The kernel of truth in text-based personality assessment

Koutsoumpis, Antonis; Oostrom, Janneke; Holtrop, Djurre; Van Breda, Ward; Ghassemi, Sina; de Vries, R.E.

Published in:
Psychological Bulletin

DOI:
10.1037/bul0000381

Publication date:
2022

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

2022, Vol. 148, Nos. 11–12, 843–868
https://doi.org/10.1037/bul0000381

The Kernel of Truth in Text-Based Personality Assessment: A Meta-Analysis of the Relations Between the Big Five and the Linguistic Inquiry and Word Count (LIWC)

Antonis Koutsoumpis[1], Janneke K. Oostrom[2], Djurre Holtrop[3], Ward van Breda[1], Sina Ghassemi[1], and Reinout E. de Vries[1]
[1] Department of Experimental and Applied Psychology, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam
[2] Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam
[3] Department of Social Psychology, Tilburg School of Social and Behavioral Sciences, Tilburg University

The Linguistic Inquiry and Word Count (LIWC) is a popular closed-vocabulary text analysis software program that is used to understand whether individuals' use of linguistic categories (i.e., word categories, such as negative affect) depends on their personality traits. Here, we present the first meta-analysis of the relations between the Big Five personality traits and 52 linguistic categories of the English language. Across 31 eligible samples (n = 85,724), the results showed that (a) self-reported personality traits are significantly correlated with linguistic categories, but the effect sizes are relatively small (the strongest effect sizes between the Big Five and linguistic categories ranged from $|\rho| = .08$ to .14, and the 52 LIWC categories explained on average 5.1% of personality variance); (b) observer-reported personality traits are significantly correlated with linguistic categories, with the effect sizes being small-to-medium ($|\rho| = .18–.39$, explaining 38.5% of personality variance); (c) 20 linguistic categories (out of 260; 5 Personality Traits × 52 LIWC Categories) correlated both with self- and observer-reported personality traits (the "kernel of truth" in linguistic markers of personality); and (d) 10 study, sample, and task characteristics significantly moderated the correlations of the linguistic categories with personality traits, showing that the effect sizes were mainly stronger for longer texts and older LIWC versions, among others.

Public Significance Statement

This meta-analysis identifies the linguistic categories (i.e., word categories, such as negative affect) that individuals use depending on their personality traits, as well as the linguistic categories that other people use to draw personality inferences. Individuals indeed use specific linguistic categories depending on their personality traits and others use specific linguistic categories to draw personality inferences, but those relations are dependent on study and tasks characteristics (e.g., text length, Linguistic Inquiry and Word Count version).

Keywords: meta-analysis, LIWC, Big Five personality, cue validity, cue utilization

Supplemental materials: https://doi.org/10.1037/bul0000381.supp

Antonis Koutsoumpis https://orcid.org/0000-0001-9242-4959

The authors would like to gratefully thank Drago Savic, Elmi Zwaan, Julia Houben, and Loet Coenen for their help on coding the data.

This research is part of a PhD project funded by LTP Business Psychologists. The authors have no conflicts of interest to disclose.

Raw and processed data, scripts, and Supplemental Material are available at https://osf.io/7vszn/.

Correspondence concerning this article should be addressed to Antonis Koutsoumpis, Department of Experimental and Applied Psychology, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081BT Amsterdam, The Netherlands. Email: a.koutsoumpis@vu.nl

Language is a unique feature of the human race. Through language, we express our feelings, ideas, and emotions. As such, our personality might also be revealed in our language, through the words we use (Tausczik & Pennebaker, 2010). Assessing personality from language is something that many people tacitly do, for instance, when trying to understand the personality of a writer they have never met. It is also something that professionals explicitly do, for instance, when trying to understand the personality of a job candidate from their resume. In the past few decades, the advent of automatic text analysis software allowed researchers to systematically study the relations between personality traits and language in a more objective and quantitative way. The scientific study of personality traits from language has led to the emergence of text-based personality assessment methods (see Eichstaedt et al., 2021; Park et al., 2015) with two main goals: to *predict* personality traits from text (in predictive models, personality is the dependent variable) and *explain* why and how words relate to personality traits (see Mehl et al., 2006; Yarkoni & Westfall, 2017).

The goal of the present article is to contribute to the second goal. Already, research has concluded that, indeed, personality traits can be accurately predicted from text (Moreno et al., 2021). However, many text analysis methods operate as "black boxes," making it unclear how words relate to personality traits (Yarkoni & Westfall, 2017). This has significant repercussions both for researchers interested in the behavioral indices of personality, as well as for practitioners who are often incentivized to make text-based personality assessment more transparent. Therefore, the present meta-analysis focuses on the explainability of text-based personality assessment by uncovering the most important linguistic markers of personality. We use the term "linguistic" (instead of "verbal") markers to stress the fact that we will focus on linguistic categories rather than on individual words. To this end, we focus on the most popular text analysis tool in personality research, the Linguistic Inquiry and Word Count software program (LIWC; pronounced "Luke"; Boyd et al., 2022; Pennebaker & Francis, 1999; Pennebaker et al., 2001, 2007, 2015), which has been extensively used over the past 20 years to explore the relations between personality traits and linguistic categories.

Building on Brunswik's (1955) lens model and the self–other asymmetry model (Vazire, 2010), and focusing on 52 LIWC categories and the Big Five personality traits, the present meta-analysis has four main goals. First, to identify the linguistic markers (or linguistic categories/cues; throughout the article, we use the terms interchangeably) through which individuals may express their self-reported personality traits (cue validity). Second, to identify the linguistic markers that other people use to infer personality traits (cue utilization). Third, to explore whether there is a "kernel of truth" in the linguistic markers of personality, that is, whether the linguistic markers that other people use to make personality inferences are in line with the linguistic markers that individuals use as a manifestation of their (self-reported) personality. Fourth, to account for possible heterogeneities in effect sizes, by testing the effect of 10 moderators related to task, study, and sample characteristics.

Text-Based Personality Assessment and the LIWC

Text-based personality assessment has a relatively short history. It kicked off in the middle of the 20th century with tedious hand-coded methods and has progressively moved to automatic text-processing software packages. Currently, it is being taken over by artificial intelligence methods trained on big data (Banks et al., 2018; Iliev et al., 2015; Tausczik & Pennebaker, 2010). There are two popular automatic text analysis approaches: the *open*- and *closed*-vocabulary approach, each one occupying a different niche and serving different purposes (for a comprehensive theoretical review and quantitative comparison of the two methods, see Eichstaedt et al., 2021).

The open-vocabulary approach (Schwartz et al., 2013) originated in the field of computer science with the goal to maximize predictive accuracy. This approach is data-driven, bottom-up, atheoretical, and typically applies machine learning techniques to find the optimal—often nonlinear—relations between text and personality traits (Banks et al., 2018; Eichstaedt et al., 2021). The open-vocabulary approach can employ *nontransparent* or *transparent* methods (see also Eichstaedt et al., 2021). Nontransparent methods, such as word embedding models (e.g., Word2Vector; Mikolov et al., 2013) and deep learning language models (e.g., Bidirectional Encoder Representations from Transformers; Devlin et al., 2019), take as input single or multiwords (e.g., *n*-grams) or even sentences and automatically generate the desired outcome (i.e., personality traits). These methods are considered "black boxes" because they often contain a large number of parameters (millions in the case of deep learning models) and the functions that transform these text parameters into personality traits can be difficult to explain. Hence, these nontransparent methods are typically preferred when someone is interested in maximizing explained variance instead of interpretability. On the other hand, transparent methods (e.g., Latent Dirichlet Allocation, Differential Language Analysis, Bags of Words) use the same input as the nontransparent methods, but they combine this information in a straightforward and accessible way, making clear what are the individual words (but also *n*-grams, emoticons, punctuation, topics) used to predict each personality trait (Banks et al., 2018; Eichstaedt et al., 2021). Although this latter method provides granular insights and is suitable to study the verbal manifestations of personality, a meta-analysis of studies using the transparent method is currently unrealistic since the number of published articles with this method is still very limited (for two notable exceptions, see Kern et al., 2014 and Yarkoni, 2010).

The closed-vocabulary approach is top-down and theoretically driven (Eichstaedt et al., 2021). It starts with predefined sets of keywords ("closed dictionaries") that, according to experts and/or previous research, sufficiently describe a construct (e.g., positive emotions). Then, a text is processed, and every time, the software comes across a word that is part of the internal dictionary it counts a "hit." The output score indicates the extent to which the text contains the construct of interest (e.g., the number of positive emotion words). Examples of the closed-vocabulary approach are the General Inquirer (Stone et al., 1966), the Medical Research Council psycholinguistic database (MRC; Wilson, 1988), the Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, Openness to Experience (HEXACO) text-to-personality technique (HTTP; Holtrop et al., 2022), but by far the most popular approach is LIWC, which, at the time of writing this article, counts a little more than 10,000 citations in Google Scholar. Since the goal of the present meta-analysis is to identify the linguistic markers of personality, and since a large number of LIWC studies has already been accumulated, we opted to meta-analyze the results of LIWC research.

LIWC was originally published in 1999 by Pennebaker and Francis. It is a text analysis software that automatically processes the words in a given text and calculates the percentage (or number) of words that fall into a number of predefined linguistic categories, ranging from psychological (e.g., emotions, motivation, cognitions) and biological (e.g., seeing, hearing, eating) processes to grammatical structures (e.g., pronouns, verbs) and punctuation. Since its introduction, LIWC has seen five versions (LIWC1999, LIWC2001, LIWC2007, LIWC2015, and LIWC2022; Boyd et al., 2022; Pennebaker & King, 1999; Pennebaker et al., 2001, 2007, 2015), each subsequent version having an updated (and richer) dictionary and a restructure of its linguistic categories. For instance, 2,300, 4,500, 6,400, 12,000 words and word stems, and 84, 80, 93, and 109 output categories are included in LIWC2001, LIWC2007, LIWC2015, and LIWC2022, respectively, which cover 80%–90% of all words people commonly use in writing and speech.

LIWC and Personality Traits

The purported relation between self-reported personality traits and verbal behavior is based on the assumption that individuals do not pick their words at random. Instead, words are assumed to be the behavioral manifestation of one's personality, similar to other behavioral cues, including one's voice (Borkenau & Liebler, 1992; Stern et al., 2021), facial expressions (Biel et al., 2013; Hickman et al., 2022), daily behavior (Mehl et al., 2006; Tackman et al., 2020), online behavior (Azucar et al., 2018), or selection of environments (Gosling et al., 2002; Zettler et al., 2020). In the case of LIWC, the linguistic categories are considered verbal behavioral manifestations; therefore, previous research has used these categories as indicators of self-reported personality.

Empirical findings seem to support the idea that LIWC categories can be used as linguistic markers for each of the Big Five personality traits. For example, individuals high on self-reported Extraversion express their social disposition by using more positive emotion and social-related words (Chen et al., 2020), and individuals high on self-reported Conscientiousness express their prudence by restricting themselves from profanity or negative emotions (Tackman et al., 2020). Further examples include the negative relation between Agreeableness and the use of swear words (Schwartz et al., 2013), Openness to Experience and more complex language use (Baddeley & Singer, 2008; Gill et al., 2006), and Emotional Stability and less frequent use of negative emotions (Biel et al., 2013; Hirsh & Peterson, 2009).

One explanation for the relation between self-reported personality traits and linguistic categories is that word frequency reflects what individuals are paying attention to (Baldwin, 1942; Boyd & Schwartz, 2021; Stone et al., 1966). That is, individuals thinking, for example, about "death, sex, money, or friends will refer to them in their writing or conversation" (Tausczik and Pennebaker, 2010, p. 30; see same reference for a summary of empirical findings on "words as attention"). Some more support for the idea that words reflect attention comes from studies on thought inhibition. For instance, Wegner et al. (1987) showed that when participants are requested to verbalize a stream of consciousness without thinking about a "white bear," they failed miserably. Similar experiments in the area of inhibition of cognition illustrate that words can show what participants are paying attention to (for a review of similar findings, see Purdon & Clark, 2000). Currently, the "words as attention" is one of the most commonly used language analysis approach[1] (Boyd & Schwartz, 2021; Tausczik & Pennebaker, 2010). Under this approach, individuals who self-report, for instance, high Extraversion scores will tend to use more "social" words. This happens presumably because social attention is a core element of Extraversion. Hence, extraverted individuals will tend to pay attention to social-related topics and express these thoughts through social-related words (Chen et al., 2020).

Additionally, some linguistic categories are utilized by others to draw personality inferences, even though this line of research has received less attention. For instance, individuals who produce longer texts are perceived as more extraverted (Baek & Ihm, 2020), and individuals who use more positive emotions are perceived as more agreeable[2] (Biel et al., 2013). One explanation for the relation between linguistic categories and observer reports of personality is provided by the realistic accuracy model (RAM; Funder, 1995, 2012). According to RAM, in order for observers to draw correct personality inferences (i.e., in terms of self–other agreement), they need to correctly detect and utilize (i.e., interpret) the behavioral cues (e.g., linguistic categories) of the target person. The more behavioral cues are available (and the more relevant they are to a personality trait), the more observers will draw correct personality inferences. Under this model, individuals who, for instance, produce longer texts are perceived as more extraverted because observers detect and interpret text length as indicative of Extraversion. All in all, linguistic markers of personality have been observed across different contexts and tasks, including social media (Schwartz et al., 2013; Sumner et al., 2012), essays (Abe, 2018; Pennebaker & King, 1999), text messages (Holtgraves, 2011), blogs (Yarkoni, 2010), YouTube video logs (Biel et al., 2013), reported dreams (Hawkins & Boyd, 2017), job interviews (Holtrop et al., 2019), chats (Sandy, 2013), and everyday life utterances (Mehl et al., 2006; Tackman et al., 2020).

So far, there has been one narrow meta-analysis of the relations between Extraversion and two LIWC categories, showing that individuals high on self-reported Extraversion use more positive emotion (r = .07) and social-process words (r = .08; Chen et al., 2020). However, to date, there is no comprehensive meta-analysis of the relations between a full model of personality (e.g., Big Five, HEXACO) and the entire set of LIWC categories, neither do we know whether the same linguistic categories are used by others when drawing personality inferences. If we want to know whether LIWC categories can be used as personality indicators, it is imperative to know the extent to which *all* personality traits relate to *all* LIWC categories, both when personality is measured using self-reports and observer reports. The present work is, as far as we know, the first attempt to meta-analyze the empirical evidence on the relations between 52 LIWC categories and the personality traits of Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (throughout the article, Neuroticism has been reverse-scored reflecting Emotional Stability), as described by the Big Five or five-factor model of personality (Goldberg, 1990; McCrae & John, 1992).

Cue Validity, Cue Utilization, and the Kernel of Truth of Linguistic Categories

We describe the relations between LIWC categories and personality traits following the framework of Brunswik's lens model (Brunswik, 1955; Karelaia & Hogarth, 2008; Osterholz et al., 2021). The lens model offers an explanation on why and how perception is accurate—not necessarily in relation to personality traits (Brunswik, 1955). According to the lens model, individuals (the target) express their traits through behavioral cues (e.g., words, gestures, loud voice) and other people use those cues to draw inferences about the target. In a sense, those cues operate as a "lens" through which other people draw their conclusions about the target—hence, the name "lens" model. Calculating the weight of each individual cue, it is possible to quantify the extent to which each behavioral cue contributes to accurate judgment (or, in the case of self- and observer reports of personality, self–other agreement).

[1] Note that the "words as attention" is neither the only nor the only valid approach to word frequency (Boyd & Schwartz, 2021). Other approaches suggest that words show how someone is processing information (Tausczik & Pennebaker, 2010, p. 31) or that word frequency measures "the 'intensity of an attitude' or 'amount of concern'" (Stone et al., 1966, p. 32).

[2] Note that others might make their judgment following a "holistic" and/or implicit assessment, taking into account not only the words but also the style, the structure, and other meta information of the text.

