Personality and Individual Differences 177 (2021) 110818

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Review

Can personality traits be measured analyzing written language? A meta-analytic study on computational methods

José David Moreno [a], José Á. Martínez-Huertas [a,*], Ricardo Olmos [a], Guillermo Jorge-Botana [b], Juan Botella [a]

[a] Universidad Autónoma de Madrid, Calle Iván Pavlov, 6, 28049 Madrid, Spain
[b] Universidad Nacional de Educación a Distancia, Calle de Juan del Rosal, 10, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Meta-analysis
Personality
Big five
Language
Computational models of language

ABSTRACT

In the last two decades, empirical evidence has shown that personality traits could be related to the characteristics of written language. This study describes a meta-analysis that synthesizes 23 independent estimates of the correlations between the Big Five major personality traits, and some computationally obtained indicators from written language. The results show significant combined estimates of the correlations, albeit small to moderate according to Cohen's conventions to interpret effect sizes, for the five traits (between $r = 0.26$ for agreeableness and neuroticism, and 0.30 for openness). These estimates are moderated by the type of information in the texts, the use of prediction mechanisms, and the source of publication of the primary studies. Generally, the same effective moderators operate for the five traits. It is concluded that written language analyzed through computational methods could be used to extract relevant information of personality. But further research is still needed to consider it as predictive or explanatory tool for individual differences.

1. Introduction

Over the last decades, big constructs have been proposed from several theoretical perspectives to define personality. There is a general consensus about a few constructs that can exhaustively explain personality profiles, namely the personality traits. There is a large variety of models to conceptualize personality as, for example, the 16PF (Cattell & Eber, 1950), the EPI (Eysenck, 1968), the MMPI (Hathaway & McKinley, 1951), or the HEXACO (Ashton & Lee, 2007), including models which cover alternative facets of personality like the Dark Triad or Tetrad (e.g., Paulhus & Williams, 2002; Veselka, Schermer, & Vernon, 2011), being the *Big Five* one of the most used models (McCrae & Costa Jr., 2008). All of these models measure personality traits asking people about some of their behaviors, feelings, emotional reactions, or body responses using multiple-choice tests. Complementarily to this traditional method to measure personality traits, there has been an increasing interest in predicting personality and individual differences from indirect indicators in recent years. Different studies have shown a large variety of indirect indicators related to personality, such as observing personal places like rooms or workplaces (e.g., Gosling, Ko, Mannarelli, & Morris, 2002), Facebook pages (e.g., Ivcevic & Ambady, 2012), player's behavior in games (e.g., Van Lankveld, Spronck, Van den Herik, & Arntz, 2011), physical activity intensity (e.g., Gao, Shao, & Salim, 2019), situational judgment tests (e.g., Olaru et al., 2019), eating habits and food preferences (e.g., Goldberg & Strycker, 2002), or indicators from essays of creative writing (e.g., Küfner, Back, Nestler, & Egloff, 2010). Since Pennebaker and Graybeal's (2001) findings, a number of emerging studies have focused on personality evaluation through written language. In the present study, we focus on this last set of studies, conducting a meta-analytic review about personality prediction through written language, using computational models of language, to shed some light on the state of the art of this emerging field of research.

Current improvements of the computational methods for language data and the statistical models by technological development are generating new opportunities to study personality and its relationships with language outcomes (Boyd & Pennebaker, 2017). Following Boyd and Pennebaker (2017), current research on psychological language analysis assumes that there are characteristics of personality which are embedded and reflected in the patterns of language that people use. Language data would capture lower-level personality processes that are

* Corresponding author at: Facultad de Psicología, Avda de Iván Pavlov, 6, Campus de Cantoblanco, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
E-mail addresses: josedavid.moreno@uam.es (J.D. Moreno), josea.martinez@uam.es (J.Á. Martínez-Huertas), ricardo.olmos@uam.es (R. Olmos), gdejorge@psi.uned.es (G. Jorge-Botana), juan.botella@uam.es (J. Botella).

https://doi.org/10.1016/j.paid.2021.110818
Received 16 December 2020; Received in revised form 3 March 2021; Accepted 4 March 2021

closely related to, for example, behavioral outcomes associated with personality traits. As reported in previous literature, language-based measures of personality processes would have reliable properties and relevant links with different personality traits (e.g., Boyd et al., 2015; Boyd & Pennebaker, 2015; Fast & Funder, 2008; Pennebaker & King, 1999). But the idea that linguistic terms could indicate personality traits is not new. Indeed, in the very first research about personality traits, authors used descriptions and factor analysis to determine the structure of personality (see e.g., Goldberg, 1993, for a review of the lexical hypotheses in personality; see for example that the factor structure of a matrix of adjectives used for describing different people have been used for this purpose). If two tentative traits have a high overlap in their definitions, they should only merit one tag. Nonetheless, conversely to a high structured emission of descriptions, the prediction from non-prompted texts (e.g., no directive questions and no adjectives as a choice) have a special treatment, particularly when text samples are unintentional and are retrieved informally. The proxies used to infer a personality trait are not explicit as in adjectives assignations or alternative choices. In these cases, any hidden pattern in language (combination of cues) could be useful to infer personality (e.g., subject pronouns, vocabulary expansions, adjectives, topics, causative markers, sensorimotor terms, etc.). For this reason, it needs a hard formalization of utterances in term of their cues and optionally even powerful statistical methods to merge them in a predictive model. Computational models of language can be helpful for these purposes.

Two main methods can be found in the literature analyzing language to study personality: methods based on human experts' criteria (e.g., Back, Schmukle, & Egloff, 2008; Borkenau, Mosch, Tandler, & Wolf, 2016; Hall, Goh, Mast, & Hagedorn, 2016; Hall, Pennington, & Lueders, 2014; Holleran & Mehl, 2008; Küfner et al., 2010; Marcus, Machilek, & Schütz, 2006; Vazire & Gosling, 2004), and methods based on computational models of language (e.g., Farnadi et al., 2016; Farnadi, Zoghbi, Moens, & De Cock, 2013; Gao et al., 2013; Golbeck, 2016; Golbeck, Robles, Edmondson, & Turner, 2011; Hawkins, Raymond, & Boyd, 2017; Holtgraves, 2011; Kwantes, Derbentseva, Lam, Vartanian, & Marmurek, 2016; Mairesse, Walker, Mehl, & Moore, 2007; Park et al., 2015; Qiu et al., 2017; Qiu, Lin, Ramsay, & Yang, 2012; Schwartz et al., 2013; Skowron, Tkalčič, Ferwerda, & Schedl, 2016; Thilakaratne, Weerasinghe, & Perera, 2016; Xianyu, Xu, Wu, & Cai, 2016). In the first group of methods, evaluators analyze written or oral language production, based on human experts' judgements to assess personality. For example, in the study of Borkenau et al. (2016), participants wrote essays on five domains of their life and filled in the Big Five questionnaire. Next, the essays were evaluated by judges who rated the participants' personality and other attributes. Results showed that the judges' impressions were mostly accurate when predicting some of the personality traits, as Openness to Experience, but not when they tried to predict other traits. In the second group of methods, evaluators use different types of computational linguistic models that have been trained with specific corpora to automatically assess written or transcript oral language. For example, in the study by Kwantes et al. (2016), participants completed the Big Five questionnaire and wrote five short essays in which they were asked to describe what they would do and how they would feel in five different scenarios that evoked relevant aspects for each of the five personality traits. The essays produced by the participants were converted into a semantic vector using *Latent Semantic Analysis* (LSA). The similarity between semantic vectors representing personality traits and vectors of those essays was calculated. Results showed an interesting relationship between the scores obtained by participants in the Big Five questionnaire and the predictions of their essays using the LSA.

Kwantes et al. (2016) study is just a sample of the computational methodologies we will focus on. In general terms, computational models of language represent some formal properties of the utterances and identify different personality styles or coping styles by applying some measures to that properties (e.g. pattern detection, similarity measures, supervised predictive models). In this way, utterances could be simple written productions (letters, prompt-based questions, scenario-based essays), or participation in social networks (Twitter, Facebook, blogs, microblogs). For the sake of clarity, a brief taxonomy to give a broad perspective about the computational models of language that can usually be found in the current literature shall now be described. A similar taxonomy can be found in Carvalho and Louwerse (2017). Firstly, the information of utterances can be represented by different observable cues or patterns (syntactic and semantic information) or by more abstract semantic layers of meaning representation (e.g., vector space models as LSA, LDA, Word2Vec). Secondly, these representations can be integrated into a prediction model in a machine learning environment or not.

