Article

Journal of Language and Social Psychology
2021, Vol. 40(2) 238–259

DOI: 10.1177/0261927X20944543
journals.sagepub.com/home/jls
SAGE

Word Use as an Unobtrusive Predictor of Early Departure From Organizations

Young Min Baek[1] and Jennifer Ihm[2]

Abstract

Past studies have emphasized members' personality as an important predictor of departure from organizations, but the measurement of this factor has mostly relied on self-judged personality. As alternatives to self-judged personality, our study examines how two unobtrusive measures—others-judged personality and computerized text analytic results through Linguistic Inquiry and Word Count 2015 (Pennebaker et al., 2015)—are related to members' departure from organizations ($N = 49$). Drawing from internal personnel evaluations (i.e., others-judged personality), text (i.e., self-introduction documents that applicants submitted when applying to the organization), and behavioral data (i.e., actual stay in the organization), this study indicates that unobtrusive measures significantly predict members' length of stay and that simultaneous use of both measures better predicts members' length of stay in the organization than either one separately. However, text analytic results through Linguistic Inquiry and Word Count 2015 predict members' departure more robustly. This study expands the theoretical meaning of personality and provides practical ways to predict people's organizational behaviors.

Keywords

departure from organizations, organizational turnover, unobtrusive measure, computerized text analysis, others-judged personality, Big Five personality

The departure of members imposes several costs on organizations. First, recruiting and selecting new members to replace leavers entails costs (Holtom et al., 2008). Second, low morale and eroded social capital within the organizations affect the members who

[1]Yonsei University, Seodaemun-gu, Seoul, South Korea
[2]Kwangwoon University, Nowon-gu, Seoul, South Korea

Corresponding Author:
Jennifer Ihm, School of Media and Communication, Kwangwoon University, 20 Kwangwoon-ro Nowon-gu, Seoul 01897, South Korea.
Email: ihmsy17@gmail.com

remain (Dess & Shaw, 2001; Felps et al., 2009). Departures also indicate that individuals have incurred the cost of joining an organization with which they are poorly aligned (Holtom et al., 2008). Thus, organizations that select individuals who will stay longer benefit themselves, the members, and those who as a result do not become members. Empirical analyses predicting who will stay longer can support this effort. This article, therefore, identifies a set of valid and reliable factors predicting members' length of stay.

We focus on organization members' personality, one of the most frequently executed frameworks in organizational studies (Choi et al., 2015; Holtom et al., 2008; Judge et al., 2002). Most studies examined the relationships between self-judged personality measures and members' self-reported attitudes or behavioral intentions. Measures adopted in previous studies have both strengths and limitations. For the record, self-judged personality measures are very cost-efficient and have proven reliable and valid predictive performance in organizational phenomena. However, it is also known that self-judged or self-reported measures, at least under some contexts, can be vulnerable to self-enhancement bias and impression management motives (Newman et al., 2003; Schlenker et al., 2012).

The current study takes a different route from former studies. First, we measured organizational members' actual length of stay, rather than attitude or behavioral intention, in an organization ($N = 49$). Second, to explain members' actual behavior, we examined the predictive validity of two unobtrusive measures for personality: (1) members' personality as judged by others, that is, personnel evaluations, and (2) computerized text analysis of members' self-introductions, that is, text analytic results using Linguistic Inquiry and Word Count 2015 (LIWC2015; Pennebaker et al., 2015). Organizational psychology demonstrates that others-judged personality has stronger reliability and validity for organizational members' attitude and behavioral intention, suggesting that it better predicts actual behaviors (Connelly & Ones, 2010; Watson et al., 2000). In the job interview, interviewers' ratings of applicants are basically others-judged measures, and those measures, if applicants' personality is additionally rated, are not very difficult to obtain, implying that others-judged personality measures are as cost-efficient as self-judged ones. Likewise, recent studies of computerized text analysis indicate that people's word use is a reliable and valid measure of their personality (Boyd & Pennebaker, 2017; Fast & Funder, 2008; Hall & Caton, 2017; Newman et al., 2003; Pennebaker & King, 1999; Yarkoni, 2010). When screening applicants, many organizations ask them to submit documents, such as a résumé, diploma, or written self-introduction or statement of purpose. In this study, self-introduction documents that members actually submitted when applying to an organization were analyzed using LIWC2015, an established text analysis software (Pennebaker et al., 2015). Given that automated text analysis of documents that are already submitted to organizations does not demand additional costs or manual efforts, text analytic results could be more cost-efficient than self-reported measures.

By taking advantage of both efficiency (i.e., low cost to obtain measures) and effectiveness (i.e., stronger reliability and validity) assumed in those measures, our study makes valuable contributions to predicting members' actual departure from organizations.

This study addresses three basic research questions: (1) Which dimension of others-judged personality (in our study, personnel ratings by a management team in the organization) predicts members' likelihood of leaving an organization? (2) What dimensions of word use detected from automated text analysis predict this likelihood? (3) Which set of predictors is more effective and convenient, others-judged personality or word use? To obtain empirical answers, we relied on three sources of data obtained from a nonprofit voluntary organization: (1) personnel's rating of members' personality (i.e., others-judged personality), (2) members' self-introductions with format-free submission (i.e., data for computerized text analysis), and (3) members' actual departure from organizations (i.e., length of stay).

Others-Judged Personality and Departure From Organizations

The first set of predictors for departure from organizations is others-judged personality—that is, organization members' personality based on others' judgment. Broadly speaking, previous studies have discussed two sets of predictors (Choi et al., 2015; Meyer et al., 2002) for members' organizational behaviors: (1) dispositional factors (e.g., personality) and (2) situational factors (e.g., work design, empowerment, and support from coworkers). While situational factors are strong predictors of members' psychological bond with an organization and departure from the organization (Meyer et al., 2002), two issues suggest the need to take dispositional traits (e.g., personality) seriously. First, situational factors come into play only after individuals become members. Second, it is reasonable to expect that applicants with certain dispositional traits (e.g., people who are imaginative) may be more satisfied with environments or cultures in certain types of organizations (e.g., organizations in the fashion industry) and stay longer in that type of organization.

Thus, to shape recruitment efforts, we turn to dispositional traits inferred from applicants' words and behaviors. In so doing, we use the Big Five (B5) personality, one of the most intensively studied dispositional traits in organizational psychology (Choi et al., 2015; Holtom et al., 2008; Judge et al., 2002). B5 explains personality with reference to five dimensions: extraversion, neuroticism, conscientiousness, openness to experience, and agreeableness (Gosling et al., 2006). People high in extraversion are more likely to show sociability and positive affectivity toward others and show leadership in groups. People high in neuroticism are more likely to be emotionally unstable, to perceive environments in a negative way, and to isolate themselves within organizations. High conscientiousness indicates a tendency to be careful and organized and loyal to organizations. People high in openness to experience are more likely to be imaginative, adventurous, artistic, and sensitive, and seek and suggest new opportunities in organizations. Those high in agreeableness tend to be cooperative and caring for others and will tend to maintain positive relationships with other members in organizations.

