DOES THE NEGATIVELY-WORDED ITEM EFFECT EXIST IN PERSONALITY MEASURES? A META-ANALYSIS

BY

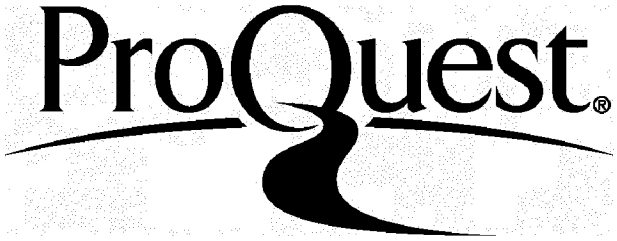
JIALIN HUANG

DEPARTMENT OF PSYCHOLOGY

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Psychology
in the Graduate College of the
Illinois Institute of Technology

Approved _____
Adviser

Chicago, Illinois
May 2015



ACKNOWLEDGEMENT

I would like to take opportunity to acknowledge and thank those who made this dissertation possible. Foremost, I would like to express my sincerest gratitude to my parents, Min and Xuefen. Their love and supports inspire me to pursue my dreams, and motivate me to make them come true.

To my advisor, Alan Mead, I wish to offer my deepest thanks. He provided guidance, advice and encouragement through the course of this dissertation. Moreover, his kindness, support and wisdom have been invaluable to me on both an academic and a personal level. I could not have imaged working with a better advisor and mentor.

Special thanks to my committee members, Dr. Scott Morris, Dr. Ron Landis, and Dr. Shlomo Argamon, for their detailed comments, helpful suggestions and great supports. Meanwhile, I am grateful to all faculty and staff at Psychology department at Illinois Institute of Technology for their hard work.

My gratuities to Dr. Michael Biderman, Dr. Alberto Maydeu-Olivares, Dr. Sam McAbee, Dr. Lewis Goldberg, Dr. AC Del Re, and Ayi Fleischer, who generously shared ideas, data and/or statistics with me, so that I was able to complete the analysis.

I am thankful to my friends, Aaron Miller, and Kevin Franke, who helped me code articles and edit the manuscript.

My supervisor and colleagues at Wonderlic also deserve my thanks for always being kind and supportive.

I appreciate encouragement, support, and friendship from my friends and peers in the U.S. and China, who make my life more pleasant and enjoyable.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Symbol	Definition
d	Effect Size
M_n	Mean of Item Discrimination on Negatively Worded Items
M_p	Mean of Item Discrimination on Positively Worded Items
SD	Pooled Standard Deviation
M_t	Mean of Item Discrimination on Negated Items
M_s	Mean of Item Discrimination on Non-Negated Items

ABSTRACT

This study meta-analyzed IRT item discrimination parameter estimates and CFA item loadings to explore the methodological effect of negatively-worded items in personality measures. We found three important moderators that determined whether the negatively-worded item effect affected a scale. The first moderator was the manner in which the scale was defined. We found a strong negatively-worded item effect for *Neuroticism* but not for *Emotional Stability*. The personality scale was also a moderator, with a negatively-worded item effect being observed for *Agreeableness*, *Extraversion*, *Neuroticism*, and *Openness* but a positively-worded item effect for *Emotional Stability* and *Conscientiousness*. Third, low-motivation samples tended to produce a larger negatively-worded item effect. Moreover, there was no statistically significant difference between personality and non-personality inventories regarding the negatively-worded item effect. Finally, item negation did not produce the expected effect. Practical implications and limitations of the study are discussed.

CHAPTER 1

INTRODUCTION

Considerable research has investigated a "negatively-worded item effect" in diverse psychological inventories, such as personality measures (Sliter & Zickar, 2014), Rosenberg Self-Esteem Scale (RSES; Greenberger, Chen, Dmitrieva, & Farruggia, 2003; DiStefano, & Motl, 2006; DiStefano, & Motl, 2009a; DiStefano, & Motl, 2009b) and many other self-report measures (Hankins, 2008; Lai, 1994; Lyrakos, Damigos, Mavreas, Georgia, & Dimoliatis, 2010; Ye, 2009). It is usually referred to as methodological bias, where negatively-worded items adversely affect some measurement properties of the scale. It is also known as the "reverse-coded items effect." For instance, negatively worded items tend to compose one unique factor in confirmatory factory analysis (CFA), indicating the items were not loaded on one factor as they were expected (Greenberger et al., 2003). Meanwhile, negatively worded items exhibit lower item discrimination and lower information based on an item response theory (IRT) analysis (Sliter & Zickar, 2014).

Although negatively-worded items have adversely affected some measurement properties of scales, there is a lack of consensus on whether the negatively-worded item effect really exists in personality measures (Campbell, Siegman, & Rees, 1967; Holden, Keen, & Jackson, 1985; Jackson & Lay, 1969; Trott & Jackson, 1967; Sliter et al., 2014). Therefore, this study used meta-analysis to investigate item properties in the considerable amount of research which provides item-level statistics. If the negatively-worded item effect happens in personality scales, the psychometric problems associated with the effect can be resolved by simply excluding negatively worded items. However, if the

negatively-worded item effect does not occur then removing negatively-worded items could damage the content validity of personality scales. There is also the question of how the items of bipolar scales, like *Neuroticism* or *Agreeableness,* know that they are worded "negatively" so that they can perform more poorly. Examining the conditions under which the effect occurs may shed light on the underlying mechanism.

The following section will define the concepts of negatively and positively worded items, the negatively-worded item effect, and negation items, review literature regarding the negatively-worded item effect, and propose the current study.

1.1 Definition of Negatively Worded Items and Operationalization of the Negatively-Worded Item Effect

1.1.1 Definition of Negatively Worded Items. Negatively-worded items indicate the opposite pole of a scale intended to measure. For example, the item "I rarely feel blue." on a *Neuroticism* scale would be negatively-worded because it indicates emotional stability. Thus, all negatively-worded items are reverse-coded during scoring and might be referred to as "reverse coded items" while positively-worded items are not reverse-coded during Likert scoring (Schmitt & Stults, 1985).

An implicit assumption in the above discussion is that scales have a direction and therefore the positivity or negativity of an item relates to the polarity of the construct and the intentions of the scale developers. Certain constructs, such as self-esteem (Rosenberg, 1965), are conceptualized as unipolar (more or less self-esteem) and the (intended) direction of measurement is unlikely to change. For example, we are unaware of anyone measuring a lack of self-esteem using Rosenberg's 1965 measure. However, personality and other constructs are conceptualized as being inherently bipolar and the nature of

negative-wording could flip, depending on the intentions of the test-developer. For instance, "I feel blue." would be negatively-worded on a measure of *Emotional Stability* but the same item would be positively-worded on a *Neuroticism* scale. Within the "Big Five" tradition (Goldberg, 1992), the poles of four of the five constructs are consistently measured as: *Extraversion, Agreeableness, Conscientiousness*, and *Openness*. The fifth factor is measured as both *Neuroticism* and *Emotional Stability*. The intentions of models other than the Big Five are generally consistent, although Cattell's five-factor model measured *Tough-Mindedness* (*Openness* reversed) and *Independence* (*Agreeableness*, reversed) (Conn & Rieke, 1994) and Eysenck's *Psychoticism* involved aspects of low *Conscientiousness*, low *Openness* and *Disagreeableness* (Eysenck & Eysenck, 1975). Please read Appendix A for more discussion of the Five-Factor Model (FFM) of personality.

For the purpose of this study, we will focus on the most common "Big Five" conceptualization of a construct and the most common (or original) conceptualization of non-personality constructs. Therefore, for personality dimensions, we will standardize on *Agreeableness*, *Consciousness*, *Extraversion*, *Emotional Stability* and *Openness*.

1.1.2 Definition of Negated Items. Previous research (Holden & Fekken, 1990; Holden, Fekken, & Jackson, 1985) has defined negated items in three ways: (1) with not (e.g., *not*, *-n't*, *never*); (2) with implicit negative (e.g., negative prefixes as *im-*, *un-*, *in-*); (3) with negative qualifiers (e.g., *seldom* or *rarely*). Accordingly, non-negated items are those that do not match any of these three rules. Those definitions will be applied for the current study.

Negative-wording is distinct from negation, and this has been a point of confusion in past research (e.g., Sliter & Zickar, 2014). An item like "I do not like to read quietly." might indicate *Extraversion*, which is generally the "intended" pole of most personality surveys, and therefore would not be negatively-worded but the item includes "not", and therefore is negated. In contrast, "I don't like loud parties." would indicate *Introversion,* and is therefore both a negatively-worded item and a negated item.

Based on the definition of negatively and positively worded items and negated items, a personality item (take *Extraversion* as an example) can be categorized into four types:

(1) A positively worded item without negation, such as "I like loud parties."

(2) A negatively worded item without negation, such as "I like quiet evenings at home."

(3) A positively worded item with negation, such as "I don't like quiet evenings."

(4) A negatively worded item with negation, such as "I don't like loud parties."

In practice, there may be some correlation between the use of negation and negative wording but the two item states are by no means perfectly correlated. In fact, they had different psychometric properties (Holden & Fekken, 1990; Holden, Fekken, & Jackson, 1985) Many negatively-worded items would not have negation and some positively-worded items may include negation.

1.1.3 Operationalization of the Negatively-Worded Item Effect. The negatively-worded item effect can be investigated in various ways, such as descriptive statistics (e.g., mean of responses), reliability or validity of scale scores, or using the parameter estimates from statistical models such as IRT and CFA.

However, some of these properties probably cannot serve as the best indicator of disparity between positively and negatively worded items. For example, the validity of scale scores is affected by the reliability of the scale scores. The maximum of validity cannot exceed the square root of reliability (Allen & Yen, 2001). Thus, a scale with lower reliability definitely has lower observed validity. Furthermore, longer scales have higher reliability than shorter scales, when everything else is equal (Spearman, 1910; Brown, 1910). Therefore, the reliability or validity of the positively- and negatively-worded items can only be compared if the numbers of such items are equal, which is rarely the case. In addition, calculating these reliability or validity coefficients may often be impossible from the data provided in most studies.

In the current study, we will focus on the effect of negatively worded items on item characteristics indicating the quality of construct-correlation of an item. In a CFA analyses, these are the item factor-loadings (in a model with simple structure), and in an IRT analysis, these are the item discrimination parameter estimates. Comparing these item-level statistics provides a straightforward and intuitive approach to assess whether negatively and positively worded items have the same psychometric quality or whether one type of item wording is more effective at measuring the underlying construct. Both CFA and IRT are popular frameworks of item/scale evaluation are widely used in psychology research (Raju, Laffitte, & Byrne, 2002). The CFA factor-loadings are closely related to IRT item discrimination parameters (see detailed discussion in Method section 2.2.2).

Other researchers investigating negatively-worded item effects have used methodologies like competing structural models (Marsh, 1996; DiStefano & Motl, 2006)

or experimental design (Barnette, 2000), which will be reviewed in later section 1.3. Although these methods have advantages, they will never be amenable to meta-analysis, because their findings may suffer from sampling errors. Furthermore, an analysis of the IRT discrimination or CFA factor loading is an appropriate empirical way to assess item quality. That is, an item's IRT discrimination (which is proportional to its loading) or its loading is the best indication of the item's relationship to the underlying trait. Consequently, directly comparing the quality of negatively- and positively-worded items is an appropriate and effective way to understand which type of item is better. The fact that such effect sizes are also amenable to meta-analysis is a second significant advantage of our approach to operationalizing the negatively-worded item effect.