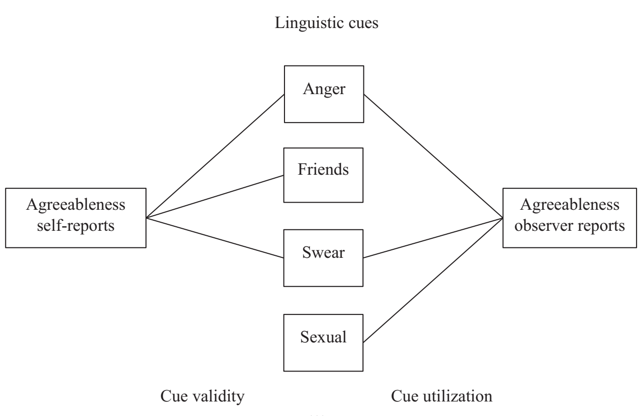
Drawing on Brunswik's lens model, we focus on three open questions. The first one focuses on the linguistic cues through which individuals express their self-reported personality traits—in Brunswikian terms, "cue validity." Cue validity describes the extent to which individuals express their self-reported personality traits through cues (in our case, linguistic categories) and is quantified through the correlation between self-reported personality traits and those cues (see Figure 1). For instance, individuals high on self-reported Conscientiousness express their self-discipline by refraining from swearing (Yarkoni, 2010), and individuals high on self-reported Extraversion express their disposition by using words that reveal positive emotions (Chen et al., 2020). Thus, the first open question is to identify the valid linguistic cues of personality traits. Note that the term "valid cues" might suggest that certain linguistic categories indicate people's "true" personality. However, according to the lens model's operationalization, valid cues simply indicate that linguistic categories are correlated with self-reported personality traits, and this is how we use it in the present article (we do not assume that valid cues capture "true" personality).

The second open question is about the linguistic cues that other people use when making personality inferences, or "cue utilization." Cue utilization describes the extent to which other people infer one's personality traits based on the available (linguistic) cues, and it is quantified through the correlation between observer reports of personality and LIWC categories. For instance, individuals who swear on YouTube vlogs are perceived by strangers as less agreeable (Biel et al., 2013), and individuals who use words longer than six letters are perceived as higher on Openness to Experience (Mehl et al., 2006). Thus, the second open question is to identify the utilized linguistic markers of personality traits.

Frequently, others infer personality traits based on the same linguistic cues that also correlate with self-reported personality traits.

Figure 1

Brunswik's Lens Model and the Linguistic Markers of Personality



Note. Linguistic markers of personality (e.g., Agreeableness) operationalized under the lens model framework; the lines between self-reports and linguistic cues represent significant correlations (cue validity); the lines between observer reports and linguistic cues represent significant correlations (cue utilization); the linguistic cues that are significantly correlated both with self- and observer reports represent the linguistic cues that constitute the "kernel of truth" of Agreeableness; "…" indicates that there might exist more than four linguistic markers of Agreeableness.

For instance, individuals who swear more in their daily life self-identify as low on Conscientiousness and they are perceived by others in the same way (Mehl et al., 2006). However, sometimes others might draw inferences about the personality traits of an individual based on cues that do not necessarily correlate with self-reported personality scores. For instance, students who use family-related words in their student essays are perceived as extraverted, even though the correlation between family-related words and students' self-reported Extraversion might be nonsignificant (Mairesse et al., 2007). The present meta-analysis allows us to explore whether there is a "core"—or a set of linguistic categories—through which individuals express their (self-reported) personality traits, while at the same time others infer those same personality traits (in Brunswikian terms, whether cue utilization is based on valid linguistic cues). We define the overlap between cue validity and cue utilization as the "kernel of truth" in linguistic markers of personality. Here, again, the "kernel of truth" does not describe "true" personality but simply the overlap between the linguistic categories that correlate both with self- and observer reports of personality. Thus, the third open question of the present meta-analysis is to identify the linguistic categories that are part of the kernel of truth in text-based personality assessment.

The importance of the kernel of truth in personality assessment has been highlighted by RAM (Funder, 1995, 2012), according to which the accuracy of personality inferences increases when others utilize valid cues (i.e., the "kernel of truth"; Letzring & Funder, 2021). Those valid cues are described as "good information" (Beer, 2021; Letzring & Funder, 2021), and they might refer to quantity and/or quality. Previous research already suggests that higher levels of information *quantity* (e.g., level of acquaintance, information richness; Borkenau & Liebler, 1992; Connelly & Ones, 2010) increase the accuracy of personality judgment (operationalized as self–other agreement). However, the role of information *quality* is less well understood (e.g., disclosing information about personal values compared to facts does not affect overall personality accuracy judgments; Beer & Brooks, 2011). Recently, Beer (2021) lamented the lack of research on the quality of information that would lead to more accurate personality judgment. To this end, if the kernel of truth idea is verified by the present meta-analytic results, it would shed some light on the linguistic cues (the quality of information; e.g., text length, affect-related words, etc.) that others use to make accurate personality judgments.

The Self–Other Knowledge Asymmetry Model

Can we make any predictions about the strength of effect sizes for cue validity and cue utilization? According to the self–other knowledge asymmetry model (SOKA; Vazire, 2010), the self (compared to others) has more information to judge *internal* traits or behavioral expressions (e.g., feelings, thoughts), compared to *external* traits or behavioral expressions (e.g., voice, gestures). This happens because the self has abundant information about their own internal thoughts and feelings but less information about (or access to) their own external behaviors (e.g., facial expressions; Beer, 2021; Nisbett et al., 1973). As a result, the self tends to privilege internal over external information (Vazire, 2010). On the other hand, others have less information about the internal thoughts and feelings of the target (the self), but they have better access to the external behaviors of the target. This asymmetry in information between self and others causes an asymmetry in accuracy when assessing internal and external traits (see also RAM; Letzring & Funder, 2021).

Indeed, previous research suggests that when the outcome variables are internal traits (e.g., self-esteem; Vazire, 2010) and less overt behaviors (e.g., dishonesty, arguing; Thielmann et al., 2017; Vazire & Mehl, 2008), they are better predicted by self- rather than observer reports of personality. However, when the outcome variables are overt behaviors—including audio features (Borkenau & Liebler, 1992), linguistic categories (Mairesse et al., 2007), daily behaviors (Mehl et al., 2006), verbal and nonverbal behavior in online interviews (Hickman et al., 2022)—those overt behaviors are better predicted by observer- rather than self-reported personality scores; and those overt behavioral cues explain larger variance in observer (rather than self-) reports of personality traits (Gifford, 1994). In the case of LIWC, the linguistic categories refer to verbal behavior, which, by definition, is overt behavior. Therefore, drawing on the SOKA model, we hypothesize that (a) the effect sizes and (b) the explained personality variance will be larger for observer reports (cue utilization) rather than for self-reports (cue validity) of personality traits.

Moderators

The research on linguistic markers of personality does not always provide consistent results. The possible role of moderators in this relation is generally understudied, something that has not escaped the attention of previous researchers (Pennebaker et al., 2003). Recently, there have been some efforts to address this gap. The meta-analysis on the relations between self-reported Extraversion and two LIWC categories (Chen et al., 2020) suggested large heterogeneity in effect sizes, which was explained by moderators related to task type (e.g., synchronous vs. asynchronous tasks) and LIWC version. Similar heterogeneity in effect sizes was also reported in a meta-analysis of the relations between the first-person singular pronoun LIWC category (e.g., I, me, mine) and self-reported depression (Edwards & Holtzman, 2017), even though none of the five tested moderators (e.g., demographics, task type) accounted for this heterogeneity.

So far, the research on heterogeneity in linguistic markers of personality has been restricted to two LIWC categories (positive emotions, social processes) and one self-reported personality trait (Extraversion). However, other relations between personality traits and linguistic categories might also be affected by moderators. For instance, in spoken language, self-reported Extraversion is more strongly correlated with the total amount of produced words (r = .29), but in written language, this relation is nonsignificant (r = .03; Mairesse et al., 2007), suggesting that the mode of text (spoken or written) might affect this relation. Therefore, the third goal of the present meta-analysis was to test the effect of several moderators on the relation between the 52 LIWC categories and the Big Five personality traits. To account for any heterogeneity in effect sizes, we considered 10 moderators: five categorical (LIWC version, written–spoken language, language formality, task synchronicity, sample composition) and five continuous (sample size, text length, age, percentage women, year of publication).

Moderator Predictions Based on Trait Activation Theory

Even though most moderators will be investigated from an exploratory perspective, we make explicit predictions for two categorical moderators. More specifically, the moderators of language formality and task synchronicity[3] describe situations where participants are expected to behave in a certain way (e.g., formal vs. informal, dialogue vs. monologue). The significance of situations and their interaction with personality traits in predicting behavior have been highlighted by the trait activation theory (TAT; Tett & Guterman, 2000; Tett et al., 2021). According to TAT, for a trait to be activated, one must be exposed to situations that provide cues for expressing the targeted trait. For example, if someone wants to observe disagreeableness, it should be in a context where disagreeable behavior is possible (e.g., an informal rather than a formal situation).

Empirical support for TAT comes from studies that employ correlational designs (e.g., Holtrop et al., 2022) mainly in the area of assessment centers (Haaland & Christiansen, 2002; Lievens et al., 2006; Tett & Guterman, 2000). When interactional designs are employed, the findings are less consistent and the effects sizes are small (Koutsoumpis & De Vries, 2022; Sherman et al., 2015). Regarding linguistic categories, the meta-analysis by Chen et al. (2020) provided some empirical support for TAT, showing that the level of task synchronicity moderated the correlation between self-reported Extraversion and positive emotion words, such that individuals high on self-reported Extraversion used more positive emotion words in asynchronous but not in synchronous tasks. In the present meta-analysis, we will retest this moderator effect.

Another variable that might moderate the relation between personality traits and linguistic categories is task formality. In particular, individuals low on self-reported Agreeableness tend to swear more (Schwartz et al., 2013), but this relation seems to be affected by the level of situation formality. In informal environments, like social media or video logs, less agreeable people have been found to use more swear words (Biel et al., 2013; Qiu et al., 2012), but in formal self-presentation vignettes—for instance, when applying for accommodation—swear words become irrelevant in predicting self-reported Agreeableness (Baek & Ihm, 2020). Therefore, drawing on the TAT framework, we hypothesized that individuals low on self-reported Agreeableness will tend to use more swear-related words in informal, rather than formal, situations.

Method

Eligibility Criteria

Studies were eligible for inclusion if they (a) used any LIWC version (except for LIWC2022 that was published after the literature search), (b) measured all personality traits of Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (or Neuroticism) using a validated measure of the Big Five or HEXACO model, (c) measured personality using self- or observer reports, (d) analyzed text > 40 words (according to LIWC's manual recommendations[4]), (e) analyzed original data sets (not a

[3] Synchronous tasks are two-way interactions, part of face-to-face or online discussions with real-time feedback (e.g., daily discussion, job interview), whereas asynchronous tasks are one-way interactions lacking the option or expectation for real-time feedback (e.g., student essays, blogs). In other words, synchronous tasks are part of a dialogue and asynchronous tasks are part of a monologue.

[4] LIWC can analyze text of any length. However, by the time of the literature search, the LIWC website suggested that a number of 40 or 50 words is necessary for meaningful analyses. We liberally used the lower threshold of 40 words. Note, however, that this information has currently changed to 25–50 words (https://liwc.app/help/howitworks) and it might further change in the future.

reanalysis of existing ones), (f) tested the relations between personality and the full set of LIWC categories (studies testing only a subset of LIWC categories were excluded unless a full data set was available, since missing values may not indicate weak but rather not tested effects), (g) included healthy, English-speaking adults, and (h) provided an effect size in the text or elsewhere.

The three LIWC versions considered in the present meta-analysis (LIWC2001, LIWC2007, LIWC2015) contain dissimilar numbers of linguistic categories and dictionary entries.[5] In the present meta-analysis, only the 52 LIWC categories that are similar across all three versions were included (see Table 1). One exception was the linguistic category of auxiliary verbs that was not included in LIWC2001. Nevertheless, we decided to include auxiliary verbs because of their significant role in language as function words (thus, for auxiliary verbs, results will be based only on LIWC2007 and LIWC2015 versions). Finally, the words per sentence (WPS) and punctuation marks categories were excluded, since this information is typically lost when text is preprocessed.

Systematic Literature Search

The present meta-analysis followed the reporting guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Page et al., 2021). A systematic literature search was conducted between September 8 and 10, 2020, in three types of databases: psychological (Science Direct, Web of Science, PLOS ONE, APA PsycNet, PubMed), computer science (IEEE Xplore, Association for Computing Machinery), and dissertation and thesis (ProQuest) databases. The search was conducted on the title, abstract, and keywords of the articles, and the terms used in the search were ("personality" OR "big-5" OR "big-five" OR "HEXACO" OR "extraversion" OR "conscientiousness" OR "openness" OR "emotionality" OR "honesty" OR "agreeableness") AND ("LIWC" OR "linguistic" OR "word*" OR "text" OR "twitter" OR "Facebook" OR "verbal" OR "narrative" OR "blog" OR "social media") AND ("Pennebaker"; searched across the entire article). All databases were searched from 1999 onward, the publication year of LIWC (Pennebaker & Francis, 1999). Extra records were searched in the reference list of eligible studies, Google Scholar, and by contacting the authors of eligible articles for unpublished data sets. After removing duplicate records, the abstracts of 429 articles were screened for eligibility by the first author, followed by screening the full text of a subset of 124 articles. This process returned 26 eligible studies (31 eligible samples) with a total sample size of 85,724 participants. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart in Figure 2 illustrates the search and screening process.

Almost all studies measured personality using either the Big Five or five-factor model. The two models greatly overlap, and hence, the two terms (Big Five, five-factor) are used interchangeably. One study (Holtrop et al., 2019) used the HEXACO model. In the HEXACO model, the domains of Openness to Experience, Conscientiousness, and Extraversion are operationalized similarly to the Big Five, but the domains of Agreeableness and Neuroticism (Emotionality in HEXACO) are operationalized somewhat differently (Ashton & Lee, 2005). However, in the present study, HEXACO's Agreeableness and Emotionality were treated as similar to Big Five's Agreeableness and (reverse) Neuroticism. Results for HEXACO's Honesty–Humility will not be reported, since they were based on a single study.

Data Coding and Data Transformation

Pearson r was used as an effect size. The correlation coefficients were first transformed to Fisher's z values, then (weighted) averaged, and then back-transformed to Pearson r coefficients, following Borenstein et al.'s (2021) recommendations. One study (Yarkoni, 2010) used Spearman's rho as an effect size, which was converted to Pearson r prior to analysis, following the formula provided by Rupinski and Dunlap (1996; Formula 3, p. 420). Coefficients for Neuroticism were reverse-coded, when necessary, such that higher scores indicate higher Emotional Stability. No other personality trait or LIWC category was reverse-coded.

Each study was coded by two coders. The first author coded all studies, and each of four research assistants coded one fourth of the studies. The four research assistants held a master's degree in psychology, and the coding process took place after training (explanation of meta-analytic goals, data coding, practice session, resolve questions; for details, see Supplemental Material). One exception was the coding of language formality, which was coded later by the first and third authors. Any disagreements during the coding process were discussed in follow-up meetings, until consensus was reached.

The studies by Mehl et al. (2006, 2012) were conducted among the same participants, but the first study analyzed only spoken and the latter only written text data. To avoid introducing data dependency, we averaged the effect sizes between the two data sets, following the recommendations by Borenstein et al. (2021). Since some effect sizes in Mehl et al. (2012) were missing (25 values), for these effect sizes, we used the values from Mehl et al. (2006). The two studies also differed in sample size (96–90 participants) and, when averaging effect sizes, the sample size was treated as 96.

Missing Values and Imputation Methods

There were two types of missing values in the data set: missing moderators (e.g., age) and missing effect sizes. To address this issue, we requested missing data from the authors. This process reduced the total number of missing values; however, 12% of the values were still missing. Hence, we decided to impute these values. For each type of missing data, we followed a different imputation process (manual imputation for moderators, automatic for effect sizes).

Regarding moderators, missing data were imputed based on the average values of available data. For age, imputed values were the average age of students in student samples and the average age of the general population in general population samples. For Cronbach's α, values were imputed per personality inventory (e.g., Big Five Inventory, Ten-Item Personality Inventory, International Personality Item Pool). When enough studies were available, the average α of the available data was used to impute the missing αs, otherwise, missing alphas were imputed from test manuals. If possible, we relied on indirect information to impute missing values. For

[5] There is also the LIWC1999 version, used in the Pennebaker and King's (1999) data set. However, Mairesse et al. (2007) reanalyzed an updated version of Pennebaker and King's data set using LIWC2001 version, and the present meta-analysis includes these updated correlations based on LIWC2001.