Meta-analytic findings have consistently reported that organizational members' commitment to or satisfaction with organizations is positively related to extraversion but negatively related to neuroticism, but the associations between the other three personality dimensions and the level of commitment or satisfaction vary across

organizational contexts (Choi et al., 2015; Judge et al., 2002; Rubenstein et al., 2018; Tett et al., 2017). However, we expect to find relationships between B5 personality dimensions and members' departure that differ from those found in prior meta-analyses for several reasons.

First, our outcome measure is actual behavior (i.e., length of stay), but most previous studies examined members' attitude or intention. Given the cumulative evidence that the correlations between psychological constructs and actual behavioral outcomes are generally high, we expect that our findings using members' actual length of stay may not be seriously different from those using members' psychological orientations adopted in previous research. Second, one set of our predictor measures is others-judged B5 personality (in our study, personnel ratings by a management team in the organization), unlike the self-judged measure past meta-analyses have used (Choi et al., 2015; Holtom et al., 2008; Judge et al., 2002; Meyer et al., 2002; Rubenstein et al., 2018; Tett et al., 2017). While self-judged personality measures may suffer from a lack of reliability and validity under certain contexts (Newman et al., 2003; Schlenker et al., 2012), previous meta-analysis confirms that the correlations between self-judged and others-judged personality measures are substantial (Connelly & Ones, 2010), implying that difference in personality measurement may not lead to substantially different results.

Last and most important, difference in organizational goal between the organization investigated and other organizations in previous research should be considered seriously. The organization in our study is a voluntary organization that does not seek profits, while other studies have looked at nonvoluntary organizations (mostly profits-oriented private companies). Depending on their goal, organizations tend to recruit members who are fitted to that goal. How members in voluntary organizations socialize, perceive the environment, seek opportunities, or interact may differ from members working in other sectors (Frumkin, 2002), implying that members committed to organizations might have different personality profiles.

Given that there are no reported studies examining the associations between others-judged personality and members' actual length of stay in nonprofit organizations, we do not attempt to derive specific hypotheses. Instead, the first research question is posited as follows:

Research Question 1: Which dimension of others-judged personality is related to members' departure from the focal organization?

Unobtrusive Computerized Text Analytic Measures and Departure From Organizations

The second set of predictors for departure from organizations is computerized text analytic measures captured from members' self-introductions. Words in self-introductions and speech in job interviews have served as among the most important criteria for organizational hiring decisions; however, apart from a few recent studies (e.g., Moore et al., 2017; Naim et al., 2015), such textual information has been used through

qualitative interpretation or manual content analysis rather than through systematic quantitative scrutinization, making it vulnerable to human subjectivity. With recent advances in text-mining tools detecting individual differences from text data, it has become possible to extract linguistic features in systematic and reliable ways (e.g., Pennebaker et al., 2015) and to examine the relationships between linguistic markers that appear in writing and speech to determine attitudes toward and behaviors in organizations.

Among a variety of text-mining approaches available, we select a word count approach using LIWC2015, developed by Pennebaker et al. (2015). LIWC2015 is a computerized text analysis program based on a dictionary (i.e., lexicon) that counts the number of words that appear in a text based on classifications of words into predetermined categories. LIWC2015's default dictionary has more than 90 categories, some of which are content related (e.g., family, friends, anger), some of which are functional words (e.g., articles, prepositions), and others of which are composite summary variables (e.g., analytic thinking, clout).

While LIWC2015's underlying logic is straightforward and simple, its reliability and predictive validity for document writers' personality have been well proven across a variety of empirical text analyses. First, people's word use is "stable across time and writing topic" (Pennebaker & King, 1999, p. 1300), meaning that results of LIWC2015 can be used as reliable indicators inferring people's personality with consistency. Even if organization members' self-introductions, crafted with special care to make a good impression, are different from other texts generated in ordinary situations (e.g., diary or formal speech), habitual use of "linguistic fingerprint[s]" still remains and reveals document writers' personality (Chung & Pennebaker, 2008; Jordan & Pennebaker, 2017; Newman et al., 2003; Pennebaker & King, 1999; Yarkoni, 2010).

Second, people's word use is a valid predictor of personal characteristics (Jordan & Pennebaker, 2017; Kacewicz et al., 2013) and behavioral outcomes, such as deception (Newman et al., 2003) or academic performance in educational institutions (Pennebaker et al., 2014; Robinson et al., 2013). Despite the presence of impression management endeavors, people's self-introductions contain linguistic markers that reveal what they are interested in and concerned about; what they are good or bad at; where they are in the existing social hierarchies; how they interact with others; what they did in the past, do now, and will do in the future; and, sometimes, whether they attempt to deceive.

Although departure from organizations is not directly investigated, Moore et al. (2017) found that LIWC2015-based text analysis of applicants' self-introductions is valid to predict whether or not an applicant receives a job offer. Specifically, an applicant with strong self-verification drive uses more words classified in the function or seeing categories, and higher appearance of such categorized words correlates with expert raters' positive judgment that an applicant is authentic and unlikely to make misrepresentations. This yields a higher likelihood of receiving a job offer (see Study 3 in Moore et al., 2017). Based on this finding, we expect that LIWC2015-based text analysis of members' self-introductions may contain a comprehensive and rich picture of members and may predict important organizational behaviors such as their length of stay in an organization.

While the predictive validity of LIWC2015 has been proven, there are virtually no studies examining the relationships between its text analytic results and members' actual length of stay, implying that theoretical derivation of the hypothesis is still premature. Thus, we investigate the second research question:

Research Question 2: What dimensions of word use detected from automated text analysis predict a departure from the focal organization?

Others-Judged Personality Versus Computerized Text Analytic Results of Self-Introductions

Unlike self-judged measures popularly adopted in organizational psychology literature, both others-judged personality and computerized text analysis of people's writings are unobtrusive measures. While it has been frequently reported that unobtrusive measures secure better validity and reliability than conventional self-judged measures, it is undeniable that self-judged measures have been acknowledged as having relatively good cost-efficiency and satisfactory level of predictive validity. Until recently, obtaining unobtrusive measures was difficult and costly. In organizational settings, however, it is easier to obtain unobtrusive measures and it is less expensive because others-judged measures and digitalized textual materials are widespread. At the stage of recruiting new members, job interviewers note many valuable observations about applicants, and a substantial number of written documents applicants submitted can be conveniently analyzed using advanced automated text analytic programs like LIWC2015 (Pennebaker et al., 2015). Differently put, obtaining unobtrusive measures may be more convenient and cheaper than self-reported measures.