1.2 Likert Scaling and Scale Construction

The inclusion of negatively and positively worded items is recommended for scale construction (Cronbach, 1950), especially in developing Likert scales (1932). The following section will review the basic method of Likert scaling, and the importance of negatively worded items in Likert scale construction.

Likert's scaling method is used in almost all self-report surveys, and contrasts with much more complex methods promoted by Thurstone (1928). In Likert scaling, items are selected as statements to which a respondent agrees or disagrees (e.g., a common scale has five points labeled "*Strongly Disagree*," "*Disagree*," "*Neutral*," "*Agree*," and "*Strongly Agree*") and items are scored using contiguous integers (e.g., 1, 2, 3, 4, and 5). When reverse-scored items have been included, such items are reverse-scored at this point. So, on an *Extraversion* scale, an *Introversion* item would be scored

5, 4, 3, 2, or 1. In this way, disagreeing with the introversion item would indicate more extraversion (4 or 5) and agreeing would indicate less extraversion (1 or 2). Reverse scoring also ensures that all items should have positive item-total correlations, and thus item-total correlations can be used to detect and remove off-track or low-quality items (or to identify and re-score incorrectly scored items). Although Likert's integer scoring is undoubtedly less precise than Thurstone's more complex procedures that incorporate stimulus scaling, in practice Likert's method is much simpler and is quite effective (Huang & Mead, 2014).

Why include items that must be reverse-scored? As described in later sections, this question has been asked by many researchers, so it is important to outline the reasons why a mixture of positively- and negatively-worded items might typically be used. One reason for balanced (between negatively and positively worded items) scales is to combat response biases and inattentive responding (Cronbach, 1950). If high scale scores are associated with agreeing (or disagreeing) with all items, then it is difficult to distinguish between respondents who obtained a high score because they are truly extreme on the underlying dimension versus those who responded in a biased manner. Acquiescence bias refers to a tendency to agree to all items regardless of content but inattentive responders could also choose the simple expediency of agreeing or disagreeing to all items. On a scale with mixed positively and negatively worded items, extreme high and low scores can unambiguously be interpreted, because those who answer with uniform agreement or disagreement will not be assigned moderate scales scores.

Respondents responding in an idiosyncratic manner (e.g., through carelessness or acquiescence) might possibly be distinguished from who genuinely hold middling

standing on the trait by the pattern of responses by constructing a validity scale composed of pairs of similar or opposite items (Meade & Craig, 2012). For example, a person responding in the intended manner who endorses "I rarely feel blue." should not endorse "I often feel blue." but should endorse "I'm usually happy." Pairs of such related items can be detected by correlating responses and picking items with strong positive or negative correlations. The validity scale is then created by scoring mismatching responses as one and matching responses as zero (or something similar) and then summing across all such pairs.

Another critically important reason for composing scales of mixed positively- and negatively-worded items is content validity, which refers to the degree to which a measure adequately samples the entire domain of the construct (Allen & Yen, 2001). If personality traits are defined by behaviors, thoughts, and feelings that characterize their two poles, then a scale composed of only characteristics sampled from one pole will necessarily suffer from a lack of content validity, and may even measure a somewhat incorrect construct.

Even more subtly, the discrimination of the scale scores may be related to adequacy of the sampling of the entire content domain. If behaviors, thoughts and feelings associated with one pole of a bipolar construct are systematically ignored, then measurement at the disfavored end of the construct, and perhaps also in the middle of the construct, will be inadequate and the range of scale scores will be severely restricted. Even though the Rosenberg Self-Esteem Scale (1965) is conceptualized as a unipolar scale measuring degree of self-esteem, the items sample both low self-esteem thoughts (e.g., "All in all, I am inclined to feel that I am a failure.") and high self-esteem thoughts

(e.g., "On the whole, I am satisfied with myself."). Therefore, the mixture of negatively and positively worded items ensures that the scale covers adequate contents, and its scale scores show sufficient discrimination, even though the underlying construct is unidimensional.

1.3 Literature on the Negatively-Worded Item Effect

Researchers have conducted a considerable amount of research on various scales (such as Rosenberg's Self-Esteem Scale, Life Orientation Test, and General Health Questionnaire), and found various effects for negatively-worded and/or negated items. The literature is summarized below.

1.3.1 Psychometric Properties of Negatively Worded Items. Research showed that scales with negatively worded items demonstrated inferior psychometric properties. Negatively worded items had lower means and scale reliability (like internal consistency reliability), and can reduce validity of a scale, where they tend to form a non-meaningful method factor in factor analysis (Barnette, 2000; Chamberlain & Cummings, 1984; Greenberger et al., 2003; Knight, Chisholm, Marsh, and Godfrey, 1988; Marsh, 1996; Pilotte & Gable, 1990; Schriesheim et al., 1995; Schriesheim et al., 1991; Schriesheim et al., 1981). Therefore, empirical evidence suggested that negatively worded items displayed inferior psychometric properties compared with positively worded items.

Negatively-worded items tend to present lower mean score than positively-worded items. For instance, Schmitt and Allik (2005) translated Rosenberg's Self-Esteem Scale (RSES) into 28 languages and collected data in 53 countries. Even though the correlations between positively and negatively worded items were moderate to high in

general (with exception of r=-0.07 in Tanzania), there was still a clear tendency that scores of positively worded items were higher than scores of negatively worded items (after reverse scoring the negatively worded items) across all samples.

Research has shown that scale scores with only negatively worded items exhibited lower reliability than scale scores with only positively worded items or mixed direction-of-wording items. Schriesheim and Hill (1981) compared three questionnaires (all items positively worded, all items negatively worded, and mixed) on internal consistency and means. They concluded that negatively worded items might impair the accuracy of measurement and internal consistency. Chamberlain and Cummings (1984) compared the internal consistency reliability of scores on two forms of a course evaluation. The results suggested that the reliability of scores was higher on the positively worded scale than the negatively worded scale.

Schriesheim and colleagues also investigated psychometric properties of different types of negatively worded items compared with positively worded items (Schriesheim et al., 1991; Schriesheim et al, 1995), regarding reliability and validity. They specified four types of items: Regular items (positively worded items, e.g., "He makes the use of uniform procedures required."), negated regular (negatively worded items, e.g., "He does not make the use of uniform procedures required."), polar opposite (negatively worded items, e.g., "He makes the use of uniform procedures optional.") and negated polar opposite items (positively worded items, e.g., "He does not make the use of uniform procedures optional."). The polar opposite items were written by using antonyms (listed in Roget's Thesaurus), which did not change the meaning or connotation. Negated

versions of the regular and polar opposite items were produced by adding either the phrase "does not" or "not" to each item.

They discovered that internal consistency reliability and response accuracy (which was defined as 5 minus the difference between script level and subject's response) decreased as listed in the previous order (Schriesheim et al., 1991). That is, scores of regular items showed the best reliability and response accuracy, followed by negated regular, polar opposite and negated polar opposite items came last.

The scores of negatively worded items do not only have lower reliability, but also lower construct validity. Schreisheim and colleagues (1995) factor analyzed the four types of items under different experimental conditions. Exploratory factor analysis (EFA) results showed that the polar opposite and negated polar opposite items can produce problematic factor loadings. Confirmatory factor analysis (CFA) results suggested that four separate method factors existed, although regular and negated polar opposite were expected to load on a single factor, while polar opposite and negated regular were expected to load on another factor, according to item wording direction. The results implied that each item type would have its own method effect. Moreover, the four item types may not have equal trait, method and error variance in CFA analysis. Regular items have superior properties over the other three, regarding higher trait variance, lower error variance and little method variance.

The findings of Schreisheim and colleagues (1995) are consistent with previous studies on exploring factor structure of negatively worded items. That is, the inclusion of negatively worded items is likely to undermine the expected dimensionality of a scale. Benson and Hocevar (1985) applied CFA to data from three forms of the same

questionnaire (all negatively worded, all positively worded, and a mix of half negatively worded and half positively worded), which were responses from elementary students between grade 4 to 6. They discovered that a two-factor solution according to item wording fit the mixed-format data better than a one-factor solution. Furthermore, they concluded that transformation from positive to negative wording changed the dimensionality of the scale. Similarly, Knight and colleagues (1988) suggested that negatively and positively worded items tended to load on different factors, regardless of their contents. So they stated the revised 20-item UCLA loneliness scale should be divided into two 10-item scales due to item polarity, and those scales still assessed unidimensional loneliness. Therefore, the evidence supported that negatively worded items lead to problematic factor structure of the underlying construct.

1.3.2 The Negatively-Worded Item Effect and RSES. There are far more studies on Rosenberg's Self-Esteem Scale (RSES) regarding the negatively-worded item effect than any other inventory in the literature, even though the effect has been broadly examined in different assessments, such as General Health Questionnaire (GHQ-12; Hankins, 2008; Ye, 2009) and Life Orientation Test (Lai, 1994; Lyrakos et al., 2010). Research on RSES and the negatively-worded item effect can provide a general representation of methodologies, results and issues that relate to the topic, because the effect has been widely and deeply investigated in RSES from various perspectives. Meanwhile, the review of the literature would provide certain insights for investigation of the negatively-worded item effect in personality measures.

Conceptually, global self-esteem is a distinct psychological construct from personality, but they are correlated at a degree. Global self-esteem was defined as an

individual' overall sense of worthiness as a person (Rosenberg, 1979), which can be a positive or negative attitude towards the self (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995). Because of its nature as an attitude, global self-esteem has features that all attitudes own: First, there is always an object of attitude. Global self-esteem is an attitude regarding the self as a totality. Second, attitudes involve in cognitive, affective and behavioral aspects. Therefore, global self-esteem has components of cognition, affect and behavior. Third, like other attitudes, global self-esteem has both positive and negative prospects. Last, self-esteem has a specific function, that is, protect and maintain one's self-image or self-confidence (Owen, 1993).

Meanwhile, personality was defined as trait ("...relatively enduring styles of thinking, feeling and acting"; McCrae & Costa, 1997, p. 509), especially Five-Factor Model (FFM), which will be further discussed in Appendix A. Personality does not have a particular object or a concrete purpose as global self-esteem does, even though it is measured as bipolar trait. However, empirical evidence suggested that self-esteem highly related with *Extraversion* and *Neuroticism* across multiple cultures (Schmitt & Allik, 2005). Therefore, personality and global self-esteem are distinct but related concepts.

RSES was developed by Rosenberg (1965) to measure self-esteem, which is the most widely used assessment of global self-esteem (Marsh, 1996). There are five positively worded items and five negatively worded items on a response scale ranging from 1 (strongly disagree) to 4 (strongly agree).

Rosenberg originally proposed that the self-esteem scale measures only one single factor. However, empirical studies challenged his argument and supported a two-factor model associated with direction-of-wording (Carmines & Zeller, 1979; Greenberger et

al., 2003; Marsh, 1996; Tomas & Oliver, 1999; Wang, Siegal, Falck, & Carlson, 2001). Carmines and Zeller (1979) were considered as the first authors to discuss a two-factor solution of RSES due to negatively worded items (Gana, Alaphilippe, & Bailly, 2005). Thus, they calculated the sums of positively and negatively worded items in RSES, and correlated them separately with external criteria. But there was no substantial difference. So they stated that the two-factor model of RSES may be caused by artificial methodological bias of negatively worded items.

