviewed in this article as offering an alternative to analyze text data in ways that align with traditional principles of quantitative research including transparency, replicability, and psychometric evidence of validity. This will make research findings based on qualitative data more palatable to traditional HRM researchers who rely mostly on quantitative methods. As such, the introduction of NLP in our science is an opportunity for more quantitative scholars to reorient their thinking about the value and utility of qualitative data. Moreover, text analysis, especially advanced summarization methods such as topic modeling, can empower HRM researchers and practitioners to be interpretive and creative in their analysis of text data. Surprising and nuanced findings are possible as the algorithm is used to uncover patterns and unexpected topics may emerge. To be clear, we are not advocating for the replacement of traditional qualitative methods. Instead, we are suggesting that qualitative data need not be viewed as antithetical to the core of our traditional quantitative analyses, but instead as a complement to it.

#### 4.1. A final note on large language models and chatbots

At present, only one paper in our review used a large language model to measure new or existing constructs: Hernandez & Nie (2023) used GPT-2 to develop scale items representing the Big Five personality traits to demonstrate the validity of this approach. Subsequently, there have been several studies exploring whether LLMs enable candidates to fake their answers to achieve higher scores on remote unproctored hiring assessments (e.g., cognitive test by Hickman et al., 2024; situational judgment tests by Harwood et al., 2024; and personality tests by Phillips & Robie, 2024). Because this was so rare in our review, there were few concrete opportunities to discuss its potential. Therefore, in this final section, we discuss the promise of LLMs and similarly advanced NLP-powered tools such as chatbots.

LLMs are deep learning algorithms built on millions or billions of parameters to do language-related tasks. Tasks relevant to HRM in addition to those cited above include, but are not limited to, developing preliminary job descriptions (Campion & Campion, 2025), summarizing open-ended data responses (topic modeling), extracting text from a corpus, providing feedback on interview questions, creating interview questions, drafting content of HR policies and practices, generating templates for communications (e.g., emails to job candidates), and the like. In each of these examples, LLMs can create efficiencies by developing initial drafts. After initial drafts are created, HRM practitioners can either continue to coach the algorithm through specific prompting to meet the needs of the task, or they can take what the LLM outputs and tweak it with organization-specific language. Using LLMs to summarize means that HRM researchers and practitioners have quicker access to inductive methods. Equipped with their existing knowledge on deductive methods, these advancements make deductive and inductive methods easier than ever. The potential efficiencies are crucial here as LLMs with user-friendly interfaces allow HRM managers to input their text data and request output such as number of topics, subtopics, direct quotes, and counts, which can be completed in minutes or hours instead of days, weeks, or months with continued experience and skill

with these tools. Again, however, we want to be clear that opportunities with LLMs do not replace the specialized knowledge by SMEs to assess the quality and relevance of the LLM output. Further, LLMs, at present, do not threaten the continued value of building one's own machine learning model for specific purposes. For scientists, LLMs can present challenges for reproducibility as the models are generally non-deterministic. Importantly, the information generated by these models can be incorrect, and there is the associated potential risk of overreliance on such tools without performing due diligence. Their overuse may also be to the detriment of learning and creativity because LLMs are trained to give the most likely response, unlike humans who are more capable, currently, of novelty, nuance, and analytical thinking. Nevertheless, at this stage in the research, the HRM domain is aware of the great promise that LLMs hold but are still trying to figure out all that they can do. Additional scholarship is needed to understand the scale of generative models' impact on HRM research and practice, though we predict that impact will be substantial. We encourage those in the HRM domain to engage with these tools and systematically assess how they can improve practice and research.