There are some similarities between these measures. First, both measures are more likely to capture a person's real self rather than the person's desired or presented self. Some studies have reported that the correlation between self-judged personality and attitude or behavioral outcome loses its statistical significance, or its size substantially shrinks, when others-judged personality is taken into account (Connelly & Ones, 2010; Schlenker et al., 2012; Watson et al., 2000). This result implies that others-judged personality may portray the person's real personality instead of the person's desired or presented personality, which self-judged personality portrays. Additionally, sociolinguists have identified linguistic markers that indicate liars' speech and writing (Gottschalk, 1997; Newman et al., 2003), showing that word use reflects the real self rather than the presented or fabricated self.

Second, empirical evidence also suggests that others-judged personality and word use often better predict people's actual behaviors than self-judged personality. For example, in a meta-analysis, Connelly and Ones (2010) found "stronger validities for other-ratings in predicting academic performance and job performance" than self-ratings (p. 1117) and concluded that "using other-reports to measure personality" offers "extraordinary value" (p. 1092). Studies have also found that text analysis can predict communication style (Jordan & Pennebaker, 2017), academic performance (Pennebaker et al., 2014; Robinson et al., 2013), and receipt of a job offer (Moore et al., 2017). These

findings imply that alternative unobtrusive measures could be better predictors of members' organizational behaviors, such as departure from organizations.

However, there are differences between the two unobtrusive measures. Past research also suggests that the two sets of predictors in our study will differ in terms of effectiveness and efficiency. First and foremost, others-judged personality is a subjective measure. Human observers may use heuristics to make their judgments and may be vulnerable to biases related to acquaintance with the targeted others. However, computerized text analytic results are consistent and not contaminated by subjective biases or human errors. In terms of reliability (i.e., subjectivity resistance), computerized text analytic results are better than others-judged personality.

Second, it is easier to analyze word use than others-judged personality. Hiring and training reliable judges costs time and money, but running reliable automated text analysis programs is not very expensive, and current technology can process hundreds of thousands of documents in a few seconds. In terms of convenience (i.e., easiness to use and cost), others-judged personality measures fall short of computerized text analytic results.

On the other hand, trained judges, armed with clear understanding of organizational vision and contexts, can undeniably make a better qualitative or holistic judgment about which members will stay longer in the organization, which a simple word-counting approach may fail to capture. Despite recent advances in computerized text analysis, there are subtleties that only trained human experts detect in human writings or speech (Zeldow & McAdams, 1993). When effectiveness is an issue (i.e., ability to catch subtle meaning in writings), personality measured by well-trained judges may generate better results than computerized text analytic results.

In sum, the two measures suggested as alternatives for self-judged personality have their own advantages and disadvantages. To the best of our knowledge, previous studies have not compared each unobtrusive measure's predictive validity for an organizational behavior. Given the paucity of empirical studies simultaneously testing the effects of others-judged personality and word use on behavioral outcomes, this study could make valuable contributions to evaluate how their influences are distinguished, in terms of both effectiveness (i.e., which one better predicts departure from the organization) and efficiency (i.e., which one is more convenient). Therefore, we ask the last question:

> **Research Question 3:** Among both others-judged personality and word use of organizational members, which set of predictors is better, in terms of both effectiveness and efficiency?

Method

Sample

This study focuses on members of a voluntary organization comprising people in their 20s and 30s who work in the third sector (e.g., community organizations,

nongovernmental organizations, or nonprofit organizations; Shumate et al., 2014). The organization's mission is to provide shared housing to early-career individuals working on social issues in the nonprofit sector. Envisioned benefits to organizational members include lowered living costs, social support, and synergy. The organization was created with the hope that such individuals would take advantage of these benefits to prepare for their next steps, including starting new social ventures or nonprofit organizations. Members live in one of two adjacent buildings (Building 1 and Building 2), sharing kitchen and laundry facilities while having individual bedrooms and bathrooms. Members in each building hold regular meetings to discuss social values, volunteer for the local community together, and strategize to generate social impact more broadly.

Text data for this study are a set of members' self-introductions. When individuals apply to live in the organization's housing, they must submit self-introducing writings about themselves, their missions, and their values. The management committee reviews these documents to determine whether the individual can be admitted. At the time of the data collection in June 2018, the committee had admitted 50 members since the organization started accepting applicants in October 2014, some of whom were still living in one of the buildings. Unfortunately, however, one member's self-introduction document was missing due to an administrative mistake, and thus, our final sample size is 49.

The management committee provided us with the self-introductions of all 49 members since the organization began, generating a sample spanning from October 2014 to June 2018, and its evaluations of the members' ratings on the B5 personality dimensions, relying on Gosling et al.'s (2006) 10-item inventories. The organization also provided us with members' gender, age at the time of the application, which of the two buildings they lived in, and how long they stayed there. The total corpus consisted of 37,738 words (127 pages).

We visited the buildings twice and discussed our findings with the management committee at these visits, as well as in one other face-to-face meeting and in several phone calls and email exchanges. Thus, we acquired additional information about the context of the case and triangulated our understanding of the organization to aid our analyses.

Measures

The Length of Stay. Our dependent measure is the number of months for which members actually stayed in the organization ($M = 17.10$, $SD = 10.15$, $Mdn = 13$, ranging from 1 to 37). As shown in the difference between mean and median statistics as well as range statistic, the distribution of our dependent measure hardly resembles the Gaussian distribution. In other words, conventional ordinary least square (OLS) regression is not a good modeling strategy, and thus, zero-truncated negative binomial (ZTNB) regression is adopted. Details and rationales of the ZTNB regression (Cameron & Trivedi, 2013) will be described in the Statistical Methods section.

Personal Features. The first set of predictors for members' length of stay is personal features, comprising two groups of variables. One group is sociodemographics, specifically, members' gender (49% women) and age (M = 29.69, SD = 4.60). Neither educational achievement nor income level is included in the analysis because all the members hold a BA degree as the final degree, and their earnings histories have been uneven because they have worked in or owned new, small nonprofit organizations.

