



# Psycho-managerial text mining (PMTM): a framework for developing and validating psychological/managerial constructs from a theory/text-driven approach

Jorge Iván Pérez Rave<sup>1,2</sup> · Gloria Patricia Jaramillo Álvarez<sup>3</sup> · Juan Carlos Correa Morales<sup>4</sup>

Revised: 25 April 2022 / Accepted: 28 July 2022 / Published online: 22 August 2022  
© The Author(s), under exclusive licence to Springer Nature Limited 2022

## Abstract

This paper addresses two concerns in the construction and validation of psychological/managerial constructs from texts: (1) The procedure for studying content validity is mainly based on human tasks, subjectivity, limited to small data, and difficult to reproduce/replicate; (2) Traditional latent variable representation (a merged list or random parcel of words) lacks theoretical rationality and does not guarantee the representativity of valuable qualities, perceptual entities, and actions of the construct. The paper proposes a framework—PMTM—for developing/validating measures of psychological/managerial constructs from texts (three stages and 16 steps) by incorporating linguistic, psychometrical, and computational resources. PMTM includes five new properties to guide content validity analysis automatically, and contains a new method for refining dictionaries, entitled “embedded voting,” which emulates expert judgments. In addition, the paper provides empirical evidence derived from the application of PMTM to the case of transformational leadership using data from organizational documents, interviews, essays, blog posts, surveys, and speeches by former US presidents. The framework can also be used as a comprehensive template with the potential to exploit customer/service texts from a theory/text-driven approach to overcome recent challenges stated by the marketing literature, including the lack of standardization and inconsistency in the definitions and measurement of marketing constructs.

**Keywords** Text mining · Text analytics · Constructs analysis · Automated dictionary creation · Theory/text-driven approach · Transformational leadership

## Introduction

Management research is mainly carried out using psychological/managerial construct analysis. This variable type is “a theoretical label that is given to some human attribute or ability that cannot be seen or touched, because it develops in the brain” (Brown 1988, p. 103). The standard inputs for empirically studying these constructs are obtained from structured questionnaires. For example, transformational leadership (TL), “a process by which one or more people influence others to pursue a commonly held objective” (Bohara and Tiwari 2015, p. 383), has been extensively measured through structured questionnaires (e.g., Avolio and Bass 1991; Carless et al. 2000; Dönmez and Toker, 2017; Rafferty and Griffin 2004). According to Edwards (2008), these instruments differ in the content representing TL (non-standardized TL content) and commonly require administration fees. Moreover, Andersen (2018) assert that more clarity in the construct operationalization is required, and

✉ Jorge Iván Pérez Rave  
investigacion@idinnov.com

Gloria Patricia Jaramillo Álvarez  
gpjarami@unal.edu.co

Juan Carlos Correa Morales  
jccorrea@unal.edu.co

<sup>1</sup> Universidad Nacional de Colombia, Medellín, Colombia

<sup>2</sup> IDINNOV Research Group, IDINNOV S.A.S., Medellín, Colombia

<sup>3</sup> Department of Computing Science, Universidad Nacional de Colombia, Medellín, Colombia

<sup>4</sup> School of Statistics, Universidad Nacional de Colombia, Medellín, Colombia



Valero and Jang (2020) warn of frequent problems concerning social desirability bias in TL instruments.

Recently, however, there has been increased scientific interest in developing and validating construct scales based on term dictionaries using text mining (TM). For example, some authors have studied entrepreneurial orientation (Short et al. 2010), communication, critical thinking, and leadership skills (Campion et al. 2016), health responsibility (Kjellström and Golino 2018), personal values (Ponizovskiy et al. 2020), and organizational culture (Pandey and Pandey 2017).

Although previous studies have generated valuable insights for organizational management research and related fields, we aim to address two concerns to encourage and efficiently facilitate the future development and validation of constructs from texts.

The first concern is the traditional procedure to study content validity, which is almost the same as that currently used in structured questionnaire data. It is based on traditional expert rater and the theory-driven approach (for construct definition and dimensionality) (Pandey and Pandey 2017; Ponizovskiy et al. 2020; Short et al. 2010). Such a standard procedure is valuable from the interpretivism paradigm; however, it is abstract and difficult to replicate on a large scale (e.g., big data) and presents complications in reproducing, replicating, and generalizing any findings. Besides, such a procedure is rarely informative about the linguistic properties of construct scales. These properties are fundamental, because the content reflected in textual data is a function of its linguistic constituents, structural relationships, and variables of the linguistic context (Espinosa 2017; Firth 1957). However, “most studies on terminology focus on nouns and do not take into account the use of terms in context.” (Campos and Castells 2010, p. 872). Moreover, considering Wüster (1979) (e.g., the importance of nouns), Cabré (1999, 2002), and Campos and Castells (2010) (e.g., verbs/adjectives as potential specific-domain units), a comprehensive dictionary construction should consider all these parts of speech (PoS).

Furthermore, the standard procedure (expert rater, theory-driven approach, and multiple/representative corpora) to examine the content validity of construct scales based on texts could be complemented by considering measures based on a combination of linguistics, computational, and psychometrical resources from a text-driven perspective. Previous studies (e.g., Champion et al. 2016; Kjellström and Golino 2018; Pandey and Pandey 2017; Pérez-Rave et al. 2021a; Ponizovskiy et al. 2020) highlight the value of adopting a mixed perspective (based on theory/deductive and data/inductive). However, the use of a text-driven approach in this psychological/managerial field has been limited to the use of multiple/representative corpora during dictionary construction to expand its vocabulary (Ponizovskiy et al. 2020).

Hence, the nature of its linguistic properties and how these properties can support construct content validity are rarely identified and discussed.

The second interest concern is the construct representation based on observable indicators. The traditional representation comprises a consolidated list of words for each construct (one-item perspective, e.g., Champion et al. 2016; Ponizovskiy et al. 2020) or a multiple-item view. In the latter, a merged list of terms is randomly split into parcels (frequently three parcels for each construct, e.g., Pandey and Pandey 2017). Although this procedure tends to produce correlated items and enables confirmatory factor analysis (CFA), the representation of construct reflected in texts is merely justified in numerical purposes, lacks theoretical rationality, and does not guarantee the representativity of valuable linguistic manifestations of the construct. For example, the traditional merged list of terms or random parcels of terms, without controlling PoS, does not guarantee the representativity of perceptual entities (e.g., nouns or noun phrases), qualities (e.g., adjectives or adjective phrases), and actions (e.g., verbs or verb phrases) in the measurement model, which involve different brain mechanisms (Fyshe et al. 2019; Haan et al. 2000; Martin et al. 1995). Providing empirical evidence on this linguistic representation of models during the construct development and validation stages enables more construct completeness and new lexical, syntactic, and semantic analyses. Moreover, this representation differentiates the construct variability and errors/PoS during the CFA from an adapted correlated-uniqueness perspective (Batista-Foguet and Coenders-Gallart 2000; Marsh et al. 1992). However, currently, it is unknown whether this representation (based on correlated PoS) could provide good performance concerning its psychometric properties.

The paper addresses these two concerns to propose a comprehensive alternative for studying psychological/managerial latent variables reflected in texts. The paper is based on and extends valuable suggestions from Pandey and Pandey (2017), Ponizovskiy et al. (2020), and the works of their predecessors (e.g., Short et al. 2010) by proposing a framework for developing and validating scales of psychological/managerial constructs from a theory/text-driven approach.

The paper provides four contributions. The first is complementary measures (coherence, polarity, PoS balance, convergence/differentiation, and commonality) to inform/guide content validity from a text-driven approach; these measures provide insights within and between dictionary construction steps. The second is a method to refine construct dictionaries, considering linguistic properties, entitled embedded voting, which uses word2vec and cosine similarity and emulates expert judgments from a text-driven perspective. The third is a dictionary of terms representing transformational leadership (TL) and original evidence on six performance criteria (content/internal/external validity, reliability, equity,



and practical value). The fourth is an integrative framework to develop and validate measures of psychological/managerial constructs from texts. It consists of three stages and 16 steps that incorporate the first three entitled contributions and employ linguistic, psychometrical, and computational resources.

In addition, the contributions of this paper allow us to extend frameworks for processing/analyzing textual data to discover and examine latent topics in marketing domains, such as those by Markham et al. (2015) in new product development (defining questions, identifying data sources, creation of dictionaries and rules exploring marketing terms, data collection/analysis/assessment, data scoring, and decision-making); Amado et al. (2018) in big data in marketing (collecting literature, dictionary creation, building a document-term matrix, and performing topic modeling); Mathaisel and Comm (2021) in political marketing (preprocessing texts, performing topic modeling methods, sentiment analysis, and data visualization); and Pérez-Rave et al. (2021a) in multi-criteria decision-making processes at the strategic level with several applications to customer analytics (multi-criteria case, structuring of the case from a data-driven approach, scoring of latent topics, alternative prioritization, and robustness of the solution). These previous works are valuable for information extraction/recovery (e.g., topic analysis) from texts but do not cover the field of developing and validating psychological/managerial constructs from texts. This field demands a psychometrical perspective, a mixed approach (theory/data; Short et al. 2010; Pandey and Pandey 2017; Pérez-Rave et al. 2021a; Ponizovskiy et al. 2020), and to transcend the traditional focus on lexical analysis, predefined computational methods (e.g., topic modeling; Latent Dirichlet Allocation: LDA, Blei et al. 2003), and typical performance indicators merely focused on prediction capabilities. The area of psychological/managerial constructs requires qualitative and quantitative psychometrical rigor to ensure the quality of the outputs (e.g., dimensionality and construct scores): theoretically supported and empirically plausible/valid/reliable (Bergkvist and Eisend 2021; Fried 2017; Lievens et al. 2002; Martínez et al. 2006). However, unlike organizational management disciplines, the consideration of psychometrical resources in computational fields (e.g., text/data analytics) and their application to management/marketing domains are still in their infancy (Strohmeier and Piazza 2013; Pérez-Rave et al. 2020, 2021a, b). Thus, we adopt a mixed perspective to provide a more comprehensive, efficient, objective, and controlled framework in the development and validation of constructs from texts considering linguistic, psychometrical, and computational resources.

In addition, marketing research is intensive in latent variables and covers psychological/managerial constructs intrinsic to customers' viewpoints, such as service quality,

satisfaction, loyalty, perceived value, service experience, and service engagement (Cronin Jr et al. 2021; Pérez-Ravnde a Muñoz-Giraldo 2014). However, the study of latent marketing variables has been confined to structured questionnaires, which are invasive and partially explore the construct manifestations by limiting the respondents' ability to freely express their opinions about an object of study (Kugbonu 2020). Indeed, recent studies warn of consistencies and a lack of standardization in the definition and measurement of marketing constructs (Bergkvist and Eisend 2021; Bergkvist and Langner 2017; Ptok et al. 2018). Furthermore, it is accepted in the literature that questionnaire data are focused on small sample sizes and are commonly affected by social desirability, acquiescence bias, blank responses, and straight-lining (Cornesse and Blom 2020; Grimm 2010; Holmes et al. 2019; Hume 2017; Kalugina et al. 2019). The developed framework is also valuable for marketing research/practice, since it is necessary to discuss and propose new/comprehensive/reproducible forms for operationalizing marketing constructs (Cronin Jr et al. 2021).

This paper is organized into five sections. "Introduction" section argues the need for the study within the literature. "Related works on psychological/managerial constructs reflected in texts" section exposes the basis of the study. "Methodology" section provides the developed framework and explains the proposed properties and methods. "Results of the empirical application of PMTM in the study of TL" section presents the results accompanied by interpretations. "General discussion" section comprises the general discussion of such results. "Conclusions and future work" section draws the conclusions and suggests future work.

## Related works on psychological/managerial constructs reflected in texts

Table 1 provides a comparative analysis between our framework and most representative frameworks on the study of psychological/managerial constructs using TM from a theory/data-driven approach.

Based on Table 1, the present study aims to:

- (i) enable the automated (or semi-automated) execution of dictionary creation and content validation concerning psychological/managerial constructs from texts controlling PoS by incorporating linguistic, psychometrical, and computational resources. This configuration of resources aims to attend to the first concern discussed in the Introduction to favor its use in large-scale studies (massive data) and reduce the subjectivity/abstraction in the traditional procedure of developing/validating construct measures from texts (expert raters, theory-driven approach,



**Table 1** Comparing frameworks suggested by representative studies on construct measurement from texts

| No. | Criteria                                  | Pandey and Pandey (2017) (based mainly on Short et al. 2010)  | Ponizovskiy et al (2020)   | Our study/framework   |
|-----|---|---|--|---|
| 1   | Previous referents                        | Mainly built upon and advances Short et al (2010), Illia et al (2014)   | Mainly built upon and advances Graham et al (2009), Pietraszkiewicz et al. (2019), Tausczik and Pennebaker (2010)  | Mainly built upon and advances Pandey and Pandey (2017), Ponizovskiy et al (2020), and their predecessors, such as Short et al. (2010)  |
| 2   | Application cases                         | Organizational culture (based on Denison and Neale 2000; Cameron and Quinn, 1999; and O'Reilly et al. 1991)   | Personal values (based on Schwartz's theory of basic human values; Schwartz, 1992; Schwartz et al. 2012)   | Transformational leadership (based on Avolio and Bass 1991, 1999, 2004)   |
| 3   | Theory-driven approach and its main scope | Used for construct meaning, dimensionality, and initial items (terms)   | Used for construct meaning, dimensionality, and initial items (terms)  | Used for construct meaning, dimensionality, and to define the seed corpus/corpora   |
| 4   | Data-driven approach and its main scope   | This is used to complement the dictionary initialization considering chunks derived from natural language processing. Furthermore, expert raters are used in the dictionary refinement stage                                  | This is used to complement the dictionary initialization considering word counts and co-occurrences in corpora. In addition, exploratory factor analysis and expert raters are used in the dictionary refinement stage | This is used to perform two automatic moments of dictionary expansion from two types of corpora (scientific and pragmatic), considering semantic similarity, distributed dictionary representations, and chunks from natural language processing  |
| 5   | Dictionary refinement orientation         | Wide-narrow: An extensive list of terms that are reduced using exploratory factor analysis and expert raters (manually)   | Wide-narrow: A large list of terms that are reduced using exploratory factor analysis and expert raters (manually)   | Narrow-wide: A small list of terms that are expanded considering semantic similarity from procedures based on distributed dictionary representations and two corpora (automatically)  |
| 6   | Content validity approach                 | Qualitatively and manually establishing construct definition, dimensionality, and initial terms using the suggestions of Short et al (2010): (a) Deductive procedure based on a priori theory; (b) reviews of content experts | Qualitatively and manually establishing construct definition, dimensionality, and initial terms using: (a) deductive procedure based on a priori theory; (b) reviews of content experts                                | (a) Qualitatively and manually establishing construct definition and dimensionality; (b) a seed corpus (exact definitions and descriptions from scientific papers and the formal guide of the construct); (c) automated expansion from representative corpora in several domains; (d) informed decision-making based on five proposed metrics automatically calculated within and between dictionary construction steps |
| 7   | Content validity metrics                  | One based on expert raters (inter-coder agreement)  | One based on expert raters   | Five are automatically obtained based on linguistics, psychometrical, and computational resources (coherence, PoS balance, polarity, communality, convergence/differentiation)  |
| 8   | Machine-learning methods                  | No. These are suggested for future work   | No. These are suggested for future work  | Yes. These are used in a new method (embedded voting, which uses embedded vectors, e.g., <i>word2vec</i> ); a new property (embedded-content convergence/differentiation), and predictive validity (e.g., classification trees, bagging) considering the microeconomic approach of TM   |
| 9   | Observable indicators                     | These are derived from a merged list of noun and verb phrases   | These are derived from a merged list of words (unigrams)   | Three indicators represent different PoS for each dimension (nouns, adjectives, and verbs)  |



Table 1 (continued)

| No. | Criteria                         | Pandey and Pandey (2017) (based mainly on Short et al. 2010)  | Ponizovskiy et al (2020)   | Our study/framework  |
|-----|----------------------------------|---|--|--|
| 10  | Representative corpora           | Multiple, offline-generated, and derived from organizations (e.g., annual reports and letters to company shareholders)                          | Multiple, offline/offline-generated, and derived from individuals (e.g., blog posts, essays) | Multiple, offline/online generated, and derived from individuals (blog posts, open-ended questions, and essays) and organizations (e.g., annual reports and letters to shareholders) |
| 11  | Representative domains           | Corpora from formal domains (strategic organizational levels)   | Corpora from everyday domains  | Corpora from formal and casual domains (strategic organizational levels and everyday environments)   |
| 12  | Internal and external validities | First/second-order CFA, contrast of reference groups (using <i>t</i> test), multiple corpora, and correlation analysis with reference variables | Correlation analysis with reference variables and multiple corpora                           | First/second-order CFA with correlated PoS, correlation analysis with reference variables, contrast of groups, and multiple corpora  |
| 13  | Criterion validity               | Traditional (OLS) and robust regression analysis  | Correlation analysis with output variables   | Correlation analysis with output variables, and three machine learning methods (logistic regression, classification trees, and bagging)  |
| 14  | Equity analysis                  | It is not empirically illustrated   | It is not empirically illustrated  | This is based on regression analysis (with and without bootstrapping) of the construct scores by gender and age  |
| 15  | Practical value                  | It is not empirically illustrated   | It is not empirically illustrated  | Considering patterns about the last five US presidents spanning the period 01/2001 – 01/2021   |

PoS (parts of speech); TM (text mining); CFA (confirmatory factor analysis); OLS (ordinary least squares)



e.g., Short et al. 2010; Pandey and Pandey 2017; Ponizovskiy et al. 2020). Thus, we develop and use five original properties (coherence, PoS balance, polarity, communality, convergence/differentiation) that guide the development and content validation of psychological/managerial constructs reflected in texts, which have not been considered in previous works.