Following the previous taxonomy, firstly, we can represent the properties of utterances by detecting the occurrence of some relevant observable cues or patterns, such as extracting ratios of the presence of linguistic features (words, *n-grams*, verbal tenses, verbal persons, etc.), thematic substantives from emotional lexicons (e.g., Mairesse et al., 2007; Pennebaker, Francis, & Booth, 2001), sublexical cues (punctuation, spelling, capitalization, number of letters, syllables), or even closed class words as personality indicators (e.g., personal pronouns to indicate the presence of social expansion in Holtzman et al., 2019; Campbell & Pennebaker, 2003). The LIWC tool (Pennebaker et al., 2001; Pennebaker & King, 1999) is the most famous example of this kind of paradigm and has demonstrated its utility at detecting coping styles, personality and health changes (Campbell & Pennebaker, 2003; Mitra, Counts, & Pennebaker, 2016; Tausczik & Pennebaker, 2010). Other options for the detection of observable cues are taken from the lexicon NRC (Mohammad, Zhu, & Martin, 2014), the psycholinguistic database MRC (e.g., Gill, Oberlander, & Austin, 2006), the SPLICE (e.g., Farnadi et al., 2014) or the rating systems as SentiStrength (e.g., Celli et al., 2014; Farnadi et al., 2014; see Farnadi et al., 2016, for a revision of all of these techniques). Furthermore, extensions of these tools or databases have been used in many of the papers included in the present meta-analysis. In addition, a more abstract semantic layer of meaning representation can be used with vector space models. Vector techniques are based on the automatic processing of large text corpora representing a language of a general or specific domain, although some studies recommend specific domain corpora (Kwantes et al., 2016). There are different computational models in which word occurrences are algebraically vectorized such as LSA, word2vec, or BEAGLE (for a revision on space models see Günther, Rinaldi, & Marelli, 2019; Jorge-Botana, Olmos, & Luzón, 2020; Jones, Willits, & Dennis, 2015; or McNamara, 2011). All of them coincide in that they represent the lexicon in a reduced dimensionality vector space. The final product of all of them is that the latent meanings of those vectors represent possible relevant topics that can be an extensional description of personality domain (e.g., social relations, illness, death, family, etc.; see for example, Kwantes et al., 2016).

Both techniques can represent the meaning of utterances at different representation levels. The main difference between them is that the latter models focus on the semantic relationships between words, phrases, paragraphs, etc., while the former models have a level of analysis that address pattern detection (usually defined as rules in regular expressions, or grammar and even supported by syntactic parsers). In this way, information from written materials can be represented through vector space models, or pattern detection techniques. These vector space models are more related to deeper or not apparent semantic characteristics of language, while pattern detection techniques are more related to explicit semantic or syntactical information, like detection of observable cues. Nonetheless, some vector space models are sensible to a pseudo-syntactical information because they consider the order into a context window (see Jorge-Botana et al., 2020 for a revision). Within the frame of the present meta-analysis, we will refer to semantic and syntactical information as a type of information.

Secondly, both types of information can also be integrated into a prediction model in a machine learning environment. Taking that information as input, it can be trained or analyzed to predict a different personality predominance (Gao et al., 2013; Park et al., 2015; Thilakaratne et al., 2016; Xianyu et al., 2016). While some studies use composite scores based on the raw semantic or syntactic information, these environments can be implemented with several techniques, such as logistic or multivariate regression, artificial neural networks, naive Bayes, or support vector machine. All these techniques allow the systems to identify the proportion in which observable patterns or vector components must co-occur to predict a trait. For instance, literal word detection and vector representations could participate as input of neural network models (e.g., Mandl, 1999; Martínez-Huertas, Jorge-Botana, Luzón, & Olmos, 2021). Throughout the present meta-analysis, we will refer to this set of techniques as prediction mechanisms.

Over the last years, several authors have remarked on relevant relationships between personality and written language (e.g., Boyd & Pennebaker, 2015, 2017; Boyd & Schwartz, 2021; Chung & Pennebaker, 2018; Stachl et al., 2020). Thus, the main goal of the present meta-analysis was to empirically test if this relationship is reflected in the existent literature in the field. Keeping this in mind, we conducted a meta-analytic review of the automatic analysis methods of utterances and their correlation with the Big Five Personality questionnaire as it is the most popular personality taxonomy where computational models of language have been applied. The main goal of this meta-analysis was to obtain combined estimates of the relationships between personality traits and written language using computational models of language, in order to analyze if written language can be considered as an alternative measure of personality. To achieve this goal, we focused on two main points: (1) to test the predictability of personality traits from written language, and (2) to test that predictability in the field of computational models of language. Although both points were tested jointly, the first point has important theoretical implications while the second one has important practical implications. Moreover, the second goal tests the actual relevance of different moderator variables to serve as a guide for empirical researchers to design higher quality research. Specifically, we tested the moderation role of the type of information and prediction mechanisms. We expected higher relationships between personality traits and written language for the combination of semantic and syntactic information, and also when using prediction mechanisms. Finally, we tested a set of relevant variables in the primary research as the source of text data from social networks, the language of texts, the publication source, the sex of the participants, or the text length of the materials of the studies (see the *Moderator variables* section for an overview of the rationale for the inclusion of these variables).

2. Method

2.1. Selection of studies

Firstly, on the 4th of October of 2019, we searched for the primary studies in the following databases: Google Scholar, PsycINFO, MEDLINE and PubMed. We used different combinations of the following descriptors: “personality”, “language”, “big five” and “computational”. The broadest search was obtained from “personality” and “language” descriptors. No publication type (journal, conference proceedings, technical reports) nor date criteria was imposed. The initial search showed that almost all relevant works evaluated the Big Five taxonomy and only rarely the Dark Triad or Tetrad. Thus, in the first selection of primary studies, the studies had to evaluate personality traits using any Big Five questionnaire version, and to evaluate written language with any kind of computational model of language. This led to an initial sample of 84 studies. Then, three exclusion criteria were imposed: (1) the study had to present Pearson correlation coefficients between personality traits and computational scores, or at least to provide enough data to calculate the correlation coefficients, as this was needed to conduct the meta-analytic analyses. (2) A minimum sample size of 30 participants to conduct the analyses with guarantees (see similar criteria in Bonett & Wright, 2000). (3) The study had to measure the five personality traits using personality questionnaires (McCrae & Costa Jr., 2008). This led to a final sample of 17 articles composed of 23 studies. The inter-rater reliability was high (93%; Cohen's Kappa = 0.86, $p <$ .001). It is worthy to point out that most of the articles we had to discard (k = 66) were rejected for not informing about the correlation coefficients, not providing enough data in order to calculate these coefficients, or for not measuring the five personality traits with personality questionnaires. This process is summarized in Fig. 1. Moreover, the characteristics of the final sample of studies are shown in Table 1.

2.2. Personality measurement

We studied the personality models used in all the primary studies analyzed in the present meta-analysis. As could be verified in our first review, the vast majority of the studies measured the personality traits using the Big Five Personality model (*Big Five*; McCrae & Costa Jr., 2008). This is one of the most validated personality questionnaires and it proposed five basic dimensions of personality (e.g., De Raad, 2000): Openness to Experience (*O*, related to intellectual curiosity and openness to try new things), Conscientiousness (*C*, related to self-discipline and to behave according to duty), Extraversion (*E*, generally characterized by pronounced engagement with the external world), Agreeableness (*A*, individuals with high scores in this trait tend to be considerate, kind, generous, and helpful with others), and Neuroticism (*N*, mostly related to the tendency to experience negative emotions like anger, anxiety or depression). The Big Five Personality model has received a lot of research among the last years, accumulating a solid body of knowledge on the stability and the universality of the personality traits proposed within it (e.g., McCrae & Costa Jr., 2008), and also showing high cross-validity coefficients compared to other popular personality inventories in recent studies (e.g., Grucza & Goldberg, 2007), which reflects bandwidth in the diversity of behavior that can be predicted, and fidelity to predict each type of behavior within its range.

2.3. Effect size calculation and statistical analysis

Given the correlational design of the studies analyzed and the statistics provided, our choice for the effect size measure was the Pearson

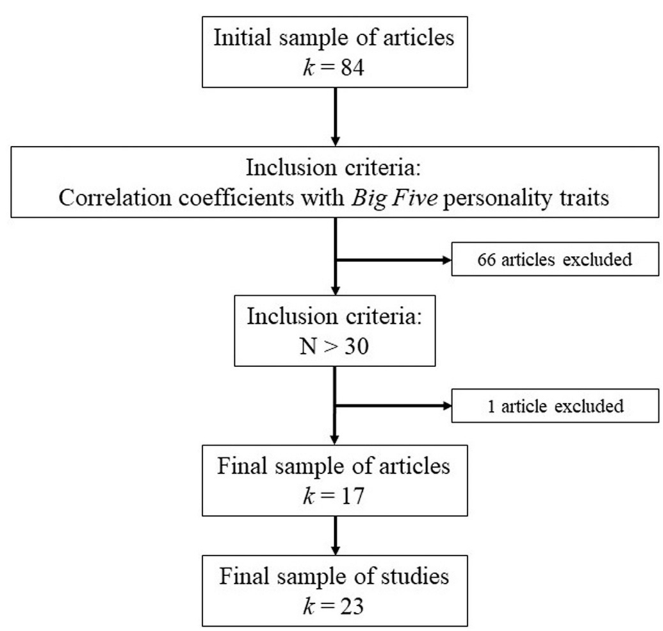


Fig. 1. Flowchart (inclusion and exclusion criteria).

Table 1
Effect size estimates and 95% CI of the studies included in the meta-analysis for each personality trait.