Negatively worded items became a focus when factor structure of RSES was analyzed by various methods and in various samples. Marsh (1996) conducted CFA (including correlated uniqueness models) to examine dimensionality of a seven-item version of RSES, which includes four positively worded items and three negatively worded items. He suggested that the method effect due to negatively worded items may contaminate one-single factor model of RSES, even though structural equation modeling (SEM) results confirmed that global self-esteem is a universal construct.

Greenberger and colleagues (2003) discovered contradictory information on factor structure of RSES when EFA and CFA were applied to the sample. They created an all negatively worded version of the scale (Revised-negative version) and an all positively worded version (Revised-positive version) based on RSES. Even though EFA analysis presented a single-factor model, CFA analysis implied a different solution for both revised versions. CFA results showed that the original RSES showed better model fit of a two-factor solution than a one-factor solution. For the Revised-negative version and the Revised-positive version, a single factor solution fit data, but a two-factor

solution fit the original RSES well. However, the one-factor solution was not ideal for both revised versions, which suggested that there might be another factor in the structure.

Tomas and Oliver (1999) and Wang and colleagues (2001) applied similar CFA methods (such as MTMM, CTCU, and CTCM) on different samples (Spanish and American samples) to examine model structure of RSES. They compared and contracted model fit of various RSES factorial specifications, including a universal model, a two-factor model, two-factor models with modifications and others. Their findings agreed that a single-factor model existed in RSES data, but it was undermined by the method effect. However, they disagreed on the source of the contamination: The first study attributed negatively worded items to the contamination (Tomas et al., 1999), but the second study suggested both positively and negatively worded items may cause the method effect (Wang et al., 2001).

Meanwhile, DiStefano and Motl (2006; 2009a; 2009b) conducted a series of studies on examining the negatively-worded item effect of RSES. Although their findings supported that a single factor model with the negatively-worded item effect model fit the data, they yielded consistent observations of the effect in different samples and different time periods.

DiStefano and Motl (2006) tested the assumption that the negatively-worded item effect was response bias by applying MTMM framework on six different instruments, including RSES, Social Physique Anxiety Scale (SPAS), Marlowe-Crowne Social Desirability scale, Behavioral Inhibition System /Behavioral Activation System (BIS/BAS), Fear of Negative Evaluation (FNE), and Self-Consciousness (SC). They discovered that the inclusion of negatively worded items (a one-factor solution with a

negatively-worded item method) generated better model fit for RSES (than one-factor solution), as well as SPAS data. Moreover, different measures shared the common method effect associated with negatively worded items. Meanwhile, the BIS/BAS and social desirability did not show a significant predictive relationship with the negatively-worded item effect, but FNE, SC and evaluations from others could predict the method effect.

DiStefano and Motel (2009a) also examined the relationships between personality traits and the negatively-worded item effect by gender. The authors applied correlated trait-correlated method (CTCM) framework to examine the role of personality in the negatively-worded item effect by using six personality instruments, such as Marlowe–Crowne Social Desirability scale (MCSD), the Lie scale from the Eysenck Personality scale (EPS-L), Behavioral Inhibition System/Behavioral Activation System (BIS/BAS), Fear of Negative Evaluation (FNE), Self-Monitoring (SM), and Self-Consciousness (SC).

The results showed that the BAS Fun Seeking scale had negative correlations with the method factor, for both female and male groups. BIS was the only personality factor that related with the method effect due to negatively worded items for men. For women, the EPS-Lie subscale, FNE scale and private self-consciousness provided significantly negative relationships with the negatively-worded item effect. But the BAS Reward Responsiveness scale showed a positive relationship with the method effect of negatively worded items. The two self-monitoring scales did not yield significant relationships with the effect for the overall sample or for the male sample. But there was a marginal significant relationship between social desirability and negatively worded items for females. The authors stated that women responded to RSES more carefully than men,

which led to less method effect associated with negatively worded items. They believed that selected personality played an important role in resulting the negatively-worded item effect in RSES, although some traits can hardly be defined as personality.

DiStefano and Motel (2006; 2009a) conclusions are similar to previous findings on the negatively-worded item effect of RSES and personality. Quilty, Oakman and Risko (2006) applied the correlated trait-correlated uniqueness (CTCU) and the correlated trait-correlated method (CTCM) model to explore the relationship between the negatively-word item effect of RSES, approach and avoidance motivation, and personality (measured by IPIP). They found that the inclusion of the effect improved model fit for RSES. Also, avoidance motivation (BIS as mentioned below) displayed a statistically significant correlation (r=-0.27) with the negatively-worded item effect. Moreover, *Conscientiousness* (r=0.14) and *Emotional Stability* (r=0.25) were significantly related to the effect. However, five personality dimensions were all significantly correlated with RSES scores, and the magnitude of correlation coefficients were larger than the correlations between the negatively-worded item effect with *Conscientiousness* and *Emotional Stability*.

On the other hand, the negatively-worded item effect on RSES demonstrated gender invariance as well as longitudinal invariance, which implied that the negatively-worded item effect is response style or response bias, rather than artifact methodological effect. Although the negatively-worded item effect was observed in both gender groups, it did not impact measurement equivalence and mean differences in global-esteem between men and women (DiStefano and Motel, 2009b). The authors specified a one-factor model of RSES with correlated uniqueness of all negatively worded items. Then,

they established measurement invariance between male and female by testing metric, scalar, latent mean, factor, and item variance equivalence accordingly. The observation suggested that the method effects of RSES showed no difference on all structural levels. That is, the magnitude of the negatively-worded item effect were invariant for men and women. It is worth noting that the factor loadings table demonstrated negatively worded items presented consistently smaller values than positively worded items.

Meanwhile, Motl and DiStefano (2002) analyzed RSES data over 6 years, which were collected across 3 waves with 2 years each wave. They established a two-factor model for a scale with four positively worded items and three negatively worded items. That is, all seven items loaded on a RSES factor, and three negatively worded items loaded on a method effect factor. Those two factors were uncorrelated with each other. The model was tested across three time periods under CTCM framework. The results indicated the negatively-worded item effect did not change across time, because the factor structure, factor loadings, item uniquenesses, factor variances, and factor covariances were equivalent in the longitudinal analysis. It suggested that the negatively-worded item effect were static and stable over time. The method effects were not noise in the data, but of substantive importance.

Both invariance studies agreed that the negatively-worded item effect existed in RSES and presented stability in subgroup and over time. So the author argued that those features fit the definition of response style (Motl & DiStefano, 2002). This explanation promoted the importance of the negatively-worded effect in RSES, and also clarified the nature of the effect, that is, more than an artificial methodological product.

In general, most RSES studies concluded that a two-factor solution due to direction-of-wording fit the data better, even though evidence supported that there is only one underlying construct (DiStefano & Motl, 2006; 2009a; 2009b). Thus, majority of scholars considered the negatively wording effect as an artificial effect. Some argued that the effect was response bias or response style.

1.3.3 Why Does the Negatively-Worded Item Effect Happen? Sufficient evidence supported presence of the negatively-worded item effect, and inferior qualities of negatively worded items. However, there are much fewer studies that attempted to investigate what factors result in the negatively-worded items effect. Item characteristics (McPherson & Mohr, 2005), item-selection strategies (Miller & Cleary, 1993), individual differences (Barnette, 1996; Benson & Hocevar, 1985; DiStefano et al., 2006; Marsh, 1996; Melnick & Gable, 1990; Tamir, 1993), and the proportion of careless responses (Schmitt & Stults, 1985; Woods, 2006) were examined for their contribution to the effect.

1.3.3.1 Item Characteristics. Item characteristics have been identified as the factors that might lead to the negatively-worded item effect. It has been shown that item extremity played important role in the negatively-worded item effect (McPherson et al., 2005). That is, items with more extreme statements are more likely to produce the negatively-worded item effect than neutral items. The authors argued that this is due to a "neither-either continuum" in parallel to extreme-moderate wording, which seems very similar to ideal point model (Chernyshenko et al., 2007; Huang & Mead, 2014). When items became more extreme, the participants tended to endorse neither positively nor negatively worded items; when items became less extreme, the participants tended to endorse either positively or negatively worded items.

1.3.3.2 Item Selection Strategies. It has been suggested that different item-selection strategies may cause emergence of the negatively-worded item effect (Miller & Cleary, 1993). The researchers selected 12 items out of the 39-item pool in the development of the revised UCLA Loneliness Scale, where 20 items were positively worded and 19 items were negatively worded. Three item-selection strategies were examined: Random selection, high alpha (high item-total correlation) and low alpha (low item-total correlation). They reported that the internal consistency reliability form high alpha selected form was the highest (0.90), then followed by random form (0.88) and low alpha selected form is the lowest coefficient alpha (0.80). Meanwhile, the correlation between positively and negatively worded items was lowest in high alpha form, compared with random form. Factor analysis showed that high alpha form tended to generate a two-factor solution due to item wording. Random form presented the similar trend, but the factor pattern was less clear. However, low alpha form was not observed the direction-of-wording effect.

1.3.3.3 Individual Differences. Individual differences, such as abilities (Barnette, 1996; Benson et al., 1985; Marsh, 1996; Melnick et al., 1990) and personality (DiStefano et al., 2006), have shown to be relevant to the negatively-worded item effect. Abilities, like reading levels, education, and cognitive requirements were reported to correlate with the negatively worded item effect.

Benson and Hocevar (1985) found that it was difficult for elementary students to indicate agreement by disagreeing with a negatively worded item. Marsh (1996) also discovered that preadolescent students had difficulty discriminating direction of wording, which was related to reading ability. Students with lower reading levels were less likely

to respond appropriately to negatively worded items compared with their peers with higher reading levels. One explanation is that individuals with higher reading abilities are better at distinguishing subtle meanings among questionnaire items (Kaufman, Rasinski, Lee, & West, 1991).

Meanwhile, Melnick and Gable (1990) reported that adult respondents with lower education levels were more likely to provide inconsistent responses when mixed item types were used in an assessment. Barnette (1996) compared distributions of positively and negatively worded items on an attitude survey completed by students and teachers. Students presented a higher proportion of different distributions due to positively and negatively worded items than teachers. Similar to reading abilities, individual with more education might be able to detect subtle distinction of meanings among items. Or those people might be more familiar with questionnaires in general.

Cognitive requirements may differ when they are needed for responding to negatively and positively worded items (Sliter & Zickar, 2014). Specifically, when negatively worded items are presented, an individual needs to process both the word (or phrase) and reverse of the item at the same time, and then endorse an option that best fits. The working memory that negatively worded items require can be twice as much as positively worded items do (Tamir, 1993). It could explain that respondents with higher reading levels or higher education levels are different than those with lower abilities.

RSES studies also suggested that different personality traits related to the negatively-worded item effect. For example, scores from self-conscientiousness, and fear of negative evaluation scales associated with RSES negatively-worded item effect (DiStefano et al., 2006). *Conscientiousness* and *Emotional Stability* were significantly

correlated with negatively-worded item effect in RSES (Quilty et al., 2006). It was argued that self-conscientiousness, and fear of negative evaluation shared a common component, that is, self-reflection. Thus, people, with greater fear of negative evaluation, higher self-conscientiousness or both, are more likely to provide accurate assessment and less sensitive to item wording (DiStefano et al., 2006). It would also apply for individuals with high *Emotional Stability*.

