



Published in final edited form as:

J Res Pers. 2008 February ; 42(1): 96–132. doi:10.1016/j.jrp.2007.04.006.

Revealing Dimensions of Thinking in Open-Ended Self-Descriptions: An Automated Meaning Extraction Method for Natural Language

Abstract

A new method for extracting common themes from written text is introduced and applied to 1,165 open-ended self-descriptive narratives. Drawing on a lexical approach to personality, the most commonly-used adjectives within narratives written by college students were identified using computerized text analytic tools. A factor analysis on the use of these adjectives in the self-descriptions produced a 7-factor solution consisting of psychologically meaningful dimensions. Some dimensions were unipolar (e.g., Negativity factor, wherein most loaded items were negatively valenced adjectives); others were dimensional in that semantically opposite words clustered together (e.g., Sociability factor, wherein terms such as shy, outgoing, reserved, and loud all loaded in the same direction). The factors exhibited modest reliability across different types of writing samples and were correlated with self-reports and behaviors consistent with the dimensions. Similar analyses with additional content words (adjectives, adverbs, nouns, and verbs) yielded additional psychological dimensions associated with physical appearance, school, relationships, etc. in which people contextualize their self-concepts. The results suggest that the meaning extraction method is a promising strategy that determines the dimensions along which people think about themselves.

Keywords

LIWC; Meaning Extraction Method; natural language; self-descriptions

In a job or clinical interview, meeting an office mate for the first time, or talking to someone at a party, we usually ask others to tell us about themselves. Directly or indirectly, we elicit people's descriptions of themselves to construct a coherent sense of them. With the exception of a small group of social and personality psychologists, most people consider free descriptions as the currency of everyday life. From the media describing politicians and celebrities, reference letters, or through casual observation and gossip, much of what we think and know about others is shaped by everyday descriptions of them.

Although free descriptions are the natural form of appraising what people are like, most formal personality assessment methods use itemized questionnaires. With statistical methods, questionnaires make it easy to summarize people's personality trait levels, to make comparisons across respondents, and to generalize across groups. With a common vocabulary of personality factors, it is easy for researchers to communicate with one another about the major dimensions of personality that are associated with a variety of behaviors. A recurring

Address correspondence to Cindy K. Chung (cindy.chung@mail.utexas.edu) or James W. Pennebaker (pennebaker@mail.utexas.edu), Department of Psychology A8000, The University of Texas at Austin, Austin, TX. 78712 USA. Preparation of this paper was aided by grants from the National Institutes of Health (MH52391) and the National Science Foundation (BCS-0528087).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

criticism of trait level questionnaires, however, is that they often present difficulties in ecological validity; responses are often constrained to numerical value judgments of experimenter-defined traits.

Here, we introduce a new method aimed at resolving the tradeoff between ease of data management and ecological validity. We describe an automated meaning extraction method to examine the self based on people's free descriptions of who they think they are. Using computerized text analytic tools, the words people use in open-ended self-descriptions can be counted, correlated, and factor analyzed in a way that makes it possible to capture the dimensions along which people think about themselves.

Distilling Personality from Language: The Lexical Approach to Personality

Following from the idea that language is important to the description of personality, Allport and Odbert (1936) embarked on the lexical approach to personality. The lexical approach assumes that all of the important ways in which individuals differ are represented by words. Allport and Odbert combed through an English language dictionary to find a total of 17,953 terms referring to individual differences. To reduce this list into a manageable set of major dimensions, continual refinements were made to the original list, culling for ambiguity, obscurity, obsolescence, redundancy, purely evaluative terms, temporary states, physical traits, and other considerations (for a review, see Goldberg, 1982). The list was further reduced statistically by performing a cluster analysis on self- and peer-ratings of the terms, with adjectives as the preferred unit of speech. The list of person descriptors came to focus primarily on adjectives, which vary in degree and kind as traits do (Saucier & Goldberg, 1996)

The most reliable orthogonal factor structure derived from the lexical approach across various word lists and across cultures is a set of factors called the Big Five which include: Extraversion, Agreeableness, Conscientiousness, Emotional Stability or Neuroticism, and Intellect or Openness to Experience (Goldberg, 1990). There are several self-report personality inventories that capture these factors such as the NEO-PIR (Costa & McCrae, 1992), the Big Five Inventory (BFI; John & Srivastava, 1999), and the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). For these scales, respondents are required to rate the degree to which a set of adjectives or attribute statements describe their personalities. These scales are reliable, and have been validated in multiple ways (Saucier & Goldberg, 2001).

However, a number of criticisms have been leveled at the Big Five. First, varying trait variable selections, by inclusion of evaluative or temporary state terms, for example, have resulted in factors that partially resemble, amalgamate, or are beyond those of the Big Five (for a review, see Paunonen & Jackson, 2000; Saucier, 1997). Second, there have been no factor analyses on trait lists compiled from actual high frequency usage in everyday language. Specific attempts have been made to include only those attributes rated as highly familiar by a series of judges, resulting in a factor structure much like the Big Five (Saucier, 1997; Saucier & Goldberg, 1997). However, it is not known how closely judges' ratings correspond to actual frequency usage. Third, the lexical approach to personality shares some of the same problems that come with any nomothetic self-report measure (for a review, see Block, 1995). With standardized check-lists and rating scales, researchers are able to define the variables of interest and have participants rate themselves on these dimensions in order to determine varying *trait levels* between individuals. However, these self-report measures have been criticized for being able to predict only some behaviors some of the time, partly due to the fact that only a subset of the measured traits are relevant to a person's self-concept (e.g Bem & Allen, 1974). *Trait relevance* has largely been ignored in the scientific study of personality (Lamiell, 1981).

Put another way, personality consists of "having" and "doing" (Allport, 1937). Traditional personality scales capture person personality as "having" particular traits but they do not

capture personality as “doing” (Cantor, 1990); most traditional personality scales do not capture the cognitive generalizations that organize and guide the processing of self-relevant information (Markus, 1977).

Capturing Dimensions of Thinking about the Self: The Cognitive Approach to Personality

One way to capture the cognitive processes at play in thinking about the self is to simply ask people to describe themselves. Although the notion that language determines thought is outdated, current experimental studies have shown that language can be considered a mechanism for “thinking” (Slobin, 1996). The way language is used tells us much about what our minds attend to, how they associate objects and events, what attributions are made, etc. The patterns of words that are used in a self-description might reveal the content and categorization of the self along various dimensions. “[George Kelly’s] plea, one shared with Gordon Allport, was that if we wanted to know what people were on about, ask them, and they might just tell us.”(Little, 2005).

Kelly’s (1955) theory of personal constructs proposed that individuals construct their interpretations of the world based on past unique experiences and use these schemas to guide them through the world. Personal constructs represent the salient, chronically activated dimensions along which an individual’s perception and behaviors are influenced. Markus (1977) built on Kelly’s work by showing that self-schemata guide information processing about the self. Measured by trait-level extremity and trait importance ratings, self-schemata enabled people to respond to personality items with shorter latencies in response to attribute relevant information than to attribute irrelevant relevant information, to generate behavioral instances related to one’s self, and to make individuals resistant to counterschematic information. More recently, Robinson (2004) found that *trait levels* do not lead to quicker judgments of trait-related words, but for judgments with long response latencies, there is a high correspondence between trait levels and judgments. His research suggests that trait levels on traditional self-reports might serve as a ‘fill-in’ belief system that kicks in when our more automatic system fails to respond (Robinson & Oishi, 2006).

Much of the cognitive approaches to personality, then, involve the measurement of *trait relevance*, either through reaction time (e.g. Robinson, 2004) or idiographic measures with some structure (e.g. Kelly, 1955; Kenrick & Stringfield, 1980). Although reaction time measures provide evidence for the cognitive processes in self-schemata and enable the ease of comparisons across individuals, what is lacking is a description of individuals (Lamiell, 1981).

Despite its intuitive appeal, the open-ended nature of idiographic assessments has kept it in the shadows of personality psychology. As with any idiographic approach, analyzing idiosyncracies in responses requires elaborate coding schemes, along with multiple trained raters. The natural variance in the degree of creativity and willingness to disclose about particular topics creates unwanted variability in the number and kinds of ideas that a researcher must analyze. There is no need for open-ended data if the goal of the researcher is simply to answer a specific question about a particular concept known to the researcher (Chaplin & John, 1989).

When self-generated adjective lists (e.g., Chaplin & John, 1989; Grice, 2004; Schiller, Tellegen, & Evens, 1995; Westen, 1996) or completely open-ended self-descriptions have been used (Donahue, 1992; Fiske & Cox, 1979, Bromley, 1977; McGuire & Padawar Padawar-Singer, 1976; Peevers & Secord, 1973; Prentice, 1990), the words or statements have been classified according to the Big Five or other imposed trait structures. These studies have shown

that the adjectives people use in ordinary self-descriptions can be classified into the Big Five categories that are used by researchers. But this does not mean that these are the salient dimensions along which lay people would naturally classify their attributes, traits, and behaviors.

Imposing a trait structure onto open-ended responses detracts from what the idiographic approach has to offer: the unique perspective of the individual. By using an idiographic approach through free response formats, subsequent behavior ratings have been shown to be more valid, presumably through the activation of relevant self-knowledge (Claeys, de Boeck, van den Bosch, Biesmans, & Bohrer, 1985). Similarly, unstructured interviews allow for 'weak' rather than 'strong' situations, allowing for behavior to vary more widely and to be more informative (Ickes, Snyder, & Garcia, 1997; Blackman & Funder, 2002). Despite its potential for illuminating lay perspectives, and for improving the validity of personality research, free response formats remain undervalued and underused.

Describing People by the Ways They Think: The Meaning Extraction Method

The lexical approach uses rigorous statistical methods to reduce personality to scores on a few broad variables. Its key strength is generalizability but its weakness is the loss of information about any given individual. On the other hand, the more open-ended nature and individualized measurement in an idiographic approach yields data that are unmanageable and difficult to generalize across individuals. In this paper, we wed these two approaches to examine how an imposed trait structure differs from a naturally-occurring system when people describe themselves. Is the structure that people naturally use in describing themselves amenable to a lexical strategy that is objective and, to some degree, nomothetic?

Our approach relies on a relatively simple factor analytic approach to people's natural language. Basically, we ask large groups of people to describe their personality in an open-ended way. The most common words used across all of the self-descriptions are then tabulated for each person's essay. Imagine, then, that we create a matrix of the 800 most commonly used words (excluding function words like pronouns prepositions etc) across the top with each person's self-descriptive data down the side. The matrix contains 1's and 0's reflecting if each word is or is not used in that particular essay. Now imagine that we compute a factor analysis on the matrix. The resulting factors will tell us which words tend to clump together. Indeed, as described in the methods and results section, these factors reflect meaningful groups of words or themes.

From a personality perspective, the meaning extraction method assesses personal concerns (McAdams, 1995). The word factors reflect how people are structuring their worlds – how they are thinking about themselves and, other salient topics of self-relevance. What distinguishes this approach from traditional self-report measures is that it is purely inductive; participants write whatever they think is relevant to themselves, and the computer simply calculates the degree to which their words form semantic clusters.

Only recently have developments in automated tools for analyzing language allowed researchers to explore natural language in efficient and reliable ways (Graesser, Gernsbacher, & Goldman, 2003). Combining these tools with statistical techniques, it is possible to extract patterns in language and relate them to theoretically meaningful psychological constructs (for a review, see Pennebaker, Mehl, & Niederhoffer, 2003). We can now efficiently analyze open-ended self-descriptions, guided by the logic and statistical methods of the lexical approach. Our meaning extraction method shares many features with Latent Semantic Analysis (LSA; Foltz, 1996; Landauer, Foltz, & Laham, 1998). LSA examines the similarity between texts with singular value decomposition on the occurrence of key content words in a text. Related methods, and especially variants of components analyses, have been instrumental in providing

intuitively comprehensible word categories in text analyses (e.g., Väyrynen & Honkela, 2005).

If the goal of the lexical approach to personality is to build a taxonomy of the major underlying personality dimensions, examining the words we use in everyday life to describe ourselves may be a good place to start to build a taxonomy of the major dimensions along which people think about themselves (Peabody, 1987; Tellegen, 1993). By sampling from the most frequently used adjectives in self-descriptions, we build on the capabilities of several automated language tools and principle components analysis to extract major themes in open-ended self-descriptions. In this paper, four broad questions will be addressed:

1) What are the basic dimensions of self-concepts based on adjectives in open-ended self-descriptions?

Since the frequency of attribute use in a language tends to correspond with the relative importance of that attribute, a persistent goal in the lexical hypothesis has been to examine only the most commonly used set of adjectives in natural language (Saucier & Goldberg, 1996). However, previous studies have not examined personality descriptions in everyday natural language to determine the important attributes that comprise the basic dimensions of self-concepts due to the limitations posed by text analyses. The present study examines actual high frequency usage of a wide selection of adjectives in open-ended self-descriptions with a word frequency ranking tool, WordSmith (Scott, 1996), and a word usage counting tool, Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001). A principal components analysis was performed on the use of each of the most commonly used adjectives to see how they naturally co-occur in self-descriptions. This text analytic procedure is called the Meaning Extraction Method.