Table 1
LIWC Categories and Example Words for LIWC2001, LIWC2007, and LIWC2015 Versions

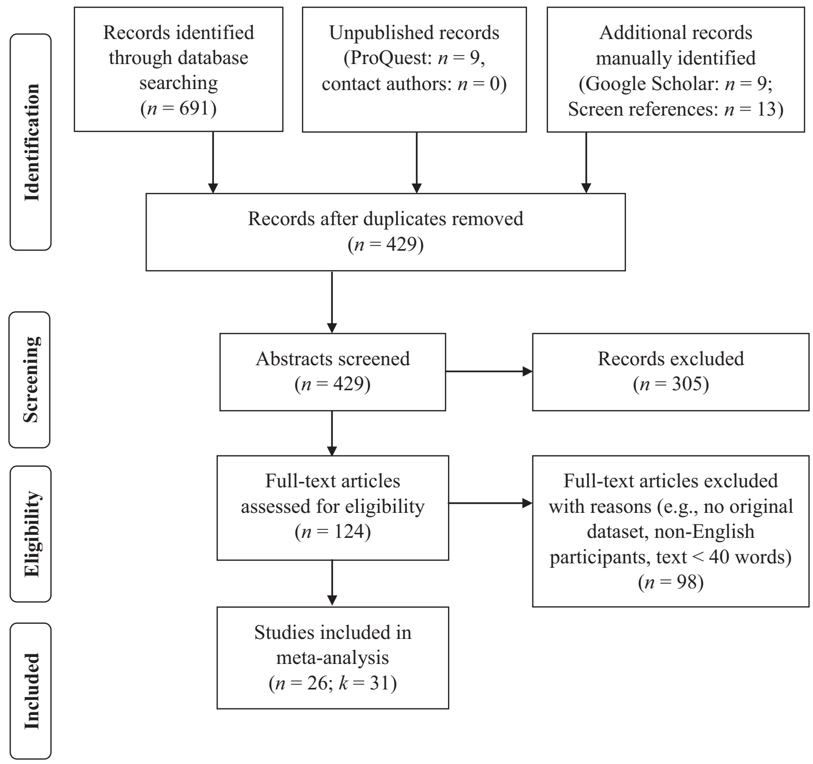
Broad LIWC category / Narrow LIWC category	Content/function words	Examples words	LIWC2001 Code name	LIWC2001 Entries	LIWC2007 Code name	LIWC2007 Entries	LIWC2015 Code name	LIWC2015 Entries
Word count			WC		WC		WC	
Words >6 letters			Sixltr		Sixltr		Sixltr	
Dictionary words			Dic		Dic		Dic	
Nonfluencies		Er, hm, umm	Nonfl	6	Nonfl	8	Nonflu	19
Fillers		Blah, I mean, you know	Fillers	6	Fillers	9	Filler	14
Total pronouns	Function words	I, them, itself	Pronoun	70	Pronoun	116	Pronoun	153
First-person singular	Function words	I, me, mine	I	9	I	12	I	24
First-person plural	Function words	We, us, our	We	11	We	12	We	12
Second-person	Function words	You, your, thou	You	14	You	20	You	30
Negations	Function words	No, not, never	Negate	31	Negate	57	Negate	62
Assent	Function words	Agree, OK, yes	Assent	18	Assent	30	Assent	36
Articles	Function words	A, an, the	Article	3	Article	3	Article	3
Prepositions	Function words	To, with, above	Preps	43	Preps	60	Preps	74
Auxiliary verbs	Function words	Am, will, have	—	—	Auxverb	144	Auxverb	141
Numbers	Content words	Second, thousand	Number	29	Number	34	Number	36
Affective processes	Content words	Happy, cried, abandon	Affect	615	Affect	915	Affect	1,393
Positive emotions	Content words	Love, nice, sweet	Posemo	43	Posemo	406	Posemo	620
Negative emotions	Content words	Hurt, ugly, nasty	Negemo	345	Negemo	499	Negemo	744
Anxiety	Content words	Worried, fearful, nervous	Anx	62	Anx	91	Anx	116
Anger	Content words	Hate, kill, annoyed	Anger	121	Anger	184	Anger	230
Sadness	Content words	Crying, grief, sad	Sad	72	Sad	101	Sad	136
Cognitive processes	Content words	Cause, know, ought	Cogmech	312	Cogmech	730	Cogproc	797
Causation	Content words	Because, effect, hence	Cause	49	Cause	108	Cause	135
Insight	Content words	Think, know, consider	Insight	116	Insight	195	Insight	259
Discrepancy	Content words	Should, would, could	Discrep	32	Discrep	76	Discrep	83
Tentative	Content words	Maybe, perhaps, guess	Tentat	79	Tentat	155	Tentat	178
Certainty	Content words	Always, never	Certain	30	Certain	83	Certain	113
Senses/perceptual processes	Content words	Observing, heard, feeling	Senses	111	Percept	273	Percept	436
See	Content words	View, saw, seen	See	31	See	72	See	126
Hear	Content words	Listen, hearing	Hear	36	Hear	51	Hear	93
Feel	Content words	Feels, touch	Feel	30	Feel	75	Feel	128
Social processes	Content words	Mate, talk, they, child	Social	314	Social	455	Social	756
Friends	Content words	Buddy, friend, neighbor	Friends	28	Friends	37	Friends	95
Family	Content words	Daughter, husband, aunt	Family	43	Family	64	Family	118
Time	Content words	End, until, season	Time	113	Time	239	Time	310
Past tense/past focus	Content words	Went, run/ago, did	Past	144	Past	145	Focuspast	341
Present tense/present focus	Content words	Is, does/today, now	Present	256	Present	169	Focuspresent	424
Future tense/future focus	Content words	gonna/will, soon	Future	14	Future	48	Focusfuture	97
Space	Content words	Down, in, thin	Space	71	Space	220	Space	360
Motion	Content words	Arrive, car, go	Motion	73	Motion	168	Motion	325
Work	Content words	Job, majors, xerox	Job	62	Work	327	Work	444
Achievement	Content words	Earn, hero, win	Achieve	60	Achieve	186	Achieve	213
Leisure	Content words	Cook, chat, movie	Leisure	102	Leisure	229	Leisure	296
Home	Content words	Apartment, kitchen	Home	26	Home	93	Home	100
Money	Content words	Audit, cash, owe	Money	75	Money	173	Money	226
Religion	Content words	Altar, church, mosque	Relig	56	Relig	159	Relig	174
Death	Content words	Bury, coffin, kill	Death	29	Death	62	Death	74
Biological processes	Content words	Eat, blood, pain	Physcal	285	Bio	567	Bio	748
Body	Content words	Cheek, hands, spit	Body	200	Body	180	Body	215
Sexual	Content words	Horny, love, incest	Sexual	49	Sexual	96	Sexual	131
Ingestion	Content words	Dish, eat, pizza	Eating	52	Ingest	111	Ingest	184
Swear words	Content words	Damn, piss, fuck	Swear	29	Swear	53	Swear	131

Note. LIWC = Linguistic Inquiry and Word Count; in the first column, indented categories represent "narrow" LIWC categories; nonindented categories represent "broad" LIWC categories; information is taken from the LIWC2001, LIWC2007, and LIWC2015 manuals (Pennebaker et al., 2001, 2007, 2015).

example, Cronbach's αs in Schwartz et al. (2013) were imputed from Kern et al. (2014) as the best approximation, since both studies analyzed the MyPersonality data set. The manual imputation processes are further detailed in the Supplemental Material.

Regarding effect sizes, missing values did not appear to be missing at random. Instead, authors did not report them when they were not statistically significant. The most conservative imputation approach would be to replace missing values with zeros, since missing effect sizes represent small, nonsignificant correlations. The limitation of this approach is that it artificially underestimates the true effect size, since nonsignificant correlations are frequently the result of relatively small sample sizes, instead of zero effect sizes. A more liberal approach would be to replace missing values with the average effect size of the available data. However, in this

Figure 2
Inclusion Criteria Through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Flowchart



approach, true effect sizes will be artificially inflated by the existing data.

To avoid a very conservative or a very liberal imputation method, we used a novel approach. For each LIWC category, we first—temporarily—replaced missing values with zeros, then summed up the available effect sizes together with the imputed zero values, and divided the sum by the total number of studies minus one, to arrive at the final imputation of the missing value. This method is neither too conservative, since it does not blindly replace all missing values with zeros nor too liberal, since it takes into account both available and zero values—although it leans toward the conservative side. In the rest of the article, we call this imputation method "mean − 1," and to understand its effect, we compared the effect sizes across different imputation methods. The magnitude of effect sizes was sorted as: nonimputed data > imputed mean > imputed "mean − 1" > imputed 0. In most cases, the effect sizes were almost identical across the (three) imputed and the nonimputed data sets or deviated only marginally. In personality self-reports, the deviation was typically found at the third decimal place or lower (the differences were larger in observer reports). As such, we decided that the "mean − 1" imputation method was a realistic representation of the missing effect sizes. Therefore, in the rest of the article, all analyses are based on the "mean − 1" imputation method, except otherwise specified.

The only manual imputation of effect sizes was the word count category in Schwartz et al. (2013). This value was imputed from Kern et al. (2014), who analyzed the same data set (MyPersonality). However, it should be noted that the effect size reported in Kern et al. (2014) was corrected for age and gender.

Correction for Attenuation

Correction for attenuation was based on the reliability of the personality measures (following the formula provided by Spearman, 1904, p. 90). For self-reports, correction was based on Cronbach's α; for observer reports, it was based on interrater agreement (i.e., intraclass correlation coefficient; ICC). One study measured interrater agreement using Krippendorff's α (Baek & Ihm, 2020), which was treated as ICC in the absence of a viable transformation to ICC; one study was based on a single observer (Sandy, 2013), and Cronbach's α was used to correct for attenuation. We did not correct for the reliability of the LIWC categories. The reasons and implications of this decision are further examined in the Discussion section. In the rest of the article, meta-analytic effect sizes refer to corrected values, except otherwise specified.

Analytical Strategy

When authors shared their data, the personality–LIWC correlation matrix was calculated only for participants with text length > 40 words (this was possible for 9 samples that were shared). When available, demographics and moderators (age, percent of women, text length, Cronbach's α) were recalculated based on the updated sample size with text length > 40 words. Therefore, the values reported here might slightly differ from the original publications.

The meta-analysis was performed after applying a random-effects model, using the R package "meta" (Schwarzer et al., 2015). We selected a random-effects, instead of fixed-effect model, as random-effects models assume that the true effect size might vary from study to study. Since the studies in the present meta-analysis differed in a number of sample and task characteristics (e.g., general population or students, written or spoken language), a random-effects model is more likely to provide accurate estimates. In contrast, fixed-effect models assume that the true effect size is exactly the same across all studies, something that is, at best, unrealistic and, at worst, produces distorted (typically inflated) meta-analytic estimates and narrower confidence intervals (see Borenstein et al., 2010; Hunter & Schmidt, 2000). The standard error was calculated by the R "meta" package, by dividing the standard deviation with the square root of the sample size (see Schwarzer et al., 2015). Heterogeneity of effect sizes was tested using Cochran's Q test, the I^2, and tau (τ) indices (Huedo-Medina et al., 2006; Veroniki et al., 2016). Publication bias was tested using Egger's test (Egger et al., 1997).

To calculate the proportion of variance that LIWC categories explained per personality trait, we conducted meta-analytic linear regressions. Personality traits—each one separately—were the dependent variables and the full set of LIWC categories of the predictors. For the relations between personality traits and LIWC categories, we relied on the effect size estimates of the present meta-analysis. To control for LIWC intercorrelations, we calculated the weighted average correlation matrix (after applying Fisher's r-to-z-to-r transformation) across the 52 LIWC categories. The LIWC intercorrelation matrix was based on a subsample of 14 studies ($n = 6{,}405$), for which we received raw LIWC scores from the authors. The regression analysis was performed in R using the "lavaan" package (Rosseel, 2012), with the harmonic mean across all analyzed cells of the correlation matrix as the sample size (Viswesvaran & Ones, 1995).

Content and Function Words

The 52 LIWC categories were further decomposed into content and function words, each of which serves a special function in language. More specifically, *content* words cover the majority of the words in a dictionary (e.g., nouns, adjectives, verbs) and convey some sort of tangible information, for instance, about actions, objects, or ideas. On the other hand, *function* words (also known as stop or style words) include mainly articles (i.e., a, an, the), pronouns (e.g., I, yours, whose), prepositions (e.g., in, at, on), and conjunctions (e.g., and, or, but) and occupy only a small proportion in a dictionary (amounting to almost 500 in English). Nevertheless, they are used at a disproportionately higher rate, and they are responsible for the personal style of the speaker or writer (Tausczik & Pennebaker, 2010). Function words are typically removed when preprocessing a text (Banks et al., 2018), despite the fact that they are typically normally distributed (unlike content words that follow a highly skewed distribution; Almodaresi et al., 2017), which makes them more suitable to be analyzed using traditional statistical techniques (Eichstaedt et al., 2021). Research using LIWC has highlighted function words' special role in language. For instance, they have been found to relate to criterion variables including personality, age, gender, political orientation (Pennebaker, 2011), and physical health (Campbell & Pennebaker, 2003), even when content words do not. Therefore, besides the full set of linguistic categories, we also calculated the percentage of personality variance explained by content and function words separately.

Furthermore, LIWC is hierarchically structured, including broad and narrow categories. For instance, the broad category of pronouns includes the narrower categories of first-person singular pronouns, first-person plural pronouns, and second-person pronouns. By including both broad and narrow categories to explain personality variance, we encounter the problem of "double-dipping" in the same category. This is not only the case for function words but also for content words. For instance, the broad category of affective processes includes the narrower categories of positive emotions, negative emotions, anxiety, anger, and sadness (see Table 1). To address this issue, we conducted the analysis of explained variance for function and content words three times, including: (a) both broad and narrow categories (thus, "double-dipping" for comparison purposes), (b) only broad categories, and (c) only narrow categories. Here, we report only the results for the "narrow" categories. We believe that this is the most optimal approach, since it includes all necessary information (e.g., first-person singular pronouns, first-person plural pronouns, and second-person pronouns), while avoiding the "double-dipping" issue. In contrary, including only broad categories might mask the true effect, since—especially for content words—categories that offer unique personality insights (e.g., positive emotions, negative emotions) might be masked under a single broad category (i.e., affective processes). For the sake of completeness, we report the results of all three analyses in Supplemental Table S21 in the supplemental material.

Transparency and Openness

Extra results and the coding scheme are available at the Supplemental Material. All data, research materials, codes, and results for other imputation methods are available at the Open Science Framework (OSF) page of the project (https://osf.io/7vszn/).

Results

Descriptive Statistics

Out of the 31 eligible samples, 26 measured personality using self-reports (cue validity) and five using observer reports (cue utilization). For cue validity, one study had a disproportionally larger sample size ($n = 72{,}809$; Schwartz et al., 2013) and one study had a disproportionally larger text length (115,423 words per participant; Yarkoni, 2010). The average sample size per study was $M = 454.6$ participants (range: 50–2,582; if Schwartz's study is included $M_{\text{with-Schwartz}} =$ 3,237 participants), the average age was $M = 27.96$ years ($SD =$ 10.50), and participants wrote or spoke $M = 3{,}058.72$ words on average ($SD = 4{,}902.93$; if Yarkoni's study is included $M_{\text{with-Yarkoni}} =$ 7,553.30 words). Since the Schwartz et al.'s (2013) study had a disproportionally larger sample size compared to the other studies, the meta-analysis was performed with and without this data set. For cue utilization, the average sample size was $M = 327.4$ participants (range: 49–942), the average age was $M = 23.66$ years ($SD = 5.52$), and the average number of words per participant was $M = 1{,}055.70$ ($SD =$ 935.64). Table 2 summarizes the studies included in the meta-analysis.

Effect Sizes

The total number of analyses was large,[6] and to facilitate readability, we only present the corrected effect sizes, after applying the

[6] 5 Personality Traits × 52 LIWC Categories × 4 Imputation Methods (no imputation, imputed 0, imputed "mean − 1", imputed mean) × 2 Values (corrected-uncorrected) × 2 Personality Assessment Methods (self-reports, observer reports) × 10 Moderators = 41,600 analyses.