1.3.3.4 Careless Responses. The proportion of careless responses in the data was related to the negatively-worded item effect. Schmitt and Stults (1985) study implied that 10% (or more) of careless respondents would create the negatively-worded item effect. Three types of correlation matrices were generated to test the hypothesis. After negatively worded items were recoded for all 400 cases (which presented no careless responses in the data), principal component analyses (PCA) were conducted. The results would represent what it is supposed to be when there are 0% careless responses. The degree of careless responses was manipulated by controlling different proportions of not reversely coded cases (e.g., 5%, 10%, 15%, and 20%) in the dataset. Results showed that when only 10% of careless responses were included then a "negative factor" would emerge. When the proportion of careless responses went up, the size of factor loadings for the negatively worded items increased as well.

Another simulation study also suggested 10% (or more) of careless respondents would lead to failure of fitting a single factor model, but fit a two-factor model, which was not supposed to (Woods, 2006). Woods applied PCA with varimax rotation on simulated data, and found that a certain amount of careless responses produced a unique methodological factor. CFA was conducted on different levels of careless responses.

When 5% of careless responses were included, a one-factor solution fit the data fairly well for all sample sizes. When 10% of careless responses were in the dataset, model fit decreased for the one-factor solution, and model fit indexes suggested researchers should explore alternative models. However, a two-factor model fit the data equally well for all sample sizes. When 20% of careless responses were in the dataset, the one-factor model performed poorly, but the two-factor model fit the data well.

1.3.4 Does the Negatively-Worded Item Effect Occur in Personality? The literature on the negatively-worded item effect in personality scales is much smaller (as compared to the literature on RSES) and the results are inconsistent. Some studies demonstrated the effect associate with negatively worded items (Campbell, Siegman, & Rees, 1967; Sliter & Zickar, 2014) but others failed to observe it (Jackson & Lay, 1969; Trott & Jackson, 1967; Holden, Keen & Jackson, 1985).

One reason to doubt the presence of a negatively-worded item effect in personality scales is the inherent bipolarity of personality dimensions so the concept of negatively-worded is relative. It seems unlikely that two researchers both using the same scale, one as *Neuroticism* and the other as *Emotional Stability*, could both find a negatively-worded item effect, because they can choose to use either polar as they intend. It seems more likely that items favoring one pole would work better consistently. If so, this may represent a social desirability effect. For example, respondents may be less willing to admit to being unconscientious, disagreeable, and close-minded, etc.

Another possibility is that the negative-wording effect is caused by inattention (Schmitt & Stults, 1985; Roszkowski & Soven, 2010), in which case personality scales that are short or where the participants are motivated to pay attention would not exhibit

the effect while longer surveys or surveys used with inattentive samples would exhibit this effect, because of the proportion of careless responses in the data.

One older study that showed distinct direction-of-wording effect in the California F Scale measure of authoritarian personality trends (Campbell et al., 1967). That is, positively and negatively worded items tended to have different properties. The authors compared the correlations between F scale, F reverse scales, Ethnocentrism (E) scale and Manifest Anxiety (MA) Scale. They found that F scale presented distinct direction-of-wording effect, according to disparate correlation patterns with other scales. Specifically, the correlations between F scale with rest dimensions were substantially larger than those between F reversal score with rest dimensions. Similarly, examination of scales form Minnesota Multiphasic Personality Inventory (MMPI) suggested "a smaller but generally consistent direction-of-wording effect".

A more recent study to find clear item wording effects was recently published by Sliter and Zickar (2014). Using IRT analysis these researchers used two different personality inventories to test the assumption that negatively worded items do not have equal psychometric properties as positively worded items under IRT framework. In the first study, pairs of words from the 100-item Goldberg Adjective Checklist (Goldberg, 1992) were analyzed. The pairs have opposite meanings because of the addition of the prefix ''un-", like happy-unhappy. Each personality dimension contains 20 items, like ''kind'' (positively worded *Agreeableness*) and ''unimaginative'' (negatively worded *Openness*). Results showed that item discrimination and item information of negatively worded items were substantially lower than those of positively worded items. The difference on item discrimination ranged from 0.85 to 3.77 (with Mean=1.34 and

SD=1.28) between positively and negatively word items. The model did not fit if item parameters estimates were fixed as equal for those items. However, mean item responses of those items were not as different as item discrimination parameters. Study 2 reached similar but weaker conclusions by administering 100 items from the International Personality Item Pool (IPIP; Goldberg, 1999). Unlike the adjective checklist, the IPIP items were more representative of typical personality scales containing both positively and negatively-worded items. Example items are "I am always prepared." (positively worded *Conscientiousness*) and "I am not interested in abstract ideas." (negatively worded *Openness*). Item discrimination and item information were lower for negatively worded items, although the difference was not as dramatic as study 1. The trend was more obvious in *Agreeableness* and *Extraversion* than *Emotional Stability*. Model fit of four subscales except emotional stability (which contains 16 negatively worded items out of 20 total items) improved after removing all negatively worded items.

Personality surveys routinely use a mixed format (Hinkin, 1995) and other research on personality measures has not detected the existence of the negatively-worded item effect. Jackson and Lay (1968) created a personality inventory with four types of items in six content domains (PRF item pool labeled Play, Social recognition, Exhibition, Cognitive structure, and Autonomy): Six positive keying statements (P), six corresponding reverse keying positively-stated reversals (R), six reverse keying negations of the six positively-stated items (Pn), and six positive keying negations of the six positively-stated reversals (Rn). For example, "P- I try to be the life of the party; Pn-I do not try to be the life of the party; R-At a party I tend to stay in the background; Rn-At a party I do not tend to stay in the background." Based on correlations and factor analysis,

they concluded that direction-of-wording execute no impact on content dimensions and that negatively- and positively-worded items exhibited similar patterns of correlations. Content dimensions were clear, unique and orthogonal. Similarly, Trott and Jackson (1967) argued that reverse-coded items that were selected based on high content saturation (high biserial correlations) did not present the direction-of-wording effect. Those items only showed loadings on the appropriate bipolar content factors. Research using the PRF (Holden et al., 1985) showed that negated items exhibited inferior properties, such as lower validity, but that no negatively-worded item effect was found.

1.3.5 Conclusions. Based on the literature review, several conclusions may be reached regarding the negatively-worded item effect. First, only a few studies examined the negatively-worded item effect in personality and they have failed to provide consistent evidence that personality items suffer from the negatively-worded item effect. Unlike RSES, there is no well-established literature on whether the method effect associated with negatively worded items exists in personality tests.

Second, most research on this topic has not distinguished negatively worded items from negated items. A majority of studies operationalized negatively worded items as reverse coded items. However, some studies mixed up negatively worded items with negated items. For instance, Sliter and Zickar (2014) used paired words from adjective checklist. They listed examples such as "happy-unhappy", which basically implied negatively worded items as the same as negated items. The misunderstanding of the concept might lead to misinterpretation, which casts doubt on the conclusions.

Studies have shown that negated items did not present decent psychometric properties, compared with other items (Holden et al., 1985; Holden & Fekken, 1990).

Holden and Fekken (1990) examined the performance of three types of negated items based on the Basic Personality Inventory (BPI): (1) Use word "not" (e.g., *not, -n't, never*); (2) implicit negatives (e.g., negative prefixes such as *im-, un-, in-*); and (3) negative qualifiers (e.g., *seldom and rarely*). Results suggested that not items had significantly negative correlations with item stability and composed goodness. However, the other two types of negated items did not have any statistically significant relationships with any external criterion (such as criterion validity, content saturation, item stability and composed goodness, which was defined a sum of standardized scores regarding previous three criteria). However, total scores of all negated items were negatively correlated with criterion validity, item stability and composite goodness. They stated that negatives in items should be avoided in scale construction, because they might confuse respondents and introduce systematic errors.

Third, social desirability (SD) was introduced in the negatively-worded item effect study simply as an external criterion (DiStefano et al., 2009; Quilty et al., 2006), but there was no attempt to examine whether SD could be a possible cause of the negatively-worded item effect. Social desirability was defined as the tendency to distort responses so that they are more desired. Two types of response distortion were identified: One is intentionally distorted, known as impression management. That is, the individuals tend to manage their impression by responding to the items in the direction, which makes them "look good". The other is unintentional distortion, namely self-deception/self-awareness. That is, the individuals distort their responses without conscientious awareness (Paulhus, 1984; Ellingson, Sackett, & Hough, 1999).

Scholars have conducted considerable research on the effects of social desirability in personality. It is a common concern when personality inventories are developed and administered. Morgeson and colleagues (2007) pointed out that response distortion cannot be avoided in personality. Personality tests are easy to fake or likely to affected by social desirability. Study also suggested the presence of SD may impair selection decisions and selection processes (Landers, Sackett, & Tuzinski, 2011).

DiStefano and Motl (2009a) showed that SD correlated with the negatively-worded item effect in a female group. In the study, they used a SD measure as an independent variable, which assumes SD is a stable trait. However, SD may be rooted in the personality dimensions. For instance, it was suggested that some personality scales are more socially desired than others (Ones, Viswesvaran, & Reiss, 1996). But there is no previous research that investigated the relationship between SD and personality dimensions on the negatively-worded item effect.

All in all, because there are mixed findings on whether the negatively-worded item effect occurs in personality, and some potential moderating factors on the negatively-worded item effect have not been examined, meta-analysis is proposed to answer these questions. Most studies on the negatively-worded item effect in personality are dated, which did not employ the most sophisticated techniques. However, considerable amount of research has been conducted on evaluate personality inventories over years, which provides a great source of applying meta-analysis regarding the effect. Meta-analysis is the appropriate approach to evaluate the stationary and stability of the negatively-worded item effect across situations as well as to explore potential factors that impact the effect. Meta-analysis will answer whether the negatively-worded item effect is

consistent across studies by utilizing previous results. Meta-analysis also provides an effective and efficient method to detect potential moderators on the effect.

The part included a description of inconsistent findings of the negatively-worded item effect in RSES and personality, common issues associated with the studies on the topic, and justification of proposing a meta-analysis for the topic. The next section will introduce how to conduct meta-analysis step-by-step.

1.4 Literature on Meta-Analysis Methodology

Meta-analysis has become a popular method that can be applied to combine existing research, estimate accurate descriptive statistics, explain inconsistent findings and discover moderators or mediators on the same topic (Rosenthal & DiMatteo, 2001). There are three main goals to conduct meta-analysis: (1) to test whether the study results are consistent across situations (2) to obtain a global index of effect size as well as confidence interval and statistical significance (3) to identify possible moderators (Huedo-Median, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Thus, meta-analysis is a suitable methodology to understand the negatively-worded item effect in personality.

1.4.1 Meta-Analysis Procedures. As a methodological framework, meta-analysis is usually followed by the steps below: (1) Select research topic; (2) Conduct literature research; (3) Code studies based on characteristics; (4) Compute effect size; (5) Analyze and interpret data; (6) Publication (Cooper, 2010; Sánchez-Meca, & Marín-Martínez, 2010). The first step of meta-analysis is to choose a clear, objective and specific research question. Meanwhile, the concepts and constructs related to the topic should be defined and operationalized.

After the research question is determined, a literature search needs to be conducted. Formulating selection criteria is essential for the search. Characteristics of studies, such as sample features, study designs, reported statistics, and publication period, should be specified during the process. However, publication bias, also known as the "file drawer" effect, is a common issue for meta-analysis. Thus, unpublished data should be included in the meta-analysis (See detailed discussion of correcting "file drawer" in Rosenthal, 1979; McDaniel, Rothstein, & Whetzel, 2006).