In order to have participants to focus less on any extraneous circumstances during the task, and to focus more on the writing topic (the self), respondents were asked to first attend to themselves through the use of a self-awareness manipulation (Duval & Wicklund, 1972).

2) What are the psychometric properties of the self-concept dimensions?

Previous attempts at deriving the basic dimensions of personality have resulted in slightly varied factor structures due to trait selection procedures. Clearly, it is crucial to ensure that our factor structure is not simply an artifact of the adjectives from our specific sample of self-descriptions and to show that our factor structure is reliable across a slightly different method of adjective selection. Therefore, a similar analysis was conducted using the most frequently used adjectives in Saucier's (1997) selection of high-familiarity adjectives. That is, the patterns of use of Saucier's high familiarity adjectives in the same sample of self-descriptions were assessed using a similar factor analytic procedure. To the extent that our resultant factor structure represents the salient, chronically activated and dispositional aspects of the self, we would expect to find reliable word usage for each factor in stream of consciousness essays written by the same he group of participants a couple of months apart A reliability analysis of the factor structure was performed on word usage in the self-descriptions and in stream of consciousness essays.

3) How are the self-concept dimensions related to other known self-report measures?

Quantification of attributes typical of personality questionnaires have allowed for them to be related to other measures with ease Idiographic approaches, on the other hand, are generally more descriptive in nature and have therefore been difficult to relate to other quantitative measures in psychology. In order to demonstrate the relations of lay self-concept dimensions derived from the meaning extraction method to established personality traits, regression-based factor scores will be calculated for each of the dimensions extracted from free response formats.

These regression-based factor scores will be correlated with scores on the Big Five dimensions and other personality and demographic measures.

4) To what degree do the adjective-based self-concept dimensions overlap with dimensions from a similar analysis based on all content words?

Both nouns and verbs have been shown to carry dispositional meaning (Hofstee & Van Heck, 1990). Nouns can clearly denote roles or types of people, and from them, we may be able to infer more about personality than from other parts of speech (de Raad & Hoskens, 1990; Saucier, 2003). Even verbs and adverbs convey dispositional meaning (de Raad & Hofstee, 1993; de Raad, Mulder, Kloosterman, & Hofstee, 1988; Semin & Marsman, 1994). In the English language, there are more person-descriptive adjectives than person-descriptive nouns, with a higher proportion of nouns carrying a negative connotation, and nouns having a higher proportion of slang words (Goldberg, 1982). As such, the bulk of lexical research on personality has focused on adjectives.

It is likely that when participants describe their self-concepts, they mention things other than their stable dispositional traits that they see as central to their personality (Epstein, 1973). For example, childhood, pets, belongings, role models, or groups have been found to be important to one's identity (Csikszentmihalyi & Rochberg-Halton, 1981; Kihlstrom, Cantor, Albright, Chew, Klein, & Niedenthal, 1988). Furthermore, behaviors (Allport, 1961), along with past and future acts or goals and selves have been shown to be important to the self and identity (Cantor, Markus, Niedenthal, & Nurius, 1986; Higgins, Klein, & Strauman, 1985; Ross & Buehler, 1994). To demonstrate what the factor structure of open-ended self-descriptions would look like with all of these content words considered, a similar analysis was carried out on all content words (i.e., adjectives, adverbs, nouns, and verbs).

Methods

Participants

College students enrolled in introductory psychology classes of one of the authors completed a variety of writing assignments over the semester as part of a class assignment. The responses of students from three consecutive years (2002, 2003 and 2004) were combined, resulting in a sample of 1430. Of these, 180 (12.6%) failed to turn in the target assignment, which was an online self-description essay. An additional 73 (5%) did not give consent to allow the authors to analyze the essays with all identifiers removed. Finally, six of the self-descriptive essays did not have a total word count of at least 100 words, suggesting that they had not taken the 20-minute self-description writing task seriously, and so were excluded from the sample. The final sample was composed of 1165 students, with a mean age of 18.7 ($SD = 1.64$) of whom 61.0% were women.

Measures

In addition to the target self-description essay, participants also completed a stream of consciousness essay as well as various questionnaires over the course of the semester. The stream of consciousness essay was a class exercise that was available online wherein students were asked to track their thoughts as they occurred for 20 minutes (see Pennebaker & King, 1999 for detailed description and psychometrics of the method). Of the students who completed the self-description essay, 1104 completed the stream of consciousness essay. The mean word count for the stream of consciousness essays was 798.3 words ($SD=248.7$) Note that the stream of consciousness assignment was due during the first three weeks of class and the target self-description writing assignment was due approximately 2 months later.

Demographic information, grades for the class, self-reported SAT scores, the Big Five Inventory (BFI; John & Srivastava, 1999) and the short form of the Beck Depression Inventory (Beck, Rial, & Ricketts, 1974) were collected as part of their introduction to the various topics in psychology. Most questionnaires were available online and, when participant participants completed them, they received feedback about their score and a description of the scales and how they worked. Because different students were absent or chose not to complete specific questionnaires on different days, sample sizes for the different measures vary.

Procedure

As part of a class exercise on self-reflection, students completed a 20-minute writing exercise on the class webpage during the last two weeks of the semester. The writing instructions elicited attention to important aspects previously theorized and found to be important to self-concepts including possible selves and others' perceptions of the self (see Cantor et al., 1986; Epstein, 1973; Markus & Nurius, 1986). In order to emphasize the focus of respondents to the self as an object of reflection, a self-awareness manipulation using a mirror was used (Duval & Wicklund, 1972). The online instructions read:

For this writing assignment, you will need a mirror. If you can, put the mirror next to your computer. If this isn't possible, go find a mirror and stand in front of it. Before you begin writing, gaze at your image for several minutes and think about who you are. The goal of this writing assignment, then, is to look at yourself and to think about who you are, who you have been in the past, and who you would like to be.

Gaze into your own eyes, look at your face. See yourself as others see you and as how you see yourself. While looking at your image, think about where you are in your life, your connections to others, and who you really are. After closely examining yourself in the mirror for several minutes, return to your computer and write about who you are.

When they were ready, participants clicked a link on the screen that started a 20 minute timer presented on the writing webpage, which included a large blank text field for typing. Individuals were told that they must write for the full 20 minutes and that they should write during the entire time. An automated message flashed onto the screen indicating when 20 minutes had passed and that participants could finish typing or at continue if they wanted. Participants used a randomly-assigned unique class code as an identifier, which linked their individual responses to the measures noted above. Their grade was dependent only on completing the assignment; content was not considered.

A sample of the self-descriptions can be read in Table 1. The self-descriptions were personal and greatly varied in style. Many commented on their appearance, their worries, their past and future roles, the effects their behaviors had on their social networks, and others' appraisals of themselves among other concerns. From this small sample of self-descriptions, it is clear that students took the assignment seriously and felt free to disclose even highly sensitive topics.

Text Analytic Strategy

Each of the self-descriptions collected from the class webpage was formatted as a single plain text file. The analyses were initially conducted for each class separately (2002, 2003, and 2004). However, because the most frequently used adjectives in each of the years were so similar, the final analyses presented below were conducted with data collapsed across classes.

Use of text-based adjectives—In order to determine the most frequently used adjectives in the self-descriptions, frequency counts were taken of all words, excluding closed-class or function words (e.g., articles, auxiliary verbs, prepositions, pronouns, etc.), using a

computerized word counter, WordSmith (Scott, 1996). Adjectives were chosen as a basis for deriving structures of self-concepts as prescribed by the traditional lexical approach (Saucier & Goldberg, 1996). Only those person-descriptor adjectives used in at least 3.0% of all self-description text files were considered in order to roughly meet the multiple variables to cases ratio guidelines for factor analyses (For reviews, see Arrindell & van der Ende, 1985; MacCallum, Widaman, Zhang, & Hong, 1999). From this list, any words that were not adjectival person descriptors, including quantitative modifiers (e.g., *few*, *some*), intensifying adjectives or adverbs (e.g., *much*, *very*), and spatial or temporal adjectives (e.g., *low*, *past*) were excluded. Also, proper adjectives, capitalized adjectives derived from proper nouns. (e.g., *American*, *Asian*, *Christian*), were excluded. A total of 88 adjectival person descriptors were retained for further analyses.

Once the adjectives were identified, each was counted separately within each essay using a feature of the text analysis program Linguistic Inquiry and Word Count. (LIWC, Pennebaker et al., 2001). LIWC is a software program that assesses the occurrence of a word or a category of words in text files. A user-defined dictionary directs LIWC as to which words or categories of words to search for. A dictionary containing each selected adjective was compiled for LIWC. Each adjective (e.g., *funny*) and, if possible, associated comparative (e.g., *funnier*) and superlative (e.g. *funniest*) forms comprised its own LIWC category. The self-descriptions were assessed for use (coded as 1) or absence (coded as 0) of each of the adjective categories in the LIWC dictionary. The final data summary, then, can be thought of as an 88 (adjectives) by 1165 (participant essays) matrix with each entry referring to the presence or absence of each adjective within each essay.

Use of Saucier high-familiarity adjectives—As noted in the introduction, there are at least two concerns of using the text-derived adjectives. The first is the subjectivity of defining an appropriate set of adjectives in the first place. The second is that the domain of adjectives ultimately selected is limited to the texts themselves. As a comparison to this approach, separate counts of Saucier's (1997) high-familiarity adjectives were made. Saucier's (1997) adjective list consisted of 500 adjectives rated by judges as frequently being used to describe a person, with an additional 25 to represent the major Big Five markers. Of these 525 adjectives, only those adjectives used in at least 3.0% of all the self-descriptions were considered for inclusion as a category in a separate LIWC dictionary. Overall, 119 Saucier adjectives were retained for further analyses. For this dictionary, each adjective (e.g., *lazy*) and, if possible, associated comparative (e.g. *lazier*) and superlative (e.g., *laziest*) forms comprised its own LIWC category.) self-descriptions were assessed for use or absence of each of the adjective categories in the LIWC dictionary. The final data summary, then, can be thought of as a 119 (adjectives) by 1165 (participant essays) matrix with each entry referring to the presence or absence of each adjective within each essay.

Use of all content words: Nouns, regular verbs, adjectives, and adverbs—In order to determine the most frequently used words in the self-descriptions, frequency counts were taken of all words, excluding closed-class words (i.e., pronouns, prepositions, conjunction, articles, and auxiliary verbs) using WordSmith. Only those root words used in at least 3.0% of all the self-description essays were considered for inclusion as a category in a separate LIWC dictionary. For this dictionary, each category consisted of a word (e.g., *support*), and all forms of its root word that could be produced using an alternate prefix (e.g. *supportable*, *supported*, *supporter*, *supporting*, *supportive*, *supports*), whether or not the alternate form appeared in at least 3.0% of all the self-descriptions. The self-descriptions were assessed for use (coded as 1) or absence (coded as 0) of each of the root word categories in the LIWC dictionary. The final data summary, then, can be thought of as a 501 (words) by 1165 (participant essays) matrix with each entry referring to the presence or absence of each word category within each essay.

The meaning extraction method—As noted above, our language-analytic strategy assumes that groups of words naturally co-occur in meaningful ways. A person with weight as a prominent part of his self-concept is more likely to habitually attend to, elaborate on, and easily categorize and access thoughts related to weight (e.g. Cantor, 1990; Markus, 1977). Accordingly, we assume that this will be reflected in more spontaneous uses of weight-related words like diet, fat, and chubby than someone who never thinks about weight. A particularly efficient way to determine the degree to which groups of words cluster together is to rely on a factor analytic approach. For each of the three word by essay matrixes, simple principal components analyses using varimax rotation were performed.

It should be emphasized that the analysis of word usage is fundamentally different from correlations and factor analyses of self-reports. For example, virtually all words have a modal use of zero. In the meaning extraction method, because all matrix entries are binary, the numbers that are analyzed simply reflect whether or not each participant used each of the words in the dictionary. Consequently, in the text-based adjective analyses, the principle components analyses were based on the correlation matrix of the 88 adjective occurrences across the 1165 participants' essays. Note that the assumptions underlying this approach are congruent with those in latent semantic analysis (Landauer, Foltz, & Laham, 1998), computational modeling of semantic spaces (e.g., Buchanan, Westbury, & Burgess, 2001), and other approaches to explore natural language patterns (e.g., Graesser, Cai, Louwerse, & Daniel, 2006; Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; Graesser, McNamara, Louwerse, & Cai, 2004). Although we report the results from principal components analyses with varimax rotation, virtually identical results were obtained using principal axis analyses and with promax, oblique, and equamax rotations.

Results

The goal of this research was to determine if the meaning extraction method yields a meaningful, coherent, and useful way of thinking about the self. Each of the four broad questions raised in the introduction are addressed separately.

Question #1: What are the basic dimensions of self-concepts based on adjectives in open-ended self-descriptions?