- (ii) represent and confirm constructs reflected in texts using a correlated PoS model and considering internal validity, external validity, reliability, equity, and practical value. Thus, we argue and validate a new psychometrical representation of constructs reflected in texts to ensure the representativity of valuable linguistic manifestations of the construct (perceptual entities, qualities, and actions); hence, this representation transcends the traditional representation based on word parcels (Campion et al. 2016; Ponizovskiy et al. 2020), which is merely justified in numerical purposes. Moreover, our methodological framework empirically validates this construct representation (measurement model based on correlated PoS) using a more comprehensive psychometrical perspective that considers internal/external validity, reliability, equity, and practical value.
- (iii) provide original evidence from a systematic application of such proposals from a theory/text-driven approach to develop and validate a new scale of TL. Thus, we extend the uses of text analytics by considering a construct (TL) not addressed in the reference works, which has been at the top of leadership research for more than three decades with support from structured questionnaires (Alqatawenh, 2018; Jackson, 2020; Kotamena et al. 2020). “MLQ (Multifactor Leadership Questionnaire, Avolio and Bass 1991, 1999, 2004) is the most frequently utilized TL instrument; its name has become synonymous with TL” (Brown and Keeping 2005, p. 249) and still requires effort to increase its operationalization quality (Andersen 2018; Edwards 2008). Our application case represents TL dimensionality from the four dimensions put forth by Avolio and Bass (1999, 2004, 1991): individual consideration; idealized influence; intellectual stimulus; and inspirational motivation.

## Methodology

### Developed framework

Based on the considerations presented in Table 1, Fig. 1 details the proposed framework, entitled Psycho-Managerial

Text Mining (PMTM). This comprises three integrative stages and 16 steps. Stage “A” refers to dictionary development and content validation, assuming that content validity can be “ensured by the plan and procedures of construction” (Nunnally 1978, p. 92, also cited by Pandey and Pandey 2017, p. 17) by adopting a theory/text-driven approach. Stage “B” focuses on confirming internal validity by developing a CFA representation comprising PoS and the correlated-uniqueness approach (Batista-Foguet and Coenders-Gallart 2000). This stage also examines scale reliability and confirms external validity based on a correlation analysis with reference/output variables and group contrasts (Martínez et al. 2006; Pandey and Pandey 2017). Moreover, this stage examines scale equity by considering personal factors, such as gender and age, usually implied in the diversity–validity dilemma (Martínez et al. 2006; Pérez-Rave et al. 2021a). Stage “C” focuses on exploring potential practical uses derived from the scale, which is essential in the implementation science that underlies TM.

From a text-driven approach, PMTM (see Fig. 1) introduces five proposed linguistic properties, one method to emulate expert voting in lexical decisions (entitled embedded voting), and a representation of the measurement model based on correlated PoS. These are argued and explained in the following section.

### Proposed properties for exploring content validity

We conceptualize and operationalize the five properties (content commonality, content polarity, PoS balance, embedded-content convergence/differentiation, and content coherence). The ideation and development of these properties emerge from a creative process utilizing an inductive perspective based on the authors’ knowledge of theoretical/methodological resources in management, psychometry and data science; and recognizing: (a) conceptualizations and methods for studying psychological/managerial constructs from psychometrical (Fried 2017; Lievens et al. 2002; Martínez et al. 2006) and data science approaches based on structured questionnaires (Pérez-Rave et al. 2021b, 2022); (b) the value of PoS to understand speaking styles and the thinking of individuals and organizations (Cabré, 1999, 2002; Ghosh and Mishra 2020; Wüster 1979); (c) the fact that nouns, adjectives, and verbs involve different brain mechanisms during text production (Fyshe et al. 2019; Haan et al. 2000; Martin et al. 1995); (d) the lack of understanding/exploitation of the linguistic context during tasks involved in the development and content validation of constructs reflected in texts, and the high subjectivity and limited reproducibility underlying these tasks (see “[Related works on psychological/managerial constructs reflected in texts](#)” section); and (e) the need to incorporate a data-driven approach to complement the theory-driven approach



**PSYCHO-MANAGERIAL TEXT MINING (PMTM)**

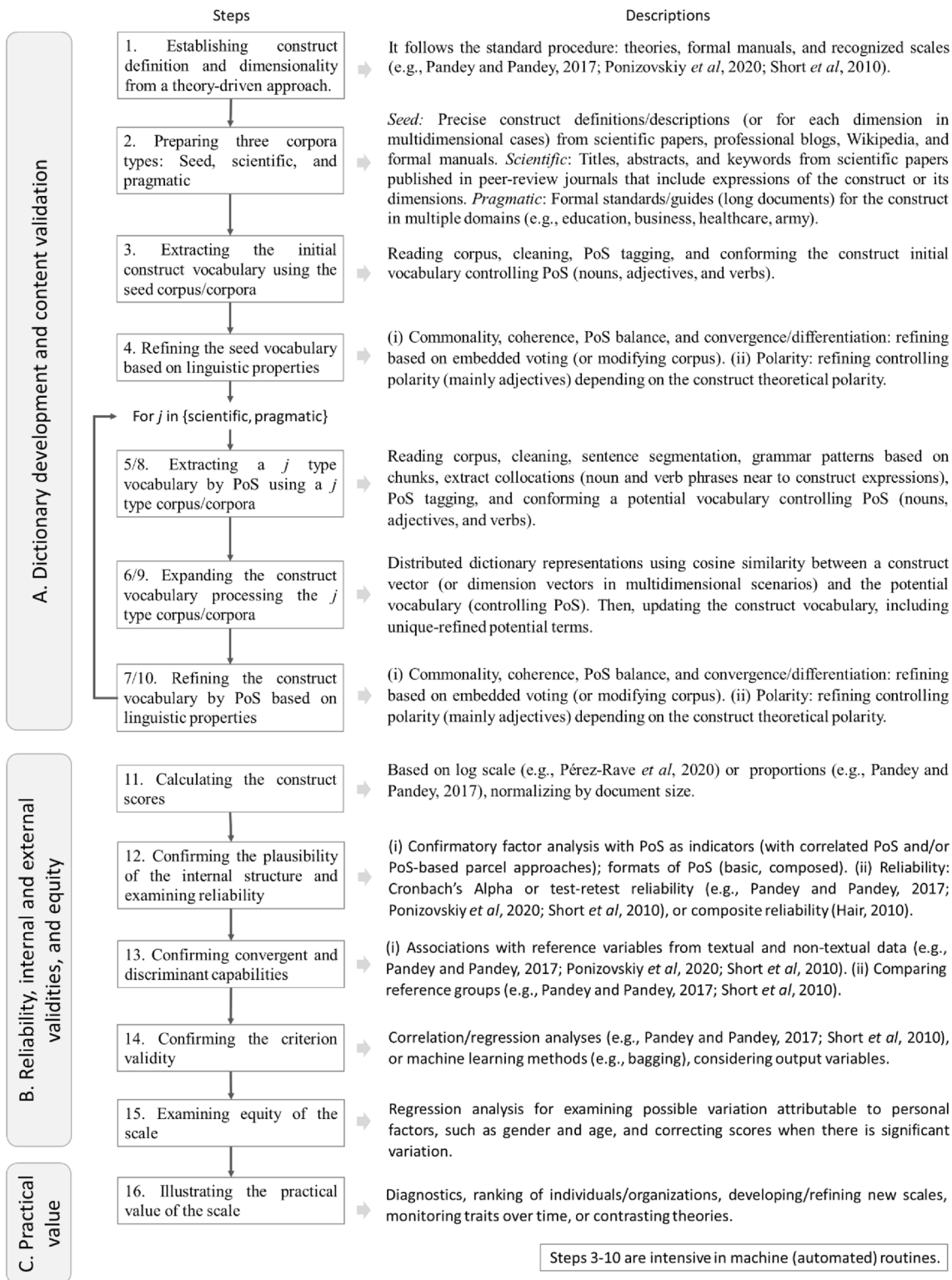


Fig. 1 PMTM framework to constructs' dictionary creation/validation. Note PoS (parts of speech)



in the study and use of psychological/managerial constructs reflected in texts (Illia et al. 2014; Nunnally 1978; Short et al. 2010; Pandey and Pandey 2017; Ponizovskiy et al. 2020). In "Results of the empirical application of PMTM in the study of TL" section (Results), we empirically examine the proposed properties in the case of TL.

**Content commonality**

A frequent practice in multidimensional construct dictionary development is to remove shared words (not necessarily stopwords), arguing that this prevents lexical correlation during the dictionary design (Pandey and Pandey 2017; Ponizovskiy et al. 2020). However, using shared words is a natural situation that should be considered in multidimensional constructs. For example, based on attentional homogeneity, it is possible to assume that certain writers with an interest in a psychological/managerial construct may share aspects (e.g., a common body of knowledge, events, trends) about the construct, which are reflected in words used by them (Abrahamson and Hambrick 1997). In addition, according to semantic contextualism, the meaning of a word is defined by the context surrounding it (e.g., "you shall know a word by the company it keeps"; Firth 1957, p. 11). Hence, two lexically similar words can have different meanings depending on their uses/contexts. Likewise, the role of a word in a sentence can vary (e.g., quality, perceptual entity, or action) depending on the PoS that it is representing (e.g., adjective, noun, or verb) (Cheng et al. 2020). The proposed framework introduces content commonality as the construct's global content reflected on both PoS and the construct dimensions. Table 2 illustrates the data visualization proposal to explore such a commonality considering *d* dimensions.

When  $f_{ijk}$  represents the number of allusive terms to the intersection between the lists of type *k* (PoS) terms for *i* and *j* dimensions (D); if  $j=i$  then  $f_{ijk}$  is the total terms of *i* dimension (diagonal to each matrix in Table 2) by PoS (*k*). From Table 2, it is possible to calculate content commonality (*C*) metrics (by pairs of dimensions:  $C_{ijk}$ , and PoS:  $C_k$ ) based on classic probability (see Eq. 1):

$$C_{ijk} = \frac{2f_{ijk}}{f_{ii} + f_{jj}}, \quad C_k = \frac{fc_k}{\sum_i f_{iik}} \quad (1)$$

$\forall k; i \neq j; i, j \in \{1, \dots, d\}, \quad k \in \{a, n, v\}$

Based on Maier (1993), who studied commonality in the design of systems/structures in operation management contexts (also cited in Stake 2001), a rule of thumb for commonality between components based on proportions is: zero (0) deduces divergent systems (e.g., structures, family products, programs), 0.25 similar designs, and one (1) deduces identical designs. Stegmann (2014) and Sari and Adriani (2019) assume that similarity scores greater than 0.25 are a minimally acceptable representation of similarity between words. Thus, a multidimensional construct from texts should share certain global aspects between its dimensions but at the same time does not cover the entire content of each one of its dimensions. Hence, based on Maier (1993), Stegmann (2014), and Sari and Adriani (2019), the suggested *C* value (and its derived measures, see Eq. 1) should not be near zero but nor should it be high. The study of content commonality based on the proposed procedure can provide insights to improve decision-making processes from a text-driven approach during the development and content validation of psychological/managerial constructs reflected in texts.

**Content polarity**

In sentiment analysis in computational fields, polarity is the positive, negative, or neutral score representing the opinion of a linguistic expression (e.g., word, phrase, sentence, or document) (Korayem et al. 2012; Taboada et al. 2011). Several dictionaries enable this purpose; for example, SentiWordNet (derived from WordNet) is a domain-independent accessible resource for studying polarity (Esuli and Sebastiani 2006). Another valuable resource for the study of polarity is AFINN, which is based on the Affective Norms for English Words (Nielsen 2011) and comprises 2477 words scored between -5 (very negative) to +5 (very positive) (Sharma et al. 2015).

The importance of dictionary polarity detection in providing complementary evidence for content validity is that it is sensitive to the context the construct is used. For example,

**Table 2** Frequency matrix of common terms by PoS in exploring content commonality

|                | Adjectives (a) |                |                | Nouns (n)      |                |                | Verbs (v)      |                |                |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                | D <sub>1</sub> | D <sub>i</sub> | D <sub>d</sub> | D <sub>1</sub> | D <sub>i</sub> | D <sub>d</sub> | D <sub>1</sub> | D <sub>i</sub> | D <sub>d</sub> |
| D <sub>1</sub> | $f_{11a}$      |                |                | $f_{11n}$      |                |                | $f_{11v}$      |                |                |
| D <sub>i</sub> | $f_{i1a}$      | $f_{iia}$      |                | $f_{i1n}$      | $f_{iin}$      |                | $f_{i1v}$      | $f_{iiv}$      |                |
| D <sub>d</sub> | $f_{d1a}$      | $f_{dia}$      | $f_{dda}$      | $f_{d1n}$      | $f_{din}$      | $f_{ddn}$      | $f_{d1v}$      | $f_{div}$      | $f_{ddv}$      |
| Commons        | $fc_a$         |                |                | $fc_n$         |                |                | $fc_v$         |                |                |

*D* (dimension); *f* (number of terms); *fc* (number of common terms)





whether theoretically, a particular construct represents a favorable attitude or behavior in organizations or individuals (e.g., TL; teamwork, labor motivation), such a construct is expected to be reflected in texts through more positive than negative expressions (e.g., “those satisfied with life use more positive than negative words,” Quercia et al. 2012, p. 965). Hence, if a researcher is designing a dictionary on a favorable construct and its global polarity is negative, it signals that something may be wrong. Likewise, whether research addresses a construct, such as *laissez-faire* or burn-out syndrome, the researcher should be warned when global polarity is positive. Thus, polarity detection can be a useful measure to inform researchers of possible content validity problems in construct dictionaries early and automatedly, supporting better decision-making processes during the development and content validation stages.

The proposed framework suggests the appraisal of two types of measures: one numerical (e.g., aggregating  $-1$ ,  $0$ , and  $+1$  scores, or  $-5$  to  $+5$ ) and another categorical (negative, neutral, and positive). Numerical polarity facilitates the exploration of such a property based on conventional statistical summary (e.g., mean, deviation). Categorical polarity enables frequency analysis individually (negative, positive, and neutral) for each dimension (contingency tables). Several formal tests can be used analytically to contrast related hypotheses from this representation. For example, determining whether the global polarity is greater than zero in a construct dictionary using mean or median tests from a numerical perspective, or the chi-squared test from a categorical viewpoint.

### PoS balance

This property is valuable when the dictionary construction is based on the narrow–wide approach (contrary to wide–narrow: an extensive list of terms evaluated through expert judgment to provide a refined list). This approach (narrow–wide) starts with a list of seed terms (e.g., theoretically supported definitions of a construct; narrow list), which is used to identify other semantically similar terms (wide list) to cover more linguistic contexts and to be more generalizable (e.g., “using in domain seed words improves the task performance over generic dictionaries,” Shakurova et al. 2019, p. 13). This narrow–wide approach is based on distributed-dictionary representations using measures, such as cosine similarity to expand the dictionary from a list of seed terms (Garten et al. 2018). In the development/validation of a construct dictionary, the representative list of seed terms should be theoretically supported and consider construct definitions, questionnaire items, and formal descriptions from construct guides (Pandey and Pandey 2017; Ponizovskiy et al. 2020; Short et al. 2010). Subsequently, whether the chosen-seed terms are a reasonable representation of a

construct, it is also expected that its linguistic manifestations in terms of PoS are maintained, at least regarding the essential components: perceptual entities (nouns), actions (verbs), and qualities (adjectives).

Considering the General Theory of Terminology (Wüster 1979) [the importance of nouns for designating a concept] and the Communicative Theory of Terminology (Cabr e 1999, 2002) [adjectives/qualities and verbs/actions can become domain-specific expressions] (Campos and Castells 2010), the proposed framework introduces PoS balance as a valuable property to inform researchers about complementary measures of construct validity. Word tagging from the PoS perspective: (i) describes different components in sentences (e.g., adjectives, nouns, verbs) (Cheng et al. 2020); (ii) provides a reasonable indication of document content without having to examine the entirety of the document considering unique/specific patterns or encoding additional information about the language used in the document (Katsyshev et al. 2020; Whittaker and Tucker 2007); and (iii) allows the analyses of statistical properties of words within each class separately (Dro dz et al. 2009). Thus, when a seed dictionary of a construct presents content validity, PoS should reflect a balance according to the studied scenario: (i) In multidimensional constructs, all dimensions have the same PoS statistic distribution; and (ii) in uni/multidimensional constructs, this distribution is maintained during the dictionary expansion steps.

For example, in the first case, if a construct consists of two dimensions, it is expected that both “A” and “B” dimensions present the same PoS distribution (e.g., percentage of nouns, adjectives, and verbs). Such homogeneity ensures that variations in construct scores are mainly derived from construct content instead of considerable PoS variations (e.g., tone, structure, type of text, writing style). In the second case, a seed dictionary based on corpus is not merely a list of words. It also represents underlying linguistic patterns that the construct reflects in textual data. Hence, when PoS distribution is not reported/controlled during the dictionary expansion steps, the seed patterns could be significantly altered, affecting content validity (e.g., atypical PoS). For example, in a notably emotional construct (e.g., happiness), adjectives are expected to represent a crucial portion of PoS distribution, compared with a more rational construct, such as critical thinking. Figure 2 describes the proposed strategy for exploring the PoS balance of the construct content.

Based on Fig. 2, Pearson’s chi-squared test is useful to explore the PoS balance between dimensions within each moment (initial, final). Two contingency tables should be compared to accomplish this. Thus, since each table (initial, final) corresponds to the crossing of two response variables with the same PoS classification method, several tests can be used, such as the chi-squared test. An adaptation from Grizzle et al. (1969) is proposed in Eq. 2 to calculate the



## PoS distribution analysis

| Moment: Initial           |                 | PoS             |                 |                 |                 |                 |
|---------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Dimensions                | Frequency       |                 |                 | Proportion      |                 |                 |
| D1                        | f <sub>1n</sub> | f <sub>1a</sub> | f <sub>1v</sub> | p <sub>1n</sub> | p <sub>1a</sub> | p <sub>1v</sub> |
| ...                       | ...             | ...             | ...             | ...             | ...             | ...             |
| Dd                        | f <sub>dn</sub> | f <sub>da</sub> | f <sub>dv</sub> | p <sub>dn</sub> | p <sub>da</sub> | p <sub>dv</sub> |
| <b>Chisq; p-value, df</b> |                 |                 |                 |                 |                 |                 |
| Moment: Final             |                 | PoS             |                 |                 |                 |                 |
| Dimensions                | Frequency       |                 |                 | Proportion      |                 |                 |
| D1                        | f <sub>1n</sub> | f <sub>1a</sub> | f <sub>1v</sub> | p <sub>1n</sub> | p <sub>1a</sub> | p <sub>1v</sub> |
| ...                       | ...             | ...             | ...             | ...             | ...             | ...             |
| Dd                        | f <sub>dn</sub> | f <sub>da</sub> | f <sub>dv</sub> | p <sub>dn</sub> | p <sub>da</sub> | p <sub>dv</sub> |
| <b>Chisq; p-value, df</b> |                 |                 |                 |                 |                 |                 |

## PoS balance between dimensions:

$$p_j(l|r) = p_j(i|r) \quad \forall l \neq i$$

## PoS balance between moments:

$$PoS_{final} = PoS_{initial}$$

With:

$$\sum_j f_j(i|r) = k_{i(r)}; \quad \sum_j p_j(i|r) = 1$$

$\forall i \in (1, \dots, d), j \in (a, n, v), r \in (initial, final)$

$d$ : number of dimensions;  $k$ : number of terms

**Fig. 2** Proposed visualization and conditions for PoS balance. *Notes* PoS (parts of speech);  $a$  (adjectives);  $n$  (nouns);  $v$  (verbs)

expected values ( $e_{ijr}$ ) under the null hypothesis (homogeneity of the distributions of the two tables; that is, PoS balance between moments):

$$e_{ijr} = f_{+jr} \left( \frac{f_{ij1} + f_{ij2}}{f_{+j1} + f_{+j2}} \right)$$

$$\chi^2_{(i) \times (j-1)} = \sum_r \sum_i \sum_j \frac{(f_{ijk} - e_{ijr})^2}{e_{ijr}} \quad (2)$$

Here,  $f_{ijr}$  and  $e_{ijr}$  are observed and expected frequencies, respectively, allusive to dimension (row  $i$ ), PoS (column  $j$ ), and moment ( $r$ ; two tables:  $r=1$ . *Initial* and  $r=2$ . *Final*), and  $f_{+jr}$  represents the sum of all observations of column  $j$  within moment (table)  $r$ ;  $e_{ijr}$  is distributed according to Eq. 2: chi-squared ( $\chi^2$ ) with  $df:(i) \times (j-1)$  degrees of freedom. The procedure to contrast the hypothesis of homogeneous tables (moments of dictionary development) was automated in Python and R. The R code will be shared on reasonable request to the corresponding author.