Once adequate articles are collected, the next step is to code the characteristics of studies for further analysis. The characteristics of the studies can be categorized into substantive, methodological and extrinsic variables (Sánchez-Meca et al., 2010). Substantive variables refer to the factors that related to the research question, while methodology variables refer to the features associated with study design. Extrinsic variables are influential factors besides substantive and methodological variables. For instance, country of the sample can be an extrinsic variable. As subjective judgments are made during this phase, two or more coders should work independently on the studies. Then, reliability within a coder and agreement among coders should be analyzed (e.g., Kappa coefficients; Cooper, 2010; Sánchez-Meca et al., 2010).

During the coding, an effect-size index should be calculated so that effect sizes can be compared and accumulated across studies. Effect size refers to an index of qualifying a relationship between two variables or a difference between two groups (Borenstein, Hedges, Higgins, & Rothstein, 2009). Thus, correlation coefficient (r) and standardized mean difference (d) are two common statistics of effect size. Due to the

scope of this study, the following discussion will only focus on standardized mean different (d) on independent groups.

1.4.2 Cohen's d and Hedge's g. The standard mean difference (d), also known as Cohen's d, can be calculated by:

$$d = \frac{(M_1 - M_2)}{S_{within}} \tag{1.1}$$

where M_1 and M_2 are the sample means in two groups, and S_{within} is the within groups standard deviation, pooled across groups.

$$S_{within} = \frac{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}{\sqrt{n_1+n_2-2}} \tag{1.2}$$

where n_1 and n_2 are the sample sizes in the two groups, and S_1 and S_2 are the standard deviations in the two groups.

Variance of the standardized mean difference can be computed by:

$$V_d = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)} \tag{1.3}$$

in which n_1 and n_2 are the sample sizes in the two groups, and d is the standard mean difference.

So standard error of d is equal to the square root of V_d:

$$SE_d = \sqrt{V_d} \tag{1.4}$$

However, d tends to overestimate the absolute value of δ (which represents the standardized mean difference parameter or population standardized mean difference; (Borenstein et al., 2010) in small samples. Hedge's g was proposed to correct the bias due

to sample size (Hedges, 1981). To convert from *d* to *g*, a correction *J* will be calculated based on the formula:

$$J = 1 - \frac{3}{4df - 1} \tag{1.5}$$

where *df* is the degrees of freedom used to estimate S_{within}, n_1+n_2-2.

Then, *g* equals to the product of J and *d* as below:

$$g = J \times d \tag{1.6}$$

Variance of *g* is the product of squared *J* and variance of *d*:

$$V_g = J^2 \times V_d \tag{1.7}$$

Standard error of *g* is the square root of the variance of *g*:

$$SE_g = \sqrt{V_g} \tag{1.8}$$

The correction factor (*J*) is always smaller than 1.0 according to its definition. Thus, *g* is always less than *d*, and the variance of *g* is always less than the variance of *d*. However, *J* is close to 1.0 for all but the smallest sample size (e.g., J=0.958 and n_1-n_2=10), the differences between *d* and *g* are subtle (Hedges, 1981).

Factors, such as sample size and study design, affect precision of effect size. For instance, the larger the sample size, the more accurate the effect size. Also, matched groups yield more precise estimates and clustered groups yield less precise estimates (Borenstein et al., 2010).

1.4.3 Fixed-Effect and Random-Effects Model. When effect sizes and sampling variances were calculated for all studies, a full dataset should be formed by coded characteristics (potential moderators), effect size, and variance. Then, statistical analysis is conducted based on three goals: (1) to calculate an average effect size (across all studies) and its confidence interval; (2) to evaluate homogeneity/heterogeneity of the effect sizes around the average; (3) to search for moderators that may explain the heterogeneity (Sutton & Higgins, 2008).

Calculation of an average effect size needs to be performed under either a fixed-effect or random-effects model. Fixed-effect model assumes that the true effect size is the same across all studies. The only variability of effect sizes is due to sampling error, and study weights are assigned in order to reduce the within-study error. Fixed-effect model is appropriate if two conditions are met: (1) All studies in the analysis are believed functionally identical; (2) The goal of meta-analysis is to compute the common effect size for the known population (Borenstein et al., 2009)

However, random-effects model allows that the true effect sizes differ across studies. The studies included in the meta-analysis are a random sample of studies that have been observed (Borenstein et al., 2009). Random-effects model fits better when a common effect size is not assumed and the goal is to estimate population effect based on observed effects. Because factors (more likely moderators) that influence the results sometimes may not be included in the meta-analysis, it causes the differences among observed effect sizes.

1.4.3.1 Fixed-Effect Model. A summary effect size can be computed by weighting each observed effect sizes. So it is important to assign weight to effect size of each study (Borenstein et al., 2009). Under a fixed-effect model, the weight is given by:

$$W_i = \frac{1}{V_{Y_i}} \tag{1.9}$$

where V_{Yi} is the within-study variance for study I, which is the same variance estimated in the formula 1.7.

The weighted mean (M) can be computed by:

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \tag{1.10}$$

Variance of the summary effect is the reciprocal of summed weights:

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i}, \tag{1.11}$$

and standard error is the square root of the variance,

$$SE_M = \sqrt{(V_M)} \tag{1.12}$$

Then, 95% lower and upper limits for the summary effect are:

$$LL_M = M - 1.96 \times SE_M, \tag{1.13}$$

and

$$UL_M = M + 1.96 \times SE_M \tag{1.14}$$

Finally, a Z-value to test the null hypothesis (the true effect size is the same across all studies) can be computed by:

$$Z = \frac{M}{SE_M} \tag{1.15}$$

For a one-tailed test the *p*-value is equal to:

$$p = 1 - \Phi(\pm|Z|) \tag{1.16}$$

where "+" is chosen if the direction of difference is as expected, and "–" is used otherwise.

For a two-tailed test the *p*-value is given by:

$$p = 2[1 - \Phi(|Z|)] \tag{1.17}$$

where Φ(Z) is the standard normal cumulative distribution (Borenstein et al., 2009).

1.4.3.2 Random-Effects Model. In order to calculate the variance of a study under random-effects model, the within study variance and between study variance (τ^2) have to be calculated. Between-studies variance (τ^2) can be computed as following:

$$T^2 = \frac{Q - df}{C} \tag{1.19}$$

where

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i\right)^2}{\sum_{i=1}^{k} W_i}, \tag{1.20}$$

$$df = k - 1, \tag{1.21}$$

in which k is the number of studies, and

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i} \tag{1.22}$$

A summary effect will be computed by combining weighted all effect sizes. The weights M^* assigned to each study is the inverse of its variance V^*_{Yi}:

$$W_i^* = \frac{1}{V_{Y_i}^*} \tag{1.23}$$

in which V^* is the sum of within-study variance for study I and the between-studies variance T^2. That is,

$$V_i^* = V_{Y_i} + T^2 \tag{1.24}$$

The weighted mean M^* can be computed by:

$$M^* = \frac{\sum_{i=1}^{k} W_i^* Y_i^*}{\sum_{i=1}^{k} W_i^*} \tag{1.25}$$

The variance of summary effect is estimated as:

$$V_M^* = \frac{1}{\sum_{i=1}^{k} W_i^*} \tag{1.26}$$

The estimated standard error of summary effect is then the square root of the variance:

$$SE_M^* = \sqrt{(V_M^*)} \tag{1.27}$$

The 95% lower and upper limits for the summary effect would be computed as:

$$LL_M^* = M^* - 1.96 \times SE_M^*, \tag{1.28}$$

and

$$UL_M^* = M^* + 1.96 \times SE_M^* \tag{1.29}$$

Finally, a *Z*-value should be computed to test whether the mean effect size is zero:

$$Z^* = \frac{M^*}{SE_M^*} \tag{1.30}$$

For a one-tailed test, the *p*-value is given by:

$$p^* = 1 - \Phi(\pm|Z^*|) \tag{1.31}$$

While for a two-tailed test, the *p*-value is given by:

$$p^* = 2[1 - \Phi(|Z^*|)] \tag{1.32}$$

where ϕ(Z) is the standard normal cumulative distribution (Borenstein et al., 2009).

1.4.4 Heterogeneity. One important goal of meta-analysis is to identify meaningful patterns in accumulated results. That is, explain variation around observed effect sizes (Borenstein et al., 2009). Heterogeneity was introduced to represent the degree to which the effect sizes have variability. When there is no heterogeneity, observed effect sizes should fall within some range of the common effect; however, if there is substantial heterogeneity, variability due to the true effect size (due to τ, between-studies variance) and variability due to different effect sizes should be distinguished. Thus, moderator analysis should be conducted if a large amount of heterogeneity is observed.

Q statistics (Cochran, 1954) and I^2 statistic (Higgins & Thompson, 2002) are common methods to estimate heterogeneity. Q statistics is typically treated as a significance test, comparing Q statistics to a Chi-square distribution with degree of freedom of k-1. Q statistics can be computed by any of the following three formulas (Borenstein et al., 2009):

$$Q = \sum_{i=1}^{k} W_i(Y_i - M)^2 , \tag{1.33}$$

or

$$Q = \sum_{i=1}^{k} \left(\frac{Y_i - M}{S_i}\right)^2 , \tag{1.34}$$

or

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i\right)^2}{\sum_{i=1}^{k} W_i} \quad (1.35)$$

where W_i is the study weight (which refers to the fixed-effect mode), Y_i is the study effect size, M is the summary effect and k is the number of studies.

I^2 can be estimated by (Borenstein et al., 2009):

$$I^2 = \left(\frac{Q-df}{Q}\right) \times 100\%, \quad (1.36)$$

or

$$I^2 = \left(\frac{Variance_{bet}}{Variance_{total}}\right) \times 100\% = \left(\frac{\tau^2}{\tau^2+V_Y}\right) \times 100\% \quad (1.37)$$

It was suggested that Q test and I^2 provides different information. Q test focuses on testing whether the between-study variance is larger than expected due to chance in the data, but I^2 checks the degree of heterogeneity. I^2 can be a complement of Q test, but both of them lack power when the sample size of studies is small. (Huedo-Medina, Sánchez-Meca, Marin-Martinez, & Botella, 2006)

However, it is worth noting that degree of heterogeneity has complex relationships with choice of random-effects or fixed-effect model. Which model to use should depend on the understanding of whether there is a common effect size among studies. Test of heterogeneity attempts to examine whether between-studies variance is zero, which does not directly relate with the assumption of a true effect. Moreover, the heterogeneity test suffers from low power (Borenstein et al., 2009). It was recommended: If random-effects model is set up and heterogeneity test is not significant, random-effects analysis automatically changes into fixed-effect analysis, as T^2 is zero. However, if fixed-

effect model is used but heterogeneity test is significant, it is important to revisit the assumption about fixed-effect model (Borenstein et al., 2009).

1.4.5 Moderator Analysis. Meta-analysis framework also allows scholars to examine whether moderators exists. Subgroup analysis and meta-regression are two major approaches, both of which were used for this study, because expected moderators involved both categorical and continuous variables.

1.4.5.1 Subgroup Analysis. For moderator analysis (comparing A versus B) under random-effects model, weighted sum of squared deviation (SS) of all A studies about the mean of A (as Q^*_A) and SS of all B studies about the mean B (as Q^*_B) will be calculated. They can be computed based on formula 1.33 to 1.35.

$$Q^*_{within} = \sum_{j=1}^{p} Q^*_j = Q^*_A + Q^*_B \tag{1.38}$$

where Q^*_{within} is the within-group weighted sum of squared deviations (SS).

Q^*_{bet} is the weighted SS of the subgroup means about the grand mean, which is given by

$$Q^*_{bet} = Q^* - Q^*_{within} \tag{1.39}$$

in which Q^* is the weighted SS of all effects about the grand mean.