A principal components extraction with varimax rotation was first performed on the most frequently used adjectives from the large sample of self-descriptions. Diagnostic tests indicated that a factor model was appropriate for the data ($KMO = .55$, Bartlett's test of sphericity = 5,791.36, $p < .001$). The resulting components will be referred to here as factors. Based on a scree of eigenvalues for the principal components analysis (Cattell, 1966; Stevens, 1992), seven factors were extracted. The first seven factors accounted for 14.7% of the total variance. Although this percentage is small for most factor analyses of questionnaires, it is not the case for natural language use. Considering the generative nature of language (the multitude of ways that ideas can be constructed and the number of synonyms that can be substituted for a word), 14.7% of the total variance in adjective use across 1165 essays is remarkably high. Factor loadings of .20 or higher were retained.

As can be seen in Table 2, each of the 7 factors brings together a group of adjectives that are psychologically meaningful and coherent. Intuitively comprehensible word patterns were captured by this meaning extraction method, despite the fact that the data (actual word use in open-ended self-descriptions) have different ranges and distributions than that of typical Likert-scale ratings.

Closer inspection of the factors reveals that the language-based dimensions are constructed in one of two ways. Approximately half of the factors are reminiscent of Kelly's dimensional

thinking associated with personal constructs. That is, semantically opposite terms tend to load in comparable ways. For example, the highest loaded items on Factor 1, which we have labeled Sociability, include quiet, shy, outgoing, reserved, comfortable, open, friendly, and insecure. The data suggest that the more a person uses the word *shy*, the more likely they are to use *outgoing*. Similarly, items on Factor 2 (Evaluation) include ugly, fat, attractive, beautiful, nice, smart, stupid, special, lazy, and cute. Factor 4 (Self-Acceptance) includes a similar group of semantically opposite words.

The other factors are more congruent with traditional trait-like dimensions wherein the majority of items are scored in a similar, unipolar way. Factors 3 (Negativity), 5 (Fitting-In), 6 (Psychological Stability), and 7 (Maturity) are generally composed of similarly-valenced items that are roughly synonymous. The highest-loaded items on the Negativity factor include mad, hurt, bad, sad, horrible, upset, etc. Similarly, the Maturity items include mature, successful, caring, capable, accomplished, determined, responsible, and loving. Indeed, one could imagine that these factors might closely mirror markers of the Big Five. For example, in many ways Sociability resembles extraversion, Negativity overlaps with neuroticism, and Maturity parallels conscientiousness. However, the other factors provide no clear links to those of the Big Five.

Question #2: What are the psychometric properties of the personality dimensions?

Overall, 2.75% of the total words in the self-description essays included adjectives from the list based on the 7 primary factors. As can be seen in Table 3, the base rate of adjective use was fairly evenly distributed across the factors. Not surprisingly, not all participants used adjectives from each of the categories. For example, 33.8% of the entire sample did not make reference to a Sociability adjective. Conversely, almost 90% of writers mentioned at least one adjective from the negativity factor. Across all 7 dimensions, only 4 of 1165 participants (0.03%) did not use a single adjective from any dimension and 12.8% used adjectives in all 7 categories. The mean number of categories for which individuals mentioned at least one adjective was 4.87 ($SD = 1.47$).

In order to ensure that our factor structure was not an artifact of our means of adjective selection, separate factor analyses were conducted on the 119 Saucier adjectives. There was a high degree of overlap in factor structure with the text-based adjectives. [Please see Appendix A for the factor structure of Saucier adjectives.] The scree analysis also suggested the use of 7 factors, which accounted for 12.6% of the variance. By and large, the Saucier factors overlapped considerably with the factors of the text-based approach. Regression-based factor scores were computed. The correlations between the regression-based factor score means of similar factors averaged $r = .67$ (ranging from .53 to .80). A couple factors from the Saucier analyses overlapped somewhat with the text-based factors but their boundaries were blurred, such that Factor 4 from the Saucier analyses correlated .53 and .47 with factors 3 and 4 from the text-based analyses; Factor 5 from the Saucier analyses correlated .25 and .31 with Factors 4 and 5 from the text-based analyses. Because of the overall similarity of the findings between the text-based and Saucier analyses, the remainder of the results focuses only on the text-based adjective approach. [Please see Appendix B for a correlation table of the inductive adjective factors and the Saucier adjective factors.]

Finally, in order to see if the factors represent topics that could be observed in participants at some distant time when completing a different, more open-ended task, the same adjective categories that were derived from the self-descriptive essays were applied to stream of consciousness essays using LIWC. Such an analysis allowed us to determine the degree that participants tend to use the same groups of adjectives across time and topic. As depicted in the righthand column of Table 3, the cross-essay reliabilities were low but statistically significant (based on simple Pearson correlations, two-tailed tests). Remarkably, these analyses showed

that the factors represent topics or words that were salient to the participants in two different open-ended tasks over time.

Question #3: How are the self-concept dimensions related to other known self-report measures?

Over the course of the semester, students completed a series of questionnaires that tapped the Big Five, depression, and various demographic dimensions. Results from the questionnaires as well as test grades were correlated with the adjectives derived from the text-based adjective factor analyses. Due to the large number of comparisons, a more stringent p level was adopted for significance tests ($p \leq .001$).

Although some of the empirically derived adjective factors were conceptually similar to some of the five factor traits, they were only weakly related to the Big Five Inventory (BFI) dimensions. As can be seen in Table 4, the significant correlations ($p \leq .001$) between BFI dimensions and the text-based factor scores are conceptually related. For example, the Sociability factor was negatively related to extraversion. That is, people who think about sociability and who use words such as *friendly* and *shy* rate themselves low in extraversion. Whereas introverted people were more likely to describe whether or not they relate, extroverted people were more likely to discuss ways of relating, using more words from the Self-Acceptance factor, and marginally more words from the Fitting In and Maturity factors ($p \leq .01$).

Those high in neuroticism tended to score higher in Negativity (using words such as *sad*, *lonely*, and *angry*) and Evaluation (e.g. *ugly*, *attractive*, *fat*, and *beautiful*). Previous studies have documented the relations of neuroticism with perfectionism (Hill, McIntire, & Bacharach, 1997), greater engagement in lower visibility behaviors (e.g., negative inner thought processing), and self-evaluation (Funder & Dobroth, 1987). These findings also confirm previous research that those high on neuroticism tend to use more negative emotion words across contexts (Pennebaker and King, 1999).

It has previously been found that people high on conscientiousness tend to use more positive emotion words (e.g., Pennebaker & King, 1999). Indeed, in this study, we found that conscientiousness had a positive correlation with Maturity ($r = .13$, $p < .001$). The more that people talked about being capable, caring, and successful, the more conscientiousness they were.

Although the Big Five openness factor and our own Fitting In factor both describe unconventionality, they were not significantly correlated. None of our factors derived from adjective use were significantly correlated with openness or agreeableness.

Correlations between demographic variables (e.g., age, sex, SAT scores, grades, and depression) and the adjective factors can be seen in Table 5. Again, the correlations are modest. Whereas no variables are significantly correlated with age, sex of participants was correlated with 4 of the seven factors. Overall, women were higher in Sociability, Negativity, Self-Acceptance, and Psychological Stability. Not surprisingly, Negativity was correlated with depression. Finally, the Psychological Stability factor was correlated with higher grades in the course as well as higher self-reported SAT scores.

Summary

Taken together, the text-based adjective factor approach provides a novel method by which to think about self-concepts. The factors are coherent and are modestly linked to recent trait approaches, such as the Big Five. More striking, however, is the fact that several of the factors

reflect a Kelley-esque personality approach. That is, they suggest that individuals are thinking along certain dimensions.

Despite the promise of the adjective approach, a number of shortcomings are immediately apparent. Perhaps most troublesome is that adjectives are used at surprisingly low rates when individuals are describing themselves. The typical student wrote 544.5 words ($SD = 222.4$) in 20 minutes of which fewer than 3 percent were adjectives. In short, the average student only used 13.3 adjectives ($SD = 7.07$) from our adjective list. As Table 2 suggested, the base rate for adjective use was quite low, resulting in fairly unstable individual scores. However, the adjective factors that were derived from the self-descriptive essays were modestly but significantly correlated with their use in stream of consciousness essays that were written two months earlier.

An alternative way of thinking is to expand the meaning extraction strategy to the use of all content-relevant words in the self-descriptive essays. In doing so, we are making the assumption that the self-concept reflects the totality of word usage rather than just adjectives per se. As noted in the introduction, such an assumption is consistent with other lexical research suggesting that person descriptive attributes and traits are encoded in other parts of speech (nouns, verbs, and adverbs). It is also consistent with research that examines social roles, belongings goals or actions as important determinants or reflections of self-concepts.

Question #4: Can self-concept dimensions be inferred from the analysis of content words?

Recall that we sought to examine the factor structure of all content words used in the self descriptions rather than just adjectives alone. Diagnostic tests indicated that a factor analysis was appropriate for the data, ($KMO = .50$, Bartlett's test of sphericity = 16,619.4, $p < .001$). Based on a scree of eigenvalues for principal components analyses of content word use in the self-descriptions, nine factors were extracted. The factors were rotated with varimax rotation. All further analyses are based on these rotated factors. The first nine factors accounted for 8.97% of the total variance. Factor loadings of 0.20 or higher were retained.

As can be seen in Table 6, the nine factor solution yielded a set of coherent dimensions, some of which resembled those found in the factor analysis of adjectives. Specifically, Factors 3 and 5 of the content words analysis had similar but more elaborate themes to the Evaluation and Sociability factors from the factor analysis of adjectives. Factor 4 resembled the Sociability factor in the adjectival analysis. Adjectives describing sociability (e.g., *quiet, outgoing, shy*) loaded positively with words describing social situations and interactions such as *conversation, comfort, group, hang, meet, party, and trust*. [Please see Appendix C for a correlation table of the inductive adjective factors and the content word factors.]

Some of the factors in the content words analysis were related to specific domains. For example, Factor 1 (Appearance), 2 (Education), and 5 (Relationships). These factors largely consisted of words (i.e. colors and nouns) that were not adjective person descriptors per se, but formed coherent factors that described different domains. For example, Factor 2, labeled Education, comprised words describing school and career. Indeed, *school* had the strongest loading onto this factor. Other words included *UT* (short for University of Texas at Austin), *class, job, major, study, and work*. Interestingly, when people talked about this topic, they tended to use an assortment of words indicating level in school or time markers such as *elementary, freshman, junior, middle, old, semester, start, and summer*.

Other factors were concerned with either more mundane or more abstract themes such as 6 (Daily Activities), 7 (Ambition), and 8 (Existentialism). These factors were largely composed of nouns and verbs that spanned present and future activities and goals, which is consistent with the notion that temporally distant selves are often integrated into the self-concept (e.g.

Ross & Buehler, 1994). Factor 6 consisted of a group of words describing common activities (e.g., *eat, read, sleep, and study*), their associated objects (e.g., *class, movie, music, and test*), and proximal time markers (e.g., *day, hour, morning, and week*). Apparently, some people view their daily activities as central to the description of who they are. On the other hand, Factor 8 consisted of words indicating that participants were reflecting on defining the self on a more abstract level, and was labeled Existentialism. Words included *aware, define, experience, human, identity, question, and society*.

Factor 9 was slightly odd because it included an equal mix of both positively and negatively loaded words. The group of words that loaded positively onto this factor indicated that participants were following the task instructions precisely (e.g., *face, look, mirror, and see*). Those words that loaded negatively onto this factor described life outside of the academic domain, (e.g., *athlete, Christian, church, drink, God, and music*). This factor was labeled Reflection/Interests.

Overall, 18.1% of the total words in the self-description essays included words from the list based on the 9 content word factors. As can be seen in Table 7, there was a higher proportion of respondents in each factor who mentioned at least one word from that factor. The words in the Evaluation dimension comprised an average of 5.35% of the words in the self-descriptions. Not surprisingly then, almost all participants (99.9%) used a word from the Evaluation dimension. Although words in the Existentialism were used much less, comprising a mean rate of 0.36% of words in self-descriptions, 74.4% of respondents mentioned at least one word in the Existentialism dimension. Across all dimensions, all participants used words from at least 2 of these dimensions and 75.2% used words in all dimensions.

The reliabilities of word use across self-descriptions and stream of consciousness essays were slightly higher than those for the adjective based analysis, averaging $r(1165) = .13, p < .01$, and statistically significant except for Factor VI, $r(1165) = .06, p = .06$ based on simple Pearson correlations, two-tailed tests). Overall, the reliabilities were only slightly greater than those for the reliabilities of the adjective based factors. However, considering the unlimited ways in which we can convey our thoughts in language, it is striking that the same groups of words were used in an unrelated, more open-ended task months apart from the self-description task.

Table 8 lists the significant ($p \leq .001$) correlations between regression-based factor scores and BFI scores. Extraversion is positively associated with the Relationships dimension, such that those who rated themselves high in extraversion were more likely to mention words regarding social roles and interactions. Those high in conscientiousness were less likely to use words in the Evaluation dimension, while those high in neuroticism were more likely to score high in Evaluation. Openness was positively associated with Daily Activities and Existentialism.