### Embedded-content convergence/differentiation

In structured questionnaire data, a construct measure is attributed with convergence traits when it correlates with other measures of the same construct or with reference constructs with which such an association is expected (Martínez et al. 2006). Likewise, there are discriminant traits when the construct presents low association with constructs assumed to be less correlated or noncorrelated. In interpreting this procedure in the TM field, we consider a construct dictionary and three reference variables theoretically/logically associated with the construct under study in three categorical magnitudes: higher, medium, and low associations. Furthermore, we consider four

embedded vectors, one representing the construct dictionary (Vc) and the others the reference variables (Vh, Vm, Vl). Then, if the construct presents convergence/differentiation capabilities, several conditions are linguistically reflected, some of which are represented in the following equation:

$$\begin{aligned} S(Vc_{ik}, Vh_{ik}) &> S(Vc_{ik}, Vm_{ik}) > S(Vc_{ik}, Vl_{ik}), \\ S(Vc_{ik}, Vh_{ik}) &> 0.25 \quad \forall i \in \{1, \dots, d\}, \quad k \in \{a, n, v\} \end{aligned} \quad (3)$$

where  $S$  is cosine similarity between two embedded vectors,  $i$  is dimension, and  $k$  is PoS (nouns, adjectives, verbs). Equation 3 represents the theoretically or logically stabilized relationships in the nomological network of the construct under study. Regarding the linguistic context to explore such relationships, a pretrained word embedding model can be used, such as the standard Google pretrained Word2Vec model (<https://code.google.com/archive/p/word2vec/>), which covers a vocabulary of 3 million terms (words and phrases) trained using roughly 100 billion expressions from a Google News data set (Mikolov et al. 2013). For example, whether a researcher is studying critical thinking, a word embedding vector of the construct can be extracted from the pretrained model using the sum or average of each word embedding present in the construct dictionary (Garten et al. 2018). Likewise, three potential reference word embeddings can be extracted for representing the following expressions: “creativity” (high expected association, Vh), “extraversion” (expected medium association, Vm), and “sport” (expected low association, Vl). Consequently, the researcher can calculate cosine similarity between the embedded vectors and verify Eq. 3. The use of cosine similarity is supported by its success and popularity in TM and related fields to summarize and describe association patterns between term vectors, generally accepting values greater than 0.25 (Sari and Adriani 2019; Stegmann 2014).



### Content coherence

Liman et al. (2020) and Sekaran and Bougie (2016) both highlight that expert opinions about the constructs' items, wordings, and phrases are a reasonable representation of content validity. Patrick et al. (2011) assert that content validity aims to demonstrate that a construct measure is appropriate and comprehensive. On the other hand, based on a text-driven approach, coherence means that humans understand and interpret the patterns discovered from TM (Chang et al. 2009). Thus, the meaning/interpretation of terms included in a construct dictionary is a reasonable (albeit not sufficient) proxy of construct content validity reflected in texts.

A resource increasingly used in big data analytics (intensive in text-drive approaches) is a word cloud; this allows researchers to summarize and describe language graphically and "is more significant if it conveys more information by itself with less information shared by other word clouds" (Cui et al. 2010, p. 123). The proposed framework uses word clouds to summarize and describe the content of the construct under study, combined with surveys of  $n$  participants to independently evaluate two word clouds in the case of a unidimensional construct (one for the studied construct and another as a distractor); or  $d$  word clouds (one per dimension) in a multidimensional construct ( $d$  dimensions;  $d > 2$ ).

In any case, consider the following instructions: *Please (i) analyze each word cloud. Then, you will be presented with statements about the construct (or its dimensions); (ii) Read them in detail and afterward; (iii) link each statement to the word cloud to which it has the greatest affinity.* Here, a statement is a theoretically supported formal definition of the investigated construct (or dimension). For example, for "idealized influence" (II; a dimension of TL), a possible statement could be: a leader who is admired, respected, and conceived as a role model, thanks to their high standards of performance, ethics, health, and outstanding behavior (Avolio and Bass 2004).

To conclude whether empirical evidence in favor of the coherence of a construct is found, the proposed framework uses the contingency table exposed in Fig. 3, illustrating  $d$  dimensions.

Note in Fig. 3 that this procedure summarizes, in tabular format, categorical responses for the affinity between  $d$  word clouds (e.g., construct dimensions) and  $d$  statements (formal definitions). Based on Howe (1985), the normalized diagonal represents the proportion of coherence/agreement between the two components (word cloud and statement) from the respondent's perspective. Thus, to conclude in favor of content coherence, three simultaneous criteria are proposed: (i) each element on the diagonal (see Fig. 3) is greater than any element outside the diagonal; (ii) the sum of elements within the diagonal is greater than half of the

| Coherence analysis |                          | Statements  |     |               | Criteria:   |
|--------------------|--------------------------|-------------|-----|---------------|---|
|                    |                          | Statement 1 | ... | Statement $d$ |   |
| Word clouds        | Construct word cloud 1   | $f_{11}$    | ... | $f_{1d}$      | (i) $f_{ii} > f_{ij}$ for $i \neq j$<br>(ii) $\text{Sum}(f_{ij}) > n/2$<br>(iii) Significant Chi-Squared test |
|                    | ...                      |             |     |               |   |
|                    | Construct word cloud $d$ | $f_{d1}$    | ... | $f_{dd}$      |   |

Fig. 3 Contingence  $d \times d$  table to summarize results about content coherence. Notes  $f$  (frequency);  $n$  (total observations);  $d$  (number of dimensions)

sample consulted. These results deduce that the human interpretation of the content summarized from construct word clouds tends to converge toward correct definitions and discriminate from other possible content. Besides, when (i) and (ii) criteria are satisfied, this paper suggests employing (iii) a Pearson's  $\chi^2$  test (degrees of freedom: product between  $d - 1$  and  $d - 1$ ) to support analytical content coherence, contrasting the possible association between rows and columns. That is, levels of word clouds (rows) are associated with (or are not independent of) statement categories (columns).

### Embedded voting

In this subsection, we propose an original method entitled "Embedded voting." In "Results of the empirical application of PMTM in the study of TL", we describe the practical use of this method considering TL.

A frequent practice in the dictionary development/validation of psychological/managerial constructs from texts is to use expert evaluations to decide upon controversies in the dictionary content. For example, the framework proposed by Short et al. (2010) and also employed by Pandey and Pandey (2017) includes the notion that researchers should "validate word lists using content experts." Likewise, when there is no agreement between experts (e.g., to include a specific word in the dictionary or deciding in which dimension to include the word), a discussion process is held until a consensus is reached (Pandey and Pandey 2017; Ponizovskiy et al. 2020). We intend to emulate and refine this procedure by developing an automated-computational perspective assisted by machine learning using word embeddings. Details of the procedure of embedded voting are shown in Fig. 4.

From Fig. 4, when there is controversy concerning whether a specific word is more pertinent to one dimension ("A") or another ("B"), embedded voting receives the list of words of each dimension, among other inputs exposed in Fig. 4 (e.g., number of experts to emulate:  $n.experts$ ). Then, preprocessing tasks are performed to obtain (i) a vector of controversial words between the dimensions (*Controv*), and (ii) two vectors of unique words ( $A_{un}$ ,  $B_{un}$ ). Likewise, two additional vectors (*Vote.A*, *Vote.B*) and one output matrix

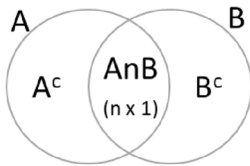


### Illustrating Embedded voting

Expressions:

- *training\_model*: using word2vec under any language.
- *param*: additional parameters for training the model.
- *mod.vocab*: obtaining vocabulary.
- *Vocab\_Vectors*: embedding vectors.
- *Cosimilar*: cosine similarity.
- *CorpAuB*: combining pre-processed corpora used to study A and B constructs.
- *kpart*: integer representing a partition for the set size allusive to common terms
- *Controv*: controversial terms; a subset (size based on *kpart*) of common terms in both A and B.
- *seed*: initial seed for reproduce results.
- *a*: increase for seed.

# A and B dictionaries



# Inputs:

- A; B; a
- kpart
- n.experts
- CorpAuB
- seed
- param

Load inputs

# Pre-processing:

```
r = round(n/kpart,0)
Controv = AnB[1:r]
A_un = A^c, B_un = B^c,
Vote.A = [rep(0,r)] (r x 1)
Vote.B = [rep(0,r)] (r x 1)
Res = [Controv, rep("A", r)] (r x 2);
```

1

For *j* from 1 to *n.experts*:

seed = seed + a

```
mod = training_model(CorpAuB, seed, param) # Training model
Vocab_Vectors = mod.vocab
```

k = 0

For *term* in *Controv*:

k = k+1

```
Vterm = Vocab_Vectors[term] # Embedding vectors
vA = Vocab_Vectors(sum(A_un))
vB = Vocab_Vectors(sum(B_un))
```

```
pA = Cosimilar(Vterm, vA) # Cosine similarity
pB = Cosimilar(Vterm, vB)
```

If (pA > pB):

```
Vote.A[k] = Vote.A[k] + 1 # Vote of current expert on term k
```

Else:

```
Vote.B[k] = Vote.B[k] + 1
```

Enf if:

Next *term*

Next *j*

# Next expert

```
Res[Vote.A < Vote.B, 2] = "B"
```

# Update total voting

Output: Res

```
# Initial
> Res <- data.frame(Controv,Assign=rep("A",r))
> Res
  Controv Assign
1 idealized  A
2 follower  A
3 others    A
4 employee  A
5 leader    A
6 stimulation A
```

1

# R outputs

```
# Final
> Res[Vote.A < Vote.B, 2] <- "B"
> Res
  Controv Assign
1 idealized  B
2 follower  B
3 others    A
4 employee  A
5 leader    B
6 stimulation B
```

Fig. 4 Developed method for emulating expert raters: General illustration of embedded voting of *n.experts* considering “A” and “B” construct dictionary options

(Res) are initialized. After that, embedded voting emulates the decision-making process of each expert in the panel (from *j* in 1 to *n.experts*). Thus, each “expert” evaluates the

association between an embedded vector of each controversial word (from *term* in *Controv*) and two aggregated-embedded vectors of unique words of both “A<sub>un</sub>” and “B<sub>un</sub>”



dimensions. That is, the expert should decide whether each controversial word (vector) is more associated with dimensions “A” or “B,” which internally is performed from a word embedding approach. Subsequently, if the “expert” vote is to include the controversial word in dimension “A,” this receives one point ( $\text{Vote.A} = \text{Vote.A}[k] + 1$ ); similarly, when dimension “B” is more related to the controversial word, it receives one point ( $\text{Vote.B} = \text{Vote.B}[k] + 1$ ). Once all “experts” have voted, embedded voting decides in which dimension each controversial word should be assigned, based on the most votes (“Res,” Fig. 4). To generate the vectors, embedded voting considers relevant content representing both “A” and “B” dimensions, using a combination of the corpora as its input (CorpAuB) from which “A” and “B” dictionaries were formed. That is, both “A” and “B” constructs provide linguistic contexts to be considered in the decision-making process. Thus, a word embedding model is trained, then the required embedded vectors ( $vA, vB$ ) are generated.

Embedded voting is one automated resource to help overcome eventual dictionary problems regarding four properties: commonality, coherence, PoS balance, and convergence/differentiation; hence, it is helpful in steps 4, 7, and 10 of the PMTM framework (see Fig. 1). The R code will be shared on reasonable request to the corresponding author.

### Representation of the measurement model based on PoS

We operationalize a new representation for psychological/managerial constructs by considering PoS. In "Results of the empirical application of PMTM in the study of TL", we empirically examine this representation in the case of TL.

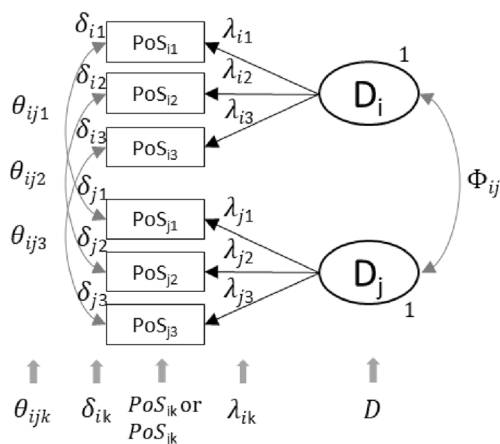
We intend to extract lexical proxies of qualities (adjectives *-a* or adjective phrases *-ap*), perceptual entities (noun *-n* or noun phrases *-np*), and actions (verbs *-v* or verb phrases

*-vp*) commonly used in texts about the psychological/managerial construct under study. This approach is consistent with the importance of nouns as the essential representation of concepts (Wüster 1979) and adjectives/verbs as lexical units with the potential for capturing qualities/actions from specific domains (Cabr e 1999, 2002). For example, according to Campos and Castells (2010), domain-specific adjectives should be treated independently during the development of dictionaries. Moreover, nouns, adjectives, and verbs involve different brain mechanisms during text production (Fyshe et al. 2019; Haan et al. 2000; Martin et al. 1995). Thus, the framework develops two formats of observable indicators based on correlated PoS:

- Basic:  $n, a,$  and  $v.$
- Composed:  $np$  (using the  $n$  scores),  $ap$  (averaging scores of  $a$  and  $n$ ), and  $vp$  (averaging scores of  $v$  and  $n$ ). This format supports phrase formation based on context-free grammar (Chomsky 1955) and the aggregated vectors of words (e.g., Garten et al. 2018).

Figure 5 illustrates the two formats considering two construct dimensions ( $i, j$ ) and a first-order model to ease understanding. However, these can be extended to more dimensions or complex structures (e.g., second-order model) or parsimonious models based on PoS parcels (e.g., averaging  $a, n,$  and  $v$  vector scores) that summarize higher order constructs.

Note in the proposed representation (see Fig. 5) that each error variance is a combination of natural errors and systematic components of each type of PoS. Analogously to questionnaire studies, one possible systematic component could be the method employed (e.g., personality measured based on a checklist of adjectives; Gough 1979). Another systematic component represented in correlated PoS can be linguistics considerations, such as the



$i, j$ : Dimension ( $D$ ); with  $i \neq j$ .  
 $k$ : PoS type: 1 ( $a$  or  $ap$ ), 2 ( $n$  or  $np$ ), 3 ( $v$  or  $vp$ ) in basic or composed format, respectively.  
 $PoS_{ik}$ : Observable indicator type  $k$  for  $D_i$ .  
 $\Phi$ : Standardized covariance (correlation between dimensions)  
 $\lambda_{ik}$  (or  $\lambda_{jk}$ ): Standardized load under  $ik$  (or  $jk$ ).  
 $\delta_{ik}$  (or  $\delta_{jk}$ ) Error variances under  $ik$  (or  $jk$ ).  
 $\theta_{ijk}$ : Error covariances.  
 Variances of  $D$  fixed to 1 for identifiability.

Fig. 5 Measurement model representation based on correlated PoS



synonymy or polysemy of words. For example, multiple adjectives (or nouns or verbs) can present the same meaning, and various meanings can be attributed to a single adjective (or noun or verb) (Lochter et al. 2016). Thus, the exposed representation can be understood as an adaptation of the correlated-uniqueness CFA (Batista-Foguet and Coenders-Gallart 2000; Marsh et al. 1992) by considering PoS. The statistical expression of an observable indicator based on PoS type  $k$  for a particular dimension ( $D$ ) is presented in the following equation:

$$\text{PoS}_k = \lambda_k D + \delta_k; \quad \delta_k = e + \gamma_k \quad (4)$$

where  $\text{PoS}_k$  is scores (e.g., sum, average, or weighted average) obtained from a merged list of type  $k$  (PoS:  $a$  or  $ap$ —representing qualities,  $n$  or  $np$ —perceptual entities, and  $v$  or  $vp$ —actions) manifestations derived from a particular dimension ( $D$ ).  $\delta_k$  is the indicator error variance comprising both random error ( $e$ ) and systematic components of the PoS effect ( $\gamma_k$ ).

To reach a conclusion concerning the plausibility of this representation (Fig. 5), the suggested metric for CFA, according to Credé and Harms (2015), is the chi-squared ( $\chi^2$ ) metric, which contrasts a hypothesized model vs. a null model (latent variables are not required to reproduce the evidence: covariance). However, due mainly to the sensitivity of  $\chi^2$  to the sample size, other complementary fit indexes (Credé and Harms 2015; Lévy and Varela 2006) are: RMSEA (root mean square error approximation), SRMR (standardized root mean square residual), CFI (comparative fit index), and TLI (Tucker–Lewis index). To interpret these indexes, it is typical to use cutoff values (e.g., proposed by Hu and Bentler, 1999; RMSEA < 0.06; SRMR < 0.08; CFI and TLI > 0.95) in constructs based on structured questionnaires (Lévy and Varela, 2006; Lorenz et al. 2021). However, there is controversy about the usefulness of these cutoff values, especially in higher order factors, which is argued by Credé and Harms (2015), who suggest that the better models are those that present a nonsignificant  $\chi^2$ , and at the same time low values for RMSEA and SRMR, as well as high values in the others (CFI, TLI).