Table 2
Summary of Studies Included in the Present Meta-Analysis

Study	N	Personality measure	Task type	LIWC version	Language type
Cue validity					
Abe (2018), Study B	222	TIPI-10	Reflection on upcoming U.S. elections	LIWC2015	Written
Abe (2020)	92	BFI-44	Online discussion over student assignment	LIWC2015	Written
Baddeley and Singer (2008), Study 1	111	BFI-44	Bereavement narrative	LIWC2001	Written
Burdick et al. (2021)	620	BFI-44	Image caption on five pictures expressing oneself	LIWC2015	Written
Gill et al. (2006)	71	IPIP-41	Blog posts	LIWC2001	Written
Golbeck, Robles, and Edmondson (2011)	50	BFI-45	Twitter status updates	LIWC2007	Written
Golbeck, Robles, and Turner (2011)	167	BFI-45	Facebook profile	LIWC2001	Written
Graybeal et al. (2002)	52	NEO-FFI-60	Compilation of eight student essays	LIWC2001	Written
Hall and Caton (2017)	282	BFI-44	Facebook status updates	LIWC2001	Written
Hawkins and Boyd (2017), Study 1	637	TIPI-10	One recent and one important dream	LIWC2015	Written
Hawkins and Boyd, (2017), Study 2	375	TIPI-10	A recent dream	LIWC2015	Written
Hawkins and Boyd (2017), Study 3	268	BFI-44	A recent dream	LIWC2015	Written
Hirsh and Peterson (2009)	94	BFAS-100	Essay on past experiences and future goals	LIWC2007	Written
Holtrop et al. (2019)	140	HEXACO-PI-R-60; HEXACO-PI-R-200	Mock selection interview	LIWC2015	Spoken
Kahn et al. (2007)	66	BFI-44	Verbal reaction to short films	LIWC2001	Spoken
Krieger (2016)	128	NEO-PI-R-240	Essay on social skills	LIWC2015	Written
Mehl et al. (2006)	96	BFI-44	Daily utterances from the electronically activated recorder (EAR) corpus	LIWC2001	Spoken
Mehl et al. (2012)	90	BFI-44	Stream of consciousness student essay	LIWC2001	Written
Pennebaker and King (1999)	2,479	BFI-44	Stream of consciousness student essay	LIWC2001	Written
Qiu et al. (2012)	142	BFI-44	Twitter status updates	LIWC2007	Written
Sandy (2013)	942	TIPI-10	Chat platform for student assignment	LIWC2007	Written
Schwartz et al. (2013)	72,809	IPIP-20; IPIP-100	Facebook status updates (MyPersonality)	LIWC2007	Written
Sumner et al. (2011)	506	BFI-44	Facebook status updates	LIWC2007	Written
Sumner et al. (2012)	2,582	TIPI-10	Twitter status updates	LIWC2007	Written
Tackman et al. (2020)	459	BFI-44; TIPI	Daily utterances from the electronically activated recorder (EAR) corpus	LIWC2015	Spoken
Yarkoni (2010)	694	IPIP-50; IPIP-300	Blog posts	LIWC2001	Written
Cue utilization					
Baek and Ihm (2020)	49	TIPI-10	Self-presentation vignettes	LIWC2015	Written
Biel et al. (2013)	408	TIPI-10	YouTube vlogs	LIWC2007	Spoken
Mehl et al. (2006)	96	BFI-44	Daily utterances from the electronically activated recorder (EAR) corpus	LIWC2001	Spoken
Qiu et al. (2012)	142	BFI-44	Twitter	LIWC2007	Written
Sandy (2013)	942	TIPI-10	Chat platform for student assignment	LIWC2007	Written

Note. LIWC = Linguistic Inquiry and Word Count; TIPI = Ten-Item Personality Inventory (Gosling et al., 2003); BFI = Big Five Inventory (John & Srivastava, 1999); IPIP = International Personality Item Pool (https://www.ipip.ori.org; Goldberg, 1999; Goldberg et al., 2006); NEO-FFI = NEO Five-Factor Inventory (Costa & McCrae, 1992); BFAS = Big Five Aspect Scales (DeYoung et al., 2007); HEXACO-PI-R = HEXACO Personality Inventory–Revised (Ashton & Lee, 2009; Lee & Ashton, 2006); NEO-PI-R = NEO Personality Inventory–Revised (Costa & McCrae, 1992); the number next to personality instruments indicates the number of items; in Schwartz et al. (2013), sample sizes ranged from 71,968 (Emotional Stability) to 72,809 (Openness to Experience); the data for Pennebaker and King (1999) were extracted from the updated analysis of Mairesse et al. (2007).

"mean − 1" imputation method. Furthermore, following the recommendation by Funder and Ozer (2019) on meaningful effect sizes, we focus on effect sizes that are at least "very small," that is, $|\rho| \geq .05$ ("an effect size r of .05 indicates an effect that is very small for the explanation of single events but potentially consequential in the not very long run"; Funder & Ozer, 2019, p. 166).[7] However, for the sake of readability, in the tables of the article, we only report effect sizes when (a) $|\rho| \geq .10$ ("an effect size r of .10 indicates an effect that is still small at the level of single events but potentially more ultimately consequential"; Funder & Ozer, 2019, p. 166) and (b) the

[7] Naturally, one might argue that, when interpreting effect sizes, any cutoff threshold can be somewhat arbitrary, as there cannot be a one-fit-all interpretation regardless of the context or the type of variables (Bosco et al., 2015). For instance, in applied psychology, especially when variables include behavior, the *median* effect size is $r = .16$, ranging from .10 to .24 (Bosco et al., 2015, p. 436). Under this lens, in the present meta-analysis, effect sizes $r \geq .10$ could be interpreted as "medium." However, Bosco et al. (2015) operationalized "behavior" as employee behavior using summary items (e.g., job performance, counterproductive behavior, turnover, absenteeism). Instead, in the present meta-analysis, "behavior" was operationalized as verbal behavior, which is more granular. Thus, we preferred the more general guidelines of Funder and Ozer (2019), in the absence of a more optimal approach.

lower 95% CI |ρ| ≥ .05. The full set of 52 LIWC categories for cue validity and cue utilization, as well as the output for other imputation methods are available in the Supplemental Material and on the OSF page of the project, respectively.

Cue Validity

Regarding cue validity, there were 13 "very small" effect sizes for Openness to Experience, 12 for Conscientiousness, seven for Extraversion, 11 for Agreeableness, and 15 for Emotional Stability. Among the most notable findings (i.e., "small" effect sizes with at least "very small" lower confidence intervals; Table 3), individuals who scored high on Openness to Experience used more words longer than six letters (ρ = .11); individuals who scored high on Conscientiousness used fewer negative emotions (ρ = −.12), fewer swear words (ρ = −.11), and fewer anger-related words (ρ = −.10); individuals who scored high on Agreeableness used fewer anger-related words (ρ = −.14), fewer swear words (ρ = −.10), fewer negative emotions (ρ = −.10), and more positive emotions (ρ = .10); and individuals who scored high on Emotional Stability used fewer negative emotions (ρ = −.12). Regarding Extraversion, there was no effect size that exceeded the ρ ≥ .10 threshold. Results were almost identical after excluding Schwartz et al.'s (2013) data set.

Cue Utilization

Regarding cue utilization, there were 20 "very small" effect sizes for Openness to Experience, 17 for Conscientiousness, 13 for Extraversion, 14 for Agreeableness, and 24 for Emotional Stability. Among the most notable findings (Table 3, lower part), individuals who were rated by others as high on Openness to Experience used more words in general (ρ = .18) and more leisure-related words (ρ = .17); individuals who were rated by others as high on Conscientiousness

Table 3
Meta-Analytic Effect Sizes for Cue Validity and Cue Utilization (for Effect Sizes |ρ| ≥ .10 and Lower 95% CI |ρ| ≥ .05)

Personality trait LIWC category	r	ρ ($\rho_{\text{no Schwartz}}$)	95% CI lower	95% CI upper	I^2	Q	Q p value	τ	Egger's test	Egger's p value
Cue validity										
Openness to Experience										
Words >6 letters	.09	**.11** (.11)	.09	.12	22.42	30.94	.16	0.02	0.24	.81
Conscientiousness										
Negative emotions[a]	−.10	**−.12** (−.11)	−.15	−.09	60.47	60.72	<.001	0.04	3.85	<.001
Swear[a]	−.10	**−.11** (−.10)	−.14	−.08	65.35	69.26	<.001	0.05	1.11	.28
Anger[a]	−.09	**−.10** (−.09)	−.13	−.07	71.56	84.38	<.001	0.05	3.42	.002
Agreeableness										
Anger	−.11	**−.14** (−.13)	−.17	−.11	66.59	71.84	<.001	0.05	3.77	<.001
Swear[a]	−.09	**−.10** (−.10)	−.14	−.07	78.30	110.61	<.001	0.07	2.36	.03
Negative emotions	−.08	**−.10** (−.10)	−.13	−.07	61.07	61.64	<.001	0.04	4.58	<.001
Positive emotions	.09	**.10** (.09)	.07	.14	74.57	94.37	<.001	0.06	−2.28	.03
Emotional Stability										
Negative emotions	−.11	**−.12** (−.12)	−.14	−.09	57.27	56.16	<.001	0.04	0.11	.92
Cue utilization										
Openness to Experience										
Word count	.14	**.18**	.13	.22	0.00	1.53	.82	0.00	3.32	.05
Leisure	.11	**.17**	.06	.28	69.14	12.96	.01	0.10	1.46	.24
Conscientiousness										
Swear[a]	−.28	**−.35**	−.59	−.10	94.54	73.28	<.001	0.27	−1.28	.29
Anger[a]	−.26	**−.32**	−.55	−.09	93.59	62.38	<.001	0.25	−1.31	.28
Negative emotions[a]	−.25	**−.30**	−.53	−.08	93.23	59.09	<.001	0.24	−1.34	.27
Sexual	−.24	**−.29**	−.48	−.11	89.90	39.60	<.001	0.19	−2.60	.08
Biological processes	−.24	**−.28**	−.47	−.09	90.39	41.63	<.001	0.20	−1.90	.15
Pronouns	−.12	**−.17**	−.28	−.06	70.02	13.34	.01	0.10	−1.19	.32
Cognitive mechanisms	.12	**.16**	.07	.24	51.35	8.22	.08	0.07	1.55	.22
Tentativeness	.11	**.13**	.08	.18	0.00	3.36	.50	0.00	1.12	.35
Extraversion										
Word count	.27	**.31**	.10	.51	92.13	50.82	<.001	0.22	1.47	.24
Sexual	.13	**.16**	.10	.21	8.98	4.40	.36	0.02	1.61	.21
Agreeableness										
Swear[a]	−.32	**−.39**	−.71	−.07	96.78	124.11	<.001	0.35	−1.18	.32
Death	−.09	**−.12**	−.16	−.07	0.00	2.81	.59	0.00	−0.45	.69
Emotional Stability										
First-person singular pronouns	−.24	**−.29**	−.49	−.10	90.84	43.69	<.001	0.20	−1.93	.15
Pronouns	−.21	**−.27**	−.41	−.12	82.04	22.28	<.001	0.14	−2.31	.10

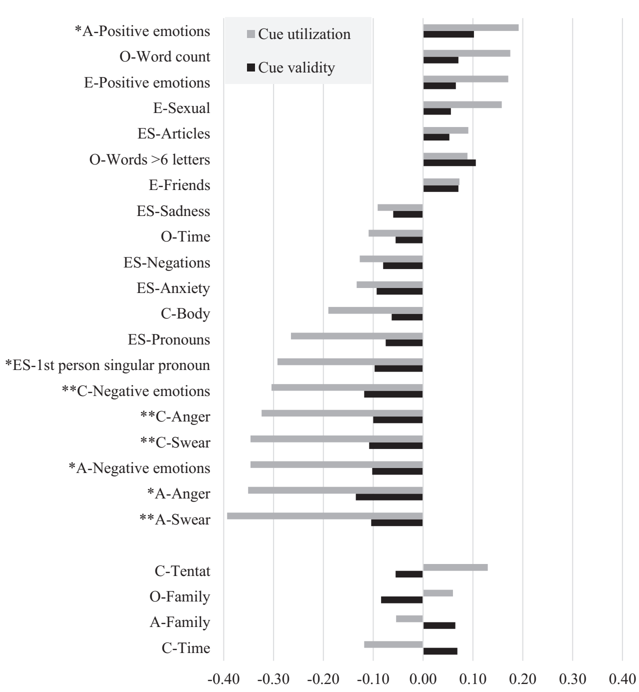
Note. LIWC = Linguistic Inquiry and Word Count; CI = confidence interval. Effect sizes after imputation with the "mean − 1" method; cue validity (self-reports): k = 25, n = 83,243–84,084; cue utilization (observer reports): k = 5, n = 1,637; r = attenuated effect sizes; ρ = disattenuated effect sizes (corrected for personality, but not for LIWC unreliability); $\rho_{\text{no Schwartz}}$ = ρ excluding Schwartz et al. (2013; this exclusion applies only to cue validity; after excluding Schwartz, n = 11,275); all values to the right of—and including—column "95% CI upper" refer to ρ; corrected effect sizes are boldfaced; all effect sizes are statistically significant.
[a] LIWC categories present both in cue validity and cue utilization.

used fewer swear words (ρ = −.35), fewer anger-related words (ρ = −.32), fewer negative emotions (ρ = −.30), fewer sexual-related words (ρ = −.29), fewer words related to biological processes (ρ = −.28), fewer pronouns (ρ = −.17), more words related to cognitive mechanisms (ρ = .16), and more tentativeness-related words (ρ = .13); individuals who were rated by others as high on Extraversion used more words in general (ρ = .31) and more sexual-related words (ρ = .16); individuals who were rated by others as high on Agreeableness used fewer swear words (ρ = −.39) and fewer death-related words (ρ = −.12); and individuals who were rated by others as high on Emotional Stability used fewer first-person singular pronouns (ρ = −.29) and fewer pronouns in general (ρ = −.27).

Kernel of Truth

There were 20 (out of 260; 5 Personality Traits × 52 LIWC Categories) valid utilized cues, that is, significant effect sizes (in the same direction) both for self- and observer reports. These effect sizes represent the "kernel of truth" in linguistic markers of personality (Figure 3). Three of them belonged to Openness to Experience, four to Conscientiousness, three to Extraversion, four to Agreeableness, and six to Emotional Stability. When we focused on at least "small" effect sizes (i.e., |ρ| ≥ .10), there were eight linguistic markers in the kernel of truth. Three of them belonged to Conscientiousness (swear words, anger, negative emotions), four to Agreeableness (swear words, anger, negative emotions, positive emotions), and one to Emotional Stability (first-person singular pronouns). When we focused on at least "small" effect sizes with at least "very small" (i.e., |ρ| ≥ .05) lower confidence intervals, there were four linguistic markers in the kernel of truth. Three of them belonged to Conscientiousness (swear words, anger, negative emotions) and one belonged to Agreeableness (swear words).

Figure 3
Kernel of Truth for Effect Sizes |ρ| ≥ .05



Note. O = Openness to Experience; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability; CI = confidence interval; effect sizes are in descending order based on cue utilization; all effect sizes are statistically significant and |ρ| ≥ .05.
* Both cue validity and cue utilization are significant, and |ρ| ≥ .10; ** For both cue validity and cue utilization |ρ| ≥ .10, and lower 95% CI |ρ| ≥ .05.

The correlation between cue validity and cue utilization for the 20 linguistic markers that constituted the kernel of truth was $r = .94$. The correlation between cue validity and cue utilization across all 260 effect sizes was $r = .52$. Interestingly, we found four additional linguistic markers where cue validity and cue utilization were in the opposite direction (lower part of Figure 3).

Explained Variance

Table 4 summarizes the percentage of explained variance per personality trait, including only the "narrow" LIWC categories. For cue validity, R^2 ranged from 3.37% (Extraversion) to 6.70% (Openness to Experience). For cue utilization, the proportion of explained variance was much larger, compared to cue utilization, ranging from 31.48% (Emotional Stability) to 47.32% (Conscientiousness).

The analysis of explained variance was repeated after selecting only function and content words (again, for the "narrow" categories). Function words included eight narrow linguistic categories (first-person singular and plural pronouns, second-person pronouns, negations, assent, articles, prepositions, and auxiliary verbs), and content words included 33 narrow linguistic categories (all narrow LIWC categories except for function words, word count, dictionary, words > 6 letters, nonfluencies, and fillers). Overall, function words explained less variance than content words, both for cue validity (average R^2 = 1.31% and 3.71% for function and content words, respectively) and cue utilization (average R^2 = 5.05% and 28.68% for function and content words, respectively).