The other group is others-judged B5 personality dimensions using Gosling et al.'s (2006) 10-item inventories. It is obvious that the short-form personality scale suffers from validity and reliability issues (Credé et al., 2012), and this will be addressed in the Discussion section. Despite the potential pitfalls, there are two reasons why we chose Gosling et al.'s (2006) short personality measures. One reason is a recommendation from a carefully executed comparison study (Credé et al., 2012). Although Credé et al. (2012) do not deny that longer personality measures (e.g., 40-item; Saucier, 1994) show better validity and reliability than Gosling et al.'s (2006) measures, they suggest that Gosling et al.'s "two-item measures represent a very substantial improvement over single-item measures in criterion validity with further moderate gains evident for scales that are slightly longer" (Credé et al., 2012, p. 885), and Gosling et al.'s (2006) measures sometimes show similar performance with some medium-length personality measures (e.g., Thalmeyer et al., 2011). The other reason is practical, which is the more important consideration for us. If we used longer personality measures, the evaluation process would be burdensome for the two project managers. Because our study could not be accomplished without cooperation from personnel ratings by a management team in the organization, we had to make the optimal, even if not the best, choice of personality inventory.

Two project managers evaluated each member relying on Gosling et al.'s (2006) 10-item inventories. Specifically, the two managers rated the extent to which the pair of traits applied to each member using a conventional 7-point Likert-type scale (1 = *strongly disagree* to 7 = *strongly agree*), and intercoder reliability between two judges was satisfactory (Krippendorff's α = .76, interval level assumed). Based on the two managers' evaluations, each dimension of B5 personality was assessed by taking the average of two items, following Gosling et al.'s scheme. For example, a member's extraversion (M = 4.11, SD = 1.19) was measured based on the two managers' averaged rating of two items: (1) extraverted, enthusiastic and (2) reserved, quiet (reversely coded). The other four dimensions were similarly measured. Two items measuring each dimension and descriptive statistics of B5 personality are reported in Table 1.

Linguistic Features. The second set of predictors for members' length of stay is computerized text analytic results extracted from members' self-introduction documents submitted when applying to the voluntary organization. After all the documents of the total 49 members were digitalized, all linguistic features were extracted using LIWC2015. Among the 93 linguistic features extracted with LIWC2015, this study focuses on 12 predictors, 10 of which were original categories in LIWC2015, but two of which, for our research purpose, we formulated based on original categories of LIWC2015.[1] Operational definitions of the 12 linguistic features are provided in Table 2 with their descriptive statistics.

Table 1. Items for Big Five Personality Dimensions.

Personality dimension	Item 1	Item 2 (reverse-coded)	*M*	*SD*
Extraversion	Extraverted, enthusiastic	Reserved, quiet	4.11	1.19
Neuroticism	Anxious, easily upset	Calm, emotionally stable	4.09	0.93
Conscientiousness	Dependable, self-disciplined	Disorganized, careless	5.07	0.84
Openness to experience	Open to new experiences, complex	Conventional, uncreative	5.26	0.88
Agreeableness	Sympathetic, warm	Critical, quarrelsome	4.38	0.93

Note. Big Five personality dimensions were measured following Gosling et al.'s (2006) 10-item inventories. Two managers' observations of each member were obtained through a conventional 7-point Likert-type scale (1 = *strongly disagree* to 7 = *strongly agree*).

A total of 10 original categories were used without any transformation. Those categories can be clustered into three groups. First, total word count is the number of words in a document, which measures the mere length of each document.

Second, six categories measuring psychological orientations revealed in members' self-introductions are included. The biological processes (e.g., eat, hands, love) category is included because it may tap members' concern over basic functions related to living in the same housing with others, such as food or health. The other five categories were considered because they are related to the mission of the voluntary organization, either directly or indirectly. The achievement (e.g., win, success) category was included because it may show a member's psychological orientation for better achievement in a society of financial fields. The affiliation (e.g., social, alliance) category was considered because the organization supports social goals and community-related tasks. The work (e.g., job, majors), leisure (e.g., cook, chat), and money (e.g., cash, audit) categories were considered because they may reveal the members' personal values.

Third, three summary variables that the developers of LIWC2015 term *analytic thinking, clout*, and *authenticity* were included. These summarize words in several categories counted in LIWC2015 to better represent the sociopsychological aspects of a document writer. Specifically, the analytic thinking variable is a composite variable distinguishing people who "think in [a] more formal, logical, and hierarchical way" (Pennebaker et al., 2014, p. 9). According to Pennebaker et al. (2014), document writers high in analytic thinking scores are more likely to show better performance in cognitive and creative works (e.g., higher GPAs [grade point averages] in educational institutions). The clout variable is a language summary score capturing document writers' collective orientations, considerate leadership, and high status within social

Table 2. Descriptive Statistics of Linguistic Features.

Linguistic markers	*M*	*SD*	Minimum	Maximum
Predictors using original categories in LIWC2015				
Word count (number of words in document)	680.39	502.36	88.00	3607.00
Biological processes[a]	1.97	1.38	0.00	7.35
Achievement[a]	2.99	0.91	1.07	4.62
Affiliation[a]	4.02	1.47	1.42	7.61
Work[a]	5.55	1.93	1.88	11.00
Leisure[a]	1.13	0.91	0.00	4.19
Money[a]	1.04	1.09	0.00	6.21
Analytic thinking[b]	78.69	11.89	47.53	95.17
Clout[b]	54.36	13.21	20.30	81.06
Authenticity[b]	55.64	21.85	8.49	93.11
Predictors transforming original categories in LIWC2015				
Emotional positivity (positive–negative emotion)	3.40	2.26	−1.06	12.50
Positive emotion[a] (original category in LIWC2015)	4.70	1.91	2.09	12.50
Negative emotion[a] (original category in LIWC2015)	1.30	0.85	0.00	4.00
Time focusing (i.e., focus past − [focus present + focus future])	−7.67	2.69	−13.35	−3.06
Focus past[a] (original category in LIWC2015)	3.29	1.40	0.86	6.97
Focus present[a] (original category in LIWC2015)	9.97	2.06	6.02	14.77
Focus future[a] (original category in LIWC2015)	0.99	0.53	0.00	2.26

Note. LIWC2015 = Linguistic Inquiry and Word Count 2015.
[a]Variables that indicate the percentage of the total words in the document that belong to the LIWC2015's original category (e.g., "5" indicates that out of the total words counted in a document, 5% of the words are classified as the designated category).
[b]Variables that are summary language variables, calculated by Pennebaker et al.'s (2015) formula, and their value, by definition, ranges from 0 to 100. Emotional positivity and time focusing are derived using the variables of the original categories in LIWC2015 based on the formula described in the table.

hierarchy (Kacewicz et al., 2014). Authenticity is a composite variable measuring how "honest or authentic" a document writer is, based on a variety of linguistic cues (e.g., more self-referencing, vivid words describing complex phenomena; Newman et al., 2003, p. 667).