As depicted in Table 9, there were several statistically significant correlations of the content word regression-based factor scores with sex, depression, SAT scores, and classroom test performance in the course. In line with previous theorizing that women have a more relational orientation than do men (Bakan, 1966), women scored higher in relationship topics, such as Sociability and Relationships. Men scored higher in Existentialism. The more depressed respondents were, the more likely they were to score higher in Evaluation, Daily Activities, and Reflection/Interests, and to score lower in Education. Higher grades were associated with Sociability and Existentialism. SAT scores were associated with Daily Activities and Existentialism.

General Discussion

In the present study, several automated language tools were used to examine everyday natural language to assess self-concepts. Like previous studies using the lexical approach, our analyses

were based on adjectives that people used to describe themselves. Consistent with an idiographic approach, the adjectives were defined by those words that people ordinarily use rather than those defined by experimenters in a pre-designed questionnaire. As an extension of this method, we examined the patterns of use of all content words (i.e., adjectives, adverbs, nouns, and verbs) in self-descriptions. Using this computer computer-based lexical strategy, we have found a way to inductively determine the lexical structure of natural text among people describing their self-concepts. Factor analyses on word use yielded coherent word clusters that suggest latent dimensions of self-concepts within the context of the writing assignment of the participants.¹

When the most commonly used adjectives were examined, seven categories appeared. These were labeled Sociability, Evaluation, Negativity, Self-Acceptance, Fitting In, Psychological Stability and Maturity. People who mentioned *ugly* were also likely to mention *attractive*; people who mentioned *shy* were also likely to mention *outgoing*; people who mentioned *weird* were also likely to mention *normal*. The use of semantically opposed words on a single factor resembled the dimension-like thinking in Kelly's personal constructs (Kelly, 1955).

In many ways, evidence of this dimensional thinking is quite significant. Consider, for example, individuals who are high on the Evaluation dimension. People who are constantly evaluating themselves and others along the dimension of thin and fat, or attractive and unattractive are approaching the world very differently from those who simply don't think about these dimensions. As the self-reports hint, these people may actually not consider themselves extremely attractive or unattractive, underweight or overweight. However, their overt behaviors may well reflect this dimensional thinking – they are likely to talk to others about these topics, buy products, watch movies, etc all oriented to this topic. In line with Kelly, then, dimensional thinking may have significant behavioral correlates. Unfortunately, the current study did not sample a broad range of behaviors that might be relevant.

When all content words were considered, nine categories appeared, some of which were represented in the adjective-based factor structure. Using all content words is recommended for future studies using this meaning extraction method since it allowed the examination of other parts of speech and domains in which the self can be expressed, and greater context with which to interpret a factor.

When asked to type their thoughts as they occurred in a stream of consciousness essay 2 months earlier, participants wrote about similar actions, objects, and attributes that they included in an open-ended self-description. The reliabilities of category use across the self-descriptions and the stream of consciousness essays were fairly low, as expected for word use in general, but most were highly significant. The factors captured salient, chronically activated concepts that were somewhat stable over time and situation. By many definitions of personality, it seems that the meaning extraction method produced factors that captured personality. More specifically, it seems that the meaning extraction method captured personality at the level of personal concerns, which McAdams (1995) describes as the level of personality at which we really begin to know a person's motivations, interests, and values.

¹Just as factor analysis is able to group questionnaire *items* that go together to summarize a concept, our method captures *words* that go together to summarize a concept. In our method, however, our respondents are the ones who define the "items" (i.e. words) that make up a factor. The fact that there are commonalities among people in the way that they use words suggests that we are capturing the common ways that people think about particular topics.

The tables in Appendix D, Appendix E, Appendix F, and Appendix G are the factor structures that result when analyzing the first half and the second half of the self-descriptive essays separately. The factor loadings are highly similar in the first half and in the second half of the essays; the general concepts (or factor labels) are the same. Taken together, these tables suggest that people are more likely to mention a word in the same semantic category as the one they are currently in, leading to correlations among the words in that domain. The fact that a group of people talk about self-relevant topics in such "runs", and that we are able to inductively characterize those common "runs" gets at the very heart of content analysis (i.e. we are characterizing a concept's properties across a corpus).

Comparisons with the Traditional Lexical Hypothesis and the Big Five

The factor structures derived from open-ended self-descriptions did not parallel that of the Big Five for two main reasons: adjective sampling and response format.

The lexical hypothesis assumes that the most important and widespread personality attributes will become encoded as a single word, which supports the study of word usage in personality research. The lexical hypothesis also assumes that the importance of a particular attribute is indicated by its representation in language (Saucier & Goldberg, 1996). The Big Five lexical approach had taken adjective codings and categorizations as the basis for determining a word's suitability for inclusion in rating scales. By counting adjective use in free descriptions of personality, we were able to determine the most representative adjectives for our sample. Our inductively selected adjective factor structure, however, was almost identical to that of the high-familiarity adjectives used in the derivation of the Big Five – but only when considering those high-familiarity adjectives that were actually used at a high frequency in our sample of free descriptions. Thus, our adjective sampling procedure, intended to follow more precisely from the assumptions of the lexical hypothesis, resulted in a slightly different factor structure than that of the Big Five.

By asking people open-ended questions about their selves, respondents were able to generate and define the categories or dimensions that were most salient, or chronically activated in them in a format (i.e., free descriptions) that was more familiar than typical Likert rating scales. The information gleaned from this 'spontaneous' approach provides information different from that of a traditional 'reactive' approach (as in the traditional rating scales assessing the Big Five dimensions), where respondents must provide ratings of stimuli supplied by the experimenter (cf. Chaplin & John, 1989). The 'spontaneous' approach, combined with the text analytic strategy presented here, produced patterns that occur in everyday natural language use (personality as "doing"). On the other hand, the factors derived from the Big Five 'reactive' approach reflect patterns in adjective ratings on self-conscious self-report scales (personality as "having").

Limitations

There are several limitations to the data used in the current study. First, the sample presents limitations on generalizing to a wider population; the sample was made up of college students from Introductory Psychology classes at a single university in the United States. How one perceives the world and one's self are based on unique experiences (Kelly, 1955). Presumably, then, the dimensions along which people carve up the world would differ based on several aspects of demographics, geographic locations, and situations among other variables. There are regularities in the way that people construct their worlds that can be captured with the technique used in this study. However, a project with a representative sampling of a wider population is warranted. Such a sample may result in a factor structure that is more similar to the Big Five because more adjectives used in the derivation of the Big Five may occur more frequently in free descriptions of the wider population, or it may uncover different domains that are considered when people describe themselves.

Another limitation of the data was that the instructions required participants to first look in the mirror as a method of self-reflection, and then to write about themselves. Depending on whether participants interpreted the instructions literally or figuratively, varying degrees of physical descriptions were included in the self-descriptions. In order for all participants to assess personality free from a physical appearance perspective, the present study should be repeated using more concise instructions that ask for an open-ended description of the self as the participant interprets it. In other words, it should be left to the participant to choose whether or not to include physical descriptors in their description of their personality.

Future research directions

Compared to taking Likert-rating scales of personality, descriptions of personality are more common in everyday life, across generations, social classes and cultures. Not only are they universally more frequent than typical personality questionnaires, but for most people, personality descriptions carry more meaning in our day to day lives than do the numbers we score on personality scales. Personality descriptions influence how we react to people, and how people make attributions about our past behaviors and make projections about our futures (cf. Donahue, 1990). For these reasons, it is important to study the structure of open-ended self-descriptions. With automated analyses, such as the meaning extraction method introduced here, it is now possible to determine the dimensions along which people think about themselves.

Could we get the same information by simply asking people to list the traits that they think are important to understanding their personality? The personality of others? Personality in general? Perceiving in general? An easy way to establish the range of perspective we are looking at in the factors derived using this technique is to ask people to describe their personality, describe the personalities of several targets, and to describe personality in general. Any commonalities or discrepancies in the various descriptions would help to determine if this method is best described as capturing a perspective of one's own unique personality vs. a perspective of personalities in general (e.g. Kenny, Mohr, & Levesque, 2001).

One area in which the application of the meaning extraction method would be useful is in cross-cultural psychology. Culture, and especially language, plays a large role in defining how one perceives the world (e.g., Slobin, 1996). People within a culture will more likely have similar views as compared to people across cultures. Large differences in the factors or personal constructs that people use to describe themselves in different languages may result. Emic tests examining salient dimensions in free descriptions can be conducted and then compared across cultures. To the degree that the rates of use of each person-descriptor differ across languages, or are only encoded in a single-language, differences in factors are expected to occur (Peabody & Goldberg, 1989). Importantly, this method can be conducted on non-English language samples, with interpretation or translation taking place only at the end of all analyses, instead of in the importation or construction of questionnaires.

This technique can be applied to other types of natural language data. For example, it can give us a broad picture at how various historical figures and important people living today have thought about themselves by examining autobiographies, blogs diaries, and lifestories without the time and labor required in traditional content coding. Beyond the expression of personality, the meaning extraction method can inform us of the patterns in any given text. For example, due to known interview biases and influences, there has been an increased need for open-ended or semi-structured assessment methods with an empirical basis (Blackman & Funder, 2002; Westen & Weinberger, 2004; 2005). For example, clinical intake interviews often begin with open-ended descriptions of symptoms, relationships, and events (Westen & Muderrisoglu, 2003). Similarly, it is recommended that witnesses and suspects in criminal cases be allowed to freely describe and elaborate on suspects, victims, and/or events before any questioning is allowed (Vrij, 2004; Vrij, Mann, & Fisher, 2006). By examining the themes used by known clinical or criminal groups using the meaning extraction method, future statements can be assessed for such themes, without the use of leading or blind questions. Using the meaning extraction method, we can compare the viewpoints of multiple observers of other meaningfully validated stimuli along common dimensions. This technique can be applied to verbal descriptions of any stimuli, then be validated and standardized for diagnostic or predictive validity.

Conclusions

For psychologists, the strength of analyzing open-ended text is that it allows us to capture ‘folk concepts’, the way people talk about themselves and others in everyday language use (Gill, 2003; Gough, 1990). On a broader level, it enables the examination of cultural patterns of thought and communication encoded in language (Slobin, 1996). With the amount of verbal material readily available through various media, especially with the growing use of the Internet (Fraley, 2004; Pew Internet & American Life Project, 2003), it is increasingly important for social scientists to consider new methods for analyzing verbal behavior as it naturally occurs. For personality and social psychologists, the Internet serves as an ecologically valid medium for examining personality (Gill, Oberlander, & Austin, 2006; Oberlander & Gill, 2006; Vazire & Gosling, 2004).

Free descriptions of personality are common in everyday life. They serve as an excellent source to examine the idiographic nature of personality. With recent advancements in text analysis, we are now presented with the ability to efficiently calculate commonalities among these idiographic samples. We can then determine how individuals differ on these commonalities in a nomothetic fashion (Tellegen, 1993). In this study of open-ended self-descriptions, it was found that self-concepts are structured in a way that does not parallel the Big Five model. Lay descriptions of self-concepts were made up of dimensions that reflect salient, chronically-activated dimensions that are reliable over time. The text analytic procedure developed from this new lexical approach of natural language use can be extended beyond the study of personality, across a multitude of languages, for a variety of research questions.

Author Note

We are grateful to Gerard Saucier, Lewis R Goldberg, and Maureen Barckley for making their high-familiarity adjective list available. Thanks to Samuel D. Gosling, Nairan Ramirez-Esparza, Richard B. Slatcher, and Ewa Kaciewicz for their comments on an earlier draft of this paper.

Appendix

Appendix A
Saucier Adjectives Factor Structure: A Varimax-Rotated Principal-Components Analysis

	I Fitting In	II Sociability	III Maturity	IV Negativity	V Evaluation	VI Honor	VII Psychological Stability
Pretty	.46	.57	.39	.47	.47	.53	.27
Kind	.41	.46	.37	.39	.44	.46	.27
Well	.36	.41	.36	.38	.40	.42	.27
Cool	.35	.38	.34	.37	.35	.26	.26
Good	.33	.33	.31	.35	.33	.24	.26
Little	.33	.31	.31	.32	.31	.24	.25
Weird	.33	.29	.29	.29	.28	.22	.24
Mean	.32	.27	.28	.26	.27	.20	.24
Funny	.31	.25	.26	.26	.25	.24	.24
Crazy	.30	.20	.26	.22	.23	.24	.24
Hard	.28	.20	.26	.21	.23	.22	.22
Normal	.27	.20	.25	.21	.22	.22	.22
Great	.25	-.23	.24	.22	.22	.22	-.29
Nice	.25		.22	.22	.22	.22	-.25
Big	.23		-.23	.21	.21	.21	-.20
Interesting	.23						
Sweet	.20						
Accomplished	.20						
Interested	.20						

Note. The above words were the most frequently used adjectives from Saucier's adjective lists that loaded greater than or equal to .20 onto a principal components analysis with varimax rotation.