Studies	N	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Farnadi et al. (2013)	250	0.20 [0.08;0.32]	0.15 [0.03;0.27]	0.16 [0.04;0.28]	0.19 [0.07;0.31]	0.20 [0.08;0.32]
Farnadi et al. (2016).S1	3731	0.02 [−0.01;0.05]	0.03 [−0.00;0.06]	0.03 [−0.00;0.06]	0.05 [0.02;0.08]	0.04 [0.01;0.07]
Farnadi et al. (2016).S2	404	0.11 [0.01;0.21]	0.22 [0.13;0.31]	0.17 [0.07;0.26]	0.20 [0.10;0.29]	0.12 [0.02;0.22]
Farnadi et al. (2016).S3	102	0.31 [0.12;0.48]	0.42 [0.25;0.57]	0.33 [0.14;0.49]	0.37 [0.19;0.53]	0.43 [0.26;0.58]
Gao et al. (2013)	1766	0.38 [0.24;0.50]	0.41 [0.27;0.52]	0.40 [0.27;0.52]	0.31 [0.17;0.44]	0.32 [0.18;0.45]
Golbeck (2016). S1	127	0.36 [0.20;0.50]	0.25 [0.08;0.41]	0.37 [0.21;0.51]	0.41 [0.26;0.55]	0.38 [0.22;0.52]
Golbeck (2016). S2	8569	0.20 [0.18;0.22]	0.20 [0.18;0.22]	0.22 [0.20;0.24]	0.24 [0.22;0.26]	0.18 [0.16;0.20]
Golbeck (2016). S3	69	0.35 [0.12;0.54]	0.06 [−0.17;0.30]	0.24 [−0.00;0.45]	0.35 [0.13;0.55]	0.18 [−0.06;0.40]
Golbeck et al. (2011)	50	0.43 [0.17;0.63]	0.37 [0.11;0.59]	0.34 [0.07;0.56]	0.36 [0.10;0.58]	0.33 [0.06;0.56]
Hawkins et al. (2017)*	731	0.11 [0.04;0.18]	0.08 [0.01;0.15]	0.15 [0.08;0.22]	0.11 [0.03;0.18]	–
Holtgraves (2011)	224	–	–	0.19 [0.06;0.31]	0.25 [0.12;0.37]	0.15 [0.02;0.28]
Kwantes et al. (2016)	87	0.31 [0.11;0.49]	0.08 [−0.13;0.29]	0.19 [−0.02;0.39]	0.02 [−0.19;0.23]	0.22 [0.01;0.41]
Mairesse et al. (2007).S1	96	0.31 [0.12;0.48]	0.30 [0.11;0.47]	0.32 [0.13;0.49]	0.30 [0.11;0.47]	0.22 [0.02;0.40]
Mairesse et al. (2007).S2	2479	0.20 [0.16;0.24]	0.14 [0.10;0.18]	0.08 [0.04;0.12]	0.16 [0.12;0.20]	0.18 [0.14;0.22]
Park et al. (2015)	4824	0.43 [0.41;0.45]	0.37 [0.35;0.39]	0.42 [0.40;0.44]	0.35 [0.32;0.37]	0.35 [0.32;0.37]
Qiu et al. (2012)	142	0.27 [0.11;0.42]	0.16 [−0.00;0.32]	0.28 [0.12;0.43]	0.20 [0.04;0.35]	0.20 [0.04;0.35]
Qiu et al. (2017).S1	470	0.13 [0.04;0.22]	0.22 [0.13;0.30]	0.15 [0.06;0.24]	0.18 [0.09;0.27]	0.15 [0.06;0.24]
Qiu et al. (2017).S2	90	0.15 [−0.06;0.35]	0.30 [0.10;0.48]	0.24 [0.03;0.43]	0.35 [0.14;0.51]	0.25 [0.08;0.46]
Schwartz et al. (2013)	71,968	0.42 [0.41;0.43]	0.35 [0.34;0.36]	0.38 [0.37;0.39]	0.31 [0.30;0.32]	0.31 [0.30;0.32]
Skowron et al. (2016)	62	0.43 [0.20;0.61]	0.29 [0.04;0.50]	0.40 [0.17;0.51]	0.30 [0.05;0.51]	0.30 [0.05;0.51]
Thilakaratne et al. (2016)	387	0.37 [0.29;0.46]	0.34 [0.25;0.42]	0.35 [0.26;0.44]	0.30 [0.20;0.38]	0.40 [0.31;0.48]
Xianyu et al. (2016)	376	0.78 [0.74;0.82]	0.71 [0.66;0.76]	0.58 [0.51;0.64]	0.59 [0.52;0.65]	0.59 [0.52;0.65]
Yarkoni (2010)	694	0.22 [0.15;0.29]	0.19 [0.12;0.26]	0.17 [0.10;0.24]	0.21 [0.14;0.28]	0.17 [0.10;0.24]

Notes. The five columns to the right show the effect size (r.) and its 95%IC calculated for each measure obtained from each of the studies included in the meta-analysis. N = Sample size; S1 = Study 1; S2 = Study 2; S3 = Study 3; * = This paper presents different studies, but only one accomplished our selection criteria.

correlation coefficient r, as all primary studies reported it directly. For statistical analysis, the coefficients were previously transformed through Fisher's formula, to have a more symmetrical distribution (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009; Botella & Sánchez-Meca, 2015). The final results were back-transformed from Fisher's value to its corresponding correlation coefficient to facilitate the interpretation of the results.

The heterogeneity of the estimations was analyzed by Q tests and I^2 indexes (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). Statistical analyses assumed a random-effects model that is generally preferred because it is more conservative than a fixed-effect model, and allows generalizing conclusions beyond the specific set of studies analyzed (Borenstein, Hedges, Higgins, & Rothstein, 2010; Hedges & Vevea, 1998; Raudenbush, 2009). The combined estimates were calculated weighting the individual studies by the inverse of their variances. The method used to estimate between-study variability was the Hartung-Knapp-Sidik-Jonkman method for random-effects meta-analysis (IntHout, Ioannidis, & Borm, 2014). All the statistical analyses were performed with R Statistical Software (R Core Team, 2019) in the 3.6.2 version, using the *metafor* package (Viechtbauer, 2010a, 2010b) for the combined estimates, the Q statistic, and the I^2 statistic estimates. In addition, we used the SPSS macros of Lipsey and Wilson (2001) in order to analyze the categorical moderator variables. To test the publication bias, we calculated Kendall's tau, the Egger's test, and also applied the trim & fill procedure and calculated the fail-safe N following Rosenberg approach (Borenstein et al., 2010; Botella & Sánchez-Meca, 2015), also through *metafor*. Separated meta-analysis were performed for each one of the five personality traits.

2.4. Moderator variables

In order to analyze the heterogeneity between the results of the studies, we conducted several moderator analyses for each of the five personality traits. Eight different moderator variables were selected based on their potential explanatory role in the results of the analyzed studies in the present meta-analysis and grouped on the basis of their methodological or theoretical nature.

On the one hand, six categorical and one numerical moderator variables were included in the analyses due to their methodological nature. Firstly, we studied the *type of information* of the input (only semantic information, only syntactic information, or a combination of both). We consider "semantic" the kind of information that is tagged in a semantic category no matter if the category is produced by pattern detection, vector space models, or a predictive model layer. We consider "syntactic" the kind of information that comes from non-semantic cues, as verbal persons, verbal tenses, discourse markers, n-grams and even sublexical cues as punctuation, spelling, capitalization, number of letters, syllables, etc. In addition, we used *prediction mechanisms* (no, yes) as a moderator variable. That is, whether a predictive model was trained with the semantic or syntactic information. As was explained previously, some studies used composite scores based on the raw semantic or syntactic information (i.e., no complex model was used). Nonetheless, some other studies used complex predictive models with that information as input, for example, multivariate regression, artificial neural networks, naive Bayes, or support vector machine.

Taking into account the high amount of written materials analyzed from social networks in the primary research (e.g., Farnadi et al., 2016; Golbeck, 2016; Park et al., 2015), we decided to include the use of *social networks* as the input in the primary studies (no, yes) as other moderator variable. In addition, to test the potential influence of the source of the results of the primary studies, we also analyzed the *language* of the texts analyzed in the primary studies (Chinese, English), and the *publication source* of the studies analyzed in the present meta-analysis (conference paper, journal paper). As a control measure, we also tested the potential influence of the instruments used to measure personality in the primary studies (*personality instrument*; BFI, others). The most of the primary studies analyzed in the present meta-analysis ($k = 18$) used a 44-items version of the BFI, but other studies ($k = 5$) used different instruments as reduced 10-items versions of the BFI, TIPI, Goldberg's 100-adjectives questionnaire, or BFI completed by indirect expert judges. Due to their low number, they were grouped on the same moderator category. Finally, the text length of the written materials analyzed in the primary studies (number of words) was included as a numerical moderator variable to test the influence of the quantity of linguistic information used as input.

On the other hand, different continuous moderator variables were included in the analyses due to their theoretical nature. Previous studies have already found sex and age differences that show different patterns in written language (e.g., Schwartz et al., 2013), which can also be reflected on personality. Thus, in the present meta-analysis, we considered both, sex and age, as potential predictors in the meta-regressions. Age of the participants was discarded as a moderator variable because of the low age range (mean age ranged from 19 to 27 years).

3. Results

3.1. Combined effect size estimates

A summary of the results obtained for the five personality traits is shown in Table 2. All the combined effect size estimates showed statistically significant, although small to moderate combined effect size estimates according to Cohen's (1988; Hemphill, 2003) conventions (r ranged from 0.26 to 0.30) for each personality trait. Significance tests were performed with the Hartung-Knapp-Sidik-Jonkman method (IntHout et al., 2014). The effect size was positive for the five measures and all of them reached p-values under 0.001. The forest plots in Fig. 2 provide a graphical overview of the effects studied for the five personality dimensions. On the other hand, all homogeneity tests showed significant values for the Q statistic ($p < .001$ for the five dimensions), leading us to reject the null hypothesis of homogeneity for all five cases. Thus, for all measures there was a large amount of heterogeneity which could be explained by the presence of moderator variables. I^2 statistic can be interpreted as follows: 25% is low, 50% is medium, and 75% is high heterogeneity (Borenstein et al., 2009; Botella & Sánchez-Meca, 2015). Then, the values of the I^2 statistic reached high values, ranging from 96.86% to 98.82%, indicating that the heterogeneity was considerably higher than expected from mere random sampling. Following the criteria proposed by Higgins and Green (2011), the heterogeneity for the five measures included in the meta-analysis should be assessed as *considerable* (>75%). This variability could be explained by the effects of moderator variables beyond random error, including the differences in the methodologies utilized by the different computational methods analyzed.