Meanwhile, *p*-value for Q is equal to CHIDIST(Q, *df*), which follows a chi-square distribution (Borenstein et al., 2009).

1.4.5.2 Robust Variance Estimation in Meta-Regression. Meta-regression is another approach to examine whether a variable moderate the outcome (See details in Borenstein et al., 2009). One crucial difference is that subgroup is recommended when the targeted

moderator is categorical, while meta-analysis is suggested when the targeted moderator is continuous (or interval; Borenstein et al., 2009).

The usage of meta-regression aimed to address another issues with personality data. As a personality measure evaluates more than one dimension, data from the measure are likely to contribute more than one effect sizes. However, those effect sizes are dependent, which violated the independence assumption of meta-analysis. Thus, Hedges, Tipton and Johnson (2010) proposed robust variance estimation (RVE) to handle dependent effect sizes in meta-regression. The following section will introduce the basic computations of RVE.

The model for RVE estimation was as below:

$$T=X\beta+\varepsilon, \tag{1.40}$$

where T is a vector of m vectors, each with k_j effect size estimation, X is a design matrix of m stacked matrices, each of dimension $k_j \times p$, β is a $p \times 1$ vector of regression coefficients, and ε_{ij} is the sampling error.

A hierarchical model of dependent effect size for study j can be written

$$\Sigma_{aj}= \tau^2 J_j+\omega^2 I_j+V_j, \tag{1.41}$$

where τ^2 is the variation in study-average effect sizes across studies, ω^2 is the within-study variation in true effect sizes, I_j is a $k_i \times k_j$ identify matrix, J_j is $k_i \times k_j$ matrix of 1s, and V_j is a $k_i \times k_j$ diagonal matrix of the estimation error variances in study j.

1.5 Current Study

The personality literature fails to reach conclusions on whether the negatively-worded item effect exists. This study was designed to meta-analyze personality data to

examine whether the negatively-worded item effect is present in personality inventories, how different the effect manifests in personality and non-personality measures, and to explore some possible moderators of the negatively-worded item effect. Also, to help interpret these findings, this study also examined the psychometric difference between negated and non-negated items in personality assessments.

The results will provide evidence on the psychometric qualities of different item types in personality, like negatively versus positively worded items, and negated versus non-negated items. The findings will also improve our understanding on whether the negatively-worded item effect occurs in assessments. If the negatively-worded item effect occurs for personality measures, it will cause problems for reliability and validity that can easily be solved by omitting negatively-worded items. If, however, the negatively-worded item effect does not occur for personality measures, then (a) telling people to avoid these items needlessly eliminates quality items, and (b) it's also interesting why the negatively-worded item effect does not occurs for personality but it has been observed often in non-personality inventories.

1.5.1 Research Questions. No consistent conclusion has been shown on whether the negatively-worded item effect exhibits in personality. Thus, research question 1 is proposed:

RQ1: *Does the negatively-worded item effect exist in personality tests*?

Research question 2 addresses the magnitude of any such effect, if any:

RQ2: *What is the magnitude of the difference between negatively and positively worded items on item discrimination (or factor loadings)*?

Specifically, a summary effect size of CFA item loadings and IRT item discrimination on negatively and positively worded items will be calculated under meta-analysis of personality measures.

As previous research mainly focuses on the model fit to investigate the negatively worded item effect, they did not place emphasis on the qualities of items. Meanwhile, mean of responses and scale reliability has been disparate due to item wording (Schmitt & Allik, 2005). However, scrutinizing on CFA item loadings and IRT item discrimination provides a brand new perspective to examine the negatively worded item effect on item-level, which is straightforward and direct for addressing the issue. Therefore, the study will analyze item loadings and IRT parameters on negatively and positively worded items to investigate the method effect associated with items.

Research has suggested that each personality dimension has its own features, and various dimensions function differently (Barrick & Mount, 1991). For instance, *Openness to experiences* was the last factor incorporated into FFM (Barrick et al., 1991), as it was difficult to define and detect. Another example, personality dimensions functioned differently regarding impression management (DeYoung, Peterson, & Higgins, 2002). Therefore, the negatively-worded item effect may be more obvious/stronger in some factors compared to others (Sliter & Zickar, 2014). Personality dimensions would moderate the negatively-worded item effect. Thus, research question 3 is:

RQ3: *Does the negatively-worded item effect occur in each personality dimension*?

1.5.2 Hypotheses. Because Holden and colleagues (1985; 1990) argued that items with negation presented inferior properties compared with positively and negatively worded

items, and direction-of-wording did not actually affect items' performance. Therefore, items with negation would present lower item characteristics from items with non-negation, due to their high demanding on cognitive processing resources. We hypothesize:

H1: *Negated items will tend to have smaller item discrimination and item loadings than non-negated items in personality measures.*

According to previous research (Schmitt & Stults, 1985; Woods, 2006), the proportion of careless responses related with emergence of the negatively-worded item effect. In practice, careless responses could result from lack of motivation. Individuals with high motivation are less likely to generate careless responses, while those with low motivation are more likely to produce careless responses. Thus, we hypothesize:

H2: *Low-motivation samples will produce larger negatively-worded item effects than high-motivation samples.*

CHAPTER 2

METHOD

2.1 Sample

Google Scholar and PsychINFO were searched for two sets of keywords: First, "negatively-worded item effect" or "reverse-coded effect" or "direction-of-wording" and "personality". This set of keywords located articles that were directly involved in the negatively-worded item effect for personality and non-personality measures (e.g. RSES). Second, "Item Response Theory" (or "IRT") and "factor analysis" (or "CFA") and "personality" were used as well. The results of this search were carefully scrutinized, because they related to the negatively-worded item effect indirectly. The articles that provided enough information were retained.

Next, top-tier journals (*Journal of Personality and Social Psychology*, *Journal of Applied Psychology*, *Personality and Individual Difference,* and *Psychological Methods*) were examined manually. After briefly reviewing all articles (e.g., titles, keywords, and/or abstracts) from the year 2005 to 2014, we identified those on personality, and retained the ones that provided enough information for meta-analysis on the negatively-worded item effect. The ten-year time period was specified for two reasons. First, some time period has to be chosen, and IRT and CFA studies seem more likely in most recent 10 years. Also, we contacted researchers whose studies did not include enough statistics, and we judged it is unlikely that the information was retained from more than 10 years.

We kept the studies that reported statistics on positively and negatively worded items (e.g., mean and standard deviation of positively and negatively worded items, if

they focused on the item wording effect directly), and/or that included item discrimination or factor loadings of all items and contents of each personality item or names of personality measures. We also added unpublished datasets on personality measures, which were obtained by literature research and by privately contacting other researchers.

In this way, we discovered 245 studies in total. After removing the ones that failed to provide item-level information (no adequate statistics for meta-analysis, N=129), unfit samples (not adult working sample, N=22), irrelevant constructs (personality traits do not fit FFM, N=10), inappropriate measures (clinical personality measures, N=15; see discussion in 2.2.1), 77 studies were retained for analysis. However, some studies contributed more than one effect sizes, because they involved multiple samples or multiple personality traits. Among the cases, thirteen studies were unpublished (two of them are dissertation, two are conference paper, and others were obtained by personal contact). For all 77 studies, we analyzed 42 datasets by applying IRT, because we obtained the original datasets. More details regarding the studies involved are presented in Table 17 of Appendix B.

2.2 Procedures

2.2.1 Coding Procedures. After selecting articles, coding was conducted by the author and another two coders. According to research questions and hypotheses, a study was coded mainly based on three aspects:

(1) Did it use a personality measure or a non-personality measure or both? Personality inventories refer to the established non-clinical scales that aimed to evaluate

Five-Factor personality model. Personality assessments, which did not follow the Five-Factor Model or which were developed to provide clinical diagnosis, were excluded from the analysis, because consumers who took clinical personality measures might have different responding process from individuals who took non-clinical personality assessments. For certain personality measures, some dimensions fit the FFM but not all. Those that fit the FFM were included, while others were excluded. Take Eysenck Personality Question (EPQ) as an example. E and Q scales were retained, but P and L scales were removed. While self-reported measures, which were designed to evaluate non-personality and non-clinical psychological traits were considered as non-personality measures. However, cognitive ability tests were not involved in the study, because the underlying function of cognitive ability and non-cognitive ability can be fundamentally different. If a personality inventory was involved, the article was coded as personality; if a non-personality measure was included, it was coded as non-personality; if the study contained both personality and non-personality assessments, it was coded as both personality and non-personality. Meanwhile, names of inventory were recorded. For personality measures, construct of each dimension was recorded as well.

(2) What type does a sample belong to, like students, employees, applicants, or general population? We focused on general population of working adults, so studies that involved in special samples were excluded, such as preschoolers, seniors or clinical consumers. Meanwhile, item statistics need to be driven from the whole sample, so the information on a particular subgroup of working population was not included in the study. For instance, analysis conducted solely on male or female group was removed,

unless it was proven that item statistics was equivalent for the samples or the statistics can be driven for the combined whole sample.

(3) What type of motivation reward is used in the study, such as research, class credit, selection/promotion, application, or personal reports? Criterion (2) and (3) combined for coding motivation. For instance, if the sample was collected from general population under research settings with no or low stake, it was coded as a low-motivation sample; if the sample was from applicants under selection or general population who attempted to obtain rewarded (e.g., personality reports), it was coded as a high-motivation sample.

(4) Is an item negatively or positively worded items, and items with negation or not? For the purpose of the study, each item in the study was coded. Negatively worded items are referred to as the items that are negative-keyed, and positively worded items are referred to as the items that are positive-keyed (Schmitt & Stults, 1985). It is worth noting that personality dimensions should be always consistent across all personality tests. Specifically, *Emotional Stability* was used instead of *Neuroticism*. If analysis in a study was conducted based on *Neuroticism* scoring, negatively and positively worded items were swapped, so that it reflected *Emotional Stability* dimension. Then, further computation was conducted accordingly.

Negated items are the items with not (e.g., *not*, *-n't*, *never*), or implicit negative (e.g., negative prefixes as *im-*, *un-*, *in-*), or negative qualifiers (e.g., *seldom* or *rarely;* Holden et al., 1985; 1990). However, negated items can be either positively or negatively worded items. Non-negated items are the items that are not negated.

Besides all criteria above, an article also was corded based on whether the construct was swapped (e.g., all *Neuroticism* items need to be coded as *Emotional Stability*), what IRT or CFA model was used. Appendix C shows the coding sheet used.

When the coding finished, reliability of each coder and agreement between coders were computed, Cohen's *Kappa*=0.54 across all variables. Where there were disparities between coders, consensus was reached before further analysis.

2.2.2 CFA and IRT. Because all research questions and hypotheses are based on item discrimination from IRT or factor loadings from CFA, item response theory (IRT) analysis was conducted on the unanalyzed datasets. Otherwise, item discrimination or factor loadings of each item were retrieved from the sampled studies. CFA factor loadings were only acquired from simple-structure model (a one-dimension solution). The loadings that were not obtained based on a single factor model of each personality dimension were excluded in the analysis.

IRT and CFA are two common approaches to understand relationships between items and the underlying psychological construct (Raju, Laffitte, & Byrne, 2002). Although they differ on assumptions, modeling, and computations, item discrimination produced from IRT estimation and factor loadings generated from CFA still share some similarities. Factor loadings represent the strength of relationships between items and the psychological construct measured by them. That is, the larger the values of factor loadings, the stronger relationships the items have with the construct. Meanwhile, item discrimination is proportion to the item's common factor loading. Thus, large values of item discrimination and factor loadings imply items with low measurement errors. And

item discrimination and factor loadings of the same items on the same dataset would have moderate to strong correlations.