Additionally, two linguistic predictors, emotional positivity and time focusing, which we formulated based on original categories of LIWC2015, were used. We

considered emotional positivity because people with positive thoughts, compared with those with negative ones, are expected to have fewer relationship conflicts with other residents. Emotional positivity is the difference between the number of words in the positive emotion category (e.g., happy, nice, sweet) and the negative emotion category (e.g., hurt, ugly, hate); a larger number indicates that a member's self-introduction is filled with positive, rather than negative, words.

Second, time focusing, defined as the dominant tense revealed in a document, is included because it can show the type of self-identity provided in members' self-introduction. Members who describe themselves using past tense verbs are more likely to have a concrete and already accomplished self-concept based on their actual actions in the past; members who use present or future tense verbs are more likely to have a desired but as-yet unaccomplished self-concept based on what they want to do now or in the future. Time focusing refers to this difference between past focus (i.e., past tense verbs) and present focus (i.e., present tense verbs) or future focus (i.e., future tense words). A larger number means that members' self-introduction focuses on what they did in the past, rather than what they do now or will do in the future. Higher time focusing indicates a greater past focus.

Control Measure. One dummy variable (Building 1 = 1, Building 2 = 0) was created and entered in the regression equations to indicate in which of the organization's two buildings members live. About 65% of the members (n = 32) were assigned to Building 1.

Statistical Methods

To estimate the effects of both members' personal and linguistic features on their length of stay, ZTNB regression was adopted. As researchers have widely noted, conventional OLS regression frequently fails to analyze a count-dependent variable appropriately (Long, 1996); while either Poisson or negative binomial (NB) regression is frequently used as an alternative to OLS regression when analyzing a count-dependent variable such as the number of people murdered or the number of awards earned by scholars; our dependent variable, length of stay, is qualitatively different from conventional count variables.

The key difference for length of stay is that it is always larger than zero, but conventional count variables fitted for Poisson or NB have several or lots of zero values. The earliest leaver from the organization has to stay for at least 1 month, and thus the value zero cannot occur because of truncation, and Poisson or NB regression cannot be used to analyze data in which zero counts are excluded. In the case of count-dependent variables whose minimum count is more than zero, ZTNB regression is recommended when overdispersion is detected, and zero-truncated Poisson (ZTP) regression is recommended when it is not (Cameron & Trivedi, 2013). After conducting an overdispersion test, we chose ZTNB regression to test the effects of both personal and linguistic features on members' length of stay.

When estimating ZTNB regression, R package VGAM (Yee, 2010) was used.

Results

Model Construction and Overdispersion Test

To address our research questions, we constructed a total of three models: P (personal) model, L (linguistic) model, and PL (personal and linguistic) model. P model examines only the effects of personal features (i.e., gender, age, and B5 personality dimensions) on members' length of stay, and L model inspects only the effects of linguistic features (i.e., 12 predictors obtained through the LIWC2015) on the length of stay. PL model, however, simultaneously estimates the influences of both personal and linguistic features on members' length of stay. Comparing the results from the three models helps us understand which predictor set would better predict members' length of stay.

Given that our dependent variable is a count variable, for which the value zero cannot occur, either ZTP or ZTNB regression would be an appropriate modeling strategy (Cameron & Trivedi, 2013). To determine which regression model is more appropriate, we conducted overdispersion tests (ZTP if the dependent variable is not overdispersed, ZTNB if overdispersed). Results confirmed that members' length of stay is clearly overdispersed in the P model ($\chi^2(1) = 99.75$, $p < .001$), the L model ($\chi^2(1) = 61.98$, $p < .001$), and the PL model ($\chi^2(1) = 39.05$, $p < .001$). Thus, ZTNB regression is estimated to address our research questions.

Results Testing Personal and Linguistic Features on Members' Length of Stay

Three models of ZTNB regression were estimated, and their results are presented in Table 3. As shown at the bottom of Table 3, results of a log-likelihood ratio test show that the PL model is statistically better than the P model ($\chi^2(12) = 21.85$, $p < .001$) or the L model ($\chi^2(7) = 14.09$, $p < .001$). Additionally, McFadden's pseudo R^2 statistics of the three models show that the PL model (0.10) has better explanatory power than the P model (0.04) or the L model (0.07). In other words, both personal and linguistic features function as statistically significant predictors that help predict members' length of stay in the voluntary organization.

Based on a model comparison test and McFadden's pseudo R^2, we choose the PL model as the final model, and its incidental risk ratio (IRR) is also reported in Table 3. Among the seven personal features, members' openness to experience—that is, one of the B5 dimensions—is the only predictor achieving statistical significance ($b = 0.17$, IRR $= 1.19$, $p < .05$), meaning that people who are more likely to enjoy new things and to be open to unfamiliar experiences and different cultures will stay longer in the organization. However, other personal features are not related to members' length of stay with statistical significance.

Among the linguistic features, a total of four predictors achieve statistical significance. First, members showing a stronger analytic thinking writing style are more likely to leave the organization earlier ($b = -0.23$, IRR $= 0.80$, $p < .05$). Given that this style indicates skill at cognitive and creative tasks (Pennebaker et al., 2014), this

Table 3. Results of Zero-Truncated Negative Binomial (ZTNB) Regression Predicting Members' Length of Stay Using Both Personal and Linguistic Features ($N = 49$).

	P model, estimated (*SE*)	L model, estimated (*SE*)	PL model, estimated (*SE*)	IRR
Intercept	2.40*** (0.17)	2.55*** (0.13)	2.28*** (0.15)	9.75
Physical context				
Location (Building 1 = 1)	0.42* (0.18)	0.33* (0.16)	0.50** (0.17)	1.65
Personal features				
Gender (female = 1)	0.23 (0.16)		0.27 (0.15)	1.31
Age, years	0.17 (0.09)		0.18 (0.10)	1.20
Extraversion	0.14 (0.10)		0.08 (0.09)	1.08
Neuroticism	−0.13 (0.10)		−0.07 (0.10)	0.93
Conscientiousness	−0.11 (0.09)		−0.10 (0.08)	0.91
Openness to experience	0.16* (0.08)		0.17* (0.08)	1.19
Agreeableness	0.13 (0.10)		−0.05 (0.10)	0.95
Linguistic features				
Word count		0.00 (0.08)	−0.10 (0.08)	0.90
Biological process		−0.09 (0.08)	−0.14 (0.08)	0.87
Achievement		−0.03 (0.08)	0.01 (0.09)	1.01
Affiliation		−0.13 (0.11)	−0.12 (0.10)	0.89
Work		−0.08 (0.10)	−0.04 (0.10)	0.97
Leisure		0.05 (0.08)	−0.10 (0.08)	0.91
Money		−0.08 (0.09)	−0.06 (0.10)	0.94
Analytic thinking		−0.25* (0.12)	−0.23* (0.11)	0.80
Clout		0.32* (0.14)	0.32* (0.14)	1.37
Authenticity		0.19 (0.11)	0.25* (0.11)	1.28
Emotional positivity		0.24** (0.09)	0.14 (0.09)	1.15
Time focusing		0.26* (0.10)	0.22* (0.09)	1.25
Overdispersion parameter	1.54*** (0.27)	1.80*** (0.28)	2.18*** (0.32)	
McFadden's pseudo R^2	0.04	0.07	0.10	
Model comparisons (LR $\chi^2(df)$)				
P model vs. PL model			21.85* (12)	
L model vs. PL model			14.09* (7)	