3.2. Publication bias

Publication bias was explored using different statistical and graphical tests. Table 3 presents the results for Kendall's tau, Egger's test, Trim & Fill, and Rosenberg fail-safe N procedure. In general, the funnel plots from Fig. 3 did not present any tendency to publication bias in our sample of studies. Only agreeableness shows some cues of bias, both in the trim & fill test and the visual inspection of the funnel plot. However, it is not credible that the studies could be excluded from publication because of this correlation while the correlations of the other traits are significant. The results of these statistical tests were clear: the meta-analytic results do not present concerns about publication bias.

3.3. Moderator analyses

Focusing on their methodological nature, the effects of six categorical variables and one numerical were analyzed for each personality trait (see Table 4). Regarding to the categorical variables, firstly, for the *type of information* (only syntactic information vs. only semantic information vs. a combination of both), a statistically significant effect was observed for Openness, Extraversion, Agreeableness and Neuroticism (the same tendency can be observed for Conscientiousness). This result can be explained by the higher performance of the combination of syntactic and semantic information versus using just one type of information. Secondly, for the use of *prediction mechanisms*, a statistically significant effect was observed favoring their use in all the personality traits. Thirdly, for the use of *social networks'* text data, no statistically significant differences were observed in any personality trait. Fourthly, for the *language* of the text analyzed in the primary study, a significantly higher prediction of Conscientiousness was observed in favor of Chinese (a similar tendency can be observed in the rest of the personality traits). Fifthly, for the *publication source*, a statistically significant effect was observed in favor of conference publications as compared with journal publications in all the personality traits. Finally, no statistically significant differences were found in any personality trait due to the different *personality instruments* used in the primary studies. Regarding to the numerical variable, a meta-regression model was fit for text length (number of words), but it did not show significant effects on the estimated effect size.

Regarding to the variables selected due to their theoretical nature, finally only a meta-regression model was fit for sex (proportion of women in the sample), showing significant effects on the estimated effect size. As observed in Table 5, agreeableness was significantly more predictable when the proportion of women in the sample was higher. That is, when the proportion of women in the sample was higher, the estimated effect size was significantly higher. The same tendency can be observed for the five personality traits, but no statistically significant effects were obtained.

As can be observed among the previous analyses, it was found that most of the differences within the moderator variables are due to methodological factors. These results can be synthetized on the statistically significant differences on the estimated effect size caused by the *type of information,* the *prediction mechanisms,* the *language* of the text analyzed and the *publication source,* in opposition to the theoretical variables analyzed where only presenting statistically significant effects on the estimated effect size caused by the proportion of women in the sample (*sex*) in one of the personality traits.

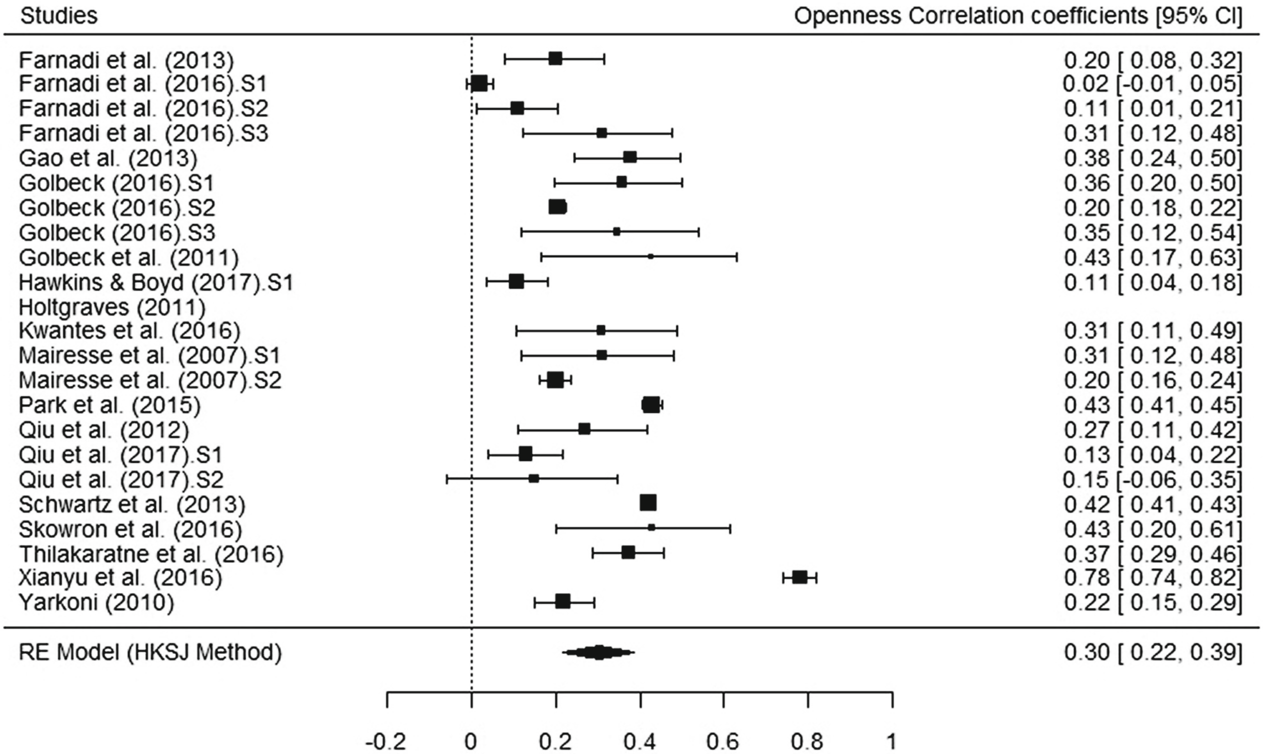
4. Discussion

In the present study, we conducted a meta-analysis from 23 primary studies, providing a synthesis of the combined effect size estimations of the predictive validity of the Big Five personality traits through computational models of language and, also exploring potential sources of heterogeneity in these effects. Our main finding was that written language shows significant relationships with the basic five personality dimensions so that it can be used as a predictor of the personality profile of the individual. These findings of the combined effect sizes are also consistent and congruent with the previous literature (e.g., Farnadi et al., 2013; Farnadi et al., 2016; Gao et al., 2013; Golbeck, 2016; Golbeck et al., 2011; Hawkins et al., 2017; Kwantes et al., 2016; Mairesse et al., 2007; Park et al., 2015; Schwartz et al., 2013; Skowron et al., 2016; Thilakaratne et al., 2016; Yarkoni, 2010). These results reinforce the relevance of personality and language relationships (Boyd & Pennebaker, 2015, 2017). Also, we found statistically significant moderating effects about the type of information used in the text materials, the

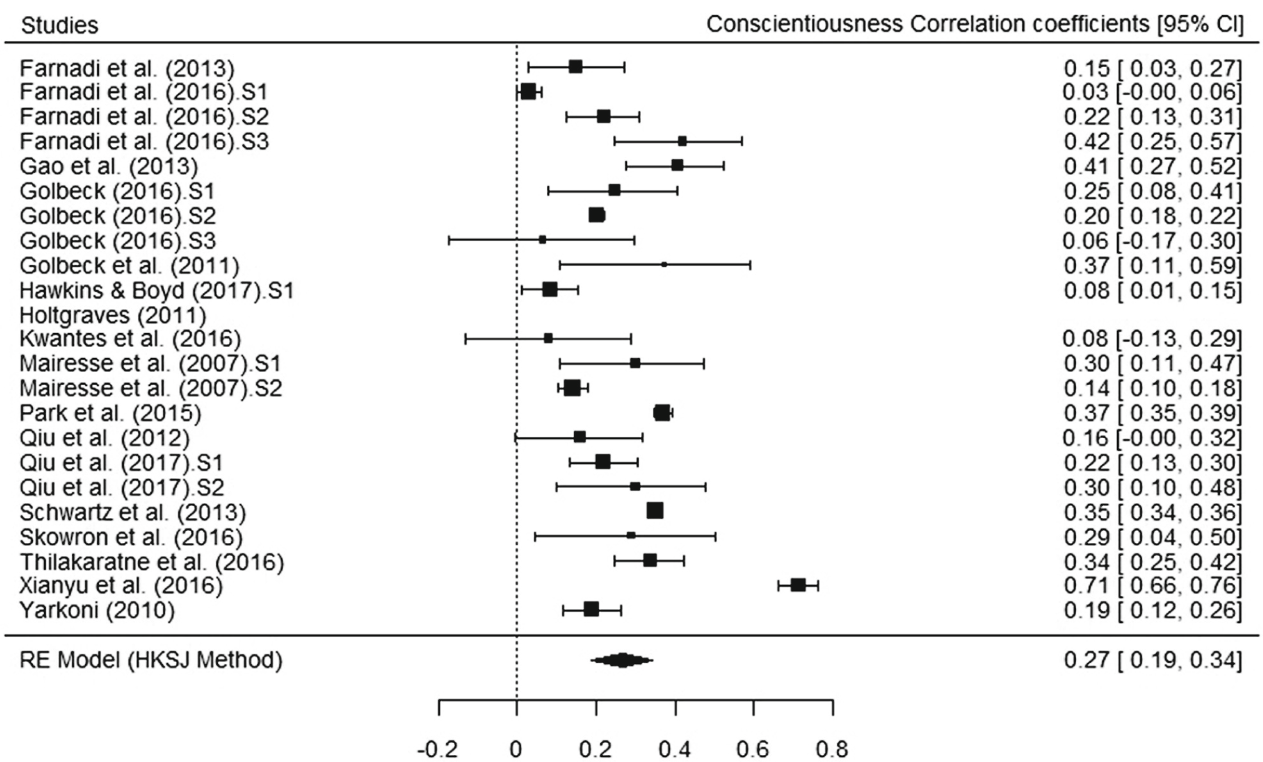
Table 2
Combined estimates for each personality trait using a random-effects model (significance tests with the Hartung-Knapp-Sidik-Jonkman method; IntHout et al., 2014).