## Results of the empirical application of PMTM in the study of TL

This section is structured in the stages of the proposed framework (see Fig. 1) to develop and validate a new scale of TL from a theory/text-driven approach using data from organizational documents, interviews, essays, blog posts, surveys, and speeches by former US presidents.

## Dictionary development and content validation (Outputs of steps 1–10 of PMTM)

Regarding “Step 1: establishing construct definition and dimensionality from a theory-driven approach” (see Fig. 1), the construct under study is TL. Considering Avolio and Bass (1991, 2004), this construct represents the capacity to inspire and guide followers towards collective goals and, simultaneously, modify their motivational basis, following their desire for achievement, self-improvement, and self-realization. Moreover, at present, their questionnaire (the MLQ™, Avolio and Bass 1991, 2004) is the most commonly used instrument to measure TL (Brown and Keeping 2005; DeDeyn 2021), which operationalizes such a construct in four dimensions: (1) Individualized consideration (IC): treating each collaborator in a differentiated way, considering their needs, capacities, goals, and expectations; (2) Intellectual stimulus (IS): stimulating followers to assume creative thoughts and behaviors and to search for innovative ideas and solutions through the promotion of reasonable doubt, the filtering of information, argumentation, and the questioning of supposed values and beliefs; (3) Idealized influence (II, two types: attributed and behavior): the leader is admired, respected, and conceived as a model to follow; followers identify with the leader and aim to emulate them. (4) Inspirational motivation (IM): fostering enthusiasm among employees, using teamwork to channel resources and capabilities to overcome the personal and organizational status quo and achieve better performance.

Concerning “Step 2: conforming three corpora types: seed, scientific, and pragmatic” (see Fig. 1); Table 3 describes the empirical data conformed to develop the dictionary.

After applying steps 3–10 of PMTM (see reproducibility in the supplementary material: “Supp1”), we present the results for the proposed properties (see “Proposed properties for exploring content validity”), considering the final TL dictionary (1073 words; nouns: 424; adjectives: 209; verbs: 440):

### PoS balance

Table 4 provides evidence to examine whether there is a PoS balance between the structure of TL dimensions.

Note, in Table 4, that the linguistic structure of the dictionary satisfies the PoS balance: chi-squared (6 degrees of freedom in each table) for the initial (based on seed corpus) and final (all corpora: seed, scientific, and pragmatic) moments are 0.919 and 0.6603, respectively. Moreover, the contrast between moments using Eq. 1 suggests that the initial dictionary expansion did not destroy the PoS structure underlying the seed corpus (formal definitions/descriptions of each TL dimension), with a chi-squared of 0.9061 and



**Table 3** Data used for steps 2–10 of PMTM

| Corpora type | Number of documents   | Document sources   |
|--------------|---|--|
| Seed         | 116 in total. These comprise 29 precise definitions or descriptions for each TL dimension | Scientific papers (e.g., from Scopus), professional blogs, Wikipedia, and the TL manual (MQL5x) defining or describing a dimension of TL   |
| Scientific   | 144 documents based on titles, abstracts, and keywords of scientific papers               | Scientific papers from peer-review journals including expressions of TL or its dimensions in title, abstract, or keywords<br>Examples of search expressions on Scopus:<br><i>TITLE (“Idealized influence” OR charism*) AND TITLE-ABS-KEY (“transformational leadership”)</i><br><i>TITLE (intellectual PRE/O (stimulus OR ability OR capability)) AND TITLE-ABS-KEY (leader)</i> |
| Pragmatic    | 25 formal guides (long documents) for leadership (general)                                | From multiple domains, such as education, army, healthcare, and business   |

Note: TL (transformational leadership)

**Table 4** PoS balance analysis considering initial (seed) and final dictionaries

| Initial dictionary (seed corpora)                    |           |    |     |     |            |       |       |       |  |
|--|-----------|----|-----|-----|------------|-------|-------|-------|--|
| PoS  | Frequency |    |     |     | Proportion |       |       |       |  |
|  | II        | IM | IS  | IC  | II         | IM    | IS    | IC    |  |
| Nouns  | 89        | 89 | 134 | 104 | 0.397      | 0.410 | 0.442 | 0.405 |  |
| Adjectives   | 38        | 41 | 52  | 49  | 0.17       | 0.189 | 0.172 | 0.191 |  |
| Verbs  | 97        | 87 | 117 | 104 | 0.433      | 0.401 | 0.386 | 0.405 |  |
| Final dictionary (seed—scientific—pragmatic corpora) |           |    |     |     |            |       |       |       |  |
| PoS  | Frequency |    |     |     | Proportion |       |       |       |  |
|  | II        | IM | IS  | IC  | II         | IM    | IS    | IC    |  |
| Nouns  | 110       | 95 | 119 | 100 | 0.425      | 0.401 | 0.398 | 0.360 |  |
| Adjectives   | 42        | 45 | 62  | 60  | 0.162      | 0.190 | 0.207 | 0.216 |  |
| Verbs  | 107       | 97 | 118 | 118 | 0.413      | 0.409 | 0.395 | 0.424 |  |

Initial:  $\chi^2$ : 2.008; *p* value: **0.919**; *df*: 6

Final:  $\chi^2$ : 4.121; *p* value: **0.6603**; *df*: 6

Comparing tables (initial and final); Eqs. 1–2:  $\chi^2$ : 3.4095; *df*: 8; *p* value: **0.9061**

II idealized influence; IM inspirational motivation; IS intellectual stimulus; IC individual consideration

8 degrees of freedom. Thus, the latent context (e.g., use, interpretation, and brain processing of words) underlying the initial content (formal definitions/descriptions) of the TL dimensions continues to keep its PoS balance, even though such content was extended based on additional contexts/corpora.

**Content commonality**

Table 5 presents evidence about content commonality in the final TL dictionary, specifying the frequency and proportion of shared terms by dimension ( $C_{ij}$ ) and PoS ( $C_k$ ).

Note, in Table 5 (considering the final TL dictionary), that in the case of nouns, 7 of the 424 terms are shared by all TL dimensions, representing 1.7% of content commonality

( $C_k$ ) at the noun level and around 6.7% between pairs of dimensions ( $C_{ij}$ ). In adjectives and verbs, these  $C_k$  are 1.9% and 1.6%, respectively, and  $C_{ij}$  range between 5.9 and 11.5%.

The evidence about content commonality (Table 5) prevents future problems of high lexical correlations that falsely show a convergence between TL dimensions. Likewise, the global terms comprising the final dictionary can be assumed as general-domain contexts, which also apply to theoretical and practical TL contexts. It is worth noting that global content commonality ( $C_k$ ) before the use of embedded voting in both the initial dictionary (from the seed corpus) and its expansions (scientific and pragmatic) presented high values. For example, considering the seed corpus, commonality for pairs of dimensions ( $C_{ij}$ ) by PoS ranging from 27.9 to 39% in nouns; 18.6–32.7% in adjectives; and



**Table 5** Content commonality analysis for the final dictionary

| $C_{ij}$ for nouns   | Frequency |    |     |     | Proportion |       |       |       |
|--|-----------|----|-----|-----|------------|-------|-------|-------|
|  | II        | IM | IS  | IC  | II         | IM    | IS    | IC    |
| II   | 110       | 7  | 7   | 7   | 1          | 0.068 | 0.061 | 0.067 |
| IM   | 7         | 95 | 7   | 7   | 0.068      | 1     | 0.065 | 0.072 |
| IS   | 7         | 7  | 119 | 7   | 0.061      | 0.065 | 1     | 0.064 |
| IC   | 7         | 7  | 7   | 100 | 0.067      | 0.072 | 0.064 | 1     |
| $C_k$ : 0.017. Global.terms: 7. {"level," "nature," "associate," "knowledge," "outcome," "accountability," "relationship"} |           |    |     |     |            |       |       |       |
| $C_{ij}$ for adjectives  | Frequency |    |     |     | Proportion |       |       |       |
|  | II        | IM | IS  | IC  | II         | IM    | IS    | IC    |
| II   | 42        | 5  | 4   | 4   | 1          | 0.115 | 0.077 | 0.078 |
| IM   | 5         | 45 | 4   | 4   | 0.115      | 1     | 0.075 | 0.076 |
| IS   | 4         | 4  | 62  | 4   | 0.077      | 0.075 | 1     | 0.066 |
| IC   | 4         | 4  | 4   | 60  | 0.078      | 0.076 | 0.066 | 1     |
| $C_k$ : 0.019. Global.terms: 4. {"best," "transformational," "next," "organizational"}                                     |           |    |     |     |            |       |       |       |
| $C_{ij}$ for verbs   | Frequency |    |     |     | Proportion |       |       |       |
|  | II        | IM | IS  | IC  | II         | IM    | IS    | IC    |
| II   | 107       | 7  | 7   | 8   | 1          | 0.069 | 0.062 | 0.071 |
| IM   | 7         | 97 | 7   | 7   | 0.069      | 1     | 0.065 | 0.065 |
| IS   | 7         | 7  | 118 | 7   | 0.062      | 0.065 | 1     | 0.059 |
| IC   | 8         | 7  | 7   | 118 | 0.071      | 0.065 | 0.059 | 1     |
| $C_k$ : 0.016. Global terms: 7. {"develop," "involve," "identify," "using," "developing," "might," "know"}                 |           |    |     |     |            |       |       |       |

II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration);  $C$  (content commonality)

18.6–30.3% in verbs. In the pragmatic corpus, these commonalities, before embedded voting, ranged between 32 and 43.6% in nouns; 25.4% and 52.2% in adjectives; and 30.3% and 38.4% in verbs (detailed outputs in Python are in “Supp2”, supplementary material). These values justify the traditional practice based on expert discussions to generate a consensus when construct dictionaries are developed based on human tasks. However, in this paper, we proposed and applied embedded voting, a method to emulate this agreement process using embedded vectors computationally (e.g., word2vec) from a machine-learning perspective.

### Embedded-content convergence/differentiation

Table 6 describes the associations (cosine similarity) between the vectors of TL dimensions (merging nouns, verbs, and adjectives) and the reference vectors with which it is expected that the dimensions have high (Vh), medium (Vm), and low (Vl) associations. Based on the conditions stated in Eq. 3, the evidence provided (Table 6) supports the content convergence/differentiation of the developed dictionary from a linguistic-computational perspective. For example, the cosine similarities between TL

dimensions ranged from 0.924 to 0.944; these are greater than the similarities between TL dimensions and reference vectors (0.835–0.352).

Table 6 shows that TL dimensions present more similarity with Vh (0.751–0.835; words used frequently by leaders rated high on TL, e.g., “team,” “members,” “provide,” “purpose,” “teach,” “information,” “concerns,” “solutions,” “creating2”; Salter et al. 2013, p. 65) than Vm (0.575–0.6; words frequently used by leaders rated low on TL, e.g., “results,” “schedule,” “tasks,” “colleagues,” “measured,” “budget;” Salter et al. 2013, p. 65) and Vl (0.352–0.461; negativity words). Associations between TL dimensions and Vl should be lower and not necessarily negative, because the TL vectors are not exclusively based on adjectives but also on nouns and verbs.

In all cases, the similarities calculated from embedded vectors (TL dimensions and reference vectors) satisfy the conditions stated in Eq. 3 in favor of content convergence/differentiation. In other words: (i) the TL dimensions share more information with them than with other measures; (ii) the TL dimensions are sensitive to changes in the measures with which they are associated; (iii) TL dimensions reflect expected association patterns with other measures (high, medium, low associations).





**Table 6** Content convergence/differentiation analysis for the final dictionary

| Similarity | II    | IM    | IS    | IC    | Vh    | Vm    | VI    |
|------------|-------|-------|-------|-------|-------|-------|-------|
| II         | 1     | 0.944 | 0.937 | 0.925 | 0.765 | 0.581 | 0.461 |
| IM         | 0.944 | 1     | 0.924 | 0.940 | 0.835 | 0.575 | 0.431 |
| IS         | 0.937 | 0.924 | 1     | 0.926 | 0.751 | 0.600 | 0.461 |
| IC         | 0.925 | 0.94  | 0.926 | 1     | 0.807 | 0.578 | 0.352 |
| Vh         | 0.765 | 0.835 | 0.751 | 0.807 | 1     | 0.486 | 0.385 |
| Vm         | 0.581 | 0.575 | 0.600 | 0.578 | 0.486 | 1     | 0.327 |
| VI         | 0.461 | 0.431 | 0.461 | 0.352 | 0.385 | 0.327 | 1     |

II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)

Vh: Unigrams used frequently by leaders rated high on TL (Salter et al. 2013, p. 65). [“team,” “members,” “provide,” “purpose,” “teach,” “information,” “concerns,” “solutions,” “creating,” “growth,” “inspire,” “quality,” “integrity,” “encourage,” “coaching,” “participation,” “workshops,” “encouraging,” “teammates,” “motivate,” “autonomy,” “relationships,” “personal,” “encouragement,” “inspiring,” “modeled,” “motivating,” “rapport,” “teamwork”]

Vm: Unigrams used frequently by leaders rated low on TL (Salter et al. 2013, p. 65). [“results,” “schedule,” “tasks,” “colleagues,” “measured,” “assessments,” “report,” “status,” “project,” “organized,” “budget,” “solution,” “success,” “boos,” “production,” “deadline,” “immediately”]

VI: Negativity unigrams from <https://thesaurus.plus/related/negativity/negativeness>. [“antagonism,” “negativeness,” “negativism,” “adversity,” “deprecation,” “disapproval,” “dubiousness,” “hostility,” “distrust,” “hesitation,” “leeriness,” “reluctance,” “suspicion,” “uncertainty,” “negation,” “pessimism”]

**Content coherence**

Figure 6 provides the word clouds used to explore the content coherence of the construct dictionary. “A,” “B,” “C,” and “D” labels are II, IM, IS, and IC, respectively.

The labels were hidden from the ten evaluators (one Ph.D. professor, six MSc. professors, and three practitioners, all in management areas). The statements (formal definitions for

TL dimensions) were presented to the evaluators in separated sections and in the following order: IM, II, IC, and IS statements (e.g., inspirational motivator: fosters enthusiasm among employees and inspires them to overcome personal/organizational status quo and achieve higher levels of performance).

Table 7 presents the contingency table that describes the affinity perceived by the evaluators, considering two



**Fig. 6** Word clouds (a random sample of words from the dictionary vocabulary) used to explore human interpretation of the TL dictionary. “A,” “B,” “C,” and “D” labels are II, IM, IS, and IC, respectively



**Table 7** Content coherence analysis considering word clouds from the final dictionary

|             | Original sample    |   |   |   |   |             | Resampled sample ( <i>n</i> : 5000) |      |      |      |      |
|-------------|--------------------|---|---|---|---|-------------|-------------------------------------|------|------|------|------|
|             | Correct statements |   |   |   |   |             | Correct statements                  |      |      |      |      |
|             | A                  | B | C | D |   |             | A                                   | B    | C    | D    |      |
| Word clouds | A                  | 9 | 1 | 0 | 0 | Word clouds | A                                   | 4515 | 485  | 0    | 0    |
|             | B                  | 1 | 6 | 2 | 1 |             | B                                   | 515  | 2967 | 1014 | 504  |
|             | C                  | 0 | 2 | 5 | 3 |             | C                                   | 0    | 995  | 2499 | 1506 |
|             | D                  | 0 | 1 | 3 | 6 |             | D                                   | 0    | 485  | 1518 | 2997 |

$\chi^2$ : 43.2; df: 9; *p* value: 1.9808e-06

$\chi^2$ : 21824.4; df: 9; *p* value: 0

**Table 8** Content polarity analysis for the final dictionary

| Polarity        | II   | IM   | IS   | IC   |
|-----------------|------|------|------|------|
| neutral         | 198  | 172  | 233  | 219  |
| positive (pos)  | 37   | 47   | 36   | 38   |
| negative (neg)  | 4    | 2    | 16   | 2    |
| p_pos           | 0.15 | 0.21 | 0.13 | 0.15 |
| p_neg           | 0.02 | 0.01 | 0.06 | 0.01 |
| ratio (pos/neg) | 9.25 | 23.5 | 2.25 | 19   |

II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration).

scenarios: original (*n*: 10 respondents) and resampled (5000 replicas).

Note that in all cases (Table 7), the correct options were chosen by the respondents in at least 50% (A–A: 9/10 respondents; C–C: 5/10) of cases; likewise, the independence hypothesis between word clouds and correct statements was rejected with a chi-squared value of 43.2 (9 degrees of freedom) and *p* value near to zero. In other words, the evidence supports that humans adequately interpret the content of the dictionary under study.

### Content polarity

It is expected that a transformational leader has/employs more positive than negative emotions/words (Diebig et al. 2017). Table 8 categorically and numerically summarizes the results.

Table 8 illustrates that the number of positive words is greater than negative ones in all cases, with ratios (pos/neg) ranging from 2.25 (IS) to 23.5 (IM). In addition, as expected, IM presented the highest positivity (21% of its terms) and IS the lowest (13% of its terms), which is consistent with the notion that (i) motivational thinking/behavior is characterized by positive emotional language; and (ii) intellectual thinking/behavior is characterized by reasonable doubt,

questions, consideration of risks and opportunities, rational persuasion, impartiality, and fewer emotional decisions (e.g., “a sense that the structures and processes are orderly and rational,” Poghosyan and Bernhardt, 2018, p. 3).

In addition, PMTM automatically controls the polarity of adjectives of all TL dimensions during the dictionary construction steps. For example, in the case of IS, adjectives such as “stupid,” “obsolete,” and “fear” [a non-adjective expression] were automatically detected and removed during the seed corpus processing; and “negative,” “poor,” and “wrong” during the final stage (pragmatic corpus). This content polarity was controlled without human intervention, and the evidence provided in Table 8 satisfies the expected results. For details of the history of these results across all steps of the dictionary construction, see the supplementary material (“Supp1”).