A final promising area for HRM researchers and practitioners are chatbots. Chatbots are NLP-trained models to provide answers to questions. Chatbots can be trained using an LLM as the engine by including additional, domain-specific, information (otherwise referred to as Retrieval Augmented Generation [RAG]; Guo et al., 2024), but this is not necessary. Chatbots can offer significant value to HRM and relevant stakeholders. For example, chatbots may be useful in answering frequently asked questions about the selection process by candidates. Similarly, they can be used to answer employee questions about where to find information about benefits, compensation, performance evaluations, and other materials employees need access to. One study in our review used chatbots: Fan et al. (2023) created a chatbot to infer personality. This is an especially promising use because this means that chatbots could help practitioners quickly measure job-related applicant characteristics. For example, engaging with a chatbot in a structured interview format may eventually act as an initial screening, especially if it serves as a realistic job preview. Information from the screening could not only be used to assess whether that candidate should continue in the selection process for the job they applied to but might also quickly give insight as to whether they are better matched for another job in the company. This latter opportunity is possible in human-only selection processes, but the efficiencies are enabled by automated scoring processes from advanced text analyses. Like LLMs, significantly more research is needed to fully explore the potential of using chatbots to improve HRM processes.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.hrmr.2025.101078>.

## Data availability

Data will be made available on request.

## References

- Ahn, Y., & Greve, H. R. (2024). Cultural spawning: Founders bringing organizational cultures to their startup. *Organization Science*. <https://doi.org/10.1287/orsc.2023.17771>
- Akinola, M., Martin, A. E., & Phillips, K. W. (2018). To delegate or not to delegate: Gender differences in affective associations and behavioral responses to delegation. *Academy of Management Journal*, 61(4), 1467–1491. <https://doi.org/10.5465/amj.2016.0662>
- Ansari, S., Garud, R., & Kumaraswamy, A. (2016). The disruptor's dilemma: TiVo and the US television ecosystem. *Strategic Management Journal*, 37(9), 1829–1853. <https://doi.org/10.1002/smj.2442>
- Baikovich, A., & Wasserman, V. (2020). Mobilizing national identity and othering practices as means of resistance. *Organization Science*, 31(5), 1220–1247. <https://doi.org/10.1287/orsc.2019.1345>
- Bailey, D. E., Leonardi, P. M., & Chong, J. (2010). Minding the gaps: Understanding technology interdependence and coordination in knowledge work. *Organization Science*, 21(3), 713–730. <https://doi.org/10.1287/orsc.1090.0473>
- Banks, G. C., Woznyj, H. M., Wesslen, R. S., Frear, K. A., Berka, G., Heggstad, E. D., & Gordon, H. L. (2019). Strategic recruitment across borders: An investigation of multinational enterprises. *Journal of Management*, 45(2), 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(4), 445–459. <https://doi.org/10.1007/s10869-017-9528-3>
- Ben-Menahem, S. M., Von Krogh, G., Erden, Z., & Schneider, A. (2016). Coordinating knowledge creation in multidisciplinary teams: Evidence from early-stage drug discovery. *Academy of Management Journal*, 59(4), 1308–1338. <https://doi.org/10.5465/amj.2013.1214>
- Bezrukova, K., Spell, C. S., Caldwell, D., & Burger, J. M. (2016). A multilevel perspective on faultlines: Differentiating the effects between group- and organizational-level faultlines. *Journal of Applied Psychology*, 101(1), 86–107. <https://doi.org/10.1037/apl0000039>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.5555/944919.944937>
- Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly*, 15(2), 211–239. <https://doi.org/10.1016/j.leaqua.2004.02.005>
- Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies*, 48, 1866–1891. <https://doi.org/10.1111/j.1467-6486.2010.00972.x>
- Bosco, F. A., Uggerslev, K. L., & Steel, P. (2017). MetaBUS as a vehicle for facilitating meta-analysis. *Human Resource Management Review*, 27(1), 237–254. <https://doi.org/10.1016/j.hrmr.2016.09.013>
- Brès, L., & Gond, J. P. (2014). The visible hand of consultants in the construction of the markets for virtue: Translating issues, negotiating boundaries and enacting responsive regulations. *Human Relations*, 67(11), 1347–1382. <https://doi.org/10.1177/0018726713519278>
- Cameron, L. D., & Rahman, H. (2022). Expanding the locus of resistance: Understanding the co-constitution of control and resistance in the gig economy. *Organization Science*, 33(1), 38–58. <https://doi.org/10.1287/orsc.2021.1557>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>

- Campion, E. D., & Campion, M. A. (2020). Using computer-assisted text analysis (CATA) to inform employment decisions: Approaches, software, and findings. *Research in Personnel and Human Resources Management*, 38, 285–325. <https://doi.org/10.1108/S0742-73012020000038010>
- Campion, E. D., & Campion, M. A. (2024). Impact of machine learning on personnel selection. *Organizational Dynamics*, 53(1). <https://doi.org/10.1016/j.orgdyn.2024.101035>
- Campion, E. D., & Campion, M. A. (2025). Text analysis using artificial intelligence as a tool for job analysis and job design. In S. Parker, F. Klonek, C. Knight, & Z. Zhang (Eds.), *SIOP Frontier Series - Transformative Work Design: Synthesis and New Directions*. Oxford University Press.