Personality trait	k	$r.$	95%CI	t	Q(df)	I^2	τ^2
Openness	22	0.30	0.22; 0.39	6.86***	1461.37(21)***	98.82%	0.04
Conscientiousness	22	0.27	0.19; 0.34	6.73***	877.45(21)***	98.42%	0.03
Extraversion	23	0.27	0.21; 0.33	8.71***	1058.14(22)***	97.42%	0.02
Agreeableness	23	0.26	0.21; 0.32	9.13***	468.28(22)***	96.86%	0.02
Neuroticism	22	0.26	0.20; 0.32	8.55***	556.53(21)***	97.20%	0.02

*** = $p < .001$; k = number of studies analyzed for each personality trait; $r.$ = mean effect size (correlation coefficient).

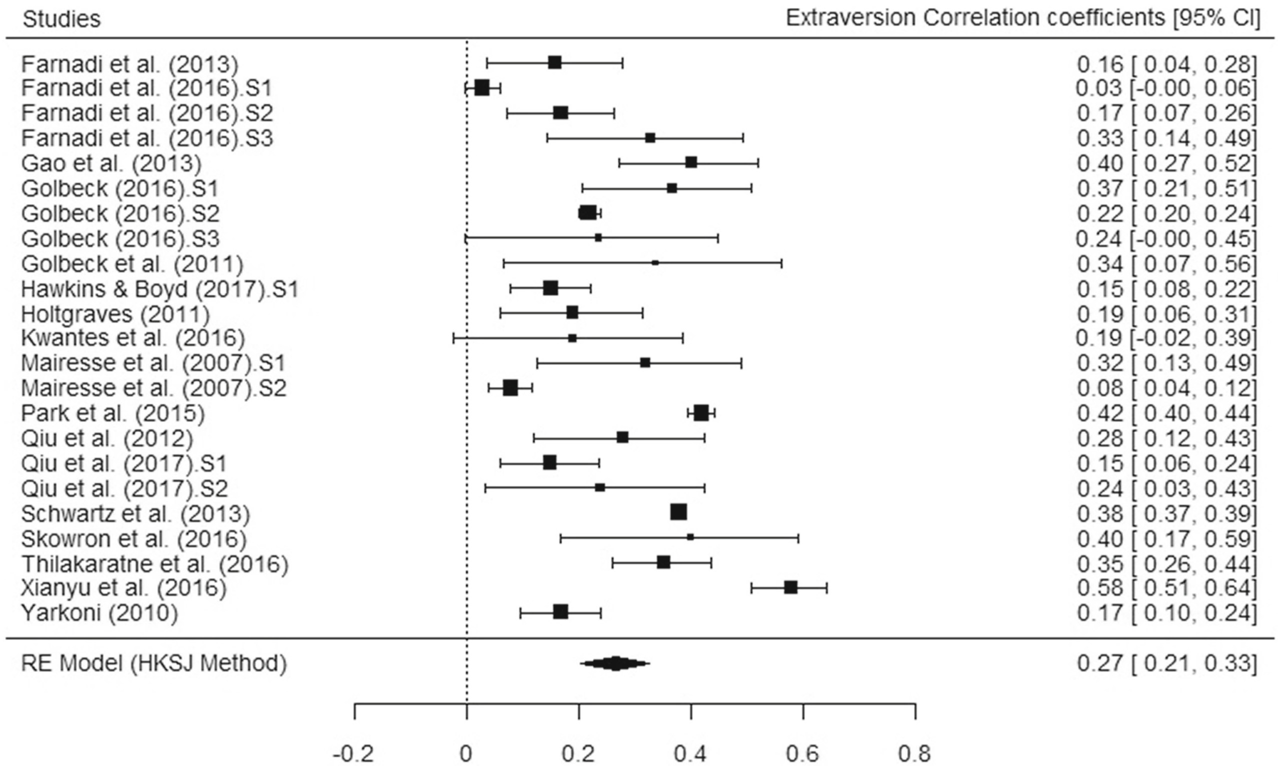


A. Forest plot for Openness

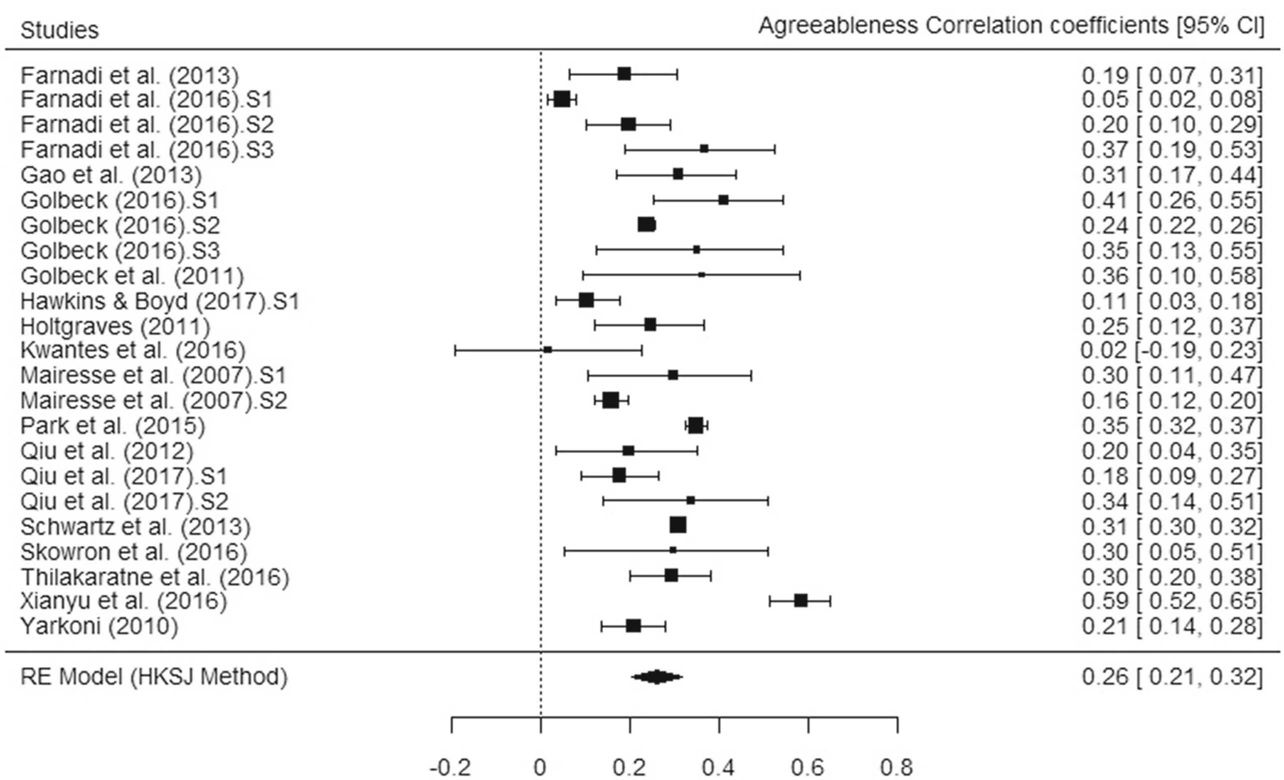


B. Forest plot for Conscientiousness

Fig. 2. A. Forest plot for Openness
B. Forest plot for Conscientiousness
C. Forest plot for Extraversion
D. Forest plot for Agreeableness
E. Forest plot for Neuroticism