Moderation Analysis

Five continuous (age, sample size, percentage of women in the sample, text length, year of publication) and five categorical moderators were considered. The categorical moderators (and their levels) were as follows: language formality (formal or informal), sample composition (general population or students), synchronicity (dialogues or monologues), text mode (written or spoken language), and LIWC version (2001, 2007, or 2015). Regarding categorical moderators, for the sake of parsimony, we report only those results for which the effect size of at least one level of the categorical moderator was |ρ| ≥ .10 and the lower 95% CI |ρ| ≥ .05. Similarly, for continuous moderators, we report only the significant moderation effects for which the effect sizes were |ρ| ≥ .10 and the lower 95% CI |ρ| ≥ .05. All other effects can be found in the Supplemental Tables S11–S20 and the OSF page of the project.

For cue validity, the Mehl et al.'s (2006, 2012) studies shared the same participants but involved different tasks. These tasks were captured by two categorical moderators of the present analysis (i.e., text mode and synchronicity). Averaging the effect sizes across different type of tasks would mask any moderation effects. Therefore, we performed the analysis twice, once excluding the 2012 and once excluding the 2006 data sets. Overall, results were very similar in both cases; however, here, we present the results after including

Table 4
Percentage of Explained Variance (R^2) per Personality Trait by "Narrow" LIWC Categories, Function, and Content Words

Personality trait	Total LIWC categories R^2 (%)	Function words R^2 (%)	Content words R^2 (%)
Cue validity			
Openness to Experience	6.70	1.49	4.36
Conscientiousness	5.13	0.88	4.44
Extraversion	3.37	0.99	2.22
Agreeableness	6.45	1.53	4.60
Emotional Stability	3.72	1.67	2.92
M	5.07	1.31	3.71
Cue utilization			
Openness to Experience	38.77	1.55	30.00
Conscientiousness	47.32	7.37	36.18
Extraversion	35.91	1.71	19.89
Agreeableness	39.14	2.70	35.47
Emotional Stability	31.48	11.93	21.88
M	38.52	5.05	28.68

Note. LIWC = Linguistic Inquiry and Word Count. Total "narrow" LIWC categories n = 46; function words n = 8 "narrow" LIWC categories; content words n = 33 "narrow" LIWC categories; values are based on the corrected effect sizes after applying the "mean − 1" imputation method.

only the Mehl et al.'s (2006) data set (full results can be found on the OSF page of the project). For cue utilization, we did not test the effect of categorical moderators because there were too few studies per level (typically one to three).

Categorical Moderators—Cue Validity

Table 5 presents the results of the categorical moderators for cue validity. Among the most prominent results, individuals high on Extraversion used more words in general, but extraverted individuals mainly spoke (ρ = .23) instead of wrote (ρ = .03) more words. Regarding language formality, the negative correlations between Agreeableness and swear words, as well as the negative correlations between Conscientiousness with negative emotions and anger-related words were especially pronounced in informal (ρ = −.14, ρ = −.14, and ρ = −.15, respectively) compared to formal settings (ρ = −.07, ρ = −.09, and ρ = −.05). Furthermore, the positive correlations between Agreeableness with social-related words and first-person plural pronouns were especially pronounced in formal (ρ = .13 and ρ = .11, respectively) compared to informal language (ρ = .05 and ρ = .03). Regarding sample composition, the negative correlations between Agreeableness and swear words, as well as the negative correlations between Conscientiousness and negative emotions were especially pronounced in general (ρ = −.14 and ρ = −.15, respectively) compared to student populations (ρ = −.07 and ρ = −.09).

Regarding the LIWC version, older versions elicited larger correlations on average (average |ρ| = .042, .039, and .030 for LIWC2001, LIWC2007, and LIWC2015, respectively). For instance, the negative correlation between home-related words and Openness to Experience was stronger in LIWC2001 (ρ = −.13), followed by LIWC2007 (ρ = −.07), and LIWC2015 (ρ = −.03). Less frequently, effect sizes were stronger for LIWC2007.

Regarding the TAT hypotheses, individuals low on self-reported Agreeableness used more swear-related words in informal (ρ = −.14), rather than formal (ρ = −.07; QM = 4.64, p = .03) situations, as hypothesized. However, we failed to replicate the moderating role of synchronicity for self-reported Extraversion and positive emotions reported in Chen et al. (2020). Despite the fact that individuals high on self-reported Extraversion used more positive emotion words in synchronous (ρ = .07) rather than asynchronous (ρ = .01) tasks, the moderation effect was not statistically significant (QM = 1.03, p = .31).

Continuous Moderators—Cue Validity and Cue Utilization

Table 6 presents the results of the continuous moderators for cue validity and cue utilization. The most prominent results were that effect sizes were typically stronger for longer texts and weaker for larger sample sizes. The relation between both self- and observer-reported Conscientiousness and Agreeableness with negatively valenced LIWC categories (i.e., anger, swear, negative emotions) was attenuated in more recent years. Regarding the continuous moderators of age and percentage of women, there were no significant moderation effects (at least, when meta-analytic effects were |ρ| ≥ .10 and lower 95% CI |ρ| ≥ .05; but see Supplemental Tables S16 and S18, for more nuanced results).

Publication Bias

Publication bias was tested using Egger's test in the uncorrected, nonimputed data sets. Out of the 260 tested effect sizes, 99 effect sizes in cue validity and 11 effect sizes in cue utilization showed significant funnel plot asymmetry. One would expect significant plot asymmetry even when there is none just out of chance (false positive) for around 13 effect sizes (5% of 260). This means that for cue validity, the 99 effect sizes with plot asymmetry well exceeded the threshold and warranted further examination.

As a first step, the Schwartz et al.'s (2013) study was omitted and the number of significant Egger's tests dropped to six. This probably means that publication bias was the result of small-study effects (i.e., all the other studies). Small-study effects do not necessarily suggest publication bias (Borenstein et al., 2021), and in the present results, we believe that publication bias was not present for the following reasons. First, whether Schwartz et al. was included or excluded in the meta-analysis, the effect sizes did not change much. If anything, after excluding Schwartz et al., the effect sizes were even smaller, contrary to the publication bias idea (see Table 3). Second, if publication bias due to small-study effects was present, we would have expected that, when sample size was used as a moderator, larger sample sizes would have elicited smaller effect sizes, and vice versa. Instead, when sample size was used as a moderator, the effect sizes did not systematically differ. Thus, it seems that the results for cue validity were not due to publication bias.

Discussion

The present meta-analysis showed that linguistic markers of personality do exist. Not only are people more likely to express their (self-reported) personality traits using specific linguistic markers but other people draw inferences about personality traits from specific linguistic cues as well. Regarding cue validity, the strongest effect sizes across the Big Five traits were relatively small and explained a small proportion of personality variance (5.07%, on average). Regarding cue utilization, the strongest effect sizes were

Table 5
Categorical Moderators

Moderator LIWC category	Personality trait	Meta-analytic ρ	*k*	ρ	95% CI lower	95% CI upper	*k*	ρ	95% CI lower	95% CI upper	*k*	ρ	95% CI lower	95% CI upper
Language formality					Formal				Informal					
Swear	A	−0.10	13	−0.07	−0.14	0.01	12	**−0.14**	−0.17	−0.11				
Negative emotions	C	−0.12		−0.09	−0.13	−0.04		**−0.14**	−0.17	−0.11				
Anger	C	−0.10		−0.05	−0.12	0.01		**−0.15**	−0.17	−0.12				
Social	A	0.08		**0.13**	0.08	0.18		0.05	0.02	0.08				
First-person plural pronouns	A	0.07		**0.11**	0.06	0.16		0.03	−0.01	0.07				
Sample composition					General population				Students					
Swear	A	−0.10	13	**−0.14**	−0.18	−0.10	12	−0.07	−0.14	−0.00				
Negative emotions	C	−0.12		**−0.15**	−0.18	−0.12		−0.09	−0.13	−0.06				
Time	C	0.07		**0.11**	0.10	0.12		0.05	0.01	0.08				
Synchronicity					Asynchronous				Synchronous					
Time	C	0.07	21	**0.10**	0.09	0.11	4	0.00	−0.05	0.05				
Tentativeness	E	−0.07		−0.07	−0.09	−0.04		**−0.13**	−0.18	−0.08				
Dictionary	C	0.03		0.02	−0.01	0.05		**0.11**	0.05	0.17				
Text mode					Written				Spoken					
Word count	E	0.03	21	0.03	0.02	0.04	4	**0.23**	0.16	0.30				
Anger	ES	−0.06		−0.06	−0.08	−0.03		**−0.13**	−0.26	−0.01				
Body	ES	−0.05		−0.04	−0.07	−0.02		**−0.15**	−0.25	−0.06				
First-person plural pronouns	C	0.04		0.03	−0.00	0.06		**0.15**	0.08	0.22				
LIWC version					LIWC2001				LIWC2007				LIWC2015	
Dictionary	O	−0.09	9	**−0.14**	−0.21	−0.08	7	−0.08	−0.09	−0.07	9	−0.07	−0.12	−0.02
Pronouns	O	−0.08		**−0.13**	−0.20	−0.07		−0.03	−0.1	0.04		−0.09	−0.13	−0.04
Home	O	−0.08		**−0.16**	−0.20	−0.12		−0.07	−0.08	−0.06		−0.03	−0.07	0.0
First-person singular pronouns	O	−0.07		**−0.15**	−0.18	−0.12		0.00	−0.06	0.05		−0.09	−0.17	−0.00
Motion	O	−0.07		**−0.17**	−0.22	−0.12		−0.03	−0.07	0.01		−0.03	−0.07	0.02
Articles	ES	0.05		**0.11**	0.08	0.14		0.03	−0.01	0.06		0.04	−0.01	0.09
Fillers	C	−0.06		−0.04	−0.07	−0.00		**−0.11**	−0.15	−0.07		−0.03	−0.08	0.01
Body	C	−0.06		−0.04	−0.07	−0.01		**−0.11**	−0.12	−0.10		−0.05	−0.09	0.00
Positive emotions	C	0.05		0.03	−0.00	0.06		**0.10**	0.07	0.14		0.02	−0.01	0.06

Note. LIWC = Linguistic Inquiry and Word Count; CI = confidence interval; A = Agreeableness; C = Conscientiousness; E = Extraversion; ES = Emotional Stability; O = Openness to Experience; the table presents significant moderation effects (categorical moderators) when the effect size in at least one level of the categorical moderator is |ρ| ≥ .10 (and lower 95% CI |ρ| ≥ .05); the group with the strongest effect size is boldfaced; all moderation effects are statistically significant.

Table 6
Continuous Moderators

Moderator LIWC category	Personality trait	Meta-analytic ρ	*b*
	Cue validity		
Sample size			
Negative emotions	A	−0.10	<−0.001
Word count			
Anger	A	−0.14	<−0.001
Swear	A	−0.10	<−0.001
Year of publication			
Anger	A	−0.14	0.006
Swear	A	−0.10	0.008
Negative emotions	ES	−0.12	0.006
Swear	C	−0.11	0.006
	Cue utilization		
Sample size			
Cognitive processes	C	0.16	<−0.001
Sexual	C	−0.29	<0.001
First-person singular pronouns	ES	−0.29	<0.001
Pronouns	C	−0.17	<0.001
Biological processes	C	−0.28	<0.001
Pronouns	ES	−0.27	<0.001
Word count			
Cognitive processes	C	0.16	<0.001
Swear	C	−0.35	<−0.001
Anger	C	−0.32	<−0.001
Negative emotions	C	−0.30	<−0.001
Sexual	C	−0.29	<−0.001
First-person singular pronouns	ES	−0.29	<−0.001
Swear	A	−0.39	<−0.001
Year of publication			
Swear	C	−0.35	0.068
Anger	C	−0.32	0.058
Negative emotions	C	−0.30	0.047
Word count	E	0.31	−0.054

Note. LIWC = Linguistic Inquiry and Word Count; A = Agreeableness; ES = Emotional Stability; C = Conscientiousness; E = Extraversion; the table presents significant moderation effects (continuous moderators) for effect sizes where |ρ| ≥ .10, and lower 95% CI |ρ| ≥ .05; all moderation effects are statistically significant.

small-to-medium and explained a medium proportion of personality variance (38.52%, on average). Furthermore, we identified 20 linguistic markers that are part of the "kernel of truth" of personality, and we also highlighted the important role of moderators in the linguistic markers of personality. Below we summarize the four main theoretical contributions of the present findings.

Theoretical Implications

Linguistic Markers of Personality

First, both for cue validity and cue utilization, the correlations between LIWC categories and personality scores were generally in line with the conceptualization of each personality trait, suggesting that linguistic categories serve as behavioral manifestations of personality traits. For instance, individuals high on self-reported Extraversion used more positive emotions and social-related words, in line with their social and positive emotions profile (Hall & Caton, 2017). Those results replicate and expand the meta-analytic findings by Chen et al. (2020) further showing that only positive emotions—but not social-related words—were significant for cue utilization.

Similar alignment between personality theory and linguistic categories was observed for all personality traits. Individuals high on self-reported Openness to Experience used more words that are longer than six letters (i.e., more complex language), in line with the positive relation between Openness to Experience and general cognitive ability (Judge et al., 2007), and they also used fewer family and home-related words (i.e., traditional values) in line with their more liberal personality profile (Yarkoni, 2010). Individuals high on self-reported Conscientiousness used more achievement and job-related words, in line with their industriousness and high job performance profile (Sackett & Walmsley, 2014), and they also used fewer negative emotion words, since individuals high on Conscientiousness seem to experience fewer negative emotions in general (Fayard et al., 2012). Individuals high on self-reported Agreeableness used fewer anger, swear, and negative emotion words, in line with previous research suggesting that high Big Five Agreeableness (but much less HEXACO Agreeableness, see Lee et al., 2013) is negatively related to the dark triad (psychopathy–narcissism–Machiavellianism; Vize et al., 2021), and they also used more positive emotions and social-related words (e.g., "we"), probably reflecting their team-oriented and cooperative disposition (Beersma et al., 2003; Bradley et al., 2013). Finally, individuals high on self-reported Emotional Stability used fewer negative emotion words (e.g., anger) and fewer words related to depressive mood (e.g., "I," anxiety, sadness), in line with previous research showing a negative relation between Emotional Stability with distressed affect (Hall & Caton, 2017) and with depressive and anxiety symptoms (Kotov et al., 2010).

Kernel of Truth

Second, we identified 20 linguistic categories that described the "kernel of truth" in the linguistic markers of personality. Those linguistic categories correlate both with self- and observer reports of personality traits. Next to these specific linguistic categories, some more support for the "kernel of truth" idea comes from the correlation between cue validity and cue utilization. Indeed, cue validity and cue utilization were in the same direction, above chance (i.e., the correlation between cue validity and cue utilization across all 260 linguistic markers was $r = .52$ and $r = .94$ for the 20 linguistic markers that describe the kernel of truth). These correlations suggest that when others make personality judgments, they utilize linguistic cues that carry valid personality information. For instance, regarding Conscientiousness and Agreeableness, both cue validity and cue utilization focused on the absence of negative emotions and swear words.

What is the theoretical implication of the kernel of truth? According to RAM (Funder, 1995, 2012), the accuracy of personality judgment increases when individuals use cues that are part of the kernel of truth—cue validity and cue utilization are in the same direction (Letzring & Funder, 2021). In RAM's terminology, those cues constitute "good information" (which together with "good judge," "good trait," and "good target" are the four moderators of accurate personality assessment; Funder, 1995, 2012). The present results showed that, when it comes to verbal behavior, the cues that are responsible for accurate personality judgment (i.e., self–other agreement) can be summarized into 20 linguistic categories.

However, we found four additional effect sizes where cue validity and cue utilization were in the opposite direction. These four additional linguistic markers do not belong to the kernel of truth. For instance, self-reported Openness to Experience was *negatively* correlated with family-related words, but observer-reported Openness to Experience was *positively* correlated with family-related words. One possible explanation for the disagreement between cue validity and cue utilization might be that self- and observer reports capture conceptually different personality information (Luft & Ingham, 1955). Self-reports capture one's identity and observer reports capture one's reputation[8] (Hogan, 1991), with each source of personality assessment capturing unique personality variance (Connelly et al., 2022; McAbee & Connelly, 2016). Therefore, cue validity and cue utilization may have been in the opposite direction for some linguistic categories presumably because self and others have unique insights into one's personality. An alternative explanation might be that the disagreement reflects measurement error either because observers drew the wrong personality inference (most observer reports were based on zero-acquaintance tasks) or because some LIWC words might have been misclassified. For instance, in the phrase "I'm anything but happy," the word "happy" would have been classified as a positive emotion, although it is actually negative (naturally, this explanation might also apply to the linguistic categories where cue validity and cue utilization were in the same direction). However, alternative explanations might also be possible, and future research could further study the disagreement between valid and utilized linguistic cues.