- Campion, E. D., Campion, M. C., & Campion, M. A. (2019). Best practices when using 360 degree feedback for performance appraisal. In A. H. Church, D. W. Bracken, J. W. Fleenor, & D. S. Rose (Eds.), *Handbook of Strategic 360 Feedback* (pp. 19–59). Oxford University Press. <https://doi.org/10.1093/oso/9780190879860.003.0003>
- Campion, E. D., Campion, M. A., Johnson, J., Carretta, T. R., Romay, S., Dirr, B., ... Mouton, A. (2024). Using natural language processing to increase prediction and reduce subgroup differences in personnel selection decisions. *Journal of Applied Psychology*, 109(3), 307–338. <https://doi.org/10.1037/apl0001144>
- Campion, E. D., & Csillag, B. (2022). Multiple jobholding motivations and experiences: A typology and latent profile analysis. *The Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000920>
- Campion, M. A., & Campion, E. D. (2023). Machine learning applications to personnel selection: Current illustrations, lessons learned, and future research. *Personnel Psychology*. <https://doi.org/10.1111/peps.12621>
- Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, 101(7), 958–975. <https://doi.org/10.1037/apl0000108>
- Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior*, 18(S1), 533–558. [https://doi.org/10.1002/\(SICI\)1099-1379\(199711\)18:1+<533::AID-JOB906>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1099-1379(199711)18:1+<533::AID-JOB906>3.0.CO;2-3)
- Caza, B. B., Moss, S., & Vough, H. (2018). From synchronizing to harmonizing: The process of authenticating multiple work identities. *Administrative Science Quarterly*, 63(4), 703–745. <https://doi.org/10.1177/0001839217733972>
- Chatman, J. A., Caldwell, D. F., O'Reilly, C. A., & Doerr, B. (2014). Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior*, 35(6), 785–808. <https://doi.org/10.1002/job.1928>
- Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, 31(1). <https://doi.org/10.1016/j.hrmr.2019.100698>
- Choudhury, P., Foroughi, C., & Larson, B. (2021). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal*, 42(4), 655–683.
- Claus, L., & Tracey, P. (2020). Making change from behind a mask: How organizations challenge guarded institutions by sparking grassroots activism. *Academy of Management Journal*, 63(4), 965–996. <https://doi.org/10.5465/amj.2017.0507>
- Cronbach, L. J., & Meehl, P. C. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cruz, D., & Meisenbach, R. (2018). Expanding role boundary management theory: How volunteering highlights contextually shifting strategies and collapsing work-life role boundaries. *Human Relations*, 71(2), 182–205. <https://doi.org/10.1177/0018726717718917>
- Curchod, C., Patriotta, G., Cohen, L., & Neysen, N. (2020). Working for an algorithm: Power asymmetries and agency in online work settings. *Administrative Science Quarterly*, 65(3), 644–676. <https://doi.org/10.1177/0001839219867024>
- DesJardine, M. R., & Shi, W. (2020). CEO temporal focus and behavioral agency theory: Evidence from mergers and acquisitions. *Academy of Management Journal*, 64(1), 265–292. <https://doi.org/10.5465/amj.2018.1470>
- Dixon, K. R., & Panteli, N. (2010). From virtual teams to virtuality in teams. *Human Relations*, 63(8), 1177–1197. <https://doi.org/10.1177/0018726709354784>
- Drisko, J., & Maschi, T. (2016). *Content Analysis*, New York, NY. Oxford University Press.