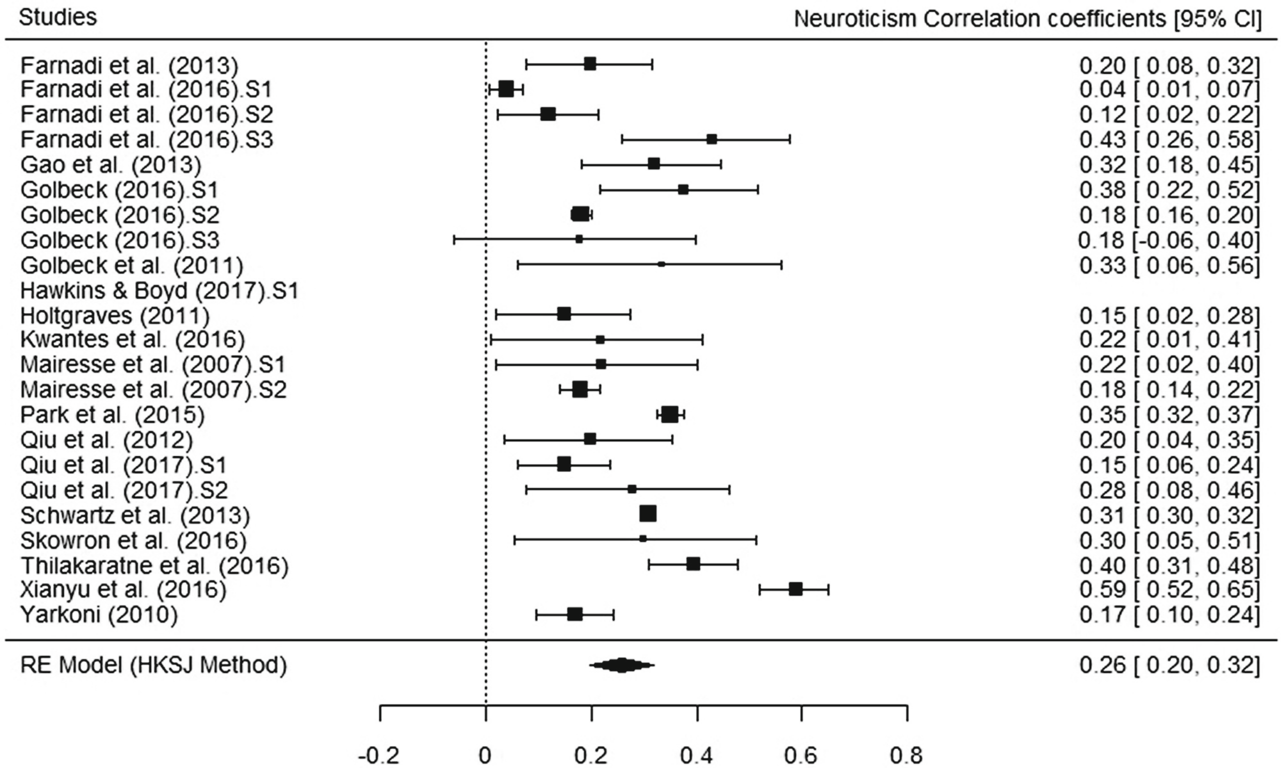


C. Forest plot for Extraversion



D. Forest plot for Agreeableness

Fig. 2. (*continued*).



E. Forest plot for Neuroticism

Fig. 2. (*continued*).

Table 3
Statistical tests for publication bias for each personality trait.

Personality trait	Kendall's tau	Egger's test	Trim & Fill	Rosenberg Fail-Safe N
Openness	0.33 (*p* = .03)	0.73 (*p* = .47)	0	86,488
Conscientiousness	0.23 (*p* = .14)	0.28 (*p* = .79)	0	59,500
Extraversion	0.27 (*p* = .07)	0.92 (*p* = .37)	0	73,478
Agreeableness	0.27 (*p* = .07)	1.07 (*p* = .30)	5[a]	51,208
Neuroticism	0.22 (*p* = .16)	0.79 (*p* = .44)	0	47,193

[a] Trim & Fill procedure estimates an effect size of 0.23 [0.16; 0.29] for agreeableness when 5 missing non-significative studies are included. Although after the "fill" the correlation is smaller, it is still significant and close to those of the other personality traits.

use of prediction mechanisms, and the publication source of the primary studies, and some interesting differences when analyzing language and sex as moderator variables. These results raise the potential of written language as a consistent and reliable indirect personality predictor, and also that computational methods are equally reliable to measure written language information to predict the Big Five personality traits (McCrae & Costa Jr., 2008). These estimates are very informative to provide an effect size reference for the predictions of personality traits using computational models of language. In addition, there was substantial heterogeneity in the effects. We consider that the moderation results we found are one of the main values of the present meta-analysis because they can explain part of this heterogeneity observed in the main results of the primary studies, and because they can give some guidelines that could improve future research. It is also noteworthy to highlight that this meta-analysis adds substantial information to what is available in the existing literature because, as far as we know, no meta-analysis with the same characteristics has been done until now.

At this point, it is also worthy to highlight results found in a previous meta-analysis conducted by Pace and Brannick (2010). In their work, the authors tested the level of similarity between the most used personality scales when measuring the same constructs, finding that the estimated mean convergent validities among all measures were almost in all the cases bellow 0.50. These estimates of convergent validity by test ranged from 0.31 to 0.54 for Agreeableness, 0.27 to 0.51 for Conscientiousness, 0.26 to 0.51 for Openness to experience, 0.37 to 0.66 for Extraversion, and 0.32 to 0.66 for Emotional stability (Neuroticism). Pace and Brannick (2010) concluded that these results are lower than could be expected, pointing significant differences between the different personality scales, even when these scales attempt to measure the same constructs. Taking into account these results, the findings of the present study (showing mean effect sizes for the five personality traits ranging from 0.26 to 0.30) can be understood as interesting and promising. Of course, the effect sizes we found about the relationship between personality and written language are small to moderate, but they are not too far from the relationships found between the most used personality scales. Thus, taking into account the combined effect sizes, results found in the present study still have to be taken with carefulness. Despite of it, starting from this promising point, we would like to propose new horizons to improve these methods.

Current results about the moderator variables can guide future primary research, indicating how to plan and design the materials and the methods with more guarantees. Specifically, these results point out that written language can be more informative if both semantic and syntactic information are analyzed when applying computational methods. This means that, if available, it is more useful to use different linguistic characteristics like semantic or syntactical information. It seems that semantic information, whether it is the result of vector representations

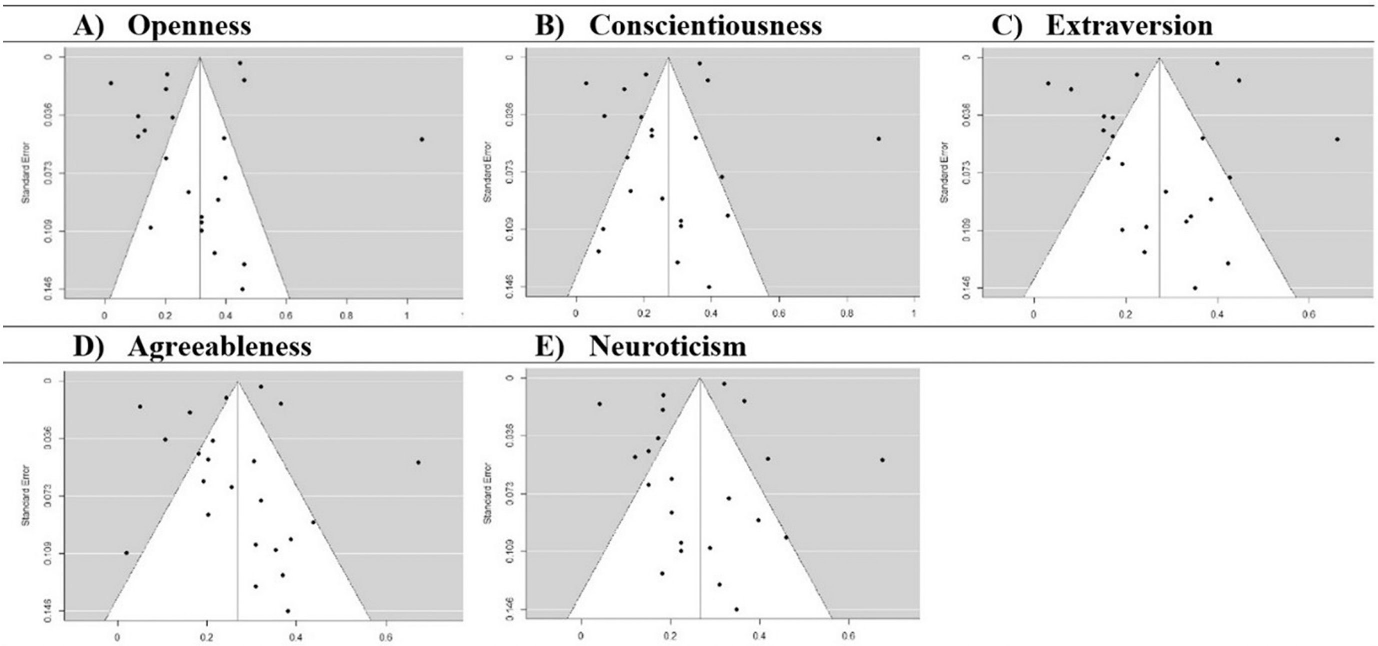


Fig. 3. Funnel plots for each personality trait.

or pattern detection techniques, can be enriched through syntactical information that can also contribute with additional and relevant information about personality. In fact, semantic information from vector space models have been proposed to detect deeper or not apparent characteristics of language, as they algebraically represent semantic regularities that capture relationships beyond the semantics, like physical characteristics about the world (e.g., distances between cities; Louwerse & Benesh, 2012; Louwerse & Zwaan, 2009) or embodiment properties like emotional valence (e.g., Hollis, Westbury, & Lefsrud, 2017; Martínez-Huertas et al., 2021). In this line, we could expect better performance if semantic information is produced by computational models, for example, vector space models (e.g., Günther et al., 2019; Jones et al., 2015; Jorge-Botana et al., 2020; McNamara, 2011).

Moreover, results also point out that using prediction mechanisms like machine learning methods, enhance the current performance of computational models of language when predicting personality. Despite the fact that we used a subsample of four studies, we would like to highlight that we found robust results for all the personality traits, and homogeneous in their effect sizes. These machine learning methods, like neural network models, are sensible to hidden patterns that are difficult to be detected and could be crucial to predict personality traits.

In addition, we can conclude that information taken from Social Networks is not more informative than other types of written materials to predict personality traits. This result highlights that less ecological laboratory research can equal the predictability of personality traits using Social Networks language, which is usually more expensive to obtain unless you use public secondary sources of information like *MyPersonality* (e.g., Kosinski & Stillwell, 2011; Park et al., 2015). Thus, giving some guidelines to improve future primary studies in this research field, taken into account the results of the present meta-analysis, we recommend future studies to collect both semantic and syntactic language data, and to complement current computational models of language measures with prediction mechanisms to increase their predictive validity of personality traits. If the quality of the collected data is high, the source of the data, whether they are from Social Networks or laboratory research, would not have an impact in the expected effect size of the study. Moreover, we would like to highlight the result of the publication source. As conference papers have fewer filters than peer reviewed journal papers, conclusions can be more biased. Nonetheless, the more conservative estimation based on journal papers is still significant and consistent.