### Validity, reliability, and equity (outputs of steps 11–15 of PMTM)

#### Confirming the plausibility of the internal structure and examining reliability

We employ three formal corpora (text data), one from organizations (letters to shareholders—2018 Fortune companies; e.g., this data type was also used in Pandey and Pandey 2017 for studying organizational culture, and Josef and Helena, 2019—effective leadership using LIWC2015, Pennebaker et al. 2015) and other texts from individual contexts (online emulated job interviews—formal, and blog posts—casual; these blog posts were also used in Ponizovskiy et al. 2020 for studying personal values). Subsequently, we calculate scores for each TL dimension using the expression (in log scale) proposed by Pérez-Rave et al. (2020). Table 9 describes the corpora processed to examine the internal validity of the dictionary.

For each format type of PoS (basic: *a*, *n*, *v*; and composed: *ap*, *np*, *vp*; see Fig. 5), we contrast three models with the four theoretical TL dimensions (II, IM, IS, and IC; Avolio and Bass 1991, 2004): single, first-order, and



**Table 9** Data used for step 11 of PMTM “Confirming the plausibility of the internal structure of the scale, and reliability”

| Data   | Sample size | Description   |
|--|-------------|---|
| 1. 2018 letters to shareholders from 1000 Fortune list | 186 obs     | Initially, 200 companies were randomly chosen from the 1000 Fortune list (100 ranked in the first 200 companies; 100 in the last 200 companies); 186 observations were available online and did not present errors when converting them from “pdf” to “txt”: 105 companies ranked in the first 250 companies and 81 ranked in the last 150 companies  |
| 2. Online job interview (emulation)                    | 128 obs     | We contract answers from a recognized private online platform for surveys. We employ traditional questions (seven in total) in job interviews, such as: “Tell me about yourself?” “Tell me about your strengths?” and “Where do you see yourself in five years?” We obtained 152 responses; 84.2% were complete and valid responses (an example of inadequate responses: “Hun b number ntmu vicci chick bob B;” “No No No.” |
| 3. Blog authorships corpus                             | 500 obs     | Randomly chosen from the training sample of blog posts used by Ponizovskiy et al (2020); <a href="https://osf.io/t9k56">https://osf.io/t9k56</a> , comprising 9960 observations   |

second-order structures (each including the described correlated PoS). The analyses were performed in R (R Core Team 2021) using lavaan (Rosseel 2012) under both maximum likelihood (ML) estimation and MLM (maximum likelihood with robust standard errors and a Satorra–Bentler scaled test statistic), considering the corpora described in Table 9. The results are provided in Table 10. Additional detailed outputs are presented in the supplementary material (“Supp2...”).

Table 10 reveals that the two formats of the observable PoS (basic: *a, n, v*; composed: *ap, np, and vp*) support the plausibility of the construct (TL). The first-order and second-order structures (the latter except in “Corpus 3: Blog Authorships Corpus”; here, it did not converge) using the basic format were notably plausible (e.g.,  $\chi^2/df$  from 1.32 to 2.32 using ML and 1.14 to 1.46 with MLM; and CFI min: 0.933, SRMR max: 0.07). However, a more parsimonious model of TL (single factor) was also plausible. Moreover, in the second-order model based on the basic format (*a, n, v*), the first-level factors (TL dimensions) presented moderated/low reliabilities (e.g., composite reliability: 0.451 to 0.7019 in “letters to shareholders”; 0.42 to 0.61 in “online job interview”), but at the second-level, the factor (TL) presented high composite reliability (e.g., 0.94 and 0.91 in “letters to shareholders” and “online job interview,” respectively).

Globally interpreting these results, the basic format of PoS (*a, n, v*) helps represent TL from texts in two cases: (i) when the interest is focused on TL globally (e.g., single-factor model) or (ii) when its dimensions will be measured as PoS parcels ensuring the representation of qualities (adjectives), perceptual entities (nouns), and actions (verbs) of the global construct (TL).

Parcels are frequently employed as indicators of multidimensional constructs in CFA (Weng 2019), such as TL (Aryee et al. 2012, averaged items into TL dimensions). In addition, these are helpful in cases with relatively small sample sizes (e.g., type 2 errors are reduced Rahaman et al. 2020; Xie 2020), high numbers of indicators for constructs (Lan and Chen 2020), or restricted correlations, among others (Kishton and Widaman 1994).

However, instead of using the traditional random parcels based on merged lists of words, we are ensuring that each parcel in each TL dimension includes (averaging) three essential elements: qualities (adjectives), perceptual entities (nouns), and actions (verbs) concerning the construct under study.

To illustrate this strategy under the basic format of PoS, we perform a CFA with four PoS parcels (2 degrees of freedom). In the corpus of letters to shareholders (*n*: 186 obs), the results were (in parentheses MLM estimations):  $\chi^2/df$ : 1.945 (1.576); CFI: 0.994 (0.995); TLI: 0.982 (0.984); RMSEA: 0.071 (0.062); SRMR: 0.021. Likewise, factor loadings were: 0.69 (II); 0.77 (IM); 0.70 (IS); and 0.89 (IC); and composite reliability for TL was 0.849. Using the corpus of online interviews (128 obs), we found:  $\chi^2/df$ : 0.1995 (0.1755); CFI: 1 (1); TLI: 1 (1); RMSEA: 0.000; SRMR: 0.009. Factor loadings were: 0.6 (II); 0.82 (IM); 0.68 (IS); and 0.66 (IC); and composite reliability of TL was 0.784.

On the other hand, concerning the composite format (*ap, np, vp*), the evidence (see Table 10) suggests that when the interest is to contrast theories comprehensively (e.g., higher order factor), such a format is preferable to the basic PoS. In the three corpora, the first-/second-order models of TL were both plausible and reliable. For example, for the second-order model using the corpus of letters to shareholders, we obtained:  $\chi^2/df$ : from 2.49 to 4.86 (with ML estimations) and 2.21 to 2.44 (with MLM); CFI min: 0.959, and SRMR max: 0.09. However, the single-factor model presented a bad fit (e.g.,  $\chi^2/df$  ranging from 18.7 to 56.6 using ML and 10.79 to 27.9 with MLM). In addition, the composite reliabilities of the second-factor model ranged from 0.89 to 0.93 for dimensions and 0.85 for the global factor (second level). In this same corpus, Cronbach’s alpha for dimensions ranged from 0.88 to 0.93 and for global TL (averaging the dimension scores) it was 0.84.

Note that, in both cases (CFA with correlated-PoS based on both basic or composed PoS formats), the multidimensionality of TL is supported; this is consistent with several works using self-report data, such as Avolio et al (1999)



**Table 10** Results of CFA with correlated PoS using basic (a, n, v) and composed (ap, np, vp) formats

| Corpus 1: 2018 letters to shareholders from 1000 Fortune list (n: 186 observations) |  |           |                 |               |                |               |      |
|---|--|-----------|-----------------|---------------|----------------|---------------|------|
| Internal structures   | $\chi^2$   | <i>df</i> | $\chi^2/df$     | CFI           | TLI            | RMSEA         | SRMR |
| <b>Basic PoS</b>  |  |           |                 |               |                |               |      |
| (A) Single  | 77.98 (50.518)   | 36        | 2.166 (1.403)   | 0.939 (0.956) | 0.888 (0.919)  | 0.079 (0.058) | 0.05 |
| (B) First-order <sup>a,b</sup>  | 63.027 (41.032)  | 31        | 2.033 (1.323)   | 0.953 (0.97)  | 0.901 (0.936)  | 0.075 (0.052) | 0.06 |
| (C) Second-order  | 64.321 (42.05)   | 33        | 1.949 (1.274)   | 0.954 (0.973) | 0.909 (0.946)  | 0.072 (0.048) | 0.06 |
| Composite.Reliab  | 1°-level: II (0.526), IM (0.677), IS (0.451), IC (0.701); 2°-level: TL (0.94)  |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.33), IM (0.68), IS (0.45), IC (0.69); TL.average: 0.84; TL.global: 0.84  |           |                 |               |                |               |      |
| <b>Composed PoS</b>   |  |           |                 |               |                |               |      |
| (A) Single  | 787.384 (388.564)  | 36        | 21.8 (10.79)    | 0.644 (0.571) | 0.348 (0.214)  | 0.336 (0.328) | 0.11 |
| (B) First-order <sup>b</sup>  | 113.499 (68.556)   | 31        | 3.661 (2.211)   | 0.961 (0.963) | 0.917 (0.921)  | 0.12 (0.104)  | 0.06 |
| (C) Second-order <sup>b</sup>   | 119.13 (72.914)  | 33        | 3.61 (2.21)     | 0.959 (0.961) | 0.918 (0.922)  | 0.119 (0.103) | 0.06 |
| Composite.Reliab  | 1°-level: II (0.89), IM (0.93), IS (0.89), IC (0.93); 2°-level: TL (0.85)      |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.88), IM (0.93), IS (0.88), IC (0.93); TL.average: 0.84; TL.global: 0.93  |           |                 |               |                |               |      |
| Corpus 2: Online job interviews (emulation) (n: 128 observations)                   |  |           |                 |               |                |               |      |
| Internal structures   | $\chi^2$   | <i>df</i> | $\chi^2/df$     | CFI           | TLI            | RMSEA         | SRMR |
| <b>Basic PoS (a, n, v)</b>  |  |           |                 |               |                |               |      |
| (A) Single  | 49.542 (47.153)  | 36        | 1.376 (1.31)    | 0.95 (0.956)  | 0.908 (0.919)  | 0.054 (0.051) | 0.06 |
| (B) First-order   | 39.59 (37.287)   | 30        | 1.32 (1.243)    | 0.965 (0.971) | 0.922 (0.935)  | 0.05 (0.045)  | 0.05 |
| (C) Second-order <sup>b</sup>   | 51.13 (48.267)   | 33        | 1.549 (1.463)   | 0.933 (0.939) | 0.866 (0.877)  | 0.066 (0.062) | 0.07 |
| Composite.Reliab  | 1°-level: II (0.5), IM (0.56), IS (0.61), IC (0.42); 2°-level: TL (0.91)       |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.51), IM (0.37), IS (0.6), IC (0.43); TL.average: 0.78; TL.global: 0.78   |           |                 |               |                |               |      |
| <b>Composed PoS</b>   |  |           |                 |               |                |               |      |
| (A) Single  | 672.980 (546.863)  | 36        | 18.694 (15.191) | 0.455 (0.452) | 0.001 (-0.005) | 0.373 (0.371) | 0.18 |
| (B) First-order   | 79.134 (76.239)  | 30        | 2.638 (2.541)   | 0.958 (0.958) | 0.908 (0.908)  | 0.114 (0.112) | 0.09 |
| (C) Second-order  | 79.804 (76.49)   | 32        | 2.494 (2.390)   | 0.959 (0.960) | 0.916 (0.917)  | 0.108 (0.107) | 0.09 |
| Composite.Reliab  | 1°-level: II (0.91), IM (0.84), IS (0.92), IC (0.89); 2°-level: TL (0.65)      |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.91), IM (0.85), IS (0.92), IC (0.89); TL.average: 0.68; TL.global: 0.87  |           |                 |               |                |               |      |
| Corpus 3: Blog authorships corpus (n: 500 observations)                             |  |           |                 |               |                |               |      |
| Internal structures   | $\chi^2$   | <i>df</i> | $\chi^2/df$     | CFI           | TLI            | RMSEA         | SRMR |
| <b>Basic PoS (a, n, v)</b>  |  |           |                 |               |                |               |      |
| (A) Single  | 93.283 (46.317)  | 36        | 2.591 (1.287)   | 0.958 (0.983) | 0.923 (0.969)  | 0.056 (0.034) | 0.04 |
| (B) First-order   | 69.604 (34.247)  | 30        | 2.32 (1.142)    | 0.971 (0.993) | 0.936 (0.985)  | 0.051 (0.024) | 0.04 |
| (C) Second-order  | The model did not converge   |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.26), IM (0.2), IS (0.46), IC (0.37); TL.average: 0.8; TL.global: 0.75    |           |                 |               |                |               |      |
| <b>Composed PoS</b>   |  |           |                 |               |                |               |      |
| (A) Single  | 2038.98 (1006)   | 36        | 56.638 (27.945) | 0.537 (0.536) | 0.151 (0.15)   | 0.334 (0.331) | 0.2  |
| (B) First-order   | 160.232 (79.227)   | 31        | 5.169 (2.556)   | 0.97 (0.977)  | 0.936 (0.951)  | 0.091 (0.079) | 0.07 |
| (C) Second-order  | 160.518 (80.517)   | 33        | 4.864 (2.44)    | 0.97 (0.978)  | 0.941 (0.955)  | 0.088 (0.076) | 0.07 |
| Composite.Reliab  | 1°-level: II (0.821), IM (0.833), IS (0.896), IC (0.866); 2°-level: TL (0.699) |           |                 |               |                |               |      |
| Cronbach's alpha  | II (0.84), IM (0.82), IS (0.89), IC (0.88); TL.average: 0.74; TL.global: 0.88  |           |                 |               |                |               |      |

<sup>a</sup>One warning (The covariance matrix of the residuals of the observed variables is not positive definite)

<sup>b</sup>Fixing a negative variance near to zero. Composite.Reliab: Composite reliability (Fornell an Larcker, 1981), loadings of the first-level dimensions, and its global factor at second-level factor (TL). Cronbach's alpha: by dimensions (averaging items within each dimension), TL.average (averaging scores of the four dimensions), TL.global: averaging all items. II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)



and Tejada et al (2001). Moreover, as expected, correlations between aggregated scores derived from basic ( $a$ ,  $n$ ,  $v$ ) and composed PoS ( $ap$ ,  $np$ ,  $vp$ ) are high (II: 0.89, IM: 0.93, IS: 0.9, IC: 0.94; and TL: 0.96), which supports the convergence between the two types of measures.

To illustrate the application of the remainder of the PMTM steps, the following sections are developed using the basic format of PoS, averaging scores of adjectives, nouns, and verbs within each TL dimension and averaging dimension scores to obtain TL scores.

### Confirming external validity (convergent and discriminant capabilities)

To analyze associations with reference variables from textual and nontextual data (e.g., Pandey and Pandey 2017; Ponizovskiy et al. 2020; Short et al. 2010), we collect and use five corpora from formal and casual contexts, which are described in Table 11.

Regarding data set “1. Blog Authorships Corpus scores for personal values, and LIWC2015 variables” ( $n$ : 8869 obs; see Table 11), the correlation analysis is shown in Table 12.

Table 12 identifies that the scores for all TL dimensions are positively correlated (most of which are significant) with variables, such as “conformity” (0.22 to 0.65), “stimulation” (0.58 to 0.76), “achievement” (0.48 to 0.56), “power” (0.22 to 0.61), “focuspresent” (0.35 to 0.67), and “work” (0.29 to 0.60). These results reinforce the scale convergence, considering measures that are expected to be positively associated with TL. Likewise, the scores of TL dimensions present negative associations (most are significant) with variables, such as “hedonism” (− 0.41 to − 0.06), “percept” (− 0.41 to − 0.35), “focuspast” (− 0.68 to − 0.47), “motion” (− 0.63 to − 0.34), “leisure” (− 0.45 to − 0.3), “home” (− 0.64 to − 0.49), “informal” (− 0.54 to − 0.22), and “swear” (− 0.52 to − 0.28). These are evidence in favor of the discriminant capability of the developed scale.

In addition, all correlations between the TL dimension scores presented high values (all significant), ranging from 0.76 (IS) to 0.89 (IS, IC), which also favor the scale convergence. Likewise, these correlations (between TL dimensions) were greater than those between TL dimensions and other variables, which advocates for the discriminant capability of the scale. Furthermore, between TL dimensions, there is also discriminant capability; for example, although all TL dimensions were negatively associated with ‘hedonism’ (− 0.41 to − 0.06), only IS was statistically significant. Similarly, IS was the only dimension that showed significant associations with “risk” (0.36) and self-direction (0.33); in addition, the highest correlation between “power” and TL dimensions was with IS (0.61, significant). Likewise, II and IM evidenced a positive (non-significant) association with “reward” (0.2

and 0.22, respectively), whereas IS was negative (non-significant; − 0.16) and IC practically nil (− 0.03). Furthermore, the aggregate measure of TL (averaging the TL dimension scores) presented significant correlations that were expected. For example, “conformity,” “universalism,” “self-direction,” “achievement,” “insight,” “focuspresent,” and “work” positively significantly are correlated with TL, but “percept,” “focuspast,” “relativ,” “motion,” “leisure,” “home,” “informal,” and “swear” are negatively and significantly correlated with TL.

Detailed correlations between the TL scale scores and the remaining 13 reference variables (described in Table 11) are in the supplementary material (“Supp3”). Next, we will summarize the main findings:

Regarding data set “2. MBTI personality including the last 50 things blog posted” ( $n$ : 8675 obs; see the data described in Table 11), we calculate the point-biserial correlation between scores of TL dimensions and variables representing personality types from MBTI (previous binarization: e.g., 1. Extraversion, 0. Introversion; 1. Sensing, 0. Intuition). Again, we found evidence from perceptual data (MBTI) supporting the convergence and discriminant capabilities of the TL scale. In all cases, correlations between the scores of TL dimensions were positive and significant (0.26 to 0.28); in addition, these were greater than correlations between these and the other variables (from − 0.04 to 0.09). Moreover, IM was negatively significantly associated with ‘think’ (− 0.04; i.e., tends toward ‘feeling’), while IS presented a positive-significant association with ‘think’ (0.09). Likewise, TL showed a tendency toward decision-making preferences instead of perceiving (correlation with “judging”: 0.07; significant).

Regarding data set “3. Managerial essay corpus and questionnaire data” ( $n$ : 179 obs for some items of MLQ; and 96 obs for items of NEO-PI-R; see the data description in Table 11), we found several logical and expected patterns: IM and TL (this latter, averaging scores) are positively and significantly correlated with ‘extraversion’ (0.21 and 0.238, respectively), and negatively correlated (non-significant) with ‘neuroticism’ (− 0.229 and − 0.134, respectively). Besides, IC and TL showed negative and significant correlations with ‘laissez faire’ (− 0.186 and − 0.147, respectively); IM, IS, and TL revealed negative/significant correlations with a reactive style of leadership (Management-by-Exception: Active, MBEA), obtaining values − 0.222, − 0.157, and − 0.181, respectively. Hence, the evidence and the correlations between TL dimensions (from 0.273 to 0.445, all significant) suggest the convergence and discriminant capability of the TL scale.