Note. IRR = incidental risk ratio; *SE* = standard error; LR = log-likelihood ratio; P model is a ZTNB regression testing personal features only; L model is a ZTNB regression testing linguistic features only; and PL model tests both personal and linguistic features simultaneously. All predictors, except dummy variables (i.e., building and gender), are mean centered. IRR is the exponential value of ZTNB regression coefficients. ZTNB is estimated using VGAM (Version 1.1-2; Yee, 2010) package in R.
*$p < .05$. **$p < .01$. ***$p < .001$.

finding implies that those people quickly leave the shared housing and become financially stable and independent because of their success in the field.

Second, members with a higher clout score tend to stay longer in the organization ($b = 0.32$, IRR = 1.37, $p < .05$). As a higher clout score indicates considerate

leadership and collective orientation (Kacewicz et al., 2014), this finding suggests that those people with a higher clout score may experience fewer relationship conflicts with other residents than others.

Third, members with a higher authenticity score are more likely to stay longer in the organization. Previous research has found that document writers showing a stronger authenticity score are honest and humble (Newman et al., 2003), suggesting that other residents are more likely to see them as easy to get along with, which can reduce relationship conflicts in the organization. Moore et al. (2017) showed a similar positive effect of words representing authenticity on likelihood of receiving a job offer.

Fourth, members with a higher time-focusing score may reside with the organization longer ($b = 0.22$, IRR $= 1.25$, $p < .05$). A higher time-focusing score means that document writers introduced themselves with more past tense verbs than those describing their present or hoped-for future state. This finding implies that members who have concrete experience in the nonprofit field, rather than vague unexperienced ideals, are more likely to align with the organizational mission.

Discussion and Conclusion

As alternatives to self-judged personality, this study focused on two unobtrusive measures (i.e., personnel evaluation as others-judged personality and computerized text analytic results through LIWC2015) and successfully showed that both measures are good predictors of organization members' actual length of stay. In organizational settings, others-judged measures based on personnel observations, if short and simple enough, do not take much effort, and obtaining automated text analytic results of documents members submitted is convenient and does not require additional cost. Additionally, previous studies have demonstrated that both unobtrusive measures show better reliability and validity than the self-judged measure (Connelly & Ones, 2010; Fast & Funder, 2008; Newman et al., 2003; Pennebaker et al., 2014; Robinson et al., 2013; Schlenker et al., 2012). Using valid and reliable sources of data that are easily accessible but not actively investigated in organizational settings, we simultaneously examined the effects of both unobtrusive measures on members' actual departure from organizations. There are, at least, four interesting findings, with theoretical and methodological implications.

First, our two alternative measures of personality measure different dimensions of personality and are related differently to length of stay. The PL model more effectively predicts length of stay than the P model or the L model, indicating that the two measures are more effective when used simultaneously and that their influences do not significantly overlap. This is consistent with previous findings that the correlations between computerized text analytic results and self-judged (Pennebaker & King, 1999) or others-judged personality (Fast & Funder, 2008) are only small to modest, and each measure (i.e., computerized text analytic measure and others-judged personality measure) touches its own dimension of people's personality. As the model comparisons in Table 3 indicate, each predictor's coefficients across models do not fluctuate much. To put it differently, though our two predictors both measure

members' personality, they measure different dimensions of personality, each of which differently influences members' departure from organizations. Such patterns imply that people's linguistic style successfully reflects some dimensions of personality that even others-judged personality fails to capture, which accords with Pennebaker and King's (1999) conclusion that "linguistic style is an independent and meaningful way of exploring personality" (p. 1296). Acquaintances' judgments or peer evaluations are relatively widespread ways to obtain others-judged personality (Connelly & Ones, 2010; Holtom et al., 2008). Our findings demonstrate that computerized text analysis captures personality dimensions that this method doesn't capture and enriches our understandings of members' likelihood of remaining in the organization. Future research should examine whether such dimensions predict other organizational behaviors like job performance, coworker support, or organizational citizenship behaviors.

The second of our key findings is that the effects of B5 personality dimensions on members' length of stay vary across organizations. Our findings are generally compatible with previous ones (Choi et al., 2015; Tett et al., 2017), except that openness to experience predicted length of stay negatively in most of those studies. This exceptional finding demands careful interpretation. We suspect that this difference arises from the fact that this organization does not seek profits and only asks organizational members to achieve their individual dreams, which represents a highly permissive organizational context that recommends that members try new projects under few organizational constraints. This may explain why the association between openness to experience and departure from the organization in our study is positive and significant, unlike the negative associations found in other types of organizations (Choi et al., 2015; Judge et al., 2002). The unique organizational context we studied may also explain why the extraversion and agreeableness dimensions show positive but no significant relationships to continuance in the organization, which is different from the significant positive associations found in other studies (Choi et al., 2015; Judge et al., 2002). In other words, despite the absence of empirical evidence, our conjecture is that the difference between our and previous findings implies the importance of organizational context, which moderates the effects of B5 personality dimensions on members' departure from organizations. Because our study does not compare multiple organizations whose contexts differ from one another, we honestly admit that our interpretation focusing on "situational specificity" (Tett et al., 2017) is merely our post hoc conjecture and should be taken tentatively until crucial empirical findings are obtained in the future.