Correlations of Regression Based Factor Scores in Text-Based Adjective Factors with Saucier Adjective Factor Structure

Appendix B

Text-based adjectives Factors

Saucier Adjective Factors	Text-based adjectives Factors						
	I-I Sociability	I-II Evaluation	I-III Negativity	I-VI Self-Acceptance	I-V Fitting In	I-VI Psychological Stability	I-VII Maturity
G-I Fitting In	.09 ⁺	.24 [*]	.14 [*]	.10 [*]	.64 [*]	-.14 [*]	.03
G-II Sociability	.80 [*]	.05	.16 [*]	-.10 [*]	-.10 [*]	-.06	.14 [*]
G-III Maturity	-.05	-.01	.06	.23 [*]	-.06 [*]	.29 [*]	.74 [*]
G-IV Negativity	.01	.17 [*]	.53 [*]	.47 [*]	-.13 [*]	.16 [*]	-.18 [*]
G-V Evaluation	-.10	.63 [*]	.01	-.05 [*]	.07	.10	.02
G-VI Honor	-.02	-.31 [*]	-.03	.25 [*]	.31 [*]	.12	.03
G-VII Psychological Stability	.25	-.21	.09	-.28 [*]	.22	.49 [*]	-.14 [*]

Note. Sample size = 1165.

* p ≤ .001

+ p ≤ .01

Correlations of Regression Based Factor Scores in Text-Based Adjective Factors with Content Word Factor Structure

Appendix C

Text-based adjectives Factors

Content Word Factors	I Sociability	II Evaluation	III Negativity	IV Self-Acceptance	V Fitting In	VI Psychological Stability	VII Maturity
I. Appearance	-.07	.32*	-.09*	-.11*	.21*	.10*	.03
II. Education	.09 ⁺	.01	-.10*	-.05 [‡]	.24*	-.06	.13
III. Evaluation	.02	.49*	.33*	.24*	.28*	.06	-.15
IV. Sociability	.66*	-.10*	.19*	.00	.18*	.15*	.07
V. Relationships	-.06 [‡]	-.01	.11*	.38*	.02	.05	.27*
VI. Daily Activities	-.10	-.04	.34*	.03	.05	-.06	-.17*
VII. Ambition	-.02	-.05	.05	.10*	-.08 ⁺	.15*	.27*
VIII. Existentialism	-.12*	-.08 ⁺	-.17*	.05	.06 [‡]	.37*	-.01
IX. Reflection/Interests	.04	.01	.13*	.04	-.16*	.12*	.03

Note. Sample size = 1165.

* $p \leq .001$

+ $p \leq .01$

Text-based Adjectives Person-Descriptive Factors: A Varimax-Rotated Principal-Components Analysis on the First Half of Self-Descriptive Essays

Appendix D

Person Descriptive Factors in the First Half of Self-Descriptive Essays

	Factor 1 Sociability	Factor 2 Evaluation	Factor 3 Self-Acceptance	Factor 4 Negativity	Factor 5 Fitting In	Factor 6 Psychological Stability	Factor 7 Maturity
Quiet	.56						
Shy	.51	Fat	.44	Tired	Upset	.51	Responsible
Reserved	.46	Ugly	.36	Angry	Mad	.35	Mature
Outgoing	.45	Attractive	.36	Sad	Scared	.30	Young
Friendly	.33	Beautiful	.36	Confused	Worried	.30	Accomplished
Crazy	.30	Horrible	.35	Lonely	Healthy	.29	Older
Comfortable	.29	Average	.27	Hurt	Sad	.28	Capable
Open	.28	Special	.27	Bad	Important	.28	Concerned
Fun	.24	Smart	.26	Depressed	Lucky	.26	Unique
Insecure	.23	Smart	.25	Stressed	Social	.24	
Nice	.20	Fun	.25	Happy	Understanding	.23	
			.24	Good	Wonderful	.21	
			.22	Horrible	Happy	.21	
			.20	Interested	Nice	-.23	
				Cute	-.22		
				Loving	-.20		

Note. The above adjectives were the most frequently used person descriptor adjectives in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the first half of self-descriptive essays.

Appendix F
Text-based Person-Descriptive Factors: A Varimax-Rotated Principal-Components Analysis on the First Half of Self-Descriptive Essays

	Person Descriptive Factors								
	Factor 1 Appearance	Factor 2 Sociability	Factor 3 Relationships	Factor 4 Education	Factor 5 Ambition	Factor 6 Daily Activities	Factor 7 Religion	Factor 8 Education II	Factor 9 Reflection Interests
Hair	.65	.42	.40	.45	.42	.43	.37	.26	.30
Eye	.62	.40	.39	.41	.33	.36	.35	.22	.29
Brown	.52	.37	.38	.39	.29	.34	.28	.21	.25
Nose	.48	.36	.37	.35	.27	.32	.28	.21	.24
Face	.43	.35	.37	.32	.26	.32	.27	.29	.21
Dark	.41	.35	.36	.31	.25	.30	.24	-.29	.20
Skin	.38	.28	.35	.29	.25	.30	.24	-.26	-.36
Eyebrows	.38	.27	.33	.26	.25	.27	.23	-.25	-.35
Teeth	.36	.27	.33	.25	.24	.27	.23	-.25	-.31
Blue	.36	.26	.31	.24	.23	.26	.23	-.24	-.23
Color	.35	.26	.30	.24	.23	.26	.23	-.24	-.22
White	.35	.26	.29	.23	.22	.25	.21	-.22	-.22
Green	.34	.25	.29	.23	.22	.25	.21	-.22	-.22
Wear	.33	.25	.27	.22	.22	.23	.21	-.22	-.22
Thin	.32	.24	.26	.22	.21	.23	-.22	-.22	-.21
Smile	.32	.23	.26	.21	.21	.22	-.22	-.22	-.21
Black	.32	.23	.26	.21	.21	.22	-.22	-.22	-.21
Features	.29	.22	.25	.22	.20	.21	-.22	-.22	-.21
Blonde	.27	.22	.25	.22	.20	.21	-.22	-.22	-.21
Look	.27	.21	.22	-.22	.20	.20	-.22	-.22	-.21
Tall	.26	.21	.21	-.22	.20	.20	-.22	-.22	-.21
Beauty	.26	.21	.21	-.22	.20	.20	-.22	-.22	-.21
Weight	.25	.21	.21	-.22	.20	.20	-.22	-.22	-.21
Short	.25	.21	.21	-.22	.20	.20	-.22	-.22	-.21
Mouth	.24	.21	.21	-.22	.20	.20	-.22	-.22	-.21
Fat	.24	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Physical Appearance	.23	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Body	.23	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Makeup	.22	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Small	.22	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Skinny	.21	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Light	.21	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Attractive	.20	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Red	.20	.20	.22	-.22	.20	.20	-.22	-.22	-.21
Life	-.21	-.22	.22	-.22	.20	.20	-.22	-.22	-.21

Note. The above words were the most frequently used content words in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the first half of self-descriptive essays.

Appendix G
Text-based Person-Descriptive Factors: A Varimax-Rotated Varimax Principal-Components Analysis on the Second Half of Self-Descriptive Essays

Person Descriptive Factors									
Factor 1 Sociability	Factor 2 Appearance	Factor 3 Relationships	Factor 4 Education	Factor 5 Daily	Factor 6 Religion	Factor 7 Existentialism	Factor 8 Evaluation	Factor 9 Reflection/Interests	
Talk	.38	.42	.51	.34	.33	.27	.42	.36	Music
Friend	.37	.42	.45	.31	.32	.26	.32	.35	Sports
Think	.34	.37	.41	.29	.29	.24	.30	.27	Car
Like	.34	.37	.37	.27	.28	.24	.28	.27	Town
Perceive	.31	.32	.35	.27	.27	.23	.26	.26	City
Feel	.31	.32	.35	.24	.25	.22	.25	.23	House
Know	.30	.32	.34	.24	.25	.22	.25	.22	Athletic
Mean	.28	.30	.31	.24	.22	.22	.24	.22	Play
Shy	.27	.29	.30	.23	.22	.22	.23	.22	Favorite
Roommate	.26	.28	.29	.22	.21	.20	.22	.22	Green
Good	.25	.27	.28	.22	.21	.20	.22	.21	Watch
Kind	.25	.27	.27	.22	.20	.20	.20	.21	Summer
Close	.24	.26	.22	.22	.20	.20	.20	.20	Change
Boyfriend	.24	.26	.22	.22	.21	.21	.20	.25	Mirror
Mood	.24	.26	.22	.22	.21	.21	.25	.24	See
Meet	.24	.25	.21	.21	.20	.20	.20	.22	Look
Fun	.24	.24	.20	.20	.20	.20	.20	.22	
Care	.24	.24	.20	.20	.20	.20	.20	.20	
Listen	.23	.23	.23	.23	.23	.23	.23	.23	
Time	.23	.23	.23	.23	.23	.23	.23	.23	
Guess	.23	.22	.21	.21	.21	.21	.21	.21	
Annoying	.22	.22	.21	.21	.21	.21	.21	.21	
Sad	.22	.22	.21	.21	.21	.21	.21	.21	
Girl	.22	.22	.21	.21	.21	.21	.21	.21	
Reason	.21	.21	.21	.21	.21	.21	.21	.21	
Tell	.21	.21	.21	.21	.21	.21	.21	.21	
Easy	.20	.20	.20	.20	.20	.20	.20	.20	
Tend	.20	.20	.20	.20	.20	.20	.20	.20	
Quiet	.20	.20	.20	.20	.20	.20	.20	.20	
Bad	.20	.20	.20	.20	.20	.20	.20	.20	
									Weight
									Lose
									Fat
									Skinny
									Glasses
									Gain
									Body
									Pounds
									Drugs
									Stopped
									Depressed
									Thin
									Change
									Goal

Note. The above words were the most frequently used content words in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the second half of self-descriptive essays.

References

- Allport, GW. *Personality: A psychological interpretation*. New York, NY: Holt; 1937.
- Allport, GW. *Pattern and growth in personality*. New York: Holt, Rinehart & Winston; 1961.
- Allport GW, Odbert HS. Trait names: A psycho-lexical study. *Psychological Monographs* 1936;47(1, Whole No. 211)
- Arrindell WA, van der Ende J. An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement* 1985;9(165):178–178.
- Bakan, D. *The duality of human existence: Isolation and communion in western man*. Beacon Press: Boston; 1996.
- Beck AT, Rial WY, Ricketts K. Short form of Depression Inventory: Cross-validation. *Psychological Reports* 1974;34:1184–1186. [PubMed: 4424377]
- Bem DJ, Allen A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review* 1974;81:506–520.
- Blackman MC, Funder DC. Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment*, 10, Special issue: Counterproductive behaviors at work 2002:109–116.
- Block J. A contrarian view of the Five-Factor approach to personality description. *Psychological Review* 1995;117:187–215.
- Bromley, DB. *Personality description in ordinary language*. New York, NY: John Wiley & Sons; 1977.
- Buchanan L, Westbury C, Burgess C. Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomics Bulletin & Review* 2001;8:531–544.
- Cantor N. From thought to behavior: 'Having' and 'doing' in the study of personality and cognition. *American Psychologist* 1990;45:735–750.
- Cantor, N.; Markus, H.; Niedenthal, P.; Nurius, P. On motivation and the self-concept. In: RM, Sorrentino; ET, Higgins, editors. *Handbook of motivation and cognition: Foundations of social behavior*. New York, NY: Guilford Press; 1986. p. 96-121.
- Cattell RB. The scree test for the number of factors. *Multivariate Behavioral Research* 1996;1:245–276.
- Chaplin WF, John OP. Structured free-descriptions of the self-concept. 1989 Unpublished manuscript
- Claeys B, de Boeck P, van den Bosch W, Biesmans R, Bohrer A. A comparison of one free-format and two fixed-format self-report personality assessment methods. *Journal of Personality and Social Psychology* 1985;49:1028–1039.
- Costa, PT., Jr; McCrae, RR. *Revised NEO Personality Inventory (NEO-PIR) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Jodessa, FL: Psychological Assessment Resources; 1992.
- Csikszentmihalyi, M.; Rochberg-Halton, E. *The meaning of things: Domestic symbols and the self*. New York, NY: Cambridge University Press; 1981.
- de Raad B, Hofstee WK. A circumplex approach to the Five Factor model: A facet structure of trait adjectives supplemented by trait verbs. *Personality and Individual Differences* 1993;15:493–505.
- de Raad B, Mulder E, Kloosterman K, Hofstee WKB. Personality-descriptive verbs. *European Journal of Personality* 1988;2:81–96.
- de Raad B, Hoskens M. Personality-descriptive nouns. *European Journal of Personality* 1990;4:131–146.
- Donahue, EM. Unpublished doctoral dissertation. Berkeley, CA: University of California, Berkeley; 1992. Bridging the gap between formal and informal assessments of personality.
- Duval, S.; Wicklund, RA. *A theory of objective self awareness*. Oxford, England: Academic Press; 1972.
- Epstein S. The self-concept revisited: Or a theory of a theory. *American Psychologist* 1973;28:404–416. [PubMed: 4703058]
- Fiske ST, Cox MG. Person concepts: The effect of target familiarity and descriptive purpose on the process of describing others. *Journal of Personality* 1979;47:136–161.
- Foltz PW. Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments & Computers* 1996;28:197–202.
- Fraley, RC. *How to conduct behavioral research over the Internet*. New York, NY: The Guilford Press; 2004.