Now, focused on the data set “4. Online job interview” ( $n$ : 128 obs; see the data described in Table 11), we calculate correlations between TL dimensions and two variables both from MLQ items (Avolio and Bass 2004): satisfaction with



**Table 11** Data used for step 12 of PMTM “Confirming external validity”

| Data   | Sample size  | Description   |
|--|--|---|
| 1. Blog authorships corpus scores for personal values, and LIWC2015 variables        | 8869 obs<br>(30 continuous reference variables from textual data)                | A sample of the training sample (9660 obs) used by Ponizovskiy et al. (2020); <a href="https://osf.io/9k56">https://osf.io/9k56</a><br>One of their corpora was “Blog Authorships Corpus” (blog post published on blogger.com in August 2004; Schler et al. 2006). This sample comprises: (a) blog posts; (b) scores from the dictionary of personal values (10 variables; developed by Ponizovskiy et al. 2020); (c) LIWC2015 scores (Pennebaker et al. 2015; we select 20 variables); and (d) the age and gender of bloggers. We automatically removed (in R) observations outside of the interval: TL scores > quantile 1% (– 5.403124) and TL scores < quantile 99% (– 1.098612), for more robust estimations |
| 2. MBTI personality including the last 50 blogs posted                               | 8675 obs. (4 reference variables –binary type—from perceptual data)              | A sample of the “MBTI-Myers-Briggs Personality Type Dataset” (8860 obs.) ( <a href="https://www.kaggle.com/datsnaek/mbti-type">https://www.kaggle.com/datsnaek/mbti-type</a> ). We automatically removed (in R) observations outside of the interval: TL scores > quantile 1% (– 5.403124) and TL scores < quantile 99% (– 1.098612), for more robust estimations. We transform the four MBTI dichotomies as binary variables (e.g., 1. Extraversion (E), 0. Introversion (I))  |
| 3. Managerial essay corpus and questionnaire data                                    | 179 obs. (2 reference continuous variables—continuous type—from perceptual data) | (a) 188 essays from students at college (8%) and technical (50%) careers, and practitioners (42%), then from psychology and managerial science fields (in Colombia); 95.2% were complete observations with the self-report “b.”<br>(b) Responses about laissez-faire leaders (mean of four items), and Management-by-Exception (Active; mean of four items), based on MLQ5x (Avolio and Bass 2004). First, participants write the essays; then they complete the self-report  |
| 4. Online job interview (emulation)  | 96 obs. (5 reference variables from perceptual data)                             | (c) Self-report of personality traits from NEO PI-R revised (Costa and McCrae 2008). Five factors: extraversion; neuroticism; openness; agreeableness; and conscientiousness  |
| 5. Champion vs. Contender companies using 2018 Annual reports from 1000 Fortune list | 128 obs. (2 reference continuous variable from perceptual data)                  | Previously described, but on this occasion, the sample considered self-report data for two constructs included in the MLQ5x: “Satisfaction with the leader” (mean of two items; Avolio and Bass 2004) and one measure derived from averaging the four items of MLQ (one item of each dimension of TL), here named TL.one  |
| 7L transformational leadership   | 60 obs. (1 comparison group: Champion vs. Contender companies)                   | This strategy was inspired by Pandey and Pandey (2017)<br><i>Champions</i> : Group of top-ranked companies on the 1000 Fortune list (we randomly chose a sample of 30 companies from rank 1 to 100)<br><i>Contenders</i> : Group of lower ranked companies on the Fortune 1000 list (a randomly chosen sample of 30 companies from rank 901 to 1000)  |

7L transformational leadership

**Table 12** Correlations based on data from the Blog Authorships Corpus and LIWC2015 (*n*: 8869 obs.)

|                             | II         | IM       | IS         | IC         | TL         |
|-----------------------------|------------|----------|------------|------------|------------|
| IM                          | 0.82****   |          |            |            |            |
| IS                          | 0.76****   | 0.71**** |            |            |            |
| IC                          | 0.83****   | 0.81**** | 0.89****   |            |            |
| TL                          | 0.91****   | 0.89**** | 0.92****   | 0.96****   |            |
| Security <sup>a</sup>       | 0.08       | - 0.01   | 0.33*      | 0.24       | 0.19       |
| Conformity <sup>a</sup>     | 0.34*      | 0.22     | 0.65****   | 0.52**     | 0.49**     |
| Tradition <sup>a</sup>      | 0.12       | 0        | 0.21       | 0.15       | 0.13       |
| Benevolence <sup>a</sup>    | 0.22       | 0.28     | 0.12       | 0.26       | 0.23       |
| Universalism <sup>a</sup>   | 0.32       | 0.19     | 0.62****   | 0.50**     | 0.46**     |
| Self-direction <sup>a</sup> | 0.61****   | 0.58***  | 0.76****   | 0.75****   | 0.74****   |
| Stimulation <sup>a</sup>    | 0.32       | 0.25     | 0.50**     | 0.40*      | 0.41*      |
| Hedonism <sup>a</sup>       | - 0.17     | - 0.06   | - 0.41*    | - 0.31     | - 0.27     |
| Achievement <sup>a</sup>    | 0.48**     | 0.48**   | 0.56****   | 0.53****   | 0.56****   |
| Power <sup>a</sup>          | 0.34*      | 0.22     | 0.61****   | 0.50**     | 0.47**     |
| Insight                     | 0.62****   | 0.63**** | 0.73****   | 0.77****   | 0.75****   |
| Percept                     | - 0.36*    | - 0.35*  | - 0.41*    | - 0.40*    | - 0.41*    |
| Sexual                      | - 0.26     | - 0.25   | - 0.27     | - 0.23     | - 0.28     |
| Reward                      | 0.2        | 0.23     | - 0.16     | - 0.03     | 0.04       |
| Risk                        | 0.21       | 0.16     | 0.36*      | 0.32       | 0.29       |
| Focuspast                   | - 0.53**** | - 0.47** | - 0.68**** | - 0.66**** | - 0.65**** |
| Focuspresent                | 0.55****   | 0.67**** | 0.35*      | 0.53****   | 0.56****   |
| Focusfuture                 | 0.05       | 0.18     | - 0.27     | - 0.13     | - 0.07     |
| Relativ                     | - 0.25     | - 0.23   | - 0.40*    | - 0.40*    | - 0.36*    |
| Motion                      | - 0.41*    | - 0.34*  | - 0.63**** | - 0.58**** | - 0.55**** |
| Space                       | - 0.18     | - 0.28   | - 0.09     | - 0.2      | - 0.2      |
| Time                        | - 0.15     | - 0.05   | - 0.41*    | - 0.34*    | - 0.28     |
| Work                        | 0.33*      | 0.29     | 0.60****   | 0.46**     | 0.47**     |
| Leisure                     | - 0.32     | - 0.3    | - 0.44**   | - 0.45**   | - 0.42*    |
| Home                        | - 0.57**** | - 0.49** | - 0.63**** | - 0.64**** | - 0.63**** |
| Money                       | 0.04       | 0        | 0.27       | 0.19       | 0.15       |
| Relig                       | 0.07       | - 0.04   | 0.14       | 0.08       | 0.07       |
| Death                       | - 0.19     | - 0.3    | - 0.04     | - 0.12     | - 0.17     |
| Informal                    | - 0.31     | - 0.22   | - 0.54**** | - 0.43**   | - 0.42**   |
| Swear                       | - 0.35*    | - 0.28   | - 0.52**** | - 0.41*    | - 0.43**   |

Notes:  $p < .0001$ \*\*\*\*,  $p < .001$ \*\*\*,  $p < .01$  \*\*,  $p < .05$  \*. <sup>a</sup> Personal values dictionary based on textual data, developed by Ponizovskiy et al. (2020). The other variables are obtained on LIWC2015 (Pennebaker et al., 2015) from a data set used by Ponizovskiy et al. (2020). II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)

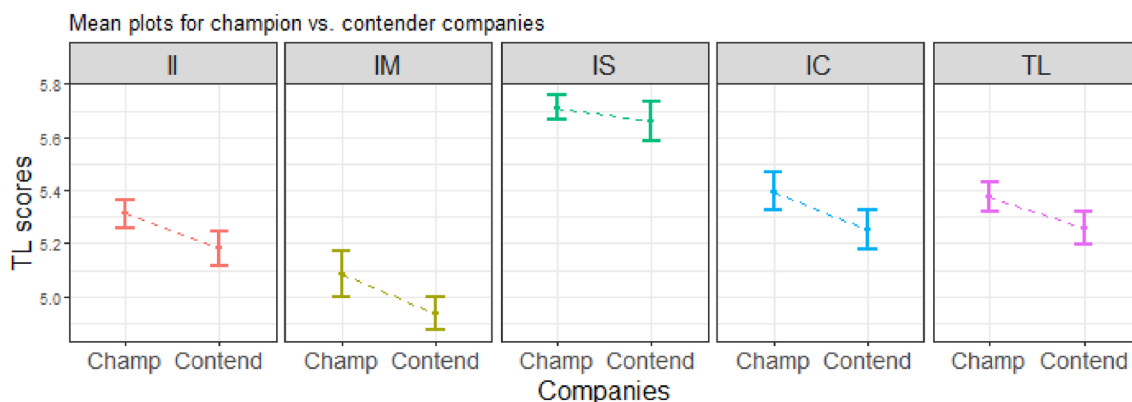
leadership ('Sat.mlq') and one TL proxy measure derived from averaging four items (one item for each TL dimension), entitled 'TL.one.mlq.'

We found that 'Sat.mlq' presented a positive and significant association with IS (0.217, at 0.05 significance) and TL (0.16, at 0.1 significance). Likewise, the proxy of TL based on self-reports ('TL.one.mlq') was positively and significantly correlated with IS (0.233, at 0.01 of significance) and TL derived from texts (0.191, at 0.05 of significance). Moreover, II and IM evidenced positive (nonsignificant) associations with 'TL.one.mlq' (0.147 and 0.132, respectively). These results are relevant, because the values of both

positive/significant correlations and positive/nonsignificant correlations are consistent with the validity criteria used by several studies, among them Ponizovskiy et al. (2020, s.p): "a typical correlation found between linguistic measures and self-reports is in the range of 0.1–0.2." Hence, we suggest that the evidence derived from online interviews and questionnaire data favors the convergence capability of the developed scale.

On the other hand, focused on the data set "5. Champion vs. Contender companies using 2018 Annual reports from 1000 Fortune list" (*n*: 60 obs, 30 for each company type; see





**Fig. 7** Mean plots of TL scores using reference groups and data derived from 2018 Annual Reports. *Notes:* II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)

the data description in Table 11), Fig. 7 provides a visual comparative analysis.

Figure 7 shows that in all dimensions and TL (averaging scores of its dimensions), except for IS, champion companies had more TL scores than contenders, which was also analytically supported using *t* tests, obtaining the following 95% confidence intervals: II (0.047, 0.211); IM (0.040, 0.255); IS (−0.034, 0.137); IC (0.045, 0.247); and TL (0.036, 0.201). This complementary evidence from a formal-organizational context reinforces the convergence (toward champion group) and discriminant capabilities (from contender group) of the developed TL scale.

Globally interpreting all evidence exposed derived from textual and nontextual data, formal and casual contexts, 43 reference variables, and one reference group (champions vs. contenders), we can suggest that the PMTM framework provided a TL dictionary with notable external validity in terms of convergence and discriminant capabilities.

### Confirming criterion validity

We employ two corpora/data sets: “MBTI personality including the last 50 things blog posted” (see the data description in Table 11) and “2018 letters to shareholders including two financial variables from 1000 Fortune list.” In the first case (8675 obs), we intend to separately predict each dichotomy of MBTI personality from TL dimension scores using three machine-learning methods. In the second, we calculate Pearson’s correlations between the scores of interest construct (TL) and two criterion-variables (Cameron and Bohannon 2000; Martínez et al. 2006) of financial type: revenues/assets and profits/assets (182 complete obs). For the first case, Fig. 8 describes the predictive capability of three machine-learning models (logistic regression, classification trees, and bagging) regarding the output variable

(preferences according to MBTI) using two samples: training (70% of obs., 5951) and validation (30%; 2550).

Figure 8 shows that models based on the developed scale can reasonably contribute to the prediction of dichotomies of personality traits (based on MBTI), with accuracies of around 77% in extraversion–introversion, 62% in sensing–intuition, 86% in thinking–feeling, and 57% in judging–perceiving.

In the second case, Table 13 provides correlations between the TL dimensions (based on letters to shareholders) and two financial indicators: revenues/assets and profitable/assets.

Table 13 reveals significant correlations between several TL dimensions and the two output variables. For example, II is positively significantly correlated with revenues/assets (0.214) and profitable/assets (0.22), and IC with profitable/assets (0.158). Likewise, TL (averaging scores of its dimensions) is also positively correlated with revenues/assets (0.123, at 0.1 of significance) and profitable/assets (0.187, at 0.05 of significance). Furthermore, IS positively correlated with profitable/assets (0.10; nonsignificant at 0.1 level).

Considering the evidence from Table 13 and Fig. 8, we can suggest that the developed scale of TL from texts presents traits of external validity based on criterion validity.

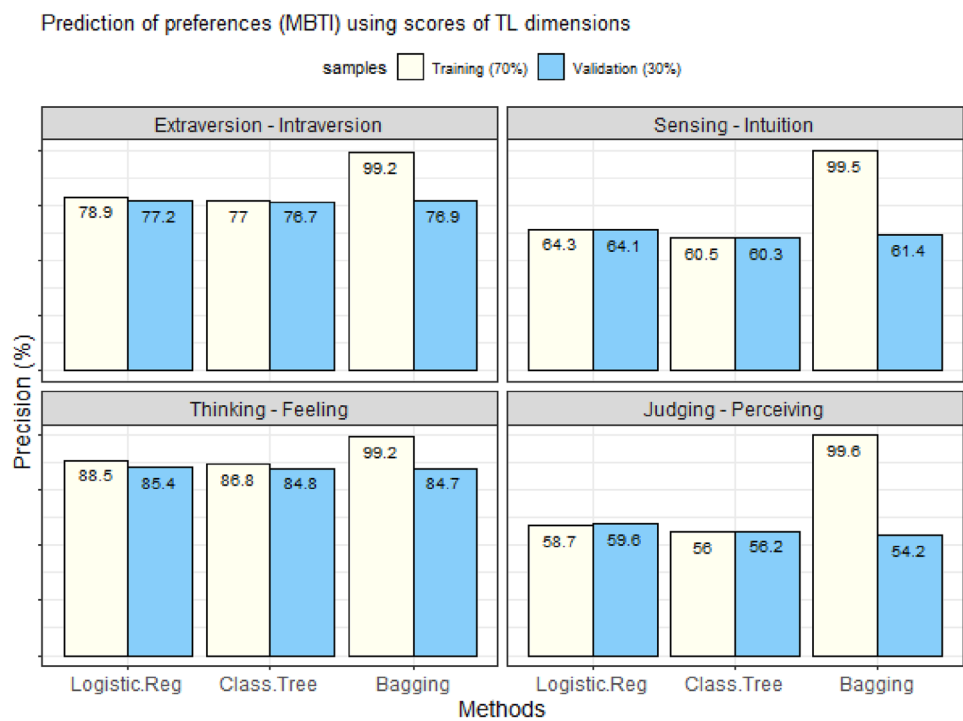
### Examining equity of the scale

We employ data sets from two different sources. The first data set consists of 93 observations (“3. Managerial essay corpus...”, described in Table 11); the second comprises two random subsamples (1000 and 4000 obs) obtained from “1. Blog authorships corpus scores...” (see the data description in Table 11). We estimate five regression models for each sample size (93, 1000, and 4000 obs) using TL dimensions and their average (TL) independently as response variables, and gender and age as regressor





**Fig. 8** Accuracy of TL dimensions based on textual data using machine learning methods. *Notes:* TL (transformational leadership)



**Table 13** Correlations with financial output variables of Fortune companies (182 obs)

|                  | II        | IM        | IS        | IC        | TL                 |
|------------------|-----------|-----------|-----------|-----------|--------------------|
| IM               | 0.431**** |           |           |           |                    |
| IS               | 0.233**   | 0.330**** |           |           |                    |
| IC               | 0.550**** | 0.535**** | 0.426**** |           |                    |
| TL               | 0.752**** | 0.790**** | 0.614**** | 0.839**** |                    |
| revenues/ assets | 0.214**   | 0.06      | -0.012    | 0.088     | 0.123 <sup>a</sup> |
| profits/ assets  | 0.220**   | 0.085     | 0.1       | 0.158*    | 0.187*             |

$p < 0.0001$ \*\*\*\*,  $p < 0.001$ \*\*\*,  $p < 0.01$  \*\*,  $p < 0.05$  \*,  $p < 0.1$ <sup>a</sup>

TL (transformational leadership); II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)

variables (two of the most frequent factors involved in the diversity–validity dilemma; Martínez et al. 2006; Pérez-Rave et al. 2021b). In all cases, gender and age were not statistically significant ( $\alpha: 0.05$ ). In addition, we carried out bootstrap regressions using 8000 replicas in each sample size scenario. The percentile intervals at the 95% level for gender coefficients in IS regressions were: (− 0.1918, 0.0061) in small, (− 0.0425, 0.0363) in medium, and (− 0.0375, 0.0008) in large scenarios. Likewise, the intervals for II were: (− 0.1142, 0.1373) in small, (− 0.0348, 0.0335) in medium, and (− 0.0309, 0.0023) in large sample sizes. In all cases, the intervals included zero.

Interpreting the regression results globally with and without bootstrapping, the evidence suggests that traditional personal factors do not significantly affect the scores produced by the developed scale. Thus, a classic problem in structured questionnaires (diversity–validity dilemma) may not be so in textual data using PMTM.