Third, our findings reconfirm the importance of composite language variables (e.g., analytic thinking, clout, and authenticity) developed by Pennebaker and his team (Jordan & Pennebaker, 2017; Kacewicz et al., 2013; Newman et al., 2013; Pennebaker et al., 2014), even when predicting people's departure from organizations. As clearly noted in Table 3, summary language variables based on Pennebaker and his team's formula using function words (e.g., pronoun, article, negation, etc.) succeeded in predicting members' length of stay with marked statistical significance. However, we found no statistically significant relationship between original LIWC2015 categories that could be related to the organizational context (e.g., achievement, affiliation, work,

etc.) and length of stay. Our findings expand the scope of the predictive validity of LIWC2015's composite language variables suggested by Pennebaker and his team (Jordan & Pennebaker, 2017; Kacewicz et al., 2013; Newman et al., 2013; Pennebaker et al., 2014) because our study demonstrates that the former findings, which are mostly derived from university students' writing while conducting laboratory experiments (e.g., Newman et al., 2003; Pennebaker et al., 2014), are similarly applicable to different demographics, to different behavioral outcomes (i.e., departure from organizations), and to different contexts (i.e., naturalistic setting).

Here, one interesting question emerges: Why do concrete context-specific word categories (e.g., achievement, affiliation, work, etc.) fail and abstract context-free composite language variables (e.g., analytic thinking, clout, authenticity) succeed to explain members' length of stay? Although contexts of textual data are not the same, two theoretical clues can be inferred from Pennebaker and his colleagues' studies (Chung & Pennebaker, 2008; Kacewicz et al., 2013; Newman et al., 2003; Pennebaker et al., 2014; Pennebaker et al., 2015; Robinson et al., 2013). On one hand, words in the original categories basically deal with a specific subject of which applicants to the organization were fully aware, implying applicants would control their words to present their desired self rather than their real self. In general, when applying to organizations, people want to present themselves as more attractive applicants and consciously try to select words related to the context of the organizations. If our afterthoughts are reasonable, the relationships between members' use of context-specific words (original categories in the LIWC2015 dictionary) are devised for their desired self, and thus, they are less likely to predict their actual behavior (i.e., organizational leave). On the other hand, composite language variables, as noted in Pennebaker and his colleagues' studies (especially Newman et al., 2003; Pennebaker et al., 2014; Robinson et al., 2013), reveal people's linguistic habits that are less likely to be consciously controlled, implying that those variables reflect the real self and better predict people's actual behaviors in the future.

Finally, time focusing, the novel composite language variable, is a novel category that could successfully explain people's behavior. The idea underlying time focusing is straightforward and simple. We believe that the presented self is clearly connected to the time on which a person focuses when presenting himself or herself. From the perspective of applicants, words in the past tense are more likely to be mobilized to describe what they have already experienced in the relevant organizational settings, but words in the future tense tend to be utilized to describe what they expect in an organizational setting not yet experienced. When applicants are recruited, members who have a concrete and realistic understanding of the field in which the organization is embedded (i.e., past time-focusing members) are less likely to feel conflict between their own self-concept and the role ascribed by the organization; however, members who have an unaccomplished and idealistic understanding of the field (i.e., future time-focusing members) are more likely to feel the chasm between their own self-concept and the behavior the organization expects from members. Although not operated in the same way for different subpopulations, Pennebaker and King (1999) reported that students who favor using past tense verbs in their writings (termed the

social past, Pennebaker & King, 1999), although not statistically significant, showed slightly higher class attendance but slightly lower activity (see Table 5 in Pennebaker & King, 1999). While tentative, our findings imply that the tense an applicant favors in their writings can be used as a linguistic marker for understanding the applicant's personality—the more past oriented, the more realistic, but the less idealistic. Differently put, past time-focusing applicants would be desirable when recruiting those who are less likely to leave the organization; however, future time-focusing applicants would be preferred when heterogeneous ideas should be introduced in the organization or the existing organizational culture should be changed. Of course, further studies should be conducted to investigate and validate our contention that self-introductions with more past tense words show that the document writer emphasizes past accomplishments by using different writing samples and different attitudinal or behavioral outcomes. However, our empirical finding deserves consideration as an interesting clue that time focusing in writings may explain people's personality and behavior in an organization.

Several limitations of this study are worth noting. First, it is questionable that Gosling et al.'s (2003) 10-item personality inventories succeed to capture members' personality. Even Gosling et al. warned that short personality measures can "not be used in place of established multi-item instruments" (p. 525). Despite empirical findings showing that Gosling et al.'s 10-item measures are obviously better than the five-item measures (Credé et al., 2012, p. 884), longer personality items are much more desirable. Additionally, given that our sample size is small ($N = 49$), readers are advised to interpret our findings regarding others-judged personality measures with special care. For example, unexpected findings discussed as the second point above would not reflect a difference in organizational contexts but could be merely by-products of poor personality measures. Uncertainty about our findings and interpretations wait for further studies comparing performance of a variety of others-judged personality measures.

Second, a variety of causes may influence short length of stay in this organization. For example, some members may leave the organization because of situational reasons, such as disappointment that they do not feel the organization's mission is being attained, or relationship conflicts with other members, while others may depart the organization for individual reasons, such as deciding to pursue a job in a for-profit company. Obviously, our study does not take these mixed motives into account. As shown in Moore et al. (2017), further mediation analyses are needed to explain the role of diverse motives mediating the relationship between both personal and linguistic features and departure from organizations.

Finally, for computerized text analysis, our study solely relies on LIWC2015 and does not examine other programs using different lexicons (e.g., General Inquirer, AFINN, EmoLex) or other supervised or unsupervised machine-learning algorithms. Like any computerized text analysis program, LIWC2015, despite its acknowledged reliability and validity, is only one method to analyze text data, although LIWC2015's performance is similar to those of other programs (Gonçalves et al., 2013; Schwartz et al., 2013). Differently put, it is reasonable to expect that other text analysis programs

or algorithms may provide different outputs, and thus different interpretations could be obtained. As other studies have shown (Gonçalves et al., 2013; Schwartz et al., 2013), a mixture of results from available text analytic tools show better predictive validity than results obtained through a single tool (or a few of them). Given that LIWC provides more diverse categories than other lexicon-based programs that report a few simple categories (e.g., sentiments like positive–negative, a set of discrete emotions), and composite language variables unique in LIWC (e.g., clout, authenticity) are very effective (see results in Table 3), we believe that other lexicon-based programs will provide more limited information than LIWC. Unfortunately, given that the corpus size in our writing sample is small, supervised or unsupervised machine-learning algorithms cannot be reliably applied. However, a combination of available computerized text analytic results would unquestionably capture the underlying dimensions of personality inferred from text data more extensively.

Despite some limitations, our findings demonstrate that unobtrusive measures, computerized text analytic results, as well as others-judged B5 personality, validly predict organization members' behavior in the form of departure from an organization. More important, we found that computerized text analysis, along with others-judged personality, helps expand the meanings of personality and its influences on organizational behavior. Despite recent advances in text-mining techniques, applications have only just begun (Kobayashi et al., 2018). As evidenced here, organizational research would provide further progress by utilizing reliable and valid measures harnessing computerized analysis of traditionally unexplored interpreted text data stored in organizations.