Furthermore, several tentative but interesting results were found for the language of text, text length and sex of the sample of the studies. For the language of the text of the primary studies, a significantly higher prediction of Conscientiousness was observed in favor of Chinese, observing a similar non-statistically significant tendency in the rest of the personality traits. This result should be taken with caution as these analyses were conducted with a subsample of four studies, but they could be explained by cultural differences in language expression (e.g., Ji, Zhang, & Nisbett, 2004), or the methods applied in these studies. Text length results show that it is more relevant to have enough and high-quality information from each participant than collecting more information from the same participant. We also found more predictability of Agreeableness as higher is the proportion of women in the sample, observing the same tendency for the rest of the personality traits. This result should also be taken with caution, but it is an interesting result that has been reported in other previous studies. For example, studies like the one conducted by Schwartz et al. (2013) point to gender as a predictive measure related to specific language production. In their study with written language from social media, the authors showed that females used more emotional words and first-person singulars, also using more psychological and social processes, while males used more swear words and object references. Taking into account these prior results, future studies should continue analyzing the effects of sex as a potential source of variability when analyzing personality through different sources of language.

4.1. Limitations of the present study

We are aware of some limitations of the present meta-analysis. On the one hand, the number of studies that are included in the analyses, and the variety of the publication sources of these studies. Whilst this is a representative sample of available studies, the number of studies published in this field is limited at the moment because predicting personality traits using different computational models of language is still an emerging field of research (e.g., Boyd & Pennebaker, 2017). This is why some subsamples of the moderator variables had a reduced number of primary studies in the analyses. Taking into account this lack of

Table 4
Moderator analysis for categorical variables (language level, SVM, social networks, language, publication source and personality instruments).

Variable	Level	Openness				Conscientiousness				Extraversion				Agreeableness				Neuroticism			
		k	$r.$	95%CI	Q_B	k	$r.$	95%CI	Q_B	k	$r.$	95%CI	Q_B	k	$r.$	95%CI	Q_B	k	$r.$	95%CI	Q_B
Type of information	Syntactic	–	–	–	6.56*	6	0.20	0.05;0.34	4.24	9	0.15	0.08;0.23	18.12***	6	0.20	0.10;0.30	9.76***	–	–	–	6.51*
	Semantic	14	0.23	0.13;0.34		9	0.23	0.11;0.35		6	0.30	0.20;0.40		10	0.22	0.14;0.30		13	0.21	0.14;0.28	
	Both	7	0.47	0.32;0.61		7	0.38	0.25;0.51		7	0.40	0.31;0.49		7	0.38	0.30;0.47		7	0.37	0.27;0.47	
Prediction mechanism	No	18	0.25	0.16;0.35	11.94***	18	0.21	0.14;0.30	14.74***	19	0.22	0.15;0.28	16.63***	19	0.24	0.18;0.30	8.82**	18	0.21	0.15;0.27	15.42***
	Yes	4	0.58	0.40;0.75		4	0.52	0.37;0.66		4	0.48	0.36;0.59		4	0.42	0.30;0.53		4	0.45	0.34;0.56	
Social Networks	No	4	0.23	0.01;0.45	0.70	4	0.15	−0.04;0.33	2.09	5	0.18	0.04;0.31	2.51	5	0.17	0.05;0.29	3.26^t	4	0.19	0.04;0.34	1.13
	Yes	18	0.33	0.23;0.43		18	0.30	0.21;0.39		18	0.30	0.23;0.37		18	0.30	0.23;0.36		18	0.28	0.21;0.35	
Language	English	18	0.30	0.18;0.41	1.81	18	0.23	0.15;0.33	6.99**	19	0.25	0.17;0.33	2.23	19	0.25	0.19;0.32	3.55^t	18	0.25	0.17;0.32	2.38
	Chinese	4	0.44	0.22;0.66		4	0.48	0.30;0.65		4	0.38	0.22;0.54		4	0.39	0.25;0.52		4	0.37	0.22;0.52	
Publication Source	Conference	6	0.50	0.34;0.66	7.09**	6	0.43	0.29;0.57	6.58*	6	0.40	0.28;0.53	5.90*	6	0.37	0.26;0.48	4.09*	6	0.40	0.28;0.51	7.11**
	Journal	16	0.25	0.15;0.34		16	0.22	0.13;0.30		17	0.23	0.16;0.30		17	0.23	0.17;0.30		16	0.22	0.16;0.28	
Personality instruments	BFI	18	0.35	0.25;0.45	2.00	18	0.29	0.20;0.38	0.51	18	0.30	0.23;0.37	2.87	18	0.29	0.22;0.35	1.23	18	0.28	0.21;0.35	0.63
	Others	4	0.18	−0.03;0.39		4	0.21	0.03;0.40		5	0.18	0.05;0.30		5	0.21	0.09;0.33		4	0.21	0.07;0.36	

t = $p < .10$; * = $p < .05$; ** = $p < .01$; *** = $p < .001$. k = number of studies analyzed for each personality trait; $r.$ = mean effect size (correlation coefficient).

Table 5
Meta-regression analysis for sex (proportion of women) for each personality trait.

Personality trait	k	b [95%IC]	t
Openness	10	0.13 [−0.66; 0.92]	0.38
Conscientiousness	10	0.48 [−0.15; 1.12]	1.76
Extraversion	11	0.30 [−0.32; 0.93]	1.09
Agreeableness	11	0.56 [0.07; 1.06]	2.58*
Neuroticism	10	0.23 [−0.41; 0.87]	0.84

* = $p < .05$.

research, we decided to also include conference papers in our study that reported the necessary data to conduct our analyses. For that reason, we controlled the publication source as a potential source of variability. Despite of it, we are aware of the conclusions we can reach about some moderator variables, taking into account the sample limitations. For that reason, we want to emphasize that these results are promising and informative, but still have to be taken cautiously. More primary research is needed in the field in order to reach more robust conclusions about the potential moderation role of these variables. In addition, we also would like to emphasize that the sizes of the results ranged from small to moderate, but they were also consistent across all the Big Five personality traits. As was previously commented, the size of these results has to be understood taking also into account previous results about the relationships found between the most used personality scales (e.g., Pace & Brannick, 2010). Thus, written language can be still considered a modest but also a useful source of information to explain personality and individual differences using computational models of language.

On the other hand, it is also noteworthy that the present results are referred to the use of Big Five personality traits (McCrae & Costa Jr., 2008) because it is one of the most commonly applied personality questionnaires with computational models of language. But future empirical studies should try to analyze if different conceptualizations of personality are differentially related to linguistic inputs employing computational models of language. Another issue to highlight is that this work is referring to personality traits, not states. Some studies have focalized their attention into changes of written styles as an indirect indicator of wellness state, including expansive behavior toward the environment or other states (e.g., Campbell & Pennebaker, 2003). Current results are also informative for the prediction of states.

5. Conclusion

To conclude, the present meta-analysis shows that written language can be considered a fruitful personality indirect indicator, and also that it is possible to accurately extract text information using computational models of language. But it is worthy to note, that the effect sizes of the present meta-analysis are small to moderate, which highlights that there is room for improvement in this research field. Thus, we found the information of this meta-analysis relevant to improve future research proposals. But it is also for health and commercial purposes, as written language can be easily collected in many different ways and situations. Some examples of written language data extraction can be social media, applications via the internet, data collecting in laboratory conditions, quick questionnaires in more ecologic contexts, or even language transcript from an oral speech during a large variety of situations, like interviews or public speeches. Also, computational models of language and machine learning algorithms are improving faster within the past years due to technological developments. Thus, we can expect an improvement in current research in the near future. Building more connections between these methods and strong psychological theories could be a key issue for future developments in the field. Finally, we would like to encourage authors to conduct future research on this promising field, taking into account some of the main points highlighted in the present meta-analysis, as data about the moderator variables analyzed in this study can be useful to design higher quality research for future research. These guides can be helpful to keep finding more robust links between language and personality.

References

References marked by a * are those that were included in the meta-analysis reported here.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. https://doi.org/10.1177/1088868306294907.

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey. bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*(4), 1116–1122. https://doi.org/10.1016/j.jrp.2008.02.001.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and spearman correlations. *Psychometrika, 65*(1), 23–28. https://doi.org/10.1007/BF02294183.

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. *John Wiley and Sons.*. https://doi.org/10.1002/9780470743386.

Borkenau, P., Mosch, A., Tandler, N., & Wolf, A. (2016). Accuracy of judgments of personality based on textual information on major life domains. *Journal of Personality, 84*(2), 214–224. https://doi.org/10.1111/jopy.12153.

Botella, J., & Sánchez-Meca, J. (2015). *Meta-Análisis en Ciencias Sociales y de la Salud [meta-analysis in social sciences and health]* (Editorial Síntesis).

Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science, 26*(5), 570–582. https://doi.org/10.1177/0956797614566658.

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

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.

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015, May). Values in words: Using language to evaluate and understand personal values. In *Proceedings of the ninth international AAAI conference on web and social media* (pp. 31–40). UK: Oxford.

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.