A prominent finding in cue utilization was that the linguistic categories that elicited the largest effect sizes ($\rho \geq .30$) were mainly negatively valenced (i.e., words related to swearing, anger, and negative emotions). This finding is in line with the notion that "bad is stronger than good," since, especially in zero-acquaintance tasks, people typically assign more weight to information that is negative, rather than positive (Baumeister et al., 2001). Thus, the results of the present meta-analysis further "unpack" this negative information, illustrating that when making personality judgments, people assign stronger weights to linguistic categories that carry information mainly related to swearing, anger, and negative emotions. This suggests that the accuracy of personality assessment on zero-acquaintance tasks might increase if the tasks require some kind of emotional response from participants, since emotional cues seem to provide the most relevant personality information even in a short period of time (e.g., first-impression tasks, like the majority of the present studies).

Self–Other Knowledge Asymmetry Model

Third, cue utilization was larger than cue validity, both in terms of effect sizes and explained variance. These results are in line with previous findings (e.g., Colvin & Funder, 1991; for a summary, see Breil et al., 2021), which show that cue utilization is generally stronger than cue validity for voice characteristics (Borkenau & Liebler, 1992), nonverbal behavior (Gifford, 1994), environmental cues in bedrooms and offices (Gosling et al., 2002), and digital footprints (Gosling et al., 2011). Those findings are explained under—and provide support to—the SOKA model (Vazire, 2010), according to which, when assessing overt behaviors (like the words people use), others tend to assign stronger weights to those overt behavioral cues, compared to the self (on the contrary, when assessing internal thoughts and feelings, the self is more accurate than others).

Moderators and TAT

Fourth, the present results highlighted the heterogeneity in effect sizes and the significant role of moderators in the linguistic markers of personality. The vast majority of effect sizes (96% in cue validity, 82% in cue utilization) showed significant heterogeneity and the 10 tested moderators explained some of this heterogeneity in 70% of the effect sizes in cue validity and 61% in cue utilization. Concretely, five categorical moderators related to task type (language formality, LIWC version, sample composition, task synchronicity, text mode) were tested only for cue validity, and five continuous moderators related to sample and study characteristics (age, sample size, percentage of women, text length, and year of publication) were tested both for cue validity and cue utilization. The main findings were that longer texts elicited stronger effect sizes, and effect sizes were typically stronger for older LIWC versions.

Regarding the cue validity moderators, the results seem somewhat at odds with previous meta-analyses that have found either none or minimum support for moderating effects on linguistic markers of self-reported Extraversion (Chen et al., 2020), Narcissism (Holtzman et al., 2019), or depression (Edwards & Holtzman, 2017). One restriction of previous meta-analyses was that they focused either on a single or two LIWC categories (with the exception of Holtzman et al., 2019). The present meta-analysis shows that, when considering multiple linguistic categories and personality traits, the linguistic markers of personality are generally affected by study characteristics. Regarding cue utilization, this is the first study to show that linguistic markers of personality are affected by sample and study characteristics, similarly to cue validity.

Two more moderators worth mentioning are the LIWC version and the year of publication. Although the stronger effect sizes for older LIWC versions are in line with previous findings (Chen et al., 2020), this finding seems counterintuitive. Richer dictionaries in more recent LIWC versions allow for more "hits" and, therefore, higher chances of registering a score for individual linguistic categories. However, the results of the present analysis showed that, even though older LIWC dictionaries have less entries per linguistic category, those entries already cover the more characteristic linguistic markers of personality and adding more words does not provide more—personality-relevant—information. One possible explanation might be that the individual words that comprise each linguistic category are not normally distributed, since quite frequently a handful of words account for the majority of "hits" within a linguistic category (see Eichstaedt et al., 2021).

Regarding the year of publication, the results showed that the correlation between some personality traits and negatively valenced words was attenuated in more recent years. This was the case for the negative correlations between (a) Conscientiousness and swear words, (b) Agreeableness and swear words, (c) Agreeableness and anger-related words, and (d) Emotional Stability and negative emotions. A closer inspection showed that the effect sizes were stronger for older LIWC versions. However, even though LIWC version elicited weaker effect sizes in general, LIWC version did not moderate these effect sizes.

[8] Note that Hogan's (1991) "reputation" is mainly operationalized as acquaintances ratings, whereas the observer reports in the present meta-analysis were a mix of acquaintance- (one study) and strangers' ratings (four studies).

One possible explanation for the attenuated correlation between personality traits and negatively valenced linguistic categories could be that, while personality traits have remained stable, the use of profanity (e.g., shit, fuck) and other negatively valenced words (e.g., hurt, hate, kill, damn, anger, angry) has increased over the last 20 years in the English language, presumably following cultural changes in Western societies (see also Twenge et al., 2017). This trend also becomes evident by searching the frequency of these words using Google Book's *n*-gram viewer (https://books.google.com/ngrams). *N*-gram viewer is a combined effort of Google and academics who have digitized "4% of all books ever printed" (Michel et al., 2011, p. 1).

Furthermore, as some authors have suggested that individuals high on honesty tend to swear more (Feldman et al., 2017; Pang et al., 2020), we tested whether negatively valenced words, and more specifically swear words, have become positively valenced over the past few decades. However, we found no evidence for such reversed effects. We believe that these previous findings are due to a misinterpretation of impression management and "lie" scales, which have been used as proxies for honesty (Feldman et al., 2017; Pang & Ring, 2020). That is, instead of a positive relation, self- and other-reported Honesty–Humility have been found to be negatively associated with self-reported profanity, which, in turn, has been found to be positively related to actual cheating (for a relevant discussion, see De Vries et al., 2018).

Regarding TAT, we were especially interested in the moderators of language formality and task synchronicity, since those two moderators may reflect trait-activating situations. According to TAT, traits are activated in relevant situations, and behavior is the result of the interplay between those traits and the relevant situation (Tett & Guterman, 2000; Tett et al., 2021). Indeed, the present results showed that individuals high on self-reported Agreeableness used more swear-related words in informal (vs. formal) situations, as hypothesized. However, we failed to replicate the moderating role of task synchronicity between self-reported Extraversion and positive emotions reported in Chen et al. (2020). One possible reason might be the methodological differences between the two studies, since Chen et al. (2020) did not correct for attenuation, did not have missing values, did not include Holtrop et al. (2019), and included two studies that did not pass the eligibility criteria of the present meta-analysis (Morgan, 2014; Pirzadeh & Pfaff, 2012).

All in all, in the present meta-analysis, the situational moderators—that would provide support for TAT—were relatively weak. This is not at odds with previous literature, where support for TAT is not always consistent (Tett et al., 2021), and effect sizes are relatively small (Koutsoumpis & De Vries, 2022; Sherman et al., 2015). The classification of situations in the present study was broad, something that might have attenuated the moderating effect of situations on verbal behavior. Nevertheless, for some linguistic categories, the moderating role of formality and synchronicity was especially prominent (e.g., the negative relation between self-reported Conscientiousness and anger-related words is better captured in informal tasks; the positive relation between self-reported Extraversion and text length is better captured in synchronous tasks), and researchers interested in the effect of situations on specific linguistic categories might want to consult Supplemental Tables S11 and S14 in the supplemental material for a case-by-case overview.

Other Findings—Function Words

One advantage of LIWC is that it includes function words, which are typically excluded in open-vocabulary approaches (Banks et al., 2018). Although previous research has highlighted their significant role in language (Campbell & Pennebaker, 2003; Chung & Pennebaker, 2007; Pennebaker, 2011; Pennebaker & Stone, 2003; Tausczik & Pennebaker, 2010), the results of the present meta-analysis showed that function words explained, on average, 1.3% of one's self-reported personality. In comparison, content words accounted for 3.7% of the variation in self-reported personality (for cue utilization, these numbers were 5.1% for function and 28.7% for content words).

Regardless of the overall small proportion of explained variance, function words do serve as valid and utilized markers for some personality traits and researchers applying open-vocabulary approaches might need to reconsider the practice of excluding them when preprocessing a text. More importantly, a case-by-case inspection of those relations poses some significant interpretability challenges. Take for instance the positive correlation between self-reported Openness to Experience and articles (ρ = .09). Previous researchers have noted that using articles in a sentence (e.g., *the* house is big vs. *some* house is big) makes a sentence more concrete (Fast & Funder, 2008), and it has been further shown that articles mainly correlate with the intellect and imagination facets of Openness to Experience (Yarkoni, 2010). The presents results showed that, in terms of effect sizes, the positive correlation between Openness to Experience and articles is, for instance, equally strong to the—much more conceptually interpretable—positive relation between self-reported Extraversion and social-related words. What is then the underlying mechanism that explains why individuals who are high on Openness to Experience use more articles? There have been some suggestions on the role of function words (e.g., processed rather automatically occupying different areas in the brain, compared to content words; see Pennebaker, 2011). However, any empirical support for those suggestions is currently missing, and the causal link between function words and personality traits is hardly understood.

LIWC Versus Other Methods

The effect sizes between LIWC and personality traits were small, especially for cue validity. We also note that LIWC has not been developed for the purpose of personality assessment. It is, however, interesting to compare the effect sizes found in this study to those obtained from human judgments or from other text-based assessment methods that were specifically designed to measure personality traits. Table 7 compares the results of the present meta-analysis with four previous meta-analyses and single studies that assessed personality traits from text (written and/or spoken). The comparison is made with the linguistic categories of the present study that showed the strongest effect size per personality trait. To facilitate cross-study comparison, all reported effect sizes are *un*corrected.

In the studies presented in Table 7, personality assessment was made either by humans (Tskhay & Rule, 2014) or automatic text analysis software (computers). In the computer-based personality assessment studies, Moreno et al. (2021) meta-analyzed 23 studies that reported correlations between personality self-reports and written language using computational methods—a blend of computer-based

Table 7
Accuracy of Personality Assessment From Humans and Computers: Cross-Studies Comparison

Assessment source	Humans	Computers			
Study	Tskhay and Rule (2014)	Moreno et al. (2021)	Hickman et al. (2019)	Holtrop et al. (2022)	The present study
Text analysis method	Human-based	Open-vocabulary	Open-vocabulary	Closed-vocabulary	Closed-vocabulary
Tool		Computational text analysis	IBM-WPI	HTTP	LIWC
Type of study	Meta-analysis	Meta-analysis	Single study	Three studies	Meta-analysis
Personality assessment source	Self-reports	Self-reports	Self-reports	Self-reports	Self-reports/observer reports
N	—[a]	97,698	166	72–155	85,724
Openness to Experience	.07	.30	.18	—	.09/.14
Conscientiousness	.11	.27	.05	.19–.22	.10/.28
Extraversion	.32	.27	.06	.36–.38	.07/.27
Agreeableness	.03	.26	.17	—	.11/.32
Emotional Stability	.07	.26	(−).20	.45	.11/.24
Honesty–Humility				.17	
Average	.12	.27	.13	.31	.10/.25

Note. IBM-WPI = International Business Machines Watson Personality Insight; HTTP = Honesty-Humiliaty, Emotionality, eXtraversion, Agreeableness, Conscientiousness, Openness to Experience text-to-personality algorithm; LIWC = Linguistic Inquiry and Word Count; scores are absolute values expressed in Pearson r; to facilitate cross-study comparison, all reported effect sizes are *un*corrected; in LIWC meta-analysis, scores reflect the linguistic category that achieved the strongest effect size per personality trait; all effect sizes are based on personality self-reports, except for LIWC's second value; for average scores, the negative correlation for Emotional Stability in the IBM-WPI measure was treated as positive, and in HTTP only, the strongest correlation coefficients for Conscientiousness and Extraversion were taken into account.
[a] Results were based on a subsample of eight to nine studies (varied per personality traits); the original sample size across 20 studies was 3,599, but the number of participants for this subsample of studies was not reported.

text analysis methods including both machine and nonmachine learning techniques. The two remaining studies applied text-based personality assessment tools specifically developed to measure either the Big Five (IBM Watson Personality Insight [IBM-WPI]; International Business Machines [IBM], 2021) or the HEXACO (HTTP; Holtrop et al., 2022) personality models. The (recently discontinued) IBM-WPI belongs to an open-vocabulary approach and produces personality scores by applying machine learning techniques trained on Twitter data. The HTTP belongs to a closed-vocabulary approach and produces personality scores after weighting individual words based on the loadings that each word has received in previous lexical studies.

When compared to human judgments, the most predictive LIWC categories per personality trait achieved higher accuracies[9] in personality assessment for almost all traits except for Extraversion and—to a lesser extent—Conscientiousness. This is in line with a growing amount of literature suggesting that computer (or mechanical) assessment outperforms humans (Kuncel et al., 2013; Meehl, 1954). When compared to other computer-based personality assessment methods, LIWC achieved lower accuracies.[10] However, LIWC performed slightly better than IBM-WPI for the traits of Conscientiousness and Extraversion (note that IBM-WPI has failed to support construct validity for Emotional Stability—indicated by a negative correlation in Table 7; Hickman et al., 2019).

Practical Implications

The present results have several practical implications, especially for practitioners interested in the generalizability of automatic text-based personality assessment. First, some researchers have already taken the LIWC–personality traits correlations a step further, building LIWC-based predictive models to assess the Big Five (Golbeck, Robles, & Edmondson, 2011; Hickman et al., 2021[11]). For example, Mairesse et al. (2007) applied machine learning techniques to two existing corpora and built the PersonalityRecognizer, a tool for automatic text-based personality assessment based on LIWC categories. Several researchers used this tool as a text-based personality assessment method, instead of directly collecting personality self-reports (e.g., Lima & de Castro, 2014; Rigby & Hassan, 2007; Tommasel et al., 2016). The present meta-analysis showed that personality traits are indeed significantly correlated with LIWC categories, but these correlations are weak, especially for cue validity. Therefore, generalizing the relatively small effect sizes of one study to

[9] Note that by focusing on the strongest LIWC predictors per personality trait (instead, e.g., on the average of all LIWC categories or the average explained variance), the comparison between LIWC and other methods might be unfair. Note, also, that such a—possibly—unfair approach is only possible post hoc. Nevertheless, the main message is that, based on the results of the present meta-analysis, future researchers interested in measuring self-reported personality traits from text, might rely only on a handful of LIWC categories that are more strongly correlated with each personality trait.

[10] Note, however, that LIWC is based on a straightforward counting method. This means that it is unqualified to understand the context and subtle nuances of language (e.g., irony, humor) like human judges do (McAuliffe et al., 2020; Pang & Ring, 2020), it does not analyze text using sophisticated machine learning approaches like the IBM-WPI does (IBM, 2021), it is not based on personality lexical studies like the HTTP (Holtrop et al., 2022), neither was it developed to measure personality in the first place.

[11] Hickman et al. (2021) also tested the relation between LIWC and personality, but it was not included in the present meta-analysis since it was published after the database search was completed.

other data sets (e.g., when making decisions in selection contexts) warrants caution.

Second, the categorical moderators of the present analysis showed that verbal behavior is not only affected by the context (see LIWC norms per text type; Pennebaker et al., 2001, 2007, 2015) but it is the result of the interaction between the person and the situation (Lewin, 1936; Tett & Guterman, 2000; Tett et al., 2021). For instance, the mode (spoken–written) and formality of language (formal–informal) interact with personality traits and affect the words people use. This might explain part of the underperformance of some text-based personality tools. Indeed, when the IBM-WPI (IBM, 2021) was applied to transcripts of job interviews, the accuracy of the text-based personality assessment was significantly lower than originally advertised or even failed to provide construct validity (Hickman et al., 2019). In light of the present results, it seems reasonable to expect that an algorithm trained on written, informal language (IBM-WPI was trained on Twitter data) does not perform as well when text comes from other contexts (e.g., spoken, formal language). Therefore, practitioners interested in using text-based personality assessment tools should be aware that applying a tool developed in a specific context to a different one carries the risk of inaccurate assessment.