- Funder DC, Drobth KM. Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology* 1987;52:409–418. [PubMed: 3559898]
- Gill, A. Personality and language: The projection and perception of personality in computer-mediated communication. Edinburgh, Scotland: University of Edinburgh; 2003. Unpublished doctoral dissertation
- Gill AJ, Oberlander J, Austin E. The perception of e-mail personality at zero-acquaintance. *Personality and Individual Differences* 2006;40:497–507.
- Goldberg, LR. From Ace to Zombie: Some explorations in the language of age personality. In: Spielberger, CD.; Butcher, JN., editors. *Advances in personality assessment*. Vol. 1. Hillsdale, NJ: Erlbaum; 1982. p. 203-234.
- Goldberg LR. An alternative “Description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology* 1990;59:1216–1229. [PubMed: 2283588]
- Gosling SD, Rentfrow PJ, Swann WB Jr. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 2003;37:504–528.
- Gough, HG. The California Personality Inventory. In: Watkins, CE., Jr; Campbell, VL., editors. *Testing in counseling practice*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; 1990. p. 37-62.
- Graesser AC, Cai Z, Louwse M, Daniel F. Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly* 2006;70:1–20.
- Graesser, AC.; Gernsbacher, MA.; Goldman, SR. Introduction to the Handbook of Discourse Processes. In: Graesser, AC.; Gernsbacher, MA.; Goldman, SR., editors. *Handbook of Discourse Processes*. Mahwah NJ: Lawrence Erlbaum Associates; 2003. p. 1-23.
- Graesser AC, Lu S, Jackson GT, Mitchell H, Ventura M, Olney A, Louwse MM. AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers* 2004;36:180–193.
- Graesser AC, McNamara DS, Louwse MM, Cai Z. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers* 2004;36:193–202.
- Grice JW. Bridging the idiographic-nomothetic divide in ratings of self and others on the Big Five. *Journal of Personality* 2004;72:203–241. [PubMed: 15016064]
- Higgins ET, Klein R, Strauman T. Self-concept discrepancy theory: A psychological model for distinguishing among different aspects of depression and anxiety. *Social Cognition* 1985;3:51–76.
- Hill RW, McIntire K, Bacharach VR. Perfectionism and the big five factors. *Journal of Social Behavior & Personality* 1997;12:257–270.
- Hofstee, WKB.; Van Heck, GL. *European Journal of Personality*. 4. 1990. Personality language [Special Issue].
- Ickes, W.; Snyder, M.; Garcia, S. Personality influences on choice of situations. In: Hogan, R.; Johnson, JA.; Briggs, SR., editors. *Handbook of personality psychology*. San Diego, CA: Academic Press; 1997. p. 165-195.
- John, OP.; Srivastava, S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, LA.; John, OP., editors. *Handbook of personality: Theory and research*. 2nd ed.. New York, NY: Guilford Press; 1999. p. 102-138.
- Kelly, GA. *The psychology of personal constructs*. New York, NY: Norton; 1955.
- Kenny DA, Mohr CD, Levesque MJ. A social relations variance partitioning of dyadic behavior. *Psychological Bulletin* 2001;127:128–141. [PubMed: 11271751]
- Kenrick DT, Stringfield DO. Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review* 1980;87:88–104.
- Kihlstrom, JF.; Cantor, N.; Albright, JS.; Chew, BR.; Klein, SB.; Niedenthal, PM. Information processing and the study of the self. In: Berkowitz, L., editor. *Advances in experimental social psychology*. Vol. 21 *Social psychological studies of the self: Perspectives and programs*. San Diego, CA: Academic Press; 1988. p. 145-178. Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Lamiell JT. Toward an idiotic psychology of personality. *American Psychologist* 1981;36:276–289.

- Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. *Discourse Processes* 1998;25:259–284.
- Little B. Personality and science projects: Six impossible things before breakfast. *Journal of Research in Personality* 2005;39:4–21.
- MacCallum RC, Widman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychological Methods* 1999;4:84–99.
- Markus H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology* 1977;35:63–78.
- Markus H, Nurius P. Possible selves. *American Psychologist* 1986;41:954–969.
- McAdams DP. What do we know when we know a person? *Journal of Personality* 1995;63:365–396.
- McGuire WJ, Padawer-Singer A. Trait salience in spontaneous self-concept. *Journal of Personality and Social Psychology* 1976;33:743–754. [PubMed: 1271234]
- Oberlander J, Gill AJ. Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes* 2006;42:239–270.
- Paunonen SV, Jackson DN. What is beyond the Big Five? Plenty! *Journal of Personality* 2000;68:821–835. [PubMed: 11001150]
- Peabody D. Selecting representative trait adjectives. *Journal of Personality and Social Psychology* 1987;52:59–71.
- Peabody D, Goldberg LR. Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology* 1989;57:552–567. [PubMed: 2778639]
- Peevers BH, Secord PF. Developmental changes in attribution of descriptive concepts to persons. *Journal of Personality and Social Psychology* 1973;57:552–567.
- Pennebaker, JW.; Francis, ME.; Booth, RJ. *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Erlbaum Publishers; 2001.
- Pennebaker JW, King LA. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 1999;77:1296–1312. [PubMed: 10626371]
- Pennebaker JW, Mehl MR, Niederhoffer K. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology* 2003;54:547–577.
- Pew Internet & American Life Project. Internet activities. 2003 oct 22.
http://www.pewinternet.org/reports/chart.asp?img=Internet_Activities_10.22.03.htm
- Prentice DA. Familiarity and differences in self- and other-representations. *Journal of Personality and Social Psychology* 1990;59:369–383. [PubMed: 2231276]
- Robinson MD. Personality as performance: Categorization tendencies and their correlates. *Current Directions in Psychological Science* 2004;13:127–129.
- Robinson MD, Oishi S. Trait self-report as a “fill-in” belief system: Categorization speed moderates the extraversion/life satisfaction relation. *Self and Identity* 2006;5:15–34.
- Ross, M.; Buehler, R. Identity through time: Constructing personal pasts and futures. In: Brewer, MB.; Hewstone, M., editors. *Self and social identity*. Malden, MA, US: Blackwell Publishing; 2004. p. 25-51.
- Saucier G. Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology* 1997;73:1296–1312. [PubMed: 9418281]
- Saucier G. Factor structure of English-language personality type-nouns. *Journal of Personality and Social Psychology* 2003;85:695–708. [PubMed: 14561123]
- Saucier, G.; Goldberg, LR. The language of personality: lexical perspectives on the five-factor model. In: Wiggins, JS., editor. *The Five-Factor Model of Personality: Theoretical perspectives*. New York, NY: Guilford; 1996. p. 21-50.
- Saucier G, Goldberg LR. Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality* 1997;10:61–77.
- Saucier G, Goldberg LR. Lexical studies of indigenous personality factors: premises, products, and prospects. *Journal of Personality* 2001;69:847–879. [PubMed: 11767821]
- Scott, M. *WordSmith*. New York, NY: Oxford University Press; 1996.

- Schiller, R.; Tellegen, A.; Evens, J. An idiographic and nomothetic study of personality description. In: Butcher, JN.; Spielberger, CD., editors. *Advances in personality assessment*. Vol. 10. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1995. p. 1-23.
- Semin GR, Marsman JG. 'Multiple-inference-inviting properties' of interpersonal verbs: Event instigation, dispositional inference and implicit causality. *Journal of Personality and Social Psychology* 1994;67:836–849.
- Slobin, D. From “thought” and “language” to “thinking” for “speaking”. In: Gumperz, JJ.; Levinson, SJ., editors. *Rethinking linguistic relativity*. New York, NY: Cambridge University Press; 1996. p. 70-96.
- Stevens, JP. *Applied multivariate statistics for the social sciences*. 2nd ed.. Hillsdale, NJ: Erlbaum; 1992.
- Tellegen A. Folk concepts and psychological concepts of personality and personality disorder. *Psychological Inquiry* 1993;4:122–130.
- Väyrynen, JJ.; Honkela, T. Comparison of independent component analysis and singular value decomposition in word context analysis. In: Honkela, T.; Könönen, V.; Pöllä, M.; Simula, O., editors. *Proceedings of AKRR'05, International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*. Finland: Espoo; 2005. p. 135-140.
- Vazire S, Gosling SD. e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology* 2004;87:123–132. [PubMed: 15250797]
- Vrij A. Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology* 2004;9:159–181.
- Vrij A, Mann S, Fisher RP. Information-gathering vs accusatory interview style: Individual differences in respondents' experiences. *Personality and Individual Differences* 2006;41:589–599.
- Westen D. A model and a method for uncovering the nomothetic from the idiographic: An alternative to the Five-Factor Model? *Journal of Research in Personality, Special Issue: The Future of Personality* 1996;30:400–413.
- Westen D, Muderrisoglu S. Reliability and validity of personality disorder assessment using a systematic clinical interview: Evaluating an alternative to structured interviews. *Journal of Personality Disorders* 2003;17:350–368.
- Westen D, Weinberger J. When clinical description becomes statistical prediction. *American Psychologist* 2004;59:595–613. [PubMed: 15491255]
- Westen D, Weinberger J. In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology*, 61, Special issue: A Great Pioneer of. *Clinical Science Remembered: Special Issues in Honor of Paul E. Meehl* 2005:1257–1276.

Table 1

Passages from Self-Descriptions

Subject	Sample of Self-Description
02524	...I know what I should wear and what I am too fat for. I know I shouldn't wear things that show my stomach. I should not try to dress too slutty. And this is because I am realistic. I think I know my boundaries. When I look in the mirror I see an average girl. I don't think I am pretty, but I am not ugly. I have red hair, not a color that attracts a lot of guys. I have pale skin (thanks to the red hair) which is also not the attractive thing nowadays.
02590	I have friends, but still enjoy alone time.. with my computer. :) I like to just sit back, relax and take in life. I don't feel the need to go out and spend money to have fun--I like the free pleasures of life (don't get me wrong, I'm not stingy by any means) :-p I like maintaining friendships with friends in the past. I've been told that the real lifelong friends you make are made in college, but I still like to have ties to my past.
03160	...Another thing I see in the mirror is a person who's life is going down the drain. My high school grades were great, I was almost in the top 5% of my class, but now in college, I'm begging and hoping for curves. This is detrimental to my self esteem. I can tell that this Christmas is going to be a very stressful one. I'll have to tell my parents why I didn't do so well in school, then I have to work in retail the entire holiday. That really sucks...
03412	...my coworkers and I have a good time. I'm glad I have opened up more because it makes life a lot more enjoyable. Unfortunately I still somewhat care what other people think, but I am working on that too. And by becoming more outgoing despite the fact that I may turn bright red out of embarrassment, my worries of my image to others are gradually vanishing...
03446	I am the kind of person who lets emotion get in the way of daily life. I see myself as someone who has been through too much for me age. I think I see myself as confused inside. I think that I am the kind of person that acts happy and wants people to not know that I am hurting inside so I just continue on like there is nothing wrong. I am an emotional person. I am the kind of person who cries at sappy movies a and who will cry if she sees others cry.
03392	...Sometimes I feel like I don't have anything to write, or say. Like I am not interesting. I have a mole on my nose, that in this Chinese face reading book it says I will die of a venereal disease. That scares me. Here in Austin I get tons of pimples, I wonder if the ones on my chin are hormone provoked. That's what they say. I have a small forehead and to me that means I am not witty or intelligent. My teeth are gapping. Could it be the smoking?...
04312	...People expect me to maintain this "perfect" image they have of me. It makes it hard to be myself a lot of the times. If I get upset, or cry, or yell, or forget something, others get extremely upset with me when if anyone else did this it wouldn't matter to them. I hate this. I feel I have to be a Barbie doll a lot of the times. Because I have always made all A's, been popular, friendly, and willing to do things for others, if I ever "slip" people look at me like something is terribly wrong. I feel constant pressure to do what others expect of me and it makes me tired a lot of the time...