Even though item location is an important element of IRT, we preferred item discrimination to item location. Item location demonstrated the difficulty of an item, rather than the relationship between the item and the underlying construct, which is a less intuitive index to evaluate the negatively-worded item effect. Meanwhile, item discrimination and factor loadings shared certain similarities, which allowed us to analyze item quality, even though item discrimination and factor loadings are not on the same metric. However, the calculation of effect size makes sure that the scale differences are removed, so it is appropriate to combine effect sizes from both IRT and CFA studies.

2.2.3 Computation Procedures. In order to avoid redundancy and confusion, this section will only use item discrimination as a reference for both item discrimination and factor loadings. Effect sizes were calculated for item discrimination difference between negatively and positively worded items on each personality dimension.

For each effect size, item discrimination difference is equal to mean of item discrimination on negatively worded items (M_N) minus mean of item discrimination on positively worded items (M_P) divided by pooled standard deviation of positively and negatively worded items. That is,

$$d = \frac{M_N - M_P}{SD}$$

Even though control group (like positively worded items) standard deviation (SD) can be used when homogeneity of variance is not supported by the data, as it is difficult to determine which SD is the control group, therefore, pooled SD were applied for the calculation. Similar justification and application were used for the computation of

all analysis in the study. It is worth noting that negative value of d indicates that negatively worded items have inferior quality, while positive value of d indicates that negatively worded items have superior quality.

Effect sizes were computed for item discrimination difference between negated and non-negated items on each personality dimension. For each effect size, item discrimination difference is equal to mean of item discrimination on negated items (M_t) minus mean of item discrimination on non-negated items (M_s) divided by pooled standard deviation of all negated and non-negated items. That is,

$$d = \frac{M_t - M_s}{SD}$$

Because Cohen's d is biased in most situations, Hedges' g was computed according to formula 1.3 to 1.5. Based on formula 1.19 to 1.32, within-study and between-studies variance, and weight of each effect size can be computed, and then a summary effect size was calculated as well as its 95% confidence intervals. Confidence interval provides the variability around estimated mean effect size due to sampling error. The procedures described above were applied to compute summary effect sizes and 95% confidence intervals for the current study.

For personality data (regarding *H1*, and *H2*), Robust Variance Estimation (RVE) in meta-regression was applied to examine moderator effect due to dependent effect sizes. But if dependent effect sizes were not an issue, subgroup analysis by Q-statistics was applied for detecting moderation.

2.3 Software

R, an open-source statistical software was used to conduct all analysis. During the process, sqldf (Grothendieck, 2012), Himsc (Harrell, 2008), MAd (Del Re & Hoyt, 2012), robumeta (Fisher & Tipton, 2010), and metafor (Viechtbauer, 2010) were involved in data manipulation, statistics computation (such as weighted mean, weighted standard deviation, Cohen's *d*, Hedges' *g*, and subgroup analysis), RVE in meta-regression and plotting (forest plots, trim-and-fill analysis, and funnel plots). Appendix F includes all the R code for this project.

CHAPTER 3

RESULTS

In order to answer research questions and hypotheses, effect sizes were calculated based on Section 2.2.3. After computing Hedges' g, frequency distributions of all effect sizes were examined. We noticed that majority of effect sizes ranged from -6 to 6. Thus, we considered outliers were effect sizes that are greater than 6 or smaller than -6. For personality measures, one effect size was larger than 6; for non-personality measures, one effect size was smaller than -6. All the following analysis was performed by including and excluding those outliers.

3.1 Research Question 1 and 2 Results

Research question 1 asked whether the negatively-worded item effect exists in personality tests. Thus, a summary effect size was estimated based on effect sizes of item discrimination difference between negatively and positively worded items. Effect size of each study involved were listed in Appendix C, which includes actual effect sizes of item discrimination differences between negatively and positively worded items in each personality dimension from all datasets. Effect sizes range from -5.11 to 10.33, and standard deviations of each effect size range from 0.0002 to 0.05.

A summary effect size of item discrimination difference were computed based on these effect sizes under random-effects model, as well as 95% confidence interval. Because one study may contribute more than one personality dimensions (for example, E and Q scales from EPQ, and OCEAN scales from IPIP50), some of effect sizes are

dependent. Thus, the summary effect size was calculated in two ways. First, an overall summary effect size was computed across all personality dimensions. As shown in Table 1, when *Emotional Stability* was involved in the calculation of a summary effect size, the negatively-worded item effect was very small, and not significant. However, when *Neuroticism* was used, the negatively-worded item effect became larger and statistically significant. Meanwhile, inclusion or exclusion of an outlier did not change the general trend, although it affected the magnitude of the summary effect sizes.

Table 1 .

A Summary Effect Size of the Effect across All Personality Dimensions

	k	Random Effects (*SE*)	*Z*	95%CI.l	95%CI.u	*p*	*Q*	*df.Q*	Q_p	I^2
1	144	-0.05 (*0.15*)	-0.295	-0.34	0.25	0.77	151826	143	0	99.9%
2	143	-0.12 (*0.12*)	-0.977	-0.35	0.12	0.33	94067	142	0	99.8%
3	144	-0.37 (*0.15*)	-2.546	-0.65	-0.09	0.01	138477	143	0	99.9%
4	143	-0.30 (*0.12*)	-2.494	-0.53	-0.06	0.01	92826	142	0	99.9%

Note. *k*=number of samples; Radom Effect= a summary effect size from the random effects meta-analysis; *Z*=standardized value of the summary effect size under normal distribution; 95% CI.l=95% confidence interval lower bound; 95% CI.u= 95% confidence interval upper bound; *p*=*p*-value under normal distribution; *Q*=chi-square test for homogeneity of observed studies; *df*= degree of freedom for *Q*-statistics; Q_p=*p*-value for *Q*-statistics; I^2=percentage of variance beyond sampling error.

Row 1 stands for a summary effect size across all personality dimensions including Emotional Stability; Row 2 represents a summary effect size across all personality dimensions including Emotional Stability, after removing outlier; Row 3 stands for a summary effect size across all personality dimensions including Neuroticism; and Row 4 represents a summary effect size across all personality dimensions including Neuroticism, after removing outlier.

Second, a summary effect size was calculated for each personality dimension in Table 2. Because not every study contributed to five personality dimensions, the number of studies in each dimension varied. *Agreeableness*, *Extraversion*, and *Openness* exhibited the negatively-worded item effect, but only the summary effect size of *Agreeableness* was statistically significant. For *Conscientiousness* and *Emotional*

Stability, there is a non-significant positively-worded item effect. Instead, item discrimination of negatively worded items was larger than that of positively worded items.

Table 2

A Summary Effect Size of the Effect on Each Personality Dimensions

	k	Random Effects (*SE*)	*Z*	95%CI.l	95%CI.u	*p*	*Q*	*df.Q*	Q_p	I^2
A	22	-0.63 (*0.28*)	-2.28	-1.16	-0.09	0.02	7893	21	0	99.7%
C	24	0.34 (*0.36*)	0.94	-0.37	1.05	0.35	18310	23	0	99.9%
E	59	-0.32 (*0.19*)	-1.69	-0.69	0.05	0.09	53140	58	0	99.9%
ES	18	1.30 (*1.01*)	1.29	-0.68	3.27	0.20	43505	17	0	99.96%
O	21	-0.25 (*0.19*)	-1.31	-0.62	0.12	0.19	2288	20	0	99.1%

Note. Dim=dimension of personality; *k*=number of samples; Radom Effect= a summary effect size from the random effects meta-analysis; *Z*=standardized value of the summary effect size under normal distribution; 95% CI.l=95% confidence interval lower bound; 95% CI.u= 95% confidence interval upper bound; *p*=*p*-value under normal distribution; *Q*=chi-square test for homogeneity of observed studies; *df*= degree of freedom for *Q*-statistics; Q_p=*p*-value for *Q*-statistics; I^2=percentage of variance beyond sampling error.

As *Neuroticism* in some studies was re-coded into *Emotional Stability*, a summary effect size of the negatively-worded item effect across all dimension was computed when all *Emotional Stability* items were coded as *Neuroticism*. The summary effect size was equal to -1.30 with a standard deviation of 1.01. However, when the outlier was removed, the summary effect size shrunk to -0.76 with a standard deviation of 0.30 (see details in Table 3).

Table 3

A Summary Effect Size of the Effect on Neuroticism after Removing an Outlier

k	Random Effects (*SE*)	*Z*	95%CI.l	95%CI.u	*p*	*Q*	*df.Q*	Q_p	I^2
17	-0.76 (*0.30*)	-2.50	-1.36	-0.16	0.01	3077	16	0	99%

Note. *k*=number of samples; Radom Effect= a summary effect size from the random effects meta-analysis; Z=standardized value of the summary effect size under normal distribution; 95% CI.l=95% confidence interval lower bound; 95% CI.u= 95% confidence interval upper bound; *p*=*p*-value under

normal distribution; Q=chi-square test for homogeneity of observed studies; df= degree of freedom for Q-statistics; Q_p=p-value for Q-statistics; I^2=percentage of variance beyond sampling error.

3.2 Research Question 3 Results

In order to answer research question 3 regarding whether the negatively-worded item effect occurs the same across personality dimensions, a meta-regression with robust variance estimation (RVE) was applied due to dependent effect sizes.

No model-level significant test has been developed yet, so it is impossible to evaluate whether meta-regression is statistically significant or not. However, two model-level statistics were provided (Fisher & Tipton, 2014): (1) τ^2 is the between-cluster variance component in the hierarchical effects model; (2) ω^2 is the between-studies-within-cluster variance component for the hierarchical effects meta-regression model.

For the meta-regression model, τ^2=0, and ω^2 =3.43. In Table 4, *Agreeableness* showed a statistically significant difference with *Conscientiousness* and *Emotional Stability* on the negatively-worded item effect. After removing an outlier, the same tendency keeps the same (τ^2=0.40, and ω^2 =1.67). However, the estimate of regression coefficient decreased from 1.92 into 1.39 in Table 5. In Tables 4 and 5, *Agreeableness* was always coded as the reference group. Regression coefficient estimates (except intercept) represented the difference between the mean of each personality dimension and the mean of the reference group. For example, regression coefficient of A vs. C equals to 0.97, which suggested that the mean difference between *Conscientiousness* and *Agreeableness* is 0.97 regarding the negatively worded item effect, and the difference is statistically significant. Therefore, *Conscientiousness* is more likely to show a positively-worded item effect rather than the negatively-worded item effect.

Table 4

Meta-Regression of Personality Dimensions by Robust Variance Estimation (RVE) with Emotional Stability

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u	Sig
Intercept	-0.63	0.27	-2.32	57	0.02	-1.17	-0.09	**
A vs. C	0.97	0.28	3.42	57	0.001	0.40	1.53	***
A vs. E	0.31	0.27	1.15	57	0.25	-0.23	0.84	
A vs. ES	1.92	0.75	2.57	57	0.013	0.43	3.42	**
A vs. O	0.38	0.20	1.88	57	0.065	-0.02	0.78	*

Note. A=Agreeableness, C=Conscientiousness, E=Extraversion, ES=Emotional Stability, and O=Openness. Estimate=Estimate of regression coefficient, SE=Standard Error, df=degree of freedom, p=p-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound.* stands for statistical significance at 0.1; ** stands for statistical significant at 0.05; and *** stands for statistical significance at 0.01.

