Summary
Desirability Ratings for 359 Turkish Personality Adjectives

Kürşad Demirutku

TED University

İrem Metin-Orta

Atılım University

Personality trait adjectives have long been used in research about a variety of topics in different subfields of psychology. In such studies, knowing the desirability norms of the words is consequential for controlling the probable confounding effects. In the past, various studies have been conducted in different cultures to establish norms for word frequencies (e.g., Thorndike & Lorge, 1944), concreteness (e.g., Eilola & Havelka, 2010), imagability (Altarriba, Bauer, & Benvenuto, 1999), valence (Võ, Jacobs, & Conrad, 2006), pleasantness (Bellezza, Greenwald, & Banaji, 1986), and desirability (Anderson, 1965, 1968). Among these studies, the first study to establish desirability norms for personality adjectives was conducted by Anderson (1968). These norms were referred to in other research which focused on self-schemas (e.g., Markus, 1977), social prototypes (Kinicki, Hom, Trost, & Wade, 1995), impression formation (e.g., Leahy, 1979), and stereotypes (e.g., Branscombe & Smith, 1990). However, there is no such database which contains the desirability norms of Turkish trait adjectives in the relevant literature. Therefore, the purpose of the present research was to construct a database to satisfy this deficiency. This objective has been achieved in two steps. First, we generated a personality adjectives pool depending on the empirical findings of research about Turkish personality adjectives (Somer, 1998; Somer & Goldberg, 1999; Goldberg & Somer, 2000). Second, we measured the desirability ratings of these adjectives to establish a normative database to be used in future studies.

We generated two hypotheses to test the validity of the desirability ratings. Firstly, depending on the relationship between meaningfulness of a word and its desirability (Anderson, 1968), and the relationship between meaningfulness and word frequency (Bowen, 1971), we hypothesized that desirability ratings of the personality adjectives would be positively correlated with the their word frequency in Turkish. Secondly, depending on the the *bad-is-stronger-than-good* hypothesis (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), we expected that lower desirability ratings of the negative personality adjectives would be stronger than the higher desirability ratings of the positive personality adjectives.

Method

In four consecutive studies, first, a Turkish adjective pool was generated. Then, after investigating the meaningfulness of these traits, a final list of 359 trait adjectives was prepared. Afterwards, this final list was presented to a sample of university students and desirability ratings were obtained for each adjective. Finally, test-retest reliability coefficients were computed for each adjective in a separate sample of university students.

Study 1: Generation of the Personality Adjectives Pool

We identified a number of studies in which the dimensionality of Turkish person-descriptive adjectives were investigated (Somer, 1998; Somer & Goldberg, 1999; Goldberg & Somer, 2000). We compiled 570 adjectives from these studies into a single database. After eliminating the redundant and synonymous adjectives, 367 adjectives remained in the initial pool.

Study 2: Measurement of Meaningfulness for Adjectives

Patterning after the design of Anderson (1968), we asked 10 male and 10 female university students to indicate the meaningfulness of each word for them. In the first step, they marked the adjectives which they were not certain about their meaning. In the second step, they rated the extent of which each of the remaining meaningful adjectives were appropriate to describe a person ("0 – Not appropriate at all", "3 – Very appropriate"). We used these evaluations to eliminate adjectives from the pool if at least three participants (6%) marked as meaningless and if appropriateness ratings was equal to 1.5 or below. This procedure resulted in a final adjective pool of 359 personality adjectives.

Address for Correspondence: Asst. Prof. Kürşad Demirutku, TED University, Department of Educational Sciences, Guidance and Psychological Counseling Program, Ziya Gökalp Caddesi, No: 47-48, 06420 Kolej / Çankaya / Ankara
E-posta: kursad.demirutku@tedu.edu.tr

Study 3: Measurement of Desirability Ratings

A total of 168 university students (63.13% women, %36.87 men, M_{age} = 21.19; SD_{age} = 1.70, *Range* = 18-27) participated in the study. Three hundred and fifty nine adjectives were randomly ordered, dispersed into eight pages with equal size, and these pages were arranged in a systematic random order for different participants. Participants were instructed to imagine a person who was described with each adjective, and then to rate the extent of which they would like or dislike this imagined person on a 7-point Likert scale ("0 – I would not like this person at all", "6 – I would very much like this person"). Participants filled out the questionnare either alone or in groups in a classroom setting after signing an informed consent form. Each participant received 1 bonus point (over 100 points) for their participation.

The data was analyzed adjective-wise by computing the means and standard deviations of the desirability ratings. A higher mean with a lower standard deviation indicated that a particular adjective was relatively more desirable with high agreement in the sample. Similarly, a lower mean with a lower standard deviation indicated that a particular adjective was relatively more undesirable with high agreement in the sample. Mean ratings around the scale midpoint 3.00 indicated relatively neutral adjectives. Turkish adjectives and their English equivalents are presented together with desirability parameters (*B*) and corresponding standard deviations (*SS*) in Table 1. Further statistical analyses were also conducted. Firstly, average desirability of 359 adjectives (*M* = 3.05, *SD* = 1.86, *Range* = 5.67-0.24) was not significantly different from the midpoint of the measurement scale (t_{358} = 0.52, *p* = .601, *two-tailed*). Therefore, it is possible to conclude that the adjectives in the pool were balanced in terms of desirability. In addition, desirable and undesirable adjectives displayed two skewed distributions fusing at the scale midpoint (see Figure 1). There were only 20 adjectives which can be labelled as neutral depending on the criterion of $2.50 \leq B \leq 3.50$. These distributional characteristics are in line with those reported by Anderson (1968). Percentile cutt-off points in terms of *B* is presented in Table 2 together with the marker adjectives.