- Dupree, C. H. (2024). Words of a leader: The importance of intersectionality for understanding women leaders' use of dominant language and how others receive it. *Administrative Science Quarterly*, 69(2), 271–323. <https://doi.org/10.1177/00018392231223142>
- Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, 32(4), 1246–1264. <https://doi.org/10.5465/amr.2007.26586086>
- Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, 26(4), 398–427. <https://doi.org/10.1037/met0000349>
- Elg, U., Ghauri, P. N., Child, J., & Collinson, S. (2017). MNE microfoundations and routines for building a legitimate and sustainable position in emerging markets. *Journal of Organizational Behavior*, 38(9), 1320–1337. <https://doi.org/10.1002/job.2214>
- Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology*, 108(8), 1277–1299. <https://doi.org/10.1037/apl0001082>
- Felps, W., Mitchell, T. R., Hekman, D. R., Lee, T. W., Holtom, B. C., & Harman, W. S. (2009). Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, 52(3), 545–561. <https://doi.org/10.5465/amj.2009.41331075>
- Gamache, D. L., McNamara, G., Mannor, M. J., & Johnson, R. E. (2015). Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. *Academy of Management Journal*, 58(4), 1261–1282. <https://doi.org/10.5465/amj.2013.0377>
- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. WH Freeman.
- Gibbs, C. B., Gibbs, J. L., Stanko, T. L., Tesluk, P., & Cohen, S. G. (2011). Including the “I” in virtuality and modern job design: Extending the job characteristics model to include the moderating effect of individual experiences of electronic dependence and copresence. *Organization Science*, 22(6), 1481–1499. <https://doi.org/10.1287/orsc.1100.0586>
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Glaser, B. G. (1999). The future of grounded theory. *Qualitative Health Research*, 9(6), 836–845. <https://doi.org/10.1177/104973299129122199>
- Glerum, D. R., Joseph, D. L., McKenny, A. F., & Fritzsche, B. A. (2021). The trainer matters: Cross-classified models of trainee reactions. *Journal of Applied Psychology*, 106(2), 281–299. <https://doi.org/10.1037/apl0000503>
- Guo, F., Gallagher, C. M., Sun, T., Tavooisi, S., & Min, H. (2024). Smarter people analytics with organizational text data: Demonstrations using classic and advanced NLP models. *Human Resource Management Journal*, 34(1), 39–54. <https://doi.org/10.1111/1748-8583.12426>
- Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V., Wang, M., ... Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586–632. <https://doi.org/10.5465/annals.2017.0099>
- Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, 40(8), 1316–1330. <https://doi.org/10.1002/smj.3023>
- Harwood, H., Roulin, N., & Iqbal, M. Z. (2024). “Anything you can do, I can do”: Examining the use of ChatGPT in situational judgement tests for professional program admission. *Journal of Vocational Behavior*, 154, Article 104013. <https://doi.org/10.1016/j.jvb.2024.104013>
- Hernandez, I., & Nie, W. (2023). The AI-IP: Minimizing the guesswork of personality scale item development through artificial intelligence. *Personnel Psychology*. <https://doi.org/10.1111/peps.12543>
- Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, 107(8), 1323–1351. <https://doi.org/10.1037/apl0000695>
- Hickman, L., Dunlop, P. D., & Wolf, J. L. (2024). The performance of large language models on quantitative and verbal ability tests: Initial evidence and implications for unproctored high-stakes testing. *International Journal of Selection and Assessment*. <https://doi.org/10.1111/ijsa.12479>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- Hiebl, M. R. (2023). Sample selection in systematic literature reviews of management research. *Organizational Research Methods*, 26, 229–261.

- Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276. <https://doi.org/10.1002/smj.2904>
- Jehn, K. A., & Bezrukova, K. (2004). A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior: Occupational and Organizational Psychology and Behavior*, 25(6), 703–729. <https://doi.org/10.1002/job.257>
- Jiang, Y. D., Straub, C., Klyver, K., & Mauer, R. (2021). Unfolding refugee entrepreneurs' opportunity-production process—Patterns and embeddedness. *Journal of Business Venturing*, 36(5). <https://doi.org/10.1016/j.jbusvent.2021.106138>
- Kang, J., & Han Kim, A. Y. (2017). The relationship between CEO media appearances and compensation. *Organization Science*, 28(3), 379–394. <https://doi.org/10.1287/orsc.2017.1128>
- Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, 61(2), 586–614. <https://doi.org/10.5465/amj.2016.1215>
- Koenig, N., Tonidandel, S., Thompson, I., Albritton, B., Koohifar, F., Yankov, G., ... Newton, C. (2023). Improving measurement and prediction in personnel selection through the application of machine learning. *Personnel Psychology*. <https://doi.org/10.1111/peps.12608>
- Landers, R. N., Auer, E. M., Dunk, L., Langer, M., & Tran, K. N. (2023). A simulation of the impacts of machine learning to combine psychometric employee selection system predictors on performance prediction, adverse impact, and number of dropped predictors. *Personnel Psychology*, 76(4), 1037–1060. <https://doi.org/10.1111/peps.12587>
- Liff, J., Mondragon, N., Gardner, C., Hartwell, C. J., & Bradshaw, A. (2024). Psychometric properties of automated video interview competency assessments. *Journal of Applied Psychology*, 109(6), 921–948. <https://doi.org/10.1037/apl0001173>
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.
- Machine Learning: What It Is and Why It Matters. (2019). SAS. Retrieved from [https://www.sas.com/en\\_us/insights/analytics/machine-learning.html](https://www.sas.com/en_us/insights/analytics/machine-learning.html).
- Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, 94(6), 1591–1599. <https://doi.org/10.1037/a0016539>
- Mathieu, J. E., Wolfson, M. A., Park, S., Luciano, M. M., Bedwell-Torres, W. L., Ramsay, P. S., ... Tannenbaum, S. I. (2021). Indexing dynamic collective constructs using computer-aided text analysis: Construct validity evidence and illustrations featuring team processes. *The Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000856>
- Mattarelli, E., Bertolotti, F., Prencipe, A., & Gupta, A. (2022). The effect of role-based product representations on individual and team coordination practices: A field study of a globally distributed new product development team. *Organization Science*, 33(4), 1423–1451. <https://doi.org/10.1287/orsc.2021.1487>
- Mavin, S., & Grandy, G. (2016). A theory of object appearance: Women elite leaders' intra-gender 'management' of bodies and appearance. *Human Relations*, 69(5), 1095–1120. <https://doi.org/10.1177/0018726715609107>
- Meinecke, A. L., & Kauffeld, S. (2019). Engaging the hearts and minds of followers: Leader empathy and language style matching during appraisal interviews. *Journal of Business and Psychology*, 34(4), 485–501. <https://doi.org/10.1007/s10869-018-9554-9>
- Menon, A., Choi, J., & Tabakovic, H. (2018, July). What you say your strategy is and why it matters: Natural language processing of unstructured text. In *Academy of Management Proceedings (Vol. 2018)* (p. 18319). Briarcliff Manor, NY: Academy of management.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Min, H., Peng, Y., Shoss, M., & Yang, B. (2021). Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic. *Journal of Applied Psychology*, 106(2), 214–229. <https://doi.org/10.1037/apl0000886>
- Min, H., Yang, B., Allen, D. G., Grandey, A. A., & Liu, M. (2024). Wisdom from the crowd: Can recommender systems predict employee turnover and its destinations? *Personnel Psychology*, 77(2), 475–496.