Finally, LIWC-based personality assessment explains smaller proportions of personality variance compared to other computer-based approaches (at least for cue validity). Researchers and practitioners interested in maximizing the explained variance of text-based personality assessment should consider open-vocabulary approaches (note that in the absence of validated off-the-shelf automatic text-based personality tools, one should first build these models before applying them). However, there is a trade-off between explained variance and "explainability" (i.e., transparency): open-vocabulary approaches (especially nontransparent ones) explain more personality variance but provide less information on the words or linguistic categories related to personality traits. On the other hand, LIWC explains less personality variance, but it is more transparent and makes clear what linguistic categories (but not what individual words) are relevant for each personality trait. Thus, practitioners might use the present results to showcase the transparency of personality text-based tools. However, given the small effect sizes for cue validity, practitioners should avoid using LIWC as a stand-alone instrument. Instead, when prudently used (e.g., when a trait has been previously activated, when enough text is available), or when someone is interested in cue utilization, LIWC can be used in combination with other methods of personality assessment.

Limitations

The present meta-analysis also has some limitations. One limitation is that the meta-analytic effect sizes are probably an underestimation of the true effects, since the corrected effect sizes were disattenuated only for personality measures but not for linguistic categories. We decided not to correct for LIWC unreliability (even though Cronbach's α scores are available in LIWC2007 and LIWC2015 manuals) mainly for two reasons. On the one hand, LIWC categories are *formative* measures, that is, the items do not have causal relationship with the latent linguistic category (unlike *reflective* measures, like personality traits, where the latent construct has a causal relation with the items; Diamantopoulos & Siguaw, 2006). On the other hand, the "items" that make each linguistic category (a mix of words and word stems) are not always normally distributed (Eichstaedt et al., 2021; Yarkoni, 2010), and there is an overlap across broad and narrow categories, with some words belonging to multiple categories (Pennebaker et al., 2001, 2007, 2015). This is different from typical questionnaires, where a single item belongs only to one construct. For both aforementioned reasons, Cronbach's α is not the optimal measure of reliability. An alternative approach would be to use test–retest reliability. However, the limitation of this approach is that it might artificially weaken reliability estimates, since participants who are requested to reply verbally to the same task twice, might respond differently because they believe they need to provide new information (Tausczik & Pennebaker, 2010).

Another limitation was that cue utilization was based on a small number of studies ($k = 5$; compared to $k = 26$ for cue validity). The small number of cue utilization studies has at least five implications. First, it was not possible to test the role of categorical moderators in cue utilization. Second, it was not possible to further explore how the level of acquaintance (rating acquaintances vs. strangers) or the type of information (personality assessment based on short interactions, video, audio, or text) might have affected cue utilization. Third, a typical finding in lens model studies, that cue utilization elicits more numerous (as well as stronger) effect sizes compared to cue validity, was not observed. Fourth, the number of linguistic markers that comprise the kernel of truth is probably an underestimation; if the number of studies for cue validity and cue utilization was more balanced, we would have probably observed a larger number of linguistic categories in the kernel of truth (even though the present results should already have captured the strongest of them). Fifth, since cue validity and cue utilization came from studies with dissimilar text data and participants (except for three studies), it was not possible to calculate the achievement index, that is, the correlation between self- and observer personality reports—but see Tskhay and Rule (2014) for self–other agreement when others assess personality from text; and Karelaia and Hogarth (2008), and Kaufmann and Athanasou (2009) for the achievement index across lens studies in general.

Finally, some LIWC versions are also available in other languages (e.g., LIWC2007 and LIWC2015 are available in Portuguese, Chinese, Dutch, French, German, Italian, and Japanese among others; see https://www.liwc.app/dictionaries). However, the present meta-analysis focused on the English LIWC. Therefore, the present results are English-specific, and it is not certain if they generalize to other languages.

Suggestions for Future Research

The present results still leave a number of unanswered questions, and we offer some suggestions for future research. First, previous research suggests that the strength of effect sizes for cue utilization seems to become stronger as the level of acquaintance between the self and others decreases, for instance, in zero-acquaintance tasks (Funder & Sneed, 1993; Sneed et al., 1998). One explanation is that raters in zero-acquaintance tasks base their judgments on a single task (e.g., thin slice of a video; Borkenau & Liebler, 1992; Borkenau et al., 2004), whereas close acquaintances base their rating on their personal relationship with the target, which might include numerous interactions across multiple contexts. Thus, strangers might assign more weight to the linguistic markers of personality because this is the only available information they have from the participant

(Colvin & Funder, 1991; Nisbett et al., 1973). So far, previous research generally focuses on a single level of acquaintance (typically strangers), and in a few studies where multiple levels of acquaintance were taken into account, behavior was based on summary items (e.g., "is talkative"; Sneed et al., 1998). LIWC offers an objective index of verbal behavior (e.g., word count instead of subjectively rated talkativeness) but whether the level of acquaintance affects the strength of cue utilization remains an open empirical question.

Another related topic for future research is the type of information on which observers base their personality assessments. Previous studies show that the accuracy of human judges when assessing personality traits from text is on average $r = .12$ (Tskhay & Rule, 2014). In the present meta-analysis, the effect sizes for cue utilization were generally stronger. However, in the present meta-analysis, observer reports were not only based on text but were collected either from acquaintances (Baek & Ihm, 2020) or from strangers after a short online interaction with the participant (Sandy, 2013), watching a video (Biel et al., 2013), listening to an audio file (Mehl et al., 2006), or from text (Qiu et al., 2012). Previous research suggests that all abovementioned types of information are sufficient for accurate personality assessment (Connelly & Ones, 2010). However, each type of information explains unique personality variance (e.g., Biel et al., 2013). For instance, in online video interviews, verbal behavior is more informative of Openness to Experience, and nonverbal behavior is more informative of Extraversion (Hickman et al., 2022). More importantly, sources with richer information (e.g., video instead of text) typically lead to more accurate personality assessment (Connelly & Ones, 2010). Thus, in the present study, cue utilization might also have been affected by—or intertwined with—nonverbal cues (e.g., facial expressions, voice loudness) and prior information (in the case of acquaintance). Future research can further explore how information richness affects the utilized linguistic markers of personality and/or explore whether utilized linguistic cues explain unique personality variance over and beyond nonverbal cues (e.g., facial expressions, voice characteristics) and other meta information of the written/spoken text (e.g., structure, style).

Furthermore, LIWC research has been studied almost exclusively under the Big Five personality model, although its successor, the HEXACO (Lee & Ashton, 2004), captures larger personality variation and is the most widely replicable cross-cultural personality structure (Ashton & Lee, 2020). HEXACO operationalizes the domains of Emotionality and Agreeableness in slightly different ways compared to the Big Five, but its main difference is the addition of the Honesty–Humility domain. Consequently, at this point, one important personality dimension in LIWC research has not been investigated. Researchers have started exploring the relation between LIWC and HEXACO (Holtrop et al., 2019), but the linguistic markers of Honesty–Humility still remain open for investigation.

Finally, even though the present meta-analysis helps to make automatic text-based personality tools more explainable, there are at least two ways in which explainability can be improved. On the one hand, LIWC makes clear what linguistic categories—but not what individual words[12]—are relevant for each personality trait. To further "unpack" the content of linguistic categories, future research could explore the theoretical links between personality traits and verbal behavior in combination with transparent open-vocabulary approaches (Eichstaedt et al., 2021). On the other hand, even though function words do serve as valid linguistic markers, as long as the underlying mechanism remains unknown, the explainability of linguistic markers of personality will be undermined. Thus, future research should focus on the reasons and the conditions under which function words relate to personality traits.

Final Remarks

In the present study, LIWC categories were correlated with self- and observer reports of personality. This approach might suggest to readers that self- and observer reports (using, for instance, personality inventories) are "ground truths" of personality traits, and that LIWC categories should be aligned to such ground truths. However, personality inventories are just one of the many ways in which personality can be assessed. For instance, in the first days of psychology, behavior had a central role in assessing personality traits (Baumeister et al., 2007). In later years, researchers started to overrely on personality inventories because they facilitated data collection—a practice that has been criticized (e.g., Baumeister et al., 2007; Gerpott & Lehmann-Willenbrock, 2022). However, besides inventories' convenience in assessing personality traits, there is no good reason to privilege them over other methods. In fact, one may only be able to increase the so-called (but still subjective) "accuracy" of personality description by combining multiple assessment sources and methods (Back & Nestler, 2016; Boyd et al., 2020; Funder, 2015). These methods might include self-reports, observer reports, life outcomes, and behavioral measures of personality (e.g., written/spoken text).

Given the results of the present meta-analysis, one question that might arise is whether text-based personality assessment methods (e.g., LIWC categories) can be part of a multimethod assessment of personality traits (and/or whether they add incremental validity over other methods). For this, text-based personality assessment tools need to demonstrate criterion validity, comparable to other methods. Until now, research on the criterion validity of these tools is scarce. Regarding LIWC, the research on its criterion validity is, so far, inconclusive for several reasons. First, LIWC has mainly been used for exploratory analyses instead of prediction. When the goal was to predict criterion variables, quite often researchers have used custom-built LIWC scales, for instance, after factor analyzing or selecting certain linguistic categories (e.g., Pennebaker et al., 2014; Robinson et al., 2013). The problem with custom-built scales is that, even though they serve the research goals of individual studies, it is difficult to generalize the results to other studies. Furthermore, these studies do not clarify what individual linguistic categories relate to the criterion variables. Second, some researchers have used Receptiviti (https://www.receptiviti.com/liwc), the commercial spin-off of LIWC (e.g., Roulin & Stronach, 2022). Receptiviti provides personality scores from text but not the correlations between individual LIWC categories with criterion variables. Finally, a number of studies are based on small sample sizes (e.g., D'Andrea et al., 2012; Pressman & Cohen, 2007, 2012; Rohrbaugh et al., 2012). The problem with small sample sizes is that only very large effect sizes reach statistical significance and are being reported.

[12] Note that all LIWC versions include a pdf version of the internal dictionary which researchers can download. This allows researchers to get access to the individual words and word stems per linguistic category. However, LIWC does not provide scores for individual words or word stems.

In conclusion, text-based personality assessment methods (in general, and LIWC in particular) are gaining momentum and have started showing promising results. For instance, one language-based assessment tool has shown test–retest reliability, over several months, comparable to that of personality inventories (Park et al., 2015). However, most current text-based personality assessment methods face two main challenges. First, they are compared against personality inventories as "ground truths." This is not a bad practice to establish construct and convergent validity. However, in the future, text-based personality methods should also be studied directly in relation with criterion variables (either circumventing personality inventories or exploring whether text-based assessment offers incremental validity over inventories). Second, thus far, research on text-based personality assessment has typically used a bottom-up approach, lacking theory, research on psychometric properties, and external validity (with notable exceptions, e.g., Eichstaedt et al., 2021; Park et al., 2015). The present meta-analysis takes steps toward addressing some of these challenges. For instance, the 20 linguistic markers that constitute the kernel of truth seem to capture core components of self- and observer-reported personality, thus providing a possible alternative text-based approach in capturing personality traits. Once future studies have addressed the abovementioned challenges, text-based personality assessment methods (e.g., LIWC categories, by virtue of their explainable relations between language use and personality traits) might be a promising tool to be included in multimethod assessment of personality traits.

Conclusion

Can we measure personality from text, and if yes, what are the linguistic markers of personality traits? The present meta-analysis summarized the empirical findings over the last 20 years on the relations between the Big Five personality traits and 52 linguistic categories measured with LIWC in the English language. The results showed that the LIWC categories serve as linguistic markers of personality traits since (a) individuals express their self-reported personality traits through specific linguistic cues, (b) others are able to infer personality traits through specific linguistic cues, and (c) these correlations are generally in line with personality theory. It was further shown that the linguistic markers of personality have a kernel of truth which consists of 20 utilized valid LIWC categories. Finally, the 10 moderators that were tested to explain the heterogeneity in effect sizes (sample size, gender, age, text length, year of publication, language mode, task synchronicity, LIWC version, sample composition, language formality) mainly suggested that task characteristics had a significant effect on the cue validity and cue utilization of linguistic markers of personality. All in all, the findings show that—at least to some extent—personality traits can indeed be measured from text, and LIWC categories help to explain how personality traits are related to spoken and written text.

References

References marked with an asterisk indicate studies included in the meta-analysis

*Abe, J. A. A. (2018). Personality and political preferences: The 2016 U.S. Presidential Election. *Journal of Research in Personality*, *77*, 70–82. https://doi.org/10.1016/j.jrp.2018.09.001

*Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *The Internet and Higher Education*, *45*, Article 100724. https://doi.org/10.1016/j.iheduc.2019.100724

Almodaresi, F., Ungar, L., Kulkarni, V., Zakeri, M., Giorgi, S., & Schwartz, H. A. (2017, July–August). *On the distribution of lexical features in social media* [Paper presentation]. Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada.

Ashton, M. C., & Lee, K. (2005). Honesty–humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1354. https://doi.org/10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Back, M. D., & Nestler, S. (2016). Judging personality. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge University Press., https://doi.org/10.1017/CBO9781316181959.005

*Baddeley, J. L., & Singer, J. A. (2008). Telling losses: Personality correlates and functions of bereavement narratives. *Journal of Research in Personality*, *42*(2), 421–438. https://doi.org/10.1016/j.jrp.2007.07.006

*Baek, Y. M., & Ihm, J. (2020). Word use as an unobtrusive predictor of early departure from organizations. *Journal of Language and Social Psychology*, *40*(2), 238–259. https://doi.org/10.1177/0261927X20944543

Baldwin, A. L. (1942). Personal structure analysis: A statistical method for investigating the single personality. *Journal of Abnormal and Social Psychology*, *37*(2), 163–183. https://doi.org/10.1037/h0061697

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*(4), 445–459. https://doi.org/10.1007/s10869-017-9528-3

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beer, A. (2021). Information as a moderator of accuracy in personality judgment. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 132–148). Oxford University Press.

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality*, *45*(2), 175–185. https://doi.org/10.1016/j.jrp.2011.01.001

Beersma, B., Hollenbeck, J. R., Humphrey, S. E., Moon, H., Conlon, D. E., & Ilgen, D. R. (2003). Cooperation, competition, and team performance: Toward a contingency approach. *Academy of Management Journal*, *46*(5), 572–590. https://doi.org/10.2307/30040650

*Biel, J. I., Tsiminaki, V., Dines, J., & Gatica-Perez, D. (2013, December 9–13). *Hi youtube!: Personality impressions and verbal content in social video* [Conference session]. Proceedings of the 15th ACM on International Conference on Multimodal Interaction, Sydney, NSW, Australia. https://doi.org/10.1145/2522848.2522894

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). Wiley. https://doi.org/10.1002/9781119558378

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*(4), 645–657. https://doi.org/10.1037/0022-3514.62.4.645

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614. https://doi.org/10.1037/0022-3514.86.4.599

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*(2), 431–449. https://doi.org/10.1037/a0038047

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin.

Boyd, R. L., Pasca, P., & Lanning, K. (2020). The personality panorama: Conceptualizing personality through big behavioural data. *European Journal of Personality*, *34*(5), 599–612. https://doi.org/10.1002/per.2254

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Bradley, B. H., Baur, J. E., Banford, C. G., & Postlethwaite, B. E. (2013). Team players and collective performance: How Agreeableness affects team performance over time. *Small Group Research*, *44*(6), 680–711. https://doi.org/10.1177/1046496413507609

Breil, S. M., Osterholz, S., Nestler, S., & Back, M. D. (2021). Contributions of nonverbal cues to the accurate judgment of personality traits. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 193–218). Oxford University. https://doi.org/10.1093/oxfordhb/9780190912529.013.13

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Burdick, L., Mihalcea, R., Boyd, R. L., & Pennebaker, J. W. (2021). Analyzing connections between user attributes, images, and text. *Cognitive Computation*, *13*(2), 241–260. https://doi.org/10.1007/s12559-019-09695-3

Campbell, R. S., & Pennebaker, J. W. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science*, *14*(1), 60–65. https://doi.org/10.1111/1467-9280.01419

Chen, J., Qiu, L., & Ho, M. H. R. (2020). A meta-analysis of linguistic markers of extraversion: Positive emotion and social process words. *Journal of Research in Personality*, *89*, Article 104035. https://doi.org/10.1016/j.jrp.2020.104035

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343–359). Psychology Press. https://doi.org/10.4324/9780203837702

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology*, *60*(6), 884–894. https://doi.org/10.1037/0022-3514.60.6.884

Connelly, B. S., McAbee, S. T., Oh, I.-S., Jung, Y., & Jung, C.-W. (2022). A multirater perspective on personality and performance: An empirical examination of the trait-reputation-identity model. *Journal of Applied Psychology*, *107*(8), 1352–1368. https://doi.org/10.1037/apl0000732

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

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEOFFI): Professional manual*. Psychological Assessment Resources.