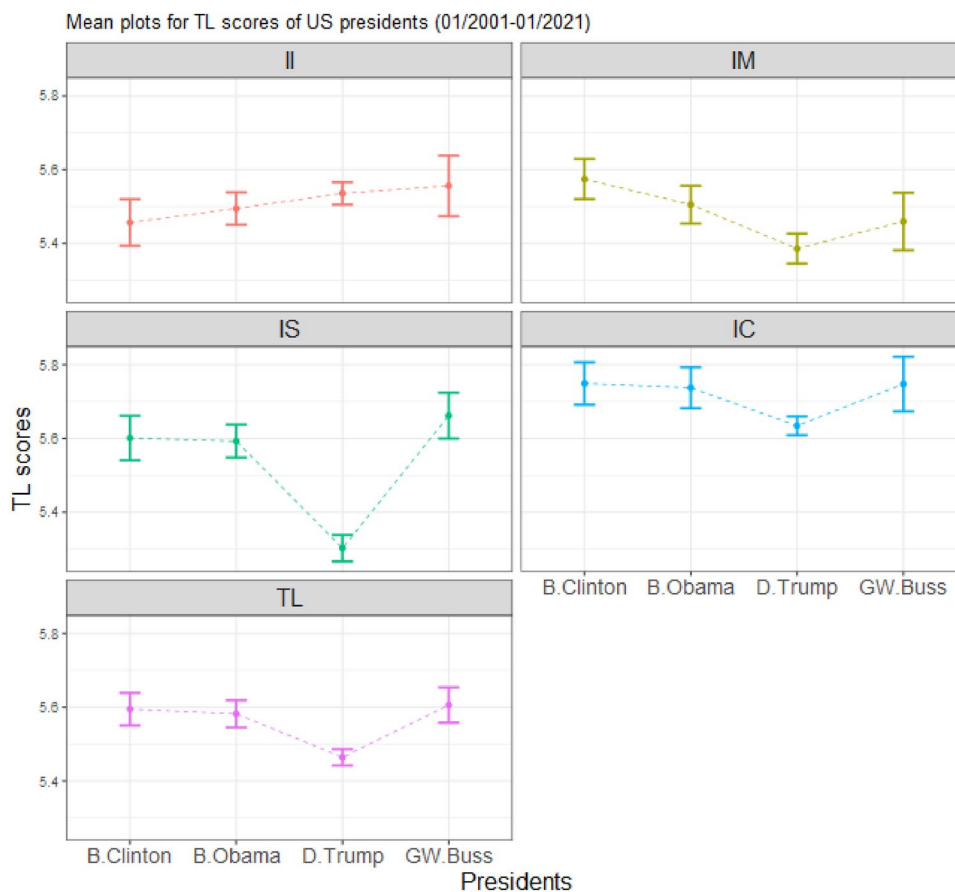
### Practical value of the scale (output of step 16 of PMTM)

To illustrate the practical value of the developed scale, we used a collection of 208 speeches by former US presidents, publicly available from Brown (2016; <http://www.thegrummarlab.com>). We chose the last four US presidents in the period 01/2001–01/2021: B.Clinton (39 obs); B.Obama (48 obs.); D.Trump (82 obs); and G.W.Bush (39 obs). Figure 9 details the TL scores (dimensions and average) for the US presidents under analysis, using mean plots with confidence intervals at the 95% level.

Figure 9 shows that the developed scale is valuable for analyzing individual differences. This analysis type is essential in describing the extent to which individuals are like one another (Loughry and McDonough 2002) and predicting several performances. Thus, the TL scale developed using PMTM allowed us to discover significant differences between President Trump and the others analyzed. Trump presented the lowest score in IM, IS, IC, and TL (average). However, regarding II, the results did not show notable differences between the presidents (“G.W.Bush–B.Clinton”



**Fig. 9** Mean plots with 95% confidence interval for scores of the last five US presidents (01/2001–01/2021). *Notes* We sum a constant (10) to scores for better visualization. TL (transformational leadership); II (idealized influence); IM (inspirational motivation); IS (intellectual stimulus); IC (individual consideration)



presented the most distance, but the  $p$  value was 0.0695 using Tukey's range test).

The results in Fig. 9 are consistent with other studies on TL in which language, behavior, and other aspects of Trump showed themselves more oriented toward transactional leadership than TL (Sternberg 2020). For example, according to Salter et al. (2017, p. 65), “Donald Trump used a greater percentage ( $M = 2.51\%$ ) of transactional words than Ted Cruz ( $M = 1.28\%$ ).” Furthermore, the high scores of Trump found in II (see Fig. 9) compared with the results in other dimensions (IM, IS, IC, and TL) are consistent with Lunbeck (2017), who highlights the fact that followers have a fascination with Trump and that he knows about this fascination. This charisma attribution is also found in Williams et al. (2020, p. 11), who use MLQ items (Avolio and Bass 2004, 1991) and conclude: “...Donald Trump [had] a higher score on the computed variable indicating perceptions that he is highly charismatic.”

In summary, this practical illustration, derived from the TL scale developed and validated using the PMTM framework, reveals that Trump presents a leadership style that differs from the other presidents considered. This distinctive feature is consistent with Fenner and Piotrowski (2018, p. 11): “the executive style of President Donald Trump has generated substantial empirical and theoretical attention.”

## General discussion

This paper considers the fundamental basis for developing and validating psychological/managerial constructs from texts, which have been systematically nourished with linguistics, psychometrical, and computational resources derived from three integrative stages and 16 steps. From this systematization, the developed framework (PMTM) extends the previous valuable works by providing four contributions:

The first, focused on the creation and content validation of measurement models of psychological/managerial constructs reflected on textual data, is a procedure comprising the initialization, expansion, and use of PoS during the creation and validation of construct dictionaries. This strategy is inspired by a logical and frequent method of construct operationalization based on structured questionnaires in sentences (items) formed by combinations of adjectives, nouns, and verbs (and auxiliary words). Thus, this strategy assumes that the study of psychological/managerial constructs is not limited to self-reports, because how a person manifests their feelings, skills, or beliefs also comprises their natural language (speaking or writing). Thus, PMTM has demonstrated that language can be used as a means for this type of analysis, which is consistent, for example, with the notion that “the ability to talk about an emotion without it being



physically present is a key component of natural language description of emotion” (Kazemzadeh et al. 2016, p. 5). According to Moulin (1992), a writer or orator chooses relevant information from the world, then builds a conceptual map (concepts and relationships) and expresses this through an oral or written discourse (linguistic level) to describe beliefs, sentiments, sensations, knowledge, behaviors, and attitudes. These expressions are highly subjective (Zhou and Zhang 2003) and, as responses to a structured questionnaire, natural language also represents underlying perceptions and diverse expressions about specific phenomenological/behavioral manifestations of constructs.

In summary, the proposed/developed framework offers more control, reproducibility, efficiency, and evidence-based decision-making during the construction and content validation of psychological/managerial constructs reflected in texts, thanks to five proposed/developed properties informing patterns of the linguistic environment of the text producer in an automated (or semi-automated) manner. In other words, after preparing the seed corpus, PMTM automatically transforms unstructured data (texts) into a structured format and carries out dictionary expansion steps whose performance (content validity) is based on and guided by five linguistic properties (commonality, polarity, coherence, convergence/differentiation, and PoS balance). This procedure expands the traditional standard for context validation in psychological/managerial constructs from texts, which is highly dependent on human tasks, subjective, limited to small data, and challenging to reproduce/replicate. Thus, PMTM facilitates the implementation of more complete and efficient tasks based on seed, scientific, and pragmatic corpora, and multiple linguistic/computational/psychometrical resources.

The second contribution is a new representation of constructs’ measurement models: a CFA with correlated-PoS. This new model representation considers three observable variables based on entities (nouns or noun phrases), qualities (adjectives or adjective phrases), and actions (verbs or verb phrases). Thus, from the semantic compositionality principle (phrase creation from word combinations; Fyshe 2015; Mitchell and Lapata 2010), a construct operationalized as a function of separated or aggregated PoS (in basic or composed formats) is more justifiable and linguistically complete than one merely comprised of a merged list of terms or random parcels of word lists. For example, Fyshe (2015) applied such a principle to advance, among others, the understanding and interpretability of phrase formation from combinations of separated nouns and adjective vectors. Haan et al. (2000), in the field of neuropsychology, provide evidence supporting a distinction in how the brain represents and processes nouns and verbs, reporting that verbs demanded more brain involvement than nouns. Martin et al. (1995) found a distinction in how actions (verbs) and

color words (adjectives) are processed in the brain. Fyshe et al. (2019) also inform the literature of differences in brain representations of nouns and verbs during phrase formation processes.

On the other hand, PMTM recognizes that the success of a TM solution in business management fields should be determined by evidence that is standard in the usage domain (Strohmeier and Piazza 2013). However, the traditional exploratory scope of text/data solutions (data-driven approach) is insufficient to confirm psychological/managerial latent variables. Therefore, PMTM adopts a theory/text-driven approach to take advantage of both construct theory (e.g., definitions, dimensionality) and data methods/technologies based on linguistic, computational, and psychometrical resources, and demonstrates (using multiple data sets; primary/secondary data; individual/organizational environments; and casual/formal domains) that a CFA representation with correlated PoS satisfies internal/external validity, reliability, equity, and practical value properties.

The third contribution is the provision of original empirical evidence derived from the framework application in creating and validating a novel scale for measuring TL based on textual data, consisting of 1073 expressions: 424 nouns, 209 adjectives, and 440 verbs. This scale is not invasive and allows studies to cover multiple empirical manifestations (entities, qualities, and actions) of the most frequently recognized dimensions of this construct: II, MI, IS, and IC. Likewise, the scale can be used in multiple types of texts, such as individual (e.g., blog posts, interview transcriptions, open questions from surveys) or organizational documents (e.g., annual reports and letters to shareholders) in several natural settings, including formal (e.g., public enterprises reports, presidential speeches, job interviews) and casual (e.g., social media or autobiographies) environments.

Comprehensively considering the discussed insights as a whole, the fourth contribution of the paper is a novel framework (PMTM) capable of inspiring, supporting, guiding, renewing, and reconfiguring future academic or business works in the creation, validation, and use of constructs (reflected in texts) involved in business management and related areas, such as marketing research. For example, PMTM provides conceptual and methodological guides to contribute to the challenge recently put forth by Cronin Jr. (2021), which comprises the need for new perspectives and proposals to comprehensively and efficiently conceptualize and operationalize marketing constructs and transcend the traditional dependency on questionnaire data.

Thus, marketing research and practice can benefit from the proposed/developed framework by exploiting service and customer texts to create measurement models (and obtaining their constructs’ scores) of psychological/managerial constructs considering scientific papers, social media, open questions from questionnaires, interview transcriptions, and



stakeholder questions/complaints/claims. Likewise, organizations can use the measurement models (and their construct scores) produced through the proposed/developed framework to (a) diagnose levels of service quality, satisfaction, trust, value, and behavioral intentions; (b) create customer hyper-segmentation strategies to transcend the traditional use of sociodemographic factors (e.g., age, gender); and (c) relate construct scores with financial/social/environmental performance measures to better understand the contributions of marketing strategies on organizational performance in the context of the digital economy.

On the other hand, from a theoretical perspective of data science, the proposed/developed framework (PMTM) can also be understood (and hence used) considering the three (compression, probabilistic, and microeconomic) theoretical perspectives of data mining argued by Mannila (2000). Applying the compression perspective to the present scenario, PMTM can be understood (and used) as a set of capabilities oriented to find, in texts from multiple domains (for dictionary creation), the underlying valuable qualities, perceptual entities, and actions (compressed data) suggested by a construct under study, thus controlling its PoS. From the lens of the probabilistic TM perspective, PMTM can be understood (and used) as a set of capabilities focused on discovering certain structures from compressed data either reflected in PoS frequencies (e.g., during content validation) or construct scores (e.g., during internal/external validation).

Therefore, linguistic and computational resources such as cleaning data, PoS tagging, collocations based on grammar patterns derived from chunks, cosine similarity, distributed dictionary representations, and word embeddings are justified and included in PMTM to discover valuable underlying patterns in the format of compressed data. Likewise, traditional procedures in structured questionnaire data analysis, such as Pearson's chi-squared tests, correlation analysis, and CFA (with correlated PoS using basic or composed PoS formats), are justified and included in PMTM to examine whether there is reasonable empirical evidence regarding certain beliefs/assumptions about the object of study (probabilistic perspective).

In addition, considering the TM microeconomic perspective (Manila 2000), PMTM can be understood (and used) as a set of capabilities to discover actionable patterns; that is, decision "x" that leads to the maximum utility  $f(x)$ . This view enabled the inclusion of machine-learning methods in PMTM, which involve splitting the sample into at least two subsets, one for training models (or patterns) and another for validation (Pérez-Rave et al. 2019). Although machine learning is one of TM's most popular approaches, its use in developing and validating psychological/managerial constructs is scarce. Hence, we consider the suggestions of Pandey and Pandey (2017) and Ponizovskiy et al (2020) concerning the need to examine forms to incorporate machine-learning

methods into the development/validation of constructs. Thus, PMTM includes one new method (embedded voting) and a new property (embedded-content convergence/differentiation), both of which are based on embedded vectors (e.g., word2vec). In addition, we also use logistic regression, classification trees, and bagging to analyze predictive capability from a supervised perspective.

## Conclusions and future work

The present study recognizes the contributions of Pandey and Pandey (2017) and Ponizovskiy et al (2020) and their predecessors (e.g., Short et al. 2010) and extends them toward a more enriched framework (PMTM) to develop and validate psychological/managerial constructs using TM considering linguistics, psychometrical, and computational resources. The evidence obtained from the application of the framework demonstrates both (a) the combination of linguistic properties and psychometrical/computational resources from a text-driven approach, effectively/efficiently guiding the development and validation of psychological/managerial constructs from texts; (b) constructing measurement models (from texts) operationalized as a function of correlated PoS emulating qualities, perceptual entities, and actions, presenting a good performance in several corpora regarding internal/external validities, reliability, equity, and practical value.

In future work, the PMTM framework should be validated considering other psychological/managerial constructs and contexts. Another future line of study is conducting a sensibility analysis of the proposed linguistic properties considering several violations and nonviolations of content validity. This opportunity will facilitate the examination of the extent to which such automated properties (and their methods, for example, embedded voting) may complement not only human tasks but also replace them. Future studies can also benefit from this paper by integrally using the PMTM framework or one or more of its stages, properties, and methods, considering the compressed, probabilistic, and microeconomic perspectives from which PMTM was interpreted/discussed.

In the marketing field, this paper can be used as a foundation on which to exploit customer/service textual data using the proposed/developed framework, which considers a theory/text-driven approach and incorporates linguistic, psychometrical, and computational resources. Thus, this paper serves as a comprehensive, reproducible, and efficient template to transcend the dependency on questionnaire data and overcome the current lack of standardization and inconsistency in the operationalization of marketing constructs.



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1057/s41270-022-00181-8>.

**Acknowledgements** Ministry of Science, Technology and Innovation (Minciencias), Colombia.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abrahamson, E., and D.C. Hambrick. 1997. Attentional homogeneity in industries: The effect of discretion. *Journal of Organizational Behavior: THE International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 18 (S1): 513–532.
- Alqatawneh, A.S. 2018. Transformational leadership style and its relationship with change management. *Verslas: Teorija Ir Praktika* 19 (1): 17–24.
- Amado, A., P. Cortez, P., Rita, S., Moro. 2018. Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics* 24(1): 1–7.
- Andersen, J.A. 2018. The old man and the sea of leadership: looking for effectiveness. *LSE Business Review* 19: 71–81.
- Aryee, S., F.O. Walumbwa, Q. Zhou, and C.A. Hartnell. 2012. Transformational leadership, innovative behavior, and task performance: Test of mediation and moderation processes. *Human Performance* 25 (1): 1–25.
- Avolio, B., and B. Bass. 1991. *The full-range of leadership development*. Binghamton, NY: Center for Leadership Studies.
- Avolio, B., and B. Bass. 2004. *Multifactor leadership Questionnaire. Manual and sampler set*, 3rd ed. Menio Park: Mind Garden.
- Avolio, B.J., B.M. Bass, and D.I. Jung. 1999. Re-examining the components of transformational and transactional leadership using the Multifactor Leadership. *Journal of Occupational and Organizational Psychology* 72 (4): 441–462.
- Base comunicativa y otros artículos. Serie Monografies 3. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Batista-Foguet, J.M., and G. Coenders-Gallart. 2000. *Modelos de ecuaciones estructurales, cuadernos de estadística*. Madrid: La Muralla.
- Bergkvist, L., and M. Eisend. 2021. The dynamic nature of marketing constructs. *Journal of the Academy of Marketing Science* 49 (3): 521–541.
- Bergkvist, L., and T. Langner. 2017. Construct measurement in advertising research. *Journal of Advertising* 46: 129–140.
- Blei, D.M., A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.
- Bohara, P., and P. Tiwari. 2015. Nurturant task leadership in relation to positive emotions. *Editorial Advisory Board* 6: 382.
- Brown, J.D. 1988. *Understanding research in second language learning: A teacher's guide to statistics and research design*. New York: Cambridge University Press.
- Brown, D.W. 2016. Corpus of presidential speeches. <http://www.thegrammarlab.com> Accessed 02 March 2021.
- Brown, D.J., and L.M. Keeping. 2005. Elaborating the construct of transformational leadership: The role of affect. *The Leadership Quarterly* 16 (2): 245–272.
- Cabré, M.T. 1999. La terminología. Representación y comunicación. Elementos para una teoría de
- Cabré, M.T. 2002. Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción. In *La traducción científico-técnica y la terminología en la sociedad de la información*, ed. A. Alcina and S. Gomero, 87–105. Castellón: Publicacions de la Universitat Jaume I.
- Cameron, D., and R.W. Bohannon. 2000. Criterion validity of lower extremity Motricity Index scores. *Clinical Rehabilitation* 14 (2): 208–211.
- Cameron, K.S., and R.E. Quinn. 1999. *Diagnosing and changing organizational culture*. Reading, MA: Addison Wesley Longman.
- Campion, M.C., M.A. Campion, E.D. Campion, and M.H. Reider. 2016. Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology* 101 (7): 958.
- Campos, A. A., and S. T. Castells. 2010. Adjectives and collocations in specialized texts: Lexicographical implications. In Proceedings of the XIV EURALEX International Congress.
- Carless, S.A., A.J. Wearing, and L. Mann. 2000. A short measure of transformational leadership. *Journal of Business and Psychology* 14 (3): 389–405.
- Chang, J., S. Gerrish, C. Wang, J. Boyd-Graber, and D. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, ed. M. Jordan, Y. LeCun, and S. Solla, 288–296. London: MIT Press.
- Chomsky, N. 1955. The logical structure of linguistic theory. Manuscript, Harvard University. Published in 1975, New York: Plenum.
- Cheng, K., Y. Yue, and Z. Song. 2020. Sentiment classification based on Part-of-Speech and self-attention mechanism. *IEEE Access* 8: 16387–16396.
- Cornesse, C., and A.G. Blom. 2020. Response quality in nonprobability- and probability-based online panels. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124120914940>.
- Costa, P.T., Jr., and R.R. McCrae. 2008. *The Revised NEO Personality Inventory (NEO-PI-R)*. London: Sage Publications Inc.
- Credé, M., and P.D. Harms. 2015. 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior* 36 (6): 845–872.
- Cronin, J.J., Jr. 2021. A commentary on “the role of services in the future.” *Journal of Marketing Theory and Practice* 29 (1): 23–36.
- Cui, W., Y. Wu, S. Liu, F. Wei, M.X. Zhou, and H. Qu. 2010. Context preserving dynamic word cloud visualization. In 2010 IEEE Pacific Visualization Symposium (PacificVis), pp. 121–128. IEEE.
- DeDeyn, R. 2021. Teacher leadership and student outcomes in a US University Intensive English Program. *TESL-EJ* 24 (4): n4.
- Denison, D.R., and W. Neale. 2000. *Denison organizational culture survey*. Ann Arbor, MI: Denison Consulting.
- Diebig, M., K.C. Bormann, and J. Rowold. 2017. Day-level transformational leadership and followers' daily level of stress: A moderated mediation model of team cooperation, role conflict, and type of communication. *European Journal of Work and Organizational Psychology* 26 (2): 234–249.
- Dönmez, S., and Y. Toker. 2017. Construction of a likert-type transformational leadership scale. *DTCF Dergisi* 57 (2): 753–775.
- Drozdź, S., J. Kwapien, and A. Orczyk. 2009. Approaching the linguistic complexity. In International Conference on Complex Sciences (pp. 1044–1050). Springer, Berlin.
- Edwards, J.R. 2008. Development of the Survey of Transformational Leadership for application to the substance abuse treatment field (Doctoral dissertation, Texas Christian University).
- Espinosa, K. 2017. An Empirical Study on End-to-End Sentence Modelling. In Proceedings of ACL 2017, Student Research Workshop (pp. 128–135).