- Mittermaier, A., Patzelt, H., & Shepherd, D. A. (2023). Motivating prosocial venturing in response to a humanitarian crisis: Building theory from the refugee crisis in Germany. *Entrepreneurship Theory and Practice*, 47(3), 924–963. <https://doi.org/10.1177/10422587211102>
- Murphy, C., Klotz, A. C., & Kreiner, G. E. (2017). Blue skies and black boxes: The promise (and practice) of grounded theory in human resource management research. *Human Resource Management Review*, 27(2), 291–305. <https://doi.org/10.1016/j.hrmr.2016.08.006>
- Nadkarni, S., & Narayanan, V. K. (2005). Validity of the structural properties of text-based causal maps: An empirical assessment. *Organizational Research Methods*, 8(1), 9–40. <https://doi.org/10.1177/1094428104271999>
- Ng, W., & Sherman, E. L. (2022). In search of inspiration: External mobility and the emergence of technology intrapreneurs. *Organization Science*, 33(6), 2300–2321. <https://doi.org/10.1287/orsc.2021.1530>
- Nunnally, J. C. (1978). An overview of psychological measurement. In *Clinical Diagnosis of Mental Disorders: A Handbook* (pp. 97–146). [https://doi.org/10.1007/978-1-4684-2490-4\\_4](https://doi.org/10.1007/978-1-4684-2490-4_4)
- Palmer, I., Kabanoff, B., & Dunford, R. (1997). Managerial accounts of downsizing. *Journal of Organizational Behavior*, 18(S1), 623–639. [https://doi.org/10.1002/\(SICI\)1099-1379\(199711\)18:1+<623::AID-JOB909>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1379(199711)18:1+<623::AID-JOB909>3.0.CO;2-T)
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, 22(3), 765–797. <https://doi.org/10.1177/1094428117745648>
- Patton, E., & Johns, G. (2007). Women's absenteeism in the popular press: Evidence for a gender-specific absence culture. *Human Relations*, 60(11), 1579–1612. <https://doi.org/10.1177/0018726707084301>
- Peltokorpi, V., & Vaara, E. (2014). Knowledge transfer in multinational corporations: Productive and counterproductive effects of language-sensitive recruitment. *Journal of International Business Studies*, 45(5), 600–622. <https://doi.org/10.1057/jibs.2014.1>
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. Retrieved from <http://www.liwc.net/LIWC2007LanguageManual.pdf>.
- Phillips, J., & Robie, C. (2024). Can a computer outfake a human? *Personality and Individual Differences*, 217, Article 112434. <https://doi.org/10.1016/j.paid.2023.112434>
- Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly*, 65(1), 1–19. <https://doi.org/10.1177/0001839219887663>
- Principles for the Validation and Use of Personnel Selection Procedures. (2018). American Psychological Association. Retrieved from <https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf>.
- Putka, D. J., Oswald, F. L., Landers, R. N., Beatty, A. S., McCloy, R. A., & Yu, M. C. (2023). Evaluating a natural language processing approach to estimating KSA and interest job analysis ratings. *Journal of Business and Psychology*, 38, 385–410. <https://doi.org/10.1007/s10869-022-09824-0>
- Rothausen, T. J., Henderson, K. E., Arnold, J. K., & Malshe, A. (2017). Should I stay or should I go? Identity and well-being in sensemaking about retention and turnover. *Journal of Management*, 43(7), 2357–2385. <https://doi.org/10.1177/0149206315569312>
- Sajjadiani, S., Daniels, M. A., & Huang, H. C. (2024). The social process of coping with work-related stressors online: A machine learning and interpretive data science approach. *Personnel Psychology*, 77(2), 321–373. <https://doi.org/10.1111/peps.12538>
- Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezee, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*, 104(10), 1207–1225. <https://doi.org/10.1037/apl0000405>
- Schifeling, T., & Soderstrom, S. (2022). Advancing reform: Embedded activism to develop climate solutions. *Academy of Management Journal*, 65(6), 1775–1803. <https://doi.org/10.5465/amj.2019.0769>
- Schinoff, B. S., Ashforth, B. E., & Corley, K. G. (2020). Virtually (in) separable: The centrality of relational cadence in the formation of virtual multiplex relationships. *Academy of Management Journal*, 63(5), 1395–1424. <https://doi.org/10.5465/amj.2018.0466>
- 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(4), 941–968. <https://doi.org/10.1177/1094428118773858>

- Short, J. C., Broberg, J. C., Coglistter, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, 13(2), 320–347. <https://doi.org/10.1177/1094428109335949>
- Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 415–435. <https://doi.org/10.1146/annurev-orgpsych-032117-104622>
- Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, 71(3), 299–333. <https://doi.org/10.1111/peps.12263>
- Speer, A. B. (2021). Scoring dimension-level job performance from narrative comments: Validity and generalizability when using natural language processing. *Organizational Research Methods*, 24(3), 572–594. <https://doi.org/10.1177/1094428120930815>
- Speer, A. B., Perrotta, J., Tenbrink, A. P., Wegmeyer, L. J., Delacruz, A. Y., & Bowker, J. (2023). Turning words into numbers: Assessing work attitudes using natural language processing. *Journal of Applied Psychology*, 108(6), 1027–1045. <https://doi.org/10.1037/apl0001061>
- Speer, A. B., Schwendeman, M. G., Reich, C. C., Tenbrink, A. P., & Siver, S. S. (2019). Investigating the construct validity of performance comments: Creation of the great eight narrative dictionary. *Journal of Business and Psychology*, 34, 747–767. <https://doi.org/10.1007/s10869-018-9599-9>
- Standards for Educational and Psychological Testing. (2014). American Educational Research Association. Retrieved from <https://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition>.
- Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: Implications for training. *Training Research Journal*, 2, 85–116.
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, 40(3), 371–384. <https://doi.org/10.2307/2393788>
- Tuncel, E., Kong, D. T., Parks, J. M., & van Kleef, G. A. (2020). Face threat sensitivity in distributive negotiations: Effects on negotiator self-esteem and demands. *Organizational Behavior and Human Decision Processes*, 161, 255–273. <https://doi.org/10.1016/j.obhdp.2020.07.004>
- Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods*, 20(2), 243–267. <https://doi.org/10.1177/1094428115610808>
- Uniform Guidelines on Employee Selection Procedures. (1978). Office of Personnel Management. Retrieved from <https://www.opm.gov/FAQs/QA.aspx?fid=a6da6c2e-e1cb-4841-b72d-53eb4adf1ab1&pid=402c2b0c-bb5c-44e9-acbc-39cc6149ad36>.
- Valtonen, L., Mäkinen, S. J., & Kirjavainen, J. (2022). Advancing reproducibility and accountability of unsupervised machine learning in text mining: Importance of transparency in reporting preprocessing and algorithm selection. *Organizational Research Methods*. <https://doi.org/10.1177/10944281221124947>
- Walker, D. D., van Jaarsveld, D. D., & Skarlicki, D. P. (2017). Sticks and stones can break my bones but words can also hurt me: The relationship between customer verbal aggression and employee incivility. *Journal of Applied Psychology*, 102(2), 163–179. <https://doi.org/10.1037/apl0000170>
- Wang, M., McAuslan, P., DiMambro, J. M., & Mięgoć, N. (2017). Impression management use in resumes and cover letters. *Journal of Business and Psychology*, 32, 727–746. <https://doi.org/10.1007/s10869-016-9470-9>
- Welch, C., & Piekkari, R. (2017). How should we (not) judge the ‘quality’ of qualitative research? A re-assessment of current evaluative criteria in international business. *Journal of World Business*, 52(5), 714–725. <https://doi.org/10.1016/j.jwb.2017.05.007>
- Wilhelmy, A., Kleinmann, M., König, C. J., Melchers, K. G., & Truxillo, D. M. (2016). How and why do interviewers try to make impressions on applicants? A qualitative study. *Journal of Applied Psychology*, 101(3), 313–332. <https://doi.org/10.1037/apl0000046>
- Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, 160, 131–148. <https://doi.org/10.1016/j.obhdp.2020.03.011>
- Zachary, M. A., Connelly, B. L., Payne, G. T., & Tribble, L. L. (2023). Virtue rhetoric in investor communications: Setting up for a letdown? *Journal of Management*, 49(2), 741–770. <https://doi.org/10.1177/01492063211002622>
- Zhang, C., Yu, M. C., & Marin, S. (2021). Exploring public sentiment on enforced remote work during COVID-19. *Journal of Applied Psychology*, 106(6), 797. <https://doi.org/10.1037/apl000093>