Table 2
Text-based adjectives Person-Descriptive Factors: A Varimax-Rotated Principal-Components Analysis

Adjective	Personality factors						
	Factor 1: Sociability	Factor 2: Evaluation	Factor 3: Negativity	Factor 4: Self-Acceptance	Factor 5: Fitting In	Factor 6: Psychological Stability	Factor 7: Maturity
Quiet	.59	.05	-.05	-.02	-.05	-.05	-.05
Shy	.54	.04	.04	-.13	.04	.06	-.03
Reserved	.48	-.01	-.10	.15	-.02	-.10	.00
Outgoing	.48	-.03	-.02	-.03	.19	.08	.08
Comfortable	.40	-.09	.03	-.03	.00	.09	-.10
Open	.31	-.18	.08	-.11	.08	.24	.04
Friendly	.28	.07	.06	-.01	.11	-.05	.10
Insecure	.20	.02	.12	.04	-.01	.12	-.20
Easy	.10	.00	-.10	.00	.01	.09	.07
Cold	.09	-.04	.06	-.06	-.01	-.08	-.02
Ugly	-.10	.51	.11	-.12	.03	.13	-.18
Fat	-.06	.49	.04	.08	.01	.06	-.18
Attractive	.12	.42	-.08	.02	.00	.01	.02
Beautiful	-.16	.41	.06	.02	.05	.04	.01
Nice	.04	.29	.22	.00	.14	-.09	.16
Smart	.07	.28	.07	.14	.03	-.14	.19
Stupid	.09	.28	.15	-.10	.15	-.03	-.11
Average	-.05	.26	-.09	-.10	.06	.15	.18
Special	.03	.25	-.13	-.15	-.05	.04	.11
Lazy	-.00	.21	.14	-.06	.10	-.12	.11
Cute	.42	.21	.09	.13	.06	-.01	-.15
Stuck	.16	.17	-.08	.01	.04	.04	.06
Innocent	.14	.16	.11	.07	-.15	-.02	-.07
Intelligent	.07	.16	-.15	-.02	.03	.05	.08
Mad	-.03	-.07	.45	-.05	-.01	-.10	.10
Hurt	.01	-.05	.36	.06	-.01	.04	.17
Bad	-.01	.09	.36	.13	.09	.03	-.08
Sad	.07	.05	.31	.14	-.03	.08	-.12
Horrible	-.02	.21	.31	-.03	.05	-.03	.03
Upset	-.06	.13	.31	.05	.15	.06	.03
Scared	.08	.03	.31	.13	-.18	-.01	.03
Good	.10	.03	.28	.11	.23	.06	-.01
Worried	.03	.02	.27	-.18	.15	.08	.11
Angry	-.01	.00	.25	-.02	-.11	.01	.07
Afraid	.20	.24	.24	-.06	-.11	-.02	.17
Tired	-.06	.14	.23	.04	-.06	-.05	-.08
Depressed	-.05	.17	.22	.08	-.02	.16	-.10
Stressed	.05	-.08	.21	.12	-.01	-.01	-.14
Honest	-.02	-.11	.11	.03	.11	.10	.01
Wonderful	-.12	.17	-.05	.47	.05	.06	.18
Loving	.04	-.05	-.02	.44	.00	-.01	.26
Lost	-.03	.13	-.06	.32	.03	.14	-.15
Blessed	-.15	-.11	-.22	.31	.15	.04	-.03
Confused	.01	.06	.10	.29	.01	-.09	-.12
Amazing	-.13	.00	-.12	.28	.21	.03	-.12
Perfect	-.02	.05	.07	.27	.02	.12	-.12
Selfish	.01	.00	.11	.26	-.05	.08	.02
Happy	.11	-.01	.19	.24	.11	.15	.07
Wrong	.18	.06	.08	.22	-.05	.02	.02

Adjective	Personality factors						
	Factor 1: Sociability	Factor 2: Evaluation	Factor 3: Negativity	Factor 4: Self-Acceptance	Factor 5: Fitting In	Factor 6: Psychological Stability	Factor 7: Maturity
Independent	.15	-.10	-.07	.21	-.03	.14	.08
Lonely	.08	-.04	.20	.21	-.11	-.10	-.14
Free	.00	.00	.08	.20	.00	-.03	.00
Sensitive	-.05	.05	.15	.20	.10	-.01	.08
Tough	.01	-.05	.00	.11	.00	.00	-.02
Interesting	-.06	-.10	-.08	.02	.44	.03	-.15
Funny	.03	.05	-.03	.07	.39	-.16	.07
Crazy	.12	.04	.02	.04	.39	-.11	-.07
Cool	.08	.14	.06	-.04	.34	-.06	.00
Fun	.20	.13	.13	.24	.33	-.13	.08
Weird	.02	.13	.03	.01	.28	-.08	.08
Conscious	.09	.03	-.11	.09	.27	-.17	-.10
Lucky	-.13	-.11	.00	.15	.27	.05	-.03
Old	-.04	.09	.00	-.05	.25	-.02	.19
Concerned	.05	-.09	.05	-.06	.25	.07	-.01
Normal	.07	.12	-.03	-.16	.23	.07	-.03
Understanding	-.07	-.16	.06	-.05	.23	.18	.02
Important	.02	.01	.01	-.01	.20	.09	.07
Aware	-.01	.03	.01	-.02	-.04	.41	.00
Healthy	-.03	.14	.10	.06	-.01	.36	-.11
Emotional	.02	-.10	.15	-.11	.08	.35	.06
Confident	.18	.00	.04	.04	-.03	.35	.01
Content	.00	-.02	-.09	.11	-.02	.32	.01
Positive	-.06	.07	-.09	.15	-.11	.32	.21
Negative	-.08	-.02	.06	-.03	-.05	.32	.15
Physical	-.02	.11	-.08	-.13	.15	.24	.04
Fake	.18	.18	-.08	.06	.03	.24	-.13
Social	.09	-.03	-.06	.03	.11	.23	-.01
Satisfied	-.01	.03	.02	.12	-.05	.20	.06
Mature	.03	-.01	.02	.03	.00	.39	.38
Successful	.03	.08	-.04	.03	-.10	-.11	.37
Caring	.10	-.06	.10	.20	.08	-.14	.30
Capable	-.12	-.02	.02	.04	-.06	.12	.28
Accomplished	-.08	.04	-.02	.04	.12	-.12	.28
Young	.02	-.02	-.01	-.09	.04	.12	.28
Determined	.04	-.10	.08	.08	-.06	.11	.26
Responsible	.11	-.04	.03	-.02	.01	.07	.23
Difficult	.04	-.12	.08	.12	.03	-.02	-.16
Unique	-.08	.08	-.11	.04	.13	.05	.15

Note. The above adjectives were the most frequently used person descriptor adjectives in open-ended self-descriptions of personality. Boldface indicates factor loadings greater than or equal to .20.

Table 3
Psychometric Properties of Personality Dimensions Based on Adjectives in Open-Ended Self-Descriptions

Factors	Mean Word Usage	<i>SD</i>	Percent > 0	Cross-essay reliability
Sociability	.27	.36	66.2	.10***
Evaluation	.31	.55	68.0	.09**
Negativity	.65	.56	89.5	.11***
Self-Acceptance	.53	.61	84.1	.09***
Fitting In	.54	.61	84.0	.06*
Psychological Stability	.21	.27	62.3	.11***
Maturity	.24	.32	67.0	.09**

Note. Mean Usage reflects the mean percentage of words in a given text that load onto a personality dimension. *SD* = standard deviation. Percent > 0 = percentage of people who use words in a given personality dimension. Cross-essay reliability = Pearson correlation between personality dimension word use in self-descriptions and in stream of consciousness essays written two months earlier

 $p \leq .001$

**
 $p \leq .01$

+
 $p = .06$.

Table 4
Correlations of Text-Based Adjective Factor Scores with BFI

Factor	Big Five Inventory Scores				
	A	E	C	N	O
Sociability	.08 ⁺	-.14*	.04	.03	-.05
Evaluation	-.04	-.08 ⁺	-.07	.13*	-.04
Negativity	.00	.04	-.04	.13*	.02
Self-Acceptance	.05	.11*	.02	.05	-.03
Fitting In	.04	.08 ⁺	-.02	-.07	.05
Psychological Stability	-.02	.03	.07	.05	.03
Maturity	.05	.09 ⁺	.13*	-.06	-.03

Note. Sample size = 868. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness. Significant correlations are in bold

* $p \leq .001$

⁺ $p \leq .01$.

Table 5
Correlations of Text-Based Adjectives Factors with Demographic Variables, Depression, and Academic Achievement

Factor	Age (n=744)	Sex (n=1151)	BDI (n=787)	Final Grade (n=1149)	SAT (n=577)
Sociability	-.07	.11*	.01	.09 ⁺	.03
Evaluation	-.02	.03	.10 ⁺	.01	.02
Negativity	.04	.11*	.13*	-.03	-.03
Self-Acceptance	-.03	.17*	.01	-.02	-.03
Fitting In	.02	.03	-.03	.02	.02
Psychological Stability	-.04	.13*	.02	.15*	.16*
Maturity	-.04	.05	-.03	.03	-.08

Note. Sex is scored, 1 = male, 2 = female; BDI = Beck Depression Inventory score; Final Grade = Final test grade in the introductory psychology course in which students were enrolled when they completed the self-description task; SAT = self-reported Scholastic Aptitude Test score. Significant correlations are in bold

* $p \leq .001$

+ $p \leq .01$.

Open-Class Words Factor Structure: A Varimax-Rotated Principal-Components Analysis (continued on page 54)

Table 6

	Factor 1: Appearance	Factor 2: Education	Factor 3: Evaluation	Factor 4: Sociability	Factor 5: Relationships
Hair	.69	School	Weight	Shy	Love
Eye	.62	College	Fat	Outgoing	Mom
Brown	.52	High	Lose	Friend	Family
Nose	.51	Year	Pretty	Quiet	Daughter
Face	.46	Texas	Bad	Open	Sister
Skin	.42	University	Boyfriend	Meet	Heart
Color	.41	Graduate	Ugly	Talk	Marry
Teeth	.40	UT	Know	Reserved	Dad
Dark	.40	Freshman	Feel	Hang	Child
Smile	.38	Grade	Guy	Close	Wife
Wear	.37	Student	Good	Comfort	Wonderful
Eyebrow	.36	Austin	Want	Tend	Woman
Black	.35	Class	Guess	Think	God
Feature	.35	Old	Like	Feel	Parents
Blue	.34	Job	Smart	Judge	Raise
White	.34	House	Stupid	Group	Brother
Green	.31	Parents	Think	Emotion	Husband
Thin	.28	Major	Stop	Perceive	Care
Appearance	.28	Start	Nice	Fun	Boyfriend
Physical	.28	Junior	Gain	Tell	Girl
Blonde	.28	Middle	Girl	Trust	Life
Notice	.27	Senior	Class	Opinion	Support
Makeup	.26	Months	Assignment	Situation	Strong
Light	.25	Business	Problem	Party	Strong
Acne	.25	Study	Body	Like	Taught
Beauty	.23	Work	Worse	Care	Hurt
Cut	.22	Play	Horrible	Express	
Picture	.22	Career	Need	Easy	
Tall	.22	Life	Attractive	Different	
Skinny	.21	Semester	Kind	Conversation	
Red	.21	Friend	Lie	Avoid	
Glasses	.21	Degree	Means	Loud	
Ugly	.21	Summer	Weird		
Asian	.20	Early	Hard		
Life	-.22	Elementary	Fun		
		Adult	Tell		

	Factor 6: Daily Activities	Factor 7: Ambition	Factor 8: Existentialism	Factor 9: Reflection/Interests
Sleep	.38	Goal	Question	Mirror
Night	.35	Achieve	Answer	See
Test	.33	Set	Aware	Face
Sit	.31	Work	Society	Look
Tire	.31	Reach	Human	Smile
Minute	.29	Determine	Identity	Happy
Hour	.28	Strive	Share	Change
Write	.28	Success	View	Wondering
Read	.27	Accomplish	Born	Music
Week	.26	Help	Define	Athlete
Morning	.26	Ability	Experience	Sport
Hate	.26	Future	Describe	Christian

et al.

	Factor 6: Daily Activities	Factor 7: Ambition	Factor 8: Existentialism	Factor 9: Reflection/Interests	
Sick	.24	Strong		Church	-.20
Death	.24	Effort	American	Drink	-.20
Study	.24	Need	Read	God	-.20
Spend	.22	Lead		Play	-.20
Movie	.22	Continue			
Watch	.21	Respect			
Pain	.21	Faith			
Cry	.21	Lack			
Late	.21	Motivate			
Class	.21	Hard			
Favorite	.20				
Stay	.20				
Day	.20				
Suppose	.20				
Eat	.20				
Time	.20				
Music	.20				

Note. The above words were the most frequently used open-class words in open-ended self-descriptions of personality that loaded greater than or equal to .20 onto a principal components analysis with varimax rotation.

Table 7
Psychometric Properties of Personality Dimensions Based on Content Words in Open-Ended Self-Descriptions

Factors	Mean Word Usage	SD	Percent > 0	Test-Retest Reliability
Appearance	1.18	1.53	87.0	.07*
Education	2.74	1.66	99.0	.21***
Evaluation	5.35	1.96	99.9	.29***
Sociability	2.39	1.05	99.7	.07*
Relationships	1.40	1.13	94.4	.26***
Daily Activities	1.20	.80	96.7	.06#
Ambition	.83	.67	91.8	.07*
Existentialism	.36	.42	74.4	.13***
Reflection	1.95	1.56	98.0	.02 ^{ns}
Interests	.68	.53	90.5	.10***

Note. Mean Word Usage reflects the mean percentage of words in a given text that load onto a personality dimension. *SD* = standard deviation. Percent > 0 = percentage of people who use words in a given personality dimension. Test-Retest Reliability = Pearson correlation between personality dimension word use in self-descriptions and in stream of consciousness essays written two months earlier

 $p \leq .001$

*
 $p \leq .05$

 $p = .06$

^{ns} = not significant. [Reflection and Interests are not separate dimensions, but were separated because their component words were negatively loaded onto the same factor].