D'Andrea, W., Chiu, P. H., Casas, B. R., & Deldin, P. (2012). Linguistic predictors of post-traumatic stress disorder symptoms following 11 September 2001. *Applied Cognitive Psychology*, *26*(2), 316–323. https://doi.org/10.1002/acp.1830

De Vries, R. E., Hilbig, B. E., Zettler, I., Dunlop, P. D., Holtrop, D., Lee, K., & Ashton, M. C. (2018). Honest people tend to use less—Not more—Profanity: Comment on Feldman et al.'s (2017) Study 1. *Social Psychological & Personality Science*, *9*(5), 516–520. https://doi.org/10.1177/1948550617714586

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies* (Vol. 1, pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282. https://doi.org/10.1111/j.1467-8551.2006.00500.x

Edwards, T., & Holtzman, N. S. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, *68*, 63–68. https://doi.org/10.1016/j.jrp.2017.02.005

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), Article 629. https://doi.org/10.1136/bmj.315.7109.629

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., Buffone, A., Iwry, J., Seligman, M. E. P., & Ungar, L. H. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, *26*(4), 398–427. https://doi.org/10.1037/met0000349

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

Fayard, J. V., Roberts, B. W., Robins, R. W., & Watson, D. (2012). Uncovering the affective core of conscientiousness: The role of self-conscious emotions. *Journal of Personality*, *80*(1), 1–32. https://doi.org/10.1111/j.1467-6494.2011.00720.x

Feldman, G., Lian, H., Kosinski, M., & Stillwell, D. (2017). Frankly, we do give a damn: The relationship between profanity and honesty. *Social Psychological & Personality Science*, *8*(7), 816–826. https://doi.org/10.1177/1948550616681055

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. https://doi.org/10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*(3), 177–182. https://doi.org/10.1177/0963721412445309

Funder, D. C. (2015). *The personality puzzle* (7th ed.). W.W. Norton.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, *64*(3), 479–490. https://doi.org/10.1037/0022-3514.64.3.479

Gerpott, F. H., & Lehmann-Willenbrock, N. (2022). Perceived and actual behaviors in research on age and work. In H. Zacher & C. Rudolph (Eds.), *Age and work: Advances in theory, methods, and practice* (pp. 135–151). Routledge. https://doi.org/10.4324/9781003089674-11

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, *66*(2), 398–412. https://doi.org/10.1037/0022-3514.66.2.398

*Gill, A. J., Nowson, S., & Oberlander, J. (2006). *Language and personality in computer-mediated communication: A cross-genre comparison*. [Unpublished manuscript]. https://nl.ijs.si/janes/wp-content/uploads/2014/09/gillothers06.pdf

*Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011, October). *Predicting personality from twitter* [Conference session]. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and in 2011 IEEE Third International Conference on Social Computing, Boston, MA, USA.

*Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CH'11 extended abstracts on human factors in computing systems* (pp. 253–262). https://doi.org/10.1145/1979742.1979614

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervield, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior and Social Networking*, *14*(9), 483–488. https://doi.org/10.1089/cyber.2010.0087

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398. https://doi.org/10.1037/0022-3514.82.3.379

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

*Graybeal, A., Sexton, J. D., & Pennebaker, J. W. (2002). The role of story-making in disclosure writing: The psychometrics of narrative. *Psychology & Health*, *17*(5), 571–581. https://doi.org/10.1080/08870440290025786

Haaland, S., & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center rating. *Personnel Psychology*, *55*(1), 137–163. https://doi.org/10.1111/j.1744-6570.2002.tb00106.x

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

*Hawkins, R. C., II, & Boyd, R. L. (2017). Such stuff as dreams are made on: Dream language, LIWC norms, and personality correlates. *Dreaming*, *27*(2), 102–121. https://doi.org/10.1037/drm0000049

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, *107*(8), 1323–1351. https://doi.org/10.1037/apl0000695

Hickman, L., Saef, R., Ng, V., Woo, S. E., Tay, L., & Bosch, N. (2021). Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews. *Human Resource Management Journal*. Advance online publication. https://doi.org/10.1111/1748-8583.12356

Hickman, L., Tay, L., & Woo, S. E. (2019). Validity evidence for off-the-shelf language-based personality assessment using video interviews: Convergent and discriminant relationships with self and observer ratings. *Personnel Assessment and Decisions*, *5*(3), Article 3. https://doi.org/10.25035/pad.2019.03.003

*Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, *43*(3), 524–527. https://doi.org/10.1016/j.jrp.2009.01.006

Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 873–919). Consulting Psychologists Press.

Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of Research in Personality*, *45*(1), 92–99. https://doi.org/10.1016/j.jrp.2010.11.015

*Holtrop, D., Van Breda, W., Oostrom, J. K., & De Vries, R. E. (2019). *Predicting faking in interviews with automated text analysis and personality* [Conference session]. Proceedings of the EAWOP Congress: European Association of Work and Organizational Psychology (EAWOP), Turin, Italy.

Holtrop, D. J., Oostrom, J. K., Van Breda, W. R. J., Koutsoumpis, A., & De Vries, R. E. (2022). Exploring the application of text-to-personality algorithms in job interviews. *European Journal of Work and Organizational Psychology*, *31*(6), 1–18. https://doi.org/10.1080/1359432X.2022.2051484

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C., Deters, F. G., Back, M. D., Donnellan, M. B., Pennebaker, J. W., Sherman, R. A., & Mehl, M. R. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *Journal of Language and Social Psychology*, *38*(5–6), 773–786. https://doi.org/10.1177/0261927X19871084

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, *8*(4), 275–292. https://doi.org/10.1111/1468-2389.00156

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, *7*(2), 265–290. https://doi.org/10.1017/langcog.2014.30

International Business Machines. (2021). *Watson PI documentation*. https://cloud.ibm.com/docs/services/personality-insights/science.html

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History. measurement. and theoretical perspectives. In L. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–13). Guilford Press.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, *92*(1), 107–127. https://doi.org/10.1037/0021-9010.92.1.107

*Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, *120*(2), 263–286. https://doi.org/10.2307/20445398

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404

Kaufmann, E., & Athanasou, J. A. (2009). A meta-analysis of judgment achievement as defined by the lens model equation. *Swiss Journal of Psychology*, *68*(2), 99–112. https://doi.org/10.1024/1421-0185.68.2.99

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., Kosinski, M., Ramones, S. M., & Seligman, M. E. (2014).

The online social self: An open vocabulary approach to personality. *Assessment*, *21*(2), 158–169. https://doi.org/10.1177/1073191113514104

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768–821. https://doi.org/10.1037/a0020327

Koutsoumpis, A., & De Vries, R. E. (2022). What does your voice reveal about you? Trait activation of voice characteristics and their relation with personality and communication styles. *Journal of Individual Differences*, *43*(3), 160–167. https://doi.org/10.1027/1614-0001/a000362

*Krieger, K. (2016). *Words of well-being: The relation of an individual's word choice to their social well-being* [Master's thesis]. Oregon State University. https://ir.library.oregonstate.edu

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060–1072. https://doi.org/10.1037/a0034156

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182–191. https://doi.org/10.1037/1040-3590.18.2.182

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the dark triad and honesty–humility. *European Journal of Personality*, *27*(2), 169–184. https://doi.org/10.1002/per.1860

Letzring, T. D., & Funder, D. C. (2021). The realistic accuracy model. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 9–22). Oxford University Press.

Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill. https://doi.org/10.1037/10019-000

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, *91*(2), 247–258. https://doi.org/10.1037/0021-9010.91.2.247

Lima, A. C. E., & de Castro, L. N. (2014). A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, *58*, 122–130. https://doi.org/10.1016/j.neunet.2014.05.020

Luft, J., & Ingham, H. (1955). The Johari window, a graphic model of interpersonal awareness. *Proceedings of the Western Training Laboratory in Group Development*, *246*, 2014–2003.

*Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, *123*(5), 569–591. https://doi.org/10.1037/rev0000035

McAuliffe, W. H., Moshontz, H., McCauley, T. G., & McCullough, M. E. (2020). Searching for prosociality in qualitative data: Comparing manual, closed-vocabulary, and open-vocabulary methods. *European Journal of Personality*, *34*(5), 903–916. https://doi.org/10.1002/per.2240

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. https://doi.org/10.1037/11281-000

*Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862

*Mehl, M. R., Robbins, M. L., & Holleran, S. E. (2012). How taking a word for a word can be problematic: Context-dependent linguistic markers of extraversion and neuroticism. *Journal of Methods and Measurement in the Social Sciences*, *3*(2), 30–50. https://doi.org/10.2458/v3i2.16477

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., Aiden, E. L., & the Google Books Team. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, May 2–3). *Efficient estimation of word representations in vector space* [Conference session]. 1st International Conference on Learning Representations, Workshop Track Proceedings, Scottsdale, AZ, USA. https://doi.org/10.48550/arXiv.1301.3781

Moreno, J. D., Martínez-Huertas, J. Á., Olmos, R., Jorge-Botana, G., & Botella, J. (2021). Can personality traits be measured analyzing written language? A meta-analytic study on computational methods. *Personality and Individual Differences*, *177*, Article 110818. https://doi.org/10.1016/j.paid.2021.110818

Morgan, J. L. (2014). *Effects of personality, communication, and cross-training on virtual team performance* [Doctoral dissertation]. The University of Alabama in Huntsville.

Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, *27*(2), 154–164. https://doi.org/10.1037/h0034779

Osterholz, S., Breil, S. M., Nestler, S., & Back, M. D. (2021). Lens and dual lens models. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 44–60). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, Article 105906. https://doi.org/10.1016/j.ijsu.2021.105906

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

Pang, J. S., & Ring, H. (2020). Automated coding of implicit motives: A machine-learning approach. *Motivation and Emotion*, *44*(4), 549–566. https://doi.org/10.1007/s11031-020-09832-8

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, *211*(2828), 42–45. https://doi.org/10.1016/S0262-4079(11)62167-2

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC2007: Linguistic inquiry and word count*. https://LIWC.net

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), Article e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., & Francis, M. E. (1999). *Linguistic inquiry and word count (LIWC): A computer-based text analysis program*. Lawrence Erlbaum.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Lawrence Erlbaum.

*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

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language. Use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, *85*(2), 291–301. https://doi.org/10.1037/0022-3514.85.2.291

Pirzadeh, A., & Pfaff, M. S. (2012, October). *Emotion expression under stress in instant messaging* [Paper presentation]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. https://doi.org/10.1177/1071181312561051

Pressman, S. D., & Cohen, S. (2007). Use of social words in autobiographies and longevity. *Psychosomatic Medicine*, *69*(3), 262–269. https://doi.org/10.1097/PSY.0b013e31803cb919

Pressman, S. D., & Cohen, S. (2012). Positive emotion word use and longevity in famous deceased psychologists. *Health Psychology*, *31*(3), 297–305. https://doi.org/10.1037/a0025339

Purdon, C., & Clark, D. A. (2000). White bears and other elusive intrusions. Assessing the relevance of thought suppression for obsessional phenomena. *Behavior Modification*, *24*(3), 425–453. https://doi.org/10.1177/0145445500243008

*Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, *46*(6), 710–718. https://doi.org/10.1016/j.jrp.2012.08.008

Rigby, P. C., & Hassan, A. E. (2007, May). *What can OSS mailing lists tell us? A preliminary psychometric text analysis of the apache developer mailing list* [Conference session]. Fourth International Workshop on Mining Software Repositories (MSR'07: ICSE Workshops 2007), Minneapolis, MN, USA. https://doi.org/10.1109/MSR.2007.35

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

Rohrbaugh, M. J., Shoham, V., Skoyen, J. A., Jensen, M., & Mehl, M. R. (2012). We-talk, communal coping, and cessation success in a couple-focused intervention for health-compromised smokers. *Family Process*, *51*(1), 107–121. https://doi.org/10.1111/j.1545-5300.2012.01388.x

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Roulin, N., & Stronach, R. (2022). LinkedIn-based assessments of applicant personality, cognitive ability, and likelihood of organizational citizenship behaviors: Comparing self-, other-, and language-based automated ratings. *International Journal of Selection and Assessment*, *30*(4), 503–525. https://doi.org/10.1111/ijsa.12396

Rupinski, M. T., & Dunlap, W. P. (1996). Approximating Pearson product-moment correlations from Kendall's tau and Spearman's rho. *Educational and Psychological Measurement*, *56*(3), 419–429. https://doi.org/10.1177/0013164496056003004

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science*, *9*(5), 538–551. https://doi.org/10.1177/1745691614543972

*Sandy, C. J. (2013). *Predicting accuracy in first impressions based on language use in computer-mediated communication environments* [Doctoral dissertation, University of Texas at Austin]. ProQuest Dissertations and Theses Global.

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

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). Springer. https://doi.org/10.1007/978-3-319-21416-0

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, *109*(5), 872–888. https://doi.org/10.1037/pspp0000036

Sneed, C. D., McCrae, R. R., & Funder, D. C. (1998). Lay conceptions of the five-factor model and its indicators. *Personality and Social Psychology Bulletin*, *24*(2), 115–126. https://doi.org/10.1177/0146167298242001

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Stern, J., Schild, C., Jones, B. C., DeBruine, L. M., Hahn, A., Puts, D. A., Zettler, I., Kordsmeyer, T. L., Feinberg, D., Zamfir, D., Penke, L., & Arslan, R. C. (2021). Do voices carry valid information about a speaker's personality? *Journal of Research in Personality*, *92*, Article 104092. https://doi.org/10.1016/j.jrp.2021.104092

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. MIT Press.

*Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). *Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets* [Conference session]. 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA. https://doi.org/10.1109/ICMLA.2012.218

*Sumner, C., Byers, A., & Shearing, M. (2011). Determining personality traits & privacy concerns from Facebook activity. *Black Hat Briefings*, *11*(7), 197–221.

*Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley, S. A., Polsinelli, A. J., & Mehl, M. R. (2020). 'Personality in its natural habitat' revisited: A pooled, multi-sample examination of the relationships between the Big Five personality traits and daily behaviour and language use. *European Journal of Personality*, *34*(5), 753–776. https://doi.org/10.1002/per.2283

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*(4), 397–423. https://doi.org/10.1006/jrpe.2000.2292

Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature and applications to five lines of personality dynamics research. *Annual Review of Organizational Psychology and Organizational Behavior*, *8*(1), 199–233. https://doi.org/10.1146/annurev-orgpsych-012420-062228

Thielmann, I., Zimmermann, J., Leising, D., Hilbig, B. E., & Back, M. (2017). Seeing is knowing: On the predictive accuracy of self–and informant reports for prosocial and moral behaviours. *European Journal of Personality*, *31*(4), 404–418. https://doi.org/10.1002/per.2112

Tommasel, A., Corbellini, A., Godoy, D., & Schiaffino, S. (2016). Personality-aware followee recommendation algorithms: An empirical analysis. *Engineering Applications of Artificial Intelligence*, *51*, 24–36. https://doi.org/10.1016/j.engappai.2016.01.016

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*, 25–30. https://doi.org/10.1016/j.jrp.2013.12.004

Twenge, J. M., VanLandingham, H., & Keith Campbell, W. (2017). The seven words you can never say on television: Increases in the use of swear words in American books, 1950–2008. *SAGE Open*, *7*(3), Article 2158244017723689. https://doi.org/10.1177/2158244017723689

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. https://doi.org/10.1037/a0017908

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*(5), 1202–1216. https://doi.org/10.1037/a0013314

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equation modeling. *Personnel Psychology*, *48*(4), 865–885. https://doi.org/10.1111/j.1744-6570.1995.tb01784.x

Vize, C. E., Miller, J. D., & Lynam, D. R. (2021). Examining the conceptual and empirical distinctiveness of Agreeableness and “dark” personality items. *Journal of Personality*, *89*(3), 594–612. https://doi.org/10.1111/jopy.12601

Wegner, D. M., Schneider, D. J., Carter, S. R., III, & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, *53*(1), 5–13. https://doi.org/10.1037/0022-3514.53.1.5

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, Version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6–10. https://doi.org/10.3758/BF03202594

*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

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science*, *15*(3), 723–760. https://doi.org/10.1177/1745691619895036

Received April 8, 2022
Revision received November 1, 2022
Accepted November 2, 2022 ■