Carvalho, M. B., & Louwerse, M. M. (2017). Grammar-based and lexicon-based techniques to extract personality traits from text. In *Proceedings of the 39th annual conference of the cognitive science society* (pp. 1727–1732). UK: London.

Cattell, R. B., & Eber, H. W. (1950). *The 16 personality factor questionnaire* (Institute for Personality and Ability Testing).

Celli, F., Lepri, B., Biel, J. I., Gatica-Perez, D., Riccardi, G., & Pianesi, F. (2014, November). The workshop on computational personality recognition 2014. In *Proceedings of the ACM international conference on multimedia* (pp. 1245–1246). USA: Orlando.

Chung, C. K., & Pennebaker, J. W. (2018). What do we know when we LIWC a person? Text analysis as an assessment tool for traits, personal concerns and life stories. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *The SAGE Handbook of Personality and Individual Differences: The Science of Personality and Individual Differences (pp. 341–360)*. https://doi.org/10.4135/9781526451163.n16. SAGE.

De Raad, B. (2000). *The big five personality factors: The Psycholexical approach to personality*. Hogrefe & Huber Publishers.

Eysenck, H. J. (1968). *Eysenck personality inventory* (Educational and Industrial Testing Service).

*Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., … De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction, 26*(2–3), 109–142. https://doi.org/10.1007/s11257-016-9171-0.

Farnadi, G., Sushmita, S., Sitaraman, G., Ton, N., De Cock, M., & Davalos, S. (2014, November). A multivariate regression approach to personality impression recognition of vloggers. In *Proceedings of the WCPR* (pp. 1–6). USA: Orlando.

*Farnadi, G., Zoghbi, S., Moens, M. F., & De Cock, M. (2013, June). Recognizing personality traits using Facebook status updates. In Seventh International AAAI (Ed.), *Conference on weblogs and social media* (pp. 14–18). MI: Ann Arbor.

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.

Gao, N., Shao, W., & Salim, F. D. (2019). Predicting personality traits from physical activity intensity. *Computer, 52*(7), 47–56. https://doi.org/10.1109/MC.2019.2913751.

*Gao, R., Hao, B., Bai, S., Li, L., Li, A., & Zhu, T. (2013, October). Improving user profile with personality traits predicted from social media content. In *Proceedings of the 7th ACM conference on recommender systems* (pp. 355–358). China: Hong Kong.

Gill, A. J., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences, 40*(3), 497–507. https://doi.org/10.1016/j.paid.2005.06.027.

*Golbeck, J. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research, 2*(2), 1–10. doi:10.17705/1atrr.00009.

*Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011, October). Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 149–156). Massachusetts, USA: Boston.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26.

Goldberg, L. R., & Strycker, L. A. (2002). Personality traits and eating habits: The assessment of food preferences in a large community sample. *Personality and Individual Differences, 32*(1), 49–65. https://doi.org/10.1016/S0191-8869(01)00005-8.

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*, 379–398. https://doi.org/10.1037/0022-3514.82.3.379.

Grucza, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment, 89*(2), 167–187. https://doi.org/10.1080/00223890701468568.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science, 14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372.

Hall, J. A., Goh, J. X., Mast, M. S., & Hagedorn, C. (2016). Individual differences in accurately judging personality from text. *Journal of Personality, 84*(4), 433–445. https://doi.org/10.1111/jopy.12170.

Hall, J. A., Pennington, N., & Lueders, A. (2014). Impression management and formation on Facebook: A lens model approach. *New Media & Society, 16*(6), 958–982. https://doi.org/10.1177/1461444813495166.

Hathaway, S. R., & McKinley, J. C. (1951). *Minnesota multiphasic personality inventory; manual (revised)*. Psychological Corporation.

*Hawkins, I. I., Raymond, C., & 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.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504. https://doi.org/10.1037/1082-989X.3.4.486.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*(1), 78–79. https://doi.org/10.1037/0003-066X.58.1.78.

Higgins, J.P.T. & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0* [updated March 2011]. The Cochrane Collaboration, 2011. Retrieved from www.handbook.cochrane.org.

Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality, 42*(3), 747–754. https://doi.org/10.1016/j.jrp.2007.07.011.

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *The Quarterly Journal of Experimental Psychology, 70*(8), 1603–1619. https://doi.org/10.1080/17470218.2016.1195417.

*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.

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C., Deters, F. G., … 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*, 193. https://doi.org/10.1037/1082-989X.11.2.193.

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*(25), 1–12. https://doi.org/10.1186/1471-2288-14-25.

Ivcevic, Z., & Ambady, N. (2012). Personality impressions from identity claims on Facebook. *Psychology of Popular Media Culture, 1*, 38–45. https://doi.org/10.1037/a0027329.

Ji, L. J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology, 87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57.

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology*. Oxford University Press.

Jorge-Botana, G., Olmos, R., & Luzón, J. M. (2020). Bridging the theoretical gap between semantic representation models without the pressure of a ranking: Some lessons learnt from LSA. *Cognitive Processing, 21*, 1–21. https://doi.org/10.1007/s10339-019-00934-x.

Kosinski, M., & Stillwell, D. J. (2011). myPersonality research wiki. myPersonality project. Retrieved from http://mypersonality.org/wiki.

Küfner, A. C., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality, 44*, 427–435. https://doi.org/10.1016/j.jrp.2010.05.003.

*Kwantes, P. J., Derbentseva, N., Lam, Q., Vartanian, O., & Marmurek, H. H. (2016). Assessing the Big Five personality traits with latent semantic analysis. *Personality and Individual Differences, 102*, 229–233. https://doi.org/10.1016/j.paid.2016.07.010.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (SAGE).

Louwerse, M. M., & Benesh, N. (2012). Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth. *Cognitive Science, 36*(8), 1556–1569. https://doi.org/10.1111/cogs.12000.

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science, 33*(1), 51–73. https://doi.org/10.1111/j.1551-6709.2008.01003.x.

*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.

Mandl, T. (1999, September). *Efficient preprocessing for information retrieval with neural networks*. Presented at the 7th European Congress on Intelligent Techniques and Soft Computing (pp.13-16). Aachen, Germany.

Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology, 90*(6), 1014–1031. https://doi.org/10.1037/0022-3514.90.6.1014.

Martínez-Huertas, J. A., Jorge-Botana, G., Luzón, J. M., & Olmos, R. (2021). Redundancy, isomorphism and propagative mechanisms between emotional and amodal representations of words: A computational study. *Memory & Cognition, 49*(2), 219–234. https://doi.org/10.3758/s13421-020-01086-6.

McCrae, R. R., & Costa, P. T., Jr. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research (3rd ed., pp. 159–181)* (Guilford).

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science, 3*(1), 3–17. https://doi.org/10.1111/j.1756-8765.2010.01117.x.

Mitra, T., Counts, S., & Pennebaker, J. W. (2016, May). Understanding anti-vaccination attitudes in social media. In Tenth International AAAI (Ed.), *Conference on web and social media* (pp. 269–278). Germany: Cologne.

Mohammad, S., Zhu, X., & Martin, J. (2014, June). Semantic role labeling of emotions in tweets. In *Proceedings of the WASSA* (pp. 32–41). Maryland, USA: Baltimore.

Olaru, G., Burrus, J., MacCann, C., Zaromb, F. M., Wilhelm, O., & Roberts, R. D. (2019). Situational judgment tests as a method for measuring personality: Development and validity evidence for a test of dependability. *PLoS One, 14*(2), Article e0211884. https://doi.org/10.1371/journal.pone.0211884.

Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the "same" construct? A meta-analytic investigation. *Personality and Individual Differences, 49*(7), 669–676. https://doi.org/10.1016/j.paid.2010.06.014.

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

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6.

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

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10*(3), 90–93. https://doi.org/10.1111/1467-8721.00123.

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.

*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.

*Qiu, L., Lu, J., Ramsay, J., Yang, S., Qu, W., & Zhu, T. (2017). Personality expression in Chinese language use. *International Journal of Psychology, 52*(6), 463–472. https://doi.org/10.1002/ijop.12259.

R Development Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07-0. URL http://www.R-project.org/.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (2nd ed.) (pp. 295–315)* (Russell Sage Foundation).

*Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., … 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.

*Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016, April). Fusing social media cues: Personality prediction from twitter and Instagram. In *Proceedings of the 25th international conference companion on World Wide Web* (pp. 107–108). Québec, Canada: Montréal.

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality.*. https://doi.org/10.1002/per.2257.

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.

*Thilakaratne, M., Weerasinghe, R., & Perera, S. (2016, October). Knowledge-driven approach to predict personality traits by leveraging social media data. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 288–295). NE, USA: Omaha.

Van Lankveld, G., Spronck, P., Van den Herik, J., & Arntz, A. (2011, September). Games as personality profiling tools. In *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)* (pp. 197–202). South Korea: Seoul.

Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*(1), 123–132. https://doi.org/10.1037/0022-3514.87.1.123.

Veselka, L., Schermer, J. A., & Vernon, P. A. (2011). Beyond the big five: The dark triad and the supernumerary personality inventory. *Twin Research and Human Genetics, 14*(2), 158–168. https://doi.org/10.1375/twin.14.2.158.

Viechtbauer, W. (2010a). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48. doi:10.18637/jss.v036.i03.

Viechtbauer, W. (2010b). Metafor: Meta-analysis package for R. R package version 1.4-0, URL. http://CRAN.R-project.org/package=metafor.

*Xianyu, H., Xu, M., Wu, Z., & Cai, L. (2016, July). Heterogeneity-entropy based unsupervised feature learning for personality prediction with cross-media data. In *2016 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6). WA, USA: Seattle.

*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.