Testing the two hypotheses of the study provided empirical support for them. As for the first hypothesis, a positive correlation was observed between the word frequencies of the Turkish trait adjectives and the desirability ratings (r_{xy} = .23, *p* < .001, *one-tailed*). Thus, as word frequency increased, desirability of the Turkish personality adjectives also increased. As for the second hypothesis, we first conducted two separate one-sample *t* tests to compare the average desirability of positive and negative adjectives to the neutral midpoint of 3.00.

effect size measures indicated that lower desirability evaluations of the negative traits were stronger than the higher desirability evaluations of the positive traits. We observed that the average desirability of the positive (*M* = 4.81, *SD* = .64, t_{74} = 24.65, *p* < .001, *one-tailed*, *d* = 2.85) and negative (*M* = 1.04, *SD* = .45, t_{74} = -38.21, *p* < .001, *one-tailed*, *d* = 4.41) adjectives was significantly different than the theoretical midpoint. The ratio of effect sizes for negative to positive adjectives was 1.55. Thus, the undesirability of negative adjectives was stronger than the desirability of positive adjectives.

Study 4: Measurement of Test-Retest Reliabilities

A total of 172 university students (86.00% women, %12.79 men, 1.21% unidentified, M_{age} = 21.8x; SD_{age} = 2.17, *Range* = 18-29) participated in the study. Adjectives which were used in Study 3 were randomized and divided into three sublists. Each sublist contained approximately 120 adjectives printed on two pages. Counterbalancing the six pages resulted in 12 forms. Each participant was administered only one form twice with a three-week interval. The forms were administered in pre-tagged envelopes to assure anonymity and confidentiality. Thus, each participant rated a different yet random set of adjectives twice. Each participant received 1 bonus point (over 100 points) for their participation. Average desirability ratings in the first and second administrations were used to compute correlations to serve as test-retest reliability measures. The size for the subsamples for each adjective ranged between 38-60.

Analyses revealed that average test-retest reliability of the adjectives was .47 (*SD* = .16, *Range* = -.02-.89). There were 24 adjectives with a reliability coefficient of .70 or above. There were 196 adjectives with a reliability coefficient equal to or greater than the average reliability. There were 308 adjectives with a reliability coefficient of .30 or above. Thirty three correlations (9.19%) were not significant. Finally, correlational analysis revelaed that desirability ratings were unrelated to test-retest reliability of the adjectives.

General Discussion

The present research is the first attempt at developing desirability norms for the Turkish personality adjectives. The resultant database contained personality adjectives with relatively low levels of test-retest reliability. It is possible that asking participants to make desirability judgments without providing specific social-contextual information might have produced instability in ratings. Perhaps, the participants have used different mental anchors. In addition, the cognitive load of evaluating so many stimuli at once might have resulted in random re-

sponding for some adjectives or participants. Nevertheless, since more than half of the adjectives had reliability coefficients above the average in positive correlation terms, it is possible to argue that there is some amount of consistency in the measurement.

The validity of the desirability ratings were assessed by testing two hypotheses. As expected, word frequency was a correlate of desirability. Taken together with Anderson's (1968) and Bowen's (1971) discussion that meaningful adjectives are more desirable, our findings pinpoint a possible measurement error in research about stereotypes, self-evaluation, or impression formation in which personality adjectives were used as stimuli. Therefore, we recommend that word frequency and desirability norms in Turkish are both taken into account when selecting adjectives for research purposes. Future studies are required for meaningfulness norms of the Turkish personality adjectives as well.

As expected, the low desirability ratings for the negative adjectives were stronger than the high desirability ratings for the positive adjectives. In other words, the negativity of the negative adjectives were stronger than the positivity of the positive. This provided another support for the bad-is-stronger-than-good hypothesis (Baumeister et al., 2001) and indicated that desirability norms in the present study have some validity. This result also implies that adjective pairs in semantic differential scaling can only be matched in terms of valence, but there will be an inherent and unavoidable bias in terms of magnitude of negativity.

There are certain limitations of the present research. First of all, an elimination of adjectives was preffered in order to provide participants a relatively managable list. In this elimination synonymous adjectives were discarded and thus the richness of language in daily usage was disregarded. Secondly, elimination process excluded adjectives which qualify properties other than personality traits. In future studies, the database we provided needs to be expanded by including the eliminated adjectives and by adding norms for controllability, modifiability, meaningfulness, valence, concreteness, and observability. Thridly, the samples of the study were composed of university students with an over-representation of women. This limits the generalizibility of other populations including children and non-student adult population.

In conclusion, the contribution of the present research is that it provided desirability norms for future studies in Turkish samples about memory, social cognition, attitudes, self-concept, and the like. To our best knowledge, our research is one of a few similar norm studies in Turkish (Göz, 2003; Tekcan & Göz, 2005), and we believe it will be a useful resource for researchers until its shortcomings are eliminated by further research.