Acknowledgments

We appreciate the editor and two anonymous reviewers who gave us insightful and productive comments on this work. We also appreciate every member of the organization and the management committee for their help.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2016S1A3A2925033).

ORCID iD

Jennifer Ihm https://orcid.org/0000-0001-7039-4162

Note

1. Other categories not mentioned in this section were dropped from our analyses. First, grammar-related categories (e.g., auxiliary verbs, common verbs, common adverbs) were

not considered in the analyses because they are mere functional words lacking socio-psychological implications. Second, several categories (e.g., personal pronouns, cognitive processes) were not included in our analyses because those categories were used to construct Pennebaker et al.s' (2015) summary language variables (e.g., analytic thinking, clout, authentic). Given that Pennebaker et al.s' summary language variables were included in our regression equations, it is reasonable to exclude those categories to avoid repetition. Third, categories whose operational definitions are not related to our research purpose (e.g., religion, death, motion, space, time) were not considered in our model. Finally, categories dealing with informal language (e.g., swear words, nonfluencies, fillers) were not considered because members' self-introductions are formal documents.

References

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, *18*, 63-68. https://doi.org/10.1016/j.cobeha.2017.07.017

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge University Press.

Choi, D., Oh, I. S., & Colbert, A. E. (2015). Understanding organizational commitment: A meta-analytic examination of the roles of the five-factor model of personality and culture. *Journal of Applied Psychology*, *100*(5), 1542-1567. https://doi.org/10.1037/apl0000014

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, *42*(1), 96-132. https://doi.org/10.1016/j.jrp.2007.04.006

Connelly, B. S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092-1122. https://doi.org/10.1037/a0021212

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874-888. https://doi.org/10.1037/a0027403

Dess, G. G., & Shaw, J. D. (2001). Voluntary turnover, social capital, and organizational performance. *Academy of Management Review*, *26*(3), 446-456. https://doi.org/10.5465/AMR.2001.4845830

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*(2), 334-346. https://doi.org/10.1037/0022-3514.94.2.334

Felps, W., Mitchell, T. R., Hekman, D. R., Lee, T. W., Holtom, B. C., & Harman, W. S. (2009). Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, *52*(3), 545-561. https://doi.org/10.5465/AMJ.2009.41331075

Frumkin, P. (2002). *On being nonprofit: A conceptual and policy primer*. Harvard University Press.

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013, October). Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on online social networks* (pp. 27-38). Association for Computing Machinery. https://doi.org/10.1145/2512938.2512951

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528. https://doi.org/10.1016/S0092-6566(03)00046-1

Gottschalk, L. A. (1997). The unobtrusive measurement of psychological states and traits. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 117-129). Lawrence Erlbaum.

Hall, M., & Caton, S. (2017). Am I who I say I am? Unobtrusive self-representation and personality recognition on Facebook. *PLOS ONE*, *12*(9), e0184417. https://doi.org/10.1371/journal.pone.0184417

Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *Academy of Management Annals*, *2*(1), 231-274. https://doi.org/10.1080/19416520802211552

Jordan, K. N., & Pennebaker, J. W. (2017). The exception or the rule: Using words to assess analytic thinking, Donald Trump, and the American presidency. *Transnational Issues in Psychological Science*, *3*(3), 312-316. https://doi.org/10.1037/tps0000125

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*(3), 530-541. https://doi.org/10.1037/0021-9010.87.3.530

Kacewicz, W., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2013). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125-143. https://doi.org/10.1177/0261927X1350265

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

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage.

Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior*, *61*(1), 20-52. https://doi.org/10.1006/jvbe.2001.1842

Moore, C., Lee, S. Y., Kim, K., & Cable, D. M. (2017). The advantage of being oneself: The role of applicant self-verification in organizational hiring decisions. *Journal of Applied Psychology*, *102*(11), 1493-1513. https://doi.org/10.1037/apl0000223

Naim, I., Tanveer, M. I., Gildea, D., & Hoque, M. E. (2015). Automated prediction and analysis of job interview performance: The role of what you say and how you say it. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *1*(1), 1-6. https://.doi.org/10.1109/FG.2015.7163127

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665-675. https://doi.org/10.1177/0146167203029005010

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296-1312. https://doi.org/10.1037/0022-3514.77.6.1296

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: An LIWC analysis. *Journal of Language and Social Psychology*, *32*(4), 469-479. https://doi.org/10.1177/0261927X13476869

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, *71*(1), 23-65. https://doi.org/10.1111/peps.12226

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, *63*(3), 506-516. https://doi.org/10.1207/s15327752jpa6303_8

Schlenker, B. R., Chambers, J. R., & Le, B. M. (2012). Conservatives are happier than liberals, but why? Political ideology, personality, and life satisfaction. *Journal of Research in Personality*, *46*(2), 127-146. https://doi.org/10.1016/j.jrp.2011.12.009

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Martin, E. P., Seligman, M. E. P, & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Shumate, M., Atouba, Y., Cooper, K. R., & Pilny, A. (2014). Two paths diverged: Examining the antecedents to social entrepreneurship. *Management Communication Quarterly*, *28*(3), 404-421. https://doi.org/10.1177/0893318914538561

Tett, R. P., Hundley, N. A., & Christiansen, N. D. (2017). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology*, *10*(3), 421-456. https://doi.org/10.1017/iop.2017.26

Thalmeyer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief and medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment*, *23*(4), 995-1009. https://doi.org/10.1037/a0024165

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*(3), 546-558. https://doi.org/10.1037//0022-3514.78.3.546

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363-373. https://doi.org/10.1016/j.jrp.2010.04.001

Yee, T. W. (2010). *VGAM: Vector generalized linear and additive models (Version 1.1-2)*. [Computer software]. https://CRAN.R-project.org/package=VGAM

Zeldow, P. B., & McAdams, D. P. (1993). On the comparison of TAT and free speech techniques in personality assessment. *Journal of Personality Assessment*, *60*(1), 181-185. https://doi.org/10.1207/s15327752jpa6001_14

Author Biographies

Young Min Baek (PhD in Communication, University of Pennsylvania) is an associate professor in the Department of Communication at Yonsei University. His research interests are the role of human texts in social interactions, relying on quantitative research methods.

Jennifer Ihm (PhD in Communication, Northwestern University) is an assistant professor in School of Media and Communication at Kwangwoon University. She uses a network approach to understand the relationships among stakeholders in the organizational context.