- Esuli, A., and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*
- Fenner, C.J., and C. Piotrowski. 2018. Communication theory in leadership research: Which theories predominate. *Psychology & Education: An Interdisciplinary Journal* 55 (2): 10–18.
- Firth, J.R. 1957. A synopsis of linguistic theory, 1930–1955. Studies in linguistic analysis, Special Vol., *Philological Society* pp. 1–32.
- Fornell, C., and D. F. Larcker. 1981. Structural equation models with unobservable variables and measurement error: Algebra and statistics. <https://doi.org/10.1177%2F002224378101800313>
- Fried, E.I. 2017. What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review* 11 (2): 130–134.
- Fyshe, A., G. Sudre, L. Wehbe, N. Rafidi, and T.M. Mitchell. 2019. The lexical semantics of adjective–noun phrases in the human brain. *Human Brain Mapping* 40 (15): 4457–4469.
- Fyshe, A., L. Wehbe, P. Talukdar, B. Murphy, and T. Mitchell. 2015. A compositional and interpretable semantic space. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human Language Technologies* (pp. 32–41).
- Garten, J., J. Hoover, K. Johnson, R. Boghrati, C. Iskiwitch, and M. Dehghani. 2018. Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. *Behavior Research Methods* 50 (1): 344–361.
- Ghosh, S., and B. K. Mishra. 2020. Parts-of-speech tagging in nlp: Utility, types, and some popular pos taggers. In *Natural Language Processing in Artificial Intelligence* (pp. 131–165). Apple Academic Press.
- Gough, H.G. 1979. A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology* 37 (8): 1398.
- Graham, J., J. Haidt, and B.A. Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology* 96: 1029–1046. <https://doi.org/10.1037/a0015141>.
- Grimm, P. 2010. *Social desirability bias*. Hoboken: Wiley international encyclopedia of marketing.
- Grizzle, J., C. Starmer, and G. Koch. 1969. Analysis of categorical data by linear models. *Biometrics* 25: 489–504.
- Haan, H., J. Streb, S. Bien, and F. Rösler. 2000. Individual cortical current density reconstructions of the semantic N400 effect: Using a generalized minimum norm model with different constraints (L1 and L2 norm). *Human Brain Mapping* 11 (3): 178–192.
- Holmes, E., S. Bourke, and C. Plumpton. 2019. Attitudes towards epilepsy in the UK population: Results from a 2018 national survey. *Seizure* 65: 12–19.
- Howe, H.L. 1985. Breast self-examination palpation skill: A methodological note. *Journal of Chronic Diseases* 38 (12): 995–1001.
- Hu, L.T., and P.M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6: 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hume, C. 2017. Enhancing questionnaire design through participant engagement to improve the outputs of evaluation (Doctoral dissertation, University of Sunderland).
- Illia, L., K. Sonpar, and M.W. Bauer. 2014. Applying co-occurrence text analysis with ALCESTE to studies of impression management. *British Journal of Management* 25: 352–372.
- Jackson, C.J. 2020. Transformational leadership and gravitas: 2000 years of no development? *Personality and Individual Differences* 156: 109760.
- Kalugina, O.B., U.V. Mikhailova, and R.R. Dema. 2019. Comparative analysis and experience of using social network analysis information systems. *Journal of Computational and Theoretical Nanoscience* 16 (8): 3166–3172.
- Katsyhev, A., A. Anikin, M. Denisov, and T. Petrova. 2020. Intelligent Approaches for the Automated Domain Ontology Extraction. In *International Congress on Information and Communication Technology* (pp. 410–417). Springer, Singapore.
- Kazemzadeh, A., J. Gibson, P. Georgiou, S. Lee, and S. Narayanan. 2016. A Socratic epistemology for verbal emotional intelligence. *PeerJ Computer Science* 2: e40.
- Kishton, J.M., and K.F. Widaman. 1994. Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement* 54 (3): 757–765.
- Kjellström, S., and H. Golino. 2018. Mining concepts of health responsibility using text mining and exploratory graph analysis. *Scandinavian Journal of Occupational Therapy*. <https://doi.org/10.1080/11038128.2018.1455896>.
- Korayem, M., D. Crandall, and M. Abdul-Mageed. 2012. Subjectivity and sentiment analysis of arabic: A survey. In *International conference on advanced machine learning technologies and applications* (pp. 128–139). Springer, Berlin, Heidelberg.
- Kotamena, F., P. Senjaya, and A.B. Prasetya. 2020. A literature review: Is transformational leadership elitist and antidemocratic? *International Journal of Social, Policy and Law* 1 (1): 36–43.
- Kugbonu, J.M. 2020. Total quality management practices and customer retention at Unity Rural Bank, Ghana (Doctoral dissertation, University of Cape Coast).
- Lan, Y., and Z. Chen. 2020. Transformational leadership, career adaptability, and work behaviors: The moderating role of task variety. *Frontiers in Psychology* 10: 2922.
- Lévy, J., and J. Varela. 2006. Modelización con estructuras de covarianzas en ciencias sociales. Temas esenciales, avanzados y aportaciones especiales. A Coruña: *Netbiblo*. <https://bit.ly/2DdwqUG>.
- Lievens, F., K. Van Dam, and N. Anderson. 2002. Recent trends and challenges in personnel selection. *Personnel Review* 31 (5): 580–601.
- Liman, A., M. Aliyu, and M. Halliru. 2020. Moderating role of job satisfaction on the relationship between multi-level marketing and distributors retention. *Ilorin Journal of Human Resource Management* 4 (1): 160–170.
- Lochter, J.V., R.F. Zanetti, D. Reller, and T.A. Almeida. 2016. Short text opinion detection using ensemble of classifiers and semantic indexing. *Expert Systems with Applications* 62: 243–249.
- Lorenz, T., K. Heinitz, C. Beer, and M. van Woerkom. 2021. Adaptation and validation of a German version of the Strengths Use and Deficit Correction (SUDCO) questionnaire. *PLoS ONE* 16 (1): e0245127.
- Loughry, W.J., and C.M. McDonough. 2002. Phenotypic variability within and between litters of nine-banded armadillos. *Southeastern Naturalist* 1 (3): 287–298.
- Lunbeck, E. 2017. The allure of Trump's narcissism. [www.lareviewofbooks.org/article/the-allure-of-trumps-narcissism/#/](http://www.lareviewofbooks.org/article/the-allure-of-trumps-narcissism/#/) Accessed 02 Apr 2021.
- Maier, T. 1993. *Similarity information through commonality analysis of a product program (in German)*. The Hague: ICED.
- Mannila, H. 2000. Theoretical frameworks for data mining. *SIGKDD Explorations* 1 (2): 30–32.
- Markham, S. K., M. Kowolenko, T. L. Michaelis. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, 58 (2): 30–39.
- Marsh, H., B. Byrne, and R. Craven. 1992. Overcoming problems in confirmatory factor analyses of MTMM data: The correlated



- uniqueness model and factorial invariance. *Multivariate Behavioral Research* 27 (4): 489–507.
- Martin, A., J.V. Haxby, F.M. Lalonde, C.L. Wiggs, and L.G. Ungerleider. 1995. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 270 (5233): 102–105.
- Martínez, R., J. Hernández, and V. Hernández. 2006. *Psicometría*. Madrid: Alianza Editorial.
- Mathaisel, D. F., and C. L. Comm. 2021. Political marketing with data analytics. *Journal of Marketing Analytics* 9 (1): 56–64
- Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. Efficient estimation of word representations in vector space. In Proceedings of Workshop at ICLR, 2013, <https://arxiv.org/pdf/1301.3781.pdf>.
- Mitchell, J., and M. Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science* 34 (8): 1388–1429.
- Moulin, B. 1992. Conceptual-graph approach for the representation of temporal information in discourse. *Knowledge-Based Systems* 5 (3): 183–192.
- Nielsen, F. Å. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint. <https://arXiv.org/abs/1103.2903>.
- Nunnally, J. 1978. *Psychometric theory*. New York, NY: McGraw-Hill.
- O'Reilly, C. A., J. Chatman, D. F. Caldwell. 1991. People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal* 34: 487–516.
- Pandey, S., and S.K. Pandey. 2017. Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*. <https://doi.org/10.1177/1094428117745648>.
- Patrick, D., L. Burke, C. Gwaltney, N. Leidy, M. Martin, E. Molsen, and L. Ring. 2011. Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: Part 2—assessing respondent understanding. *Value in Health* 14 (8): 978–988.
- Pennebaker, J., R. Boyd, K. Jordan, and K. Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.
- Pérez-Rave, J.I., J.C. Correa-Morales, and F. González-Echavarría. 2019. A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research* 36 (1): 59–96.
- Pérez-Rave, J.I., G.A.S. Figueroa, and F.G. Echavarría. 2022. A scale for measuring healthcare service quality incorporating patient-centred care and using a psychometric analytics framework. *Journal of Health Organization and Management*. <https://doi.org/10.1108/JHOM-10-2021-0387>.
- Pérez-Rave, J.I., G.P. Jaramillo-Álvarez, and J.C. Correa-Morales. 2020. Desafíos y oportunidades de la minería de datos/texto en la selección de personal. *Revista Espacios* 798: 1015.
- Pérez-Rave, J.I., G.P. Jaramillo Álvarez, and F. González Echavarría. 2021a. A psychometric data science approach to study latent variables: a case of class quality and student satisfaction. *Total Quality Management & Business Excellence*. <https://doi.org/10.1080/14783363.2021.1980381>.
- Pérez-Rave, J.I., G.P. Jaramillo-Álvarez, and J.C. Correa-Morales. 2021b. Multi-criteria decision-making leveraged by text analytics and interviews with strategists. *Journal of Marketing Analytics* 10: 30–49.
- Pérez Rave, J., and L. Muñoz Giraldo. 2014. What can't be ignored in service quality evaluation: Application contexts, tools and factors. *Revista Facultad De Ingeniería Universidad De Antioquia* 72: 145–160.
- Pietraszkiewicz, A., M. Formanowicz, M. Gustafsson Sendén, R.L. Boyd, S. Sikström, and S. Szczesny. 2019. The big two dictionaries: Capturing agency and communion in natural language. *European Journal of Social Psychology* 49: 871–887. <https://doi.org/10.1002/ejsp.2561>.
- Poghosyan, L., and J. Bernhardt. 2018. Transformational leadership to promote nurse practitioner practice in primary care. *Journal of Nursing Management* 26 (8): 1066–1073.
- Ponizovskiy, V., M. Ardag, L. Grigoryan, R. Boyd, H. Dobewall, and P. Holtz. 2020. Development and validation of the personal values dictionary: A theory-driven tool for investigating references to basic human values in text. *European Journal of Personality* 34 (5): 885–902.
- Ptok, A., R.P. Jindal, and W.J. Reinartz. 2018. Selling, general, and administrative expense (SGA)-based metrics in marketing: Conceptual and measurement challenges. *Journal of the Academy of Marketing Science* 46: 987–1011.
- Quercia, D., J. Ellis, L. Capra, and J. Crowcroft. 2012. Tracking "gross community happiness" from tweets. In Proceedings of the ACM 2012 conference on computer supported cooperative work (pp. 965–968).
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Rafferty, A.E., and M.A. Griffin. 2004. Dimensions of transformational leadership: Conceptual and empirical extensions. *The Leadership Quarterly* 15 (3): 329–354.
- Rahaman, H.S., J. Camps, S. Decoster, and J. Stouten. 2020. Ethical leadership in times of change: the role of change commitment and change information for employees' dysfunctional resistance. *Personnel Review*. <https://doi.org/10.1108/PR-03-2019-0122>.
- Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48 (2): 1–36.
- Salter, C., M. Green, M. Hodgson, and N. Joyner. 2013. The language of transformational leaders: Communicating to the needs of followers. *The Journal of American Academy of Business* 19 (1): 60–68.
- Salter, C., D. Hargrove, P.A. Duncan, D. Coleman, and M. Woodhull. 2017. Do emergent leaders speak transformational language: A study of the language and non-verbal behavior of Donald Trump and Senator Ted Cruz announcement for candidacy speeches. *Journal of Leadership, Accountability and Ethics* 14 (1): 58.
- Sari, S., and M. Adriani. 2019. Applications of tf-idf concept to improve monolingual and cross-language information retrieval based on word embeddings. In Association for Computing Machinery, AISS 2019: 2019 International Conference on Advanced Information Science and System, Singapore, 6, 1–6.
- Schler, J., M. Koppel, S. Argamon, and J. Pennebaker. 2006. Effects of age and gender on blogging. In N. Nicolov, F. Salvetti, M. Liberman, & J. Martin (Eds.), Proceedings of 2006 AAAI spring symposium on computational approaches for analyzing weblogs (pp. 199–205). Menlo Park, CA: AAAI Press.
- Schwartz, S. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, ed. M. Zanna, 1–65. Orlando, FL: Academic.
- Schwartz, S., J. Cieciuch, M. Vecchione, E. Davidov, R. Fischer, and C. Beierlein. 2012. Refining the theory of basic individual values. *Journal of Personality and Social Psychology* 103: 663–688. <https://doi.org/10.1037/a0029393>.
- Sekaran, U., and R. Bougie. 2016. *Research methods for business: A skill building approach*, 1–65. Hoboken: Wiley.
- Shakurova, L., B. Nyari, C. Li, and M. Rotaru. 2019. Best practices for learning domain-specific cross-lingual embeddings. <https://arxiv.org/abs/1907.03112>.



- Sharma, S., P. Y. K. L. Srinivas, and R. C. Balabantaray. 2015. Text normalization of code mix and sentiment analysis. In *2015 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 1468–1473). IEEE.
- Short, J., J. Broberg, C. Cogliser, and K. Brigham. 2010. Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods* 13 (2): 320–347.
- Stake, R.B. 2001. A framework for evaluating commonality. In *Design for Configuration* (pp. 169–184). Springer, Berlin
- Stegmann, J. 2014. Paradigm shifts. Part II. Reverse Transcriptase. Analysis of reference stability and word frequencies. <https://arxiv.org/abs/1412.2416>.
- Sternberg, R.J. 2020. Transformational giftedness: Rethinking our paradigm for gifted education. *Roeper Review* 42 (4): 230–240.
- Strohmeier, S., and F. Piazza. 2013. Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications* 40 (7): 2410–2420.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics* 37 (2): 267–307.
- Tausczik, Y.R., and J.W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29: 24–54. <https://doi.org/10.1177/0261927x09351676>.
- Tejeda, M., T. Scandura, and R. Pillai. 2001. The MLQ revisited: Psychometric properties and recommendations. *The Leadership Quarterly* 12 (1): 31–52.
- Valero, J.N., and H.S. Jang. 2020. The effect of transformational leadership on network performance: A study of continuum of care homeless networks. *Journal of Public and Nonprofit Affairs* 6 (3): 303–325.
- Weng, Y. 2019. The nature of multidimensional constructs represented by item parcels in structural equation modeling. *Chinese Journal of Psychology* 61 (1): 25–50.
- Whittaker, S., and J. Hirschberg. 2007. Accessing speech data using strategic fixation. *Computer Speech & Language* 21 (2): 296–324.
- Williams, E.A., R. Pillai, K. McCombs, K.B. Lowe, and B.J. Deptula. 2020. Adaptive and maladaptive narcissism, charisma, and leadership performance: A study of perceptions about the presidential leadership of Donald Trump. *Leadership* 16 (6): 661–682.
- Wüster, E. 1979. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Viena: Springer.
- Xie, L. 2020. The impact of servant leadership and transformational leadership on learning organization: A comparative analysis. *Leadership & Organization Development Journal* 41 (2): 220–236.
- Zhou, L., and D. Zhang. 2003. NLPiR: A theoretical framework for applying natural language processing to information retrieval. *Journal of the American Society for Information Science and Technology* 54 (2): 115–123.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Jorge Iván Pérez Rave, PhD.** Industrial Engineer (Universidad de Antioquia, Colombia). Specializations: (1) Statistics, and (2) Systems Engineering (Universidad Nacional de Colombia). Masters: (1) Systems Engineering (Universidad Nacional de Colombia), and (2) Visual Analytics and Big Data (UNIR, Spain). Ph.D. in Systems Engineering (Universidad Nacional de Colombia), and Ph.D candidate in Business Management (Universitat de València, Spain). Director of the IDINNOV research group, IDINNOV S.A.S.

**Gloria Patricia Jaramillo Álvarez, PhD.** Civil Engineer (Universidad Nacional de Colombia), PhD. in Planning and Management of Hydraulic Resources (Universidad Politécnica de Valencia). Professor of Decision Analysis. Universidad Nacional de Colombia, Department of Computer Science, Medellín.

**Juan Carlos Correa Morales, PhD.** Statistician (Universidad de Medellín, Colombia), Master of Statistics (University of Kentucky), PhD. in Statistics (University of Kentucky). Professor of School of Statistics, Universidad Nacional de Colombia, Medellín, Colombia.