Table 8
Correlations of Content Word Personality Dimension Scores with BFI Scores

Factor	Big Five Inventory Scores					O
	A	E	C	N	O	
Appearance	.00	.05	.01	-.01	.05	
Education	.02	.03	.08	-.08	-.10 ⁺	
Evaluation	-.06	-.02	-.13*	.18*	-.02	
Sociability	.08	-.05	.10 ⁺	.01	.03	
Relationships	.10 ⁺	.15*	.07	.07	-.04	
Daily Activities	-.01	-.01	-.05	.01	.12*	
Ambition	.01	.00	.08 ⁺	.05	-.09 ⁺	
Existentialism	-.04	-.06	.01	-.08 ⁺	.14*	
Reflection/Interests	-.03	-.02	.01	.10 ⁺	-.02	

Note. Sample size =868. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness. Significant correlations are in bold

* $p \leq .001$

+ $p \leq .01$.

Table 9
Correlations of Content Word Factors with Demographic Variables, Depression, and Academic Achievement

Factor	Age (n=744)	Sex (n=1151)	BDI (n=787)	Final grade (n=1149)	SAT (n=577)
Appearance	-.04	.03	.02	.04	.05
Education	-.01	-.07	-.14*	.03	.04
Evaluation	.02	.09+	.12*	-.01	.02
Sociability	-.09+	.22*	.01	.15*	.05
Relationships	-.03	.30*	.01	-.01	-.11+
Daily Activities	.00	-.02	.12*	-.02	.16*
Ambition	-.01	-.03	-.05	.07	-.01
Existentialism	.01	-.10*	.04	.13*	.20*
Reflection/Interests	.06	.08	.12*	-.05	-.06

Note. Sex is scored, 1 = male, 2 = female; BDI = Beck Depression Inventory score; Final Grade = Final test grade in the introductory psychology course in which students were enrolled when they completed the self-description task; SAT = self-reported Scholastic Aptitude Test score. Significant correlations are in bold

* $p \leq .001$

+ $p \leq .01$.

Appendix A
Saucier Adjectives Factor Structure: A Varimax-Rotated Principal-Components Analysis

	I Fitting In	II Sociability	III Maturity	IV Negativity	V Evaluation	VI Honor	VII Psychological Stability
Pretty	.46	.57	.39	.47	.47	.53	.27
Kind	.41	.46	.37	.39	.44	.46	.27
Well	.36	.41	.36	.38	.40	.42	.27
Cool	.35	.38	.34	.37	.35	.26	.26
Good	.33	.33	.31	.35	.33	.24	.26
Little	.33	.31	.31	.32	.31	.24	.25
Weird	.33	.29	.29	.29	.28	.22	.24
Mean	.32	.27	.28	.26	.27	.20	.24
Funny	.31	.25	.26	.26	.25	.24	.24
Crazy	.30	.20	.26	.22	.23	.24	.24
Hard	.28	.20	.26	.22	.23	.22	.22
Normal	.27	.20	.25	.21	.21	.22	.22
Great	.25	-.23	.24	.22	.22	.22	-.29
Nice	.25		.22	.22	.22	.22	-.25
Big	.23		-.23	.21	.21	.21	-.20
Interesting	.23						
Sweet	.20						
Accomplished	.20						
Interested	.20						

Note. The above words were the most frequently used adjectives from Saucier's adjective lists that loaded greater than or equal to .20 onto a principal components analysis with varimax rotation.

Correlations of Regression Based Factor Scores in Text-Based Adjective Factors with Saucier Adjective Factor Structure

Appendix B

Saucier Adjective Factors	Text-based adjectives Factors						
	I-I Sociability	I-II Evaluation	I-III Negativity	I-VI Self-Acceptance	I-V Fitting In	I-VI Psychological Stability	I-VII Maturity
G-I Fitting In	.09 ⁺	.24 [*]	.14 [*]	.10 [*]	.64 [*]	-.14 [*]	.03
G-II Sociability	.80 [*]	.05	.16 [*]	-.10 [*]	-.10 [*]	-.06	.14 [*]
G-III Maturity	-.05	-.01	.06	.23 [*]	-.06	.29 [*]	.74 [*]
G-IV Negativity	.01	.17 [*]	.53 [*]	.47 [*]	-.13 [*]	.16 [*]	-.18 [*]
G-V Evaluation	-.10 [*]	.63 [*]	.01	-.05	.07	.10 [*]	.02
G-VI Honor	-.02	-.31 [*]	-.03	.25 [*]	.31 [*]	.12	.03
G-VII Psychological Stability	.25 [*]	-.21 [*]	.09 [*]	-.28 [*]	.22	.49 [*]	-.14 [*]

Note. Sample size = 1165.

* $p \leq .001$

+ $p \leq .01$

Correlations of Regression Based Factor Scores in Text-Based Adjective Factors with Content Word Factor Structure

Appendix C

Text-based adjectives Factors

Content Word Factors	I Sociability	II Evaluation	III Negativity	IV Self-Acceptance	V Fitting In	VI Psychological Stability	VII Maturity
I. Appearance	-.07	.32*	-.09*	-.11*	.21*	.10*	.03
II. Education	.09 ⁺	.01	-.10*	-.05	.24*	-.06	.13
III. Evaluation	.02	.49*	.33*	.24*	.28*	.06	-.15
IV. Sociability	.66*	-.10*	.19*	.00	.18*	.15*	.07
V. Relationships	-.06	-.01	.11*	.38*	.02	.05	.27*
VI. Daily Activities	-.10*	-.04	.34*	.03	.05	-.06	-.17*
VII. Ambition	-.02	-.05	.05	.10*	-.08 ⁺	.15*	.27*
VIII. Existentialism	-.12*	-.08 ⁺	-.17*	.05	.06	.37*	-.01
IX. Reflection/Interests	.04	.01	.13*	.04	-.16*	.12*	.03

Note. Sample size = 1165.

* $p \leq .001$

+ $p \leq .01$

Text-based Adjectives Person-Descriptive Factors: A Varimax-Rotated Principal-Components Analysis on the First Half of Self-Descriptive Essays

Appendix D

Person Descriptive Factors in the First Half of Self-Descriptive Essays

	Factor 1 Sociability	Factor 2 Evaluation	Factor 3 Self-Acceptance	Factor 4 Negativity	Factor 5 Fitting In	Factor 6 Psychological Stability	Factor 7 Maturity
Quiet	.56						
Shy	.51	Fat	.44	Tired	Upset	.48	Responsible
Reserved	.46	Ugly	.36	Angry	Mad	.38	Mature
Outgoing	.45	Attractive	.36	Sad	Scared	.33	Young
Friendly	.33	Beautiful	.36	Confused	Worried	.32	Accomplished
Crazy	.30	Horrible	.35	Lonely	Healthy	.29	Older
Comfortable	.29	Average	.27	Hurt	Sad	.29	Capable
Open	.28	Special	.27	Bad	Important	.28	Concerned
Fun	.24	Smart	.26	Depressed	Lucky	.26	Unique
Insecure	.23	Smart	.25	Stressed	Social	.24	
Nice	.20	Fun	.25	Happy	Understanding	.23	
			.24	Good	Wonderful	.23	
			.22	Horrible	Happy	.21	
			.22	Interested	Nice	.20	
			.20	Cute	Nice	-.22	
				Loving	Loving	-.20	

Note. The above adjectives were the most frequently used person descriptor adjectives in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the first half of self-descriptive essays.

Appendix F
Text-based Person-Descriptive Factors: A Varimax-Rotated Principal-Components Analysis on the First Half of Self-Descriptive Essays

Person Descriptive Factors

	Factor 1 Appearance	Factor 2 Sociability	Factor 3 Relationships	Factor 4 Education	Factor 5 Ambition	Factor 6 Daily Activities	Factor 7 Religion	Factor 8 Education II	Factor 9 Reflection Interests
Hair	.65								
Eye	.62	Think	.42	School	.42	Test	God	Reflect	.26
Brown	.52	Friend	.39	High	.33	Psychology	Love	Aware	.22
Nose	.48	Know	.38	College	.29	Week	Handle	Perfect	.21
Face	.43	Perceive	.37	Change	.27	Sleep	Christian	Face	.24
Dark	.41	Like	.37	Older	.26	Tire	Relation	Fairly	.21
Skin	.38	Feel	.36	Grow	.25	Class	Drink	Values	.20
Eyebrows	.38	Shy	.35	Junior	.25	Study	Issues	School	
Teeth	.36	Talk	.33	Start	.25	Minute	Faith	University	
Blue	.36	Tell	.33	Mature	.24	Read	Help	Texas	
White	.35	Insecure	.31	Experience	.24	Write	Strong	Pretty	
Green	.35	Good	.30	Grade	.23	Hours	Church	Fun	
Wear	.35	Want	.29	Month	.22	Semester	Heart	Guy	
Thin	.32	Judge	.29	Senior	.22	Sit	Control	High	
Smile	.32	Quiet	.27	Job	.22	Room	Mirror	College	
Black	.32	Outgoing	.26	Early	.21	Night	Control	High	
Features	.29	Different	.26	Age	.21	Assignment	Mirror	College	
Blonde	.29	Meet	.26	Different	.21	Wait	Mirror	High	
Look	.27	Time	.25	Love	.20	Point	Mirror	High	
Tall	.26	Reserved	.25	Love	.20	Pass	Mirror	High	
Beauty	.26	Group	.22	Love	.20	Thanksgiving	Mirror	High	
Weight	.25	Guess	.21	Love	.20		Mirror	High	
Short	.25	Open	.21	Love	.20		Mirror	High	
Mouth	.24	Hard	.21	Love	.20		Mirror	High	
Fat	.24	Stupid	.21	Love	.20		Mirror	High	
Physical	.24	Fun	.21	Love	.20		Mirror	High	
Appearance	.23	Pretty	.20	Love	.20		Mirror	High	
Body	.23	Nice	.20	Love	.20		Mirror	High	
Makeup	.22	Care	.20	Love	.20		Mirror	High	
Small	.21	Flaw	.20	Love	.20		Mirror	High	
Skinny	.20	Comfort	.20	Love	.20		Mirror	High	
Light	.20	Bother	.20	Love	.20		Mirror	High	
Attractive	.20	Blue	-.22	Love	.20		Mirror	High	
Red	.20			Love	.20		Mirror	High	
Life	-.21			Love	.20		Mirror	High	

Note. The above words were the most frequently used content words in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the first half of self-descriptive essays.

Appendix G
Text-based Person-Descriptive Factors: A Varimax-Rotated Varimax Principal-Components Analysis on the Second Half of Self-Descriptive Essays

Person Descriptive Factors										
Factor 1 Sociability	Factor 2 Appearance	Factor 3 Relationships	Factor 4 Education	Factor 5 Daily	Factor 6 Religion	Factor 7 Existentialism	Factor 8 Evaluation	Factor 9 Reflection/Interests		
Talk	.38	.42	.51	.34	.33	.27	.42	.36	Music	
Friend	.37	.42	.45	.31	.32	.26	.32	.35	Sports	
Think	.34	.37	.41	.29	.29	.24	.30	.27	Car	
Like	.34	.37	.37	.27	.28	.24	.28	.27	Town	
Perceive	.31	.32	.35	.27	.27	.23	.26	.26	City	
Feel	.31	.32	.35	.24	.25	.22	.25	.23	House	
Know	.30	.32	.34	.24	.25	.22	.25	.22	Athletic	
Mean	.28	.30	.31	.24	.22	.22	.24	.22	Play	
Shy	.27	.29	.30	.23	.22	.22	.23	.22	Favorite	
Good	.26	.28	.29	.22	.21	.20	.22	.22	Green	
Roommate	.25	.27	.27	.22	.21	.20	.22	.21	Watch	
Kind	.25	.26	.24	.22	.21	.20	.20	.21	Summer	
Close	.24	.26	.22	.22	.20	.20	.20	.20	Change	
Boyfriend	.24	.26	.22	.22	.21	.20	.20	.20	Goal	
Guy	.24	.26	.21	.21	.21	.21	.21	.25	Mirror	
Mood	.24	.25	.21	.21	.21	.21	.21	.24	See	
Meet	.24	.24	.21	.20	.20	.20	.20	.24	Look	
Fun	.24	.24	.20	.20	.20	.20	.20	.22		
Care	.24	.24	.20	.20	.20	.20	.20	.22		
Listen	.23	.23	.23	.23	.23	.23	.23	.22		
Time	.23	.23	.23	.23	.23	.23	.23	.22		
Guess	.23	.22	.21	.21	.21	.21	.21	.22		
Annoying	.22	.22	.21	.21	.21	.21	.21	.22		
Sad	.22	.22	.21	.21	.21	.21	.21	.22		
Girl	.22	.22	.21	.21	.21	.21	.21	.22		
Reason	.21	.21	.21	.21	.21	.21	.21	.22		
Tell	.21	.21	.21	.21	.21	.21	.21	.22		
Easy	.20	.20	.20	.20	.20	.20	.20	.22		
Tend	.20	.20	.20	.20	.20	.20	.20	.22		
Quiet	.20	.20	.20	.20	.20	.20	.20	.22		
Bad	.20	.20	.20	.20	.20	.20	.20	.22		

Note. The above words were the most frequently used content words in open-ended self-descriptions of personality that loaded greater than .20 in a factor analysis of their use in the second half of self-descriptive essays.