Table 5

Meta-Regression of Personality Dimensions by Robust Variance Estimation (RVE) with Emotional Stability after Removing an Outlier

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u	Sig
Intercept	-0.63	0.27	-2.32	57	0.02	-1.17	-0.09	**
A vs. C	0.97	0.28	3.42	57	0.001	0.40	1.53	***
A vs. E	0.31	0.27	1.15	57	0.25	-0.23	0.84	
A vs. ES	1.39	0.45	3.08	57	0.003	0.48	2.29	***
A vs. O	0.38	0.20	1.88	57	0.07	-0.02	0.78	*

Note. A=Agreeableness, C=Conscientiousness, E=Extraversion, ES=Emotional Stability, and O=Openness. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound. * stands for statistical significance at 0.1; ** stands for statistical significant at 0.05; and *** stands for statistical significance at 0.01.

Meanwhile, the same analysis was performed when *Neuroticism* was considered rather than *Emotional Stability* (See details in Table 6 and Table 7; τ^2=0.50, and ω^2 =2.54). *Agreeableness* was still significantly different from *Conscientiousness* regarding the negatively-worded item effect. The difference between *Agreeableness* and *Neuroticism* was not statistically significance. When the outlier was removed, the conclusion was still the same (τ^2=0.016, and ω^2 =2.06).

Table 6

Meta-Regression of Personality Dimensions by Robust Variance Estimation (RVE) with Neuroticism

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.L	95% CI.U	Sig
Intercept	-0.67	0.27	-2.32	57	0.02	-1.17	-0.09	**
A vs. C	0.97	0.28	3.42	57	0.001	0.40	1.53	***
A vs. E	0.31	0.27	1.15	57	0.25	-0.23	0.84	
A vs. N	-0.67	0.67	-1.01	57	0.32	-2.00	0.66	
A vs. O	0.38	0.2	1.88	57	0.07	-0.02	0.78	*

Note. A=Agreeableness, C=Conscientiousness, E=Extraversion, N=Neuroticism, and O=Openness. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound. * stands for statistical significance at 0.1; ** stands for statistical significant at 0.05; and *** stands for statistical significance at 0.01.

Table 7

Meta-Regression of Personality Dimensions by Robust Variance Estimation (RVE) with Neuroticism after Removing One Outlier

	Estimate	*SE*	*t*-value	*df*	*P*	95% CI.l	95% CI.u	Sig
Intercept	-0.63	0.27	-2.32	57	0.02	-1.17	-0.09	**
A vs. C	0.97	0.28	3.42	57	0.001	0.40	1.539	***
A vs. E	0.31	0.27	1.15	57	0.25	-0.22	0.894	
A vs. N	-0.14	0.50	-0.27	57	0.79	-1.13	0.86	
A vs. O	0.38	0.20	1.88	57	0.07	-0.02	0.78	*

Note. A=Agreeableness, C=Conscientiousness, E=Extraversion, N=Neuroticism, and O=Openness. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound. * stands for statistical significance at 0.1; ** stands for statistical significant at 0.05; and *** stands for statistical significance at 0.01.

3.3 Hypothesis 1 Results

Hypothesis 1 suggested that negated items produce smaller item discrimination than non-negated items. A summary effect size of item discrimination difference was computed based on these effect sizes under random-effects model, as well as 95% confidence interval. Meanwhile, the value of the summary effect size represents the magnitude of difference between negated and non-negated items on item discrimination.

In order to examine the hypothesis, a summary effect size was calculated for the difference between negated and non-negated items regarding item discrimination. Similar to Research Question 1, the effect sizes were calculated at overall level and dimension level. The result of overall effect size was presented in Table 8, where the summary effect size was small and non-significant.

Table 8

A Summary Effect Size of Negation Effect across All Personality Dimensions

k	Random Effects (*SE*)	Z	95%CI.l	95%CI.u	*p*	*Q*	*df.Q*	Q_p	I^2
119	-0.08 (*0.06*)	-1.44	-0.20	0.03	0.15	13586	118	0	99.1%

Note. *k*=number of samples; Radom Effect= a summary effect size from the random effects meta-analysis; *Z*=standardized value of the summary effect size under normal distribution; 95% CI.l=95% confidence interval lower bound; 95% CI.u= 95% confidence interval upper bound; *p*=*p*-value under normal distribution; *Q*=chi-square test for homogeneity of observed studies; *df*= degree of freedom for *Q*-statistics; Q_p=*p*-value for *Q*-statistics; I^2=percentage of variance beyond sampling error.

Meanwhile, the difference between negated and non-negated items was examined on each personality dimension in Table 9. The summary effect sizes varied from -0.37 to 0.24, which were fairly small magnitudes; none were statistically significant. Therefore, the results failed to support Hypothesis 1; we found no evidence that negated items had smaller item discriminations than non-negated items.

Negation types were coded according to Holden and his colleagues (1985; 1990): (1) Not (e.g., *not, never, n't*); (2) Negative prefix (e.g., *im-*, *dis-*); and (3) Negative qualifier (e.g., *rarely*, *seldom*). Subgroup analysis was conducted to examine whether negation types made difference on item discrimination difference between negated and non-negated items (see details in Table 10). The summary effect size of negation types ranged from -0.11 to 0.10, and none of them were statistically significant. Examination and discussion of publication bias of this hypothesis will be discussed in Section 3.6.

Table 9

A Summary Effect Size of Negation Effect on Each Personality Dimension

	k	Random Effects (*SE*)	*Z*	95%CI.l	95%CI.u	*p*	*Q*	*df.Q*	Q_p	I^2
A	21	-0.37 (*0.29*)	-1.26	-0.94	0.22	0.21	4781	20	0	99.6%
C	5	0.24 (*0.20*)	1.18	-0.16	0.63	0.24	192	4	0	97.9%
E	18	-0.24 (*0.23*)	-1.02	-0.69	0.22	0.31	2172	17	0	99.2%
ES	57	0.04 (*0.05*)	0.74	-0.07	0.14	0.46	3976	56	0	98.6%
O	18	-0.08 (*0.22*)	-0.35	-0.52	0.35	0.73	2107	17	0	99.2%

Note. *k*=number of samples; Radom Effect= a summary effect size from the random effects meta-analysis; Z=standardized value of the summary effect size under normal distribution; 95% CI.l=95% confidence interval lower bound; 95% CI.u= 95% confidence interval upper bound; *p*=*p*-value under normal distribution; *Q*=chi-square test for homogeneity of observed studies; *df*= degree of freedom for *Q*-statistics; Q_p=*p*-value for *Q*-statistics; I^2=percentage of variance beyond sampling error.

Table 10

Subgroup Analysis of Negation Types on Item Discrimination Difference

	k	Random Effects (*SE*)	95%CI.l	95%CI.u	*Z*	*p*	*Q*	*df*	*p.h*	I^2
Not	60	0.10 (*0.08*)	-0.06	0.25	1.22	0.22	7345	59	0	99%
Prefix	61	-0.11 (*0.08*)	-0.26	0.04	-1.44	0.15	10321	60	0	99%
Qualifier	19	-0.05 (*0.14*)	-0.32	0.22	-0.37	0.72	1544	18	0	99%
Overall	140	-0.02 (*0.05*)	-0.11	0.09	-0.29	0.78	21119	139	0	99%

Q	*Qw*	*Qw.df*	*Qw.p*	*Qb*	*Qb.df*	*Qb.p*
21119	19209	137	0	3.63	2	0.16

Note. Q= Heterogeneity Q statistic, Q_w=Within-study heterogeneity, Q_w.df= degree of freedom for within-study heterogeneity, Q_w.p= *p*-value of within-study heterogeneity, Q_b=Between-study heterogeneity, Q_b.df= degree of freedom for between-study heterogeneity, Q_b.p= *p*-value of between-study heterogeneity.

3.4 Hypothesis 2 Results

Meta-regression with RVE of item discrimination difference between negatively and positively worded items were performed on both high and low motivation samples to test hypothesis 3. That is, low motivation samples generated larger negatively-worded item effects than high motivation samples do. Whether a study used attentive or inattentive sample depends on its reward and motivation type in Section 2.2.1.

The results of moderator analysis were presented in Tables 11 and 12, which demonstrated that sample motivation moderated the negatively-worded item effect (τ^2=0, and ω^2 =3.76). Based on Table 11, regarding the negatively-worded item effect, the mean difference between low and high motivation group is -0.65, and the mean of high motivation group (the reference group) is 0.48. That is, low motivation samples are more likely to produce larger negatively-worded item effects.

When outliers were excluded, the regression coefficient decreased (τ^2=0.17, and ω^2 =1.85), which suggested that low motivation samples produced an even larger negatively-worded item effect than high motivation samples (see details in Table 12). However, the general tendency was consistent with inclusion of outlier. Therefore, Hypothesis 2 was supported.

Table 11

Meta-Regression of Sample Motivation by Robust Variance Estimation (RVE)

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u	Sig
Intercept	0.48	0.08	5.66	60	0.0000004	0.31	0.64	***
High vs. Low Motivation	-0.65	0.16	-4.12	60	0.0001	-0.97	-0.34	***

Note. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound. *** stands for statistical significance at 0.01.

Table 12

Meta-Regression of Sample Motivation by Robust Variance Estimation (RVE) after Removing an Outlier

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u	Sig
Intercept	0.47	0.08	5.64	60	0.0000005	0.31	0.64	***
High vs. Low Motivation	-0.74	0.18	-4.22	60	0.00009	-1.10	-0.39	***

Note. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound, and *** stands for statistical significance at 0.01.

3.5 Additional Analysis

Meta-regression with RVE was performed to test that non-personality measures generate a larger negatively-worded item effect than personality measures. For the overall model, τ^2=0.04, and ω^2 =3.09. Table 13 showed that there is no statistically significant difference on the negatively-worded item effect between personality and non-personality measures. After outliers were removed, the overall trend stayed the same (τ^2=0.49, and ω^2 =1.50), but the estimates changed for both intercept and coefficients (see the details in Table 14). There was no evidence that personality and non-personality measures differed on the negatively-worded item effect.

Table 13

Meta-Regression of Personality vs. Non-Personality by Robust Variance Estimation (RVE)

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u
Intercept	-0.47	0.87	-0.55	73	0.59	-2.2	1.25
Personality vs. Non-Personality	0.43	0.87	0.49	73	0.67	-1.31	2.17

Note. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound.

Table 14

Meta-Regression of Personality vs. Non-Personality by Robust Variance Estimation (RVE) after Removing Outliers

	Estimate	*SE*	*t*-value	*df*	*p*	95% CI.l	95% CI.u
intercept	0.36	0.32	1.12	72	0.27	-0.28	0.996
Personality vs. Non-Personality	-0.48	0.35	-1.37	72	0.18	-1.17	0.22

Note. Estimate=Estimate of regression coefficient, SE=Standard Error, *df*=degree of freedom, p=*p*-value of t-test, 95% CI.l=95% confidence interval lower bound, 95%CI.u=95% confidence interval upper bound.

Subgroup analysis was conducted to examine the effect of the proportion of negatively worded items on the negatively-worded item effect regarding personality

measures. Specially, if the proportion of negatively worded items was smaller than 0.33, then the case was coded as Low; if the proportion was between 0.34 and 0.66, the case was coded as Medium; if the proportion was greater than 0.67 it was coded as High.

The statistics in Table 15 showed that high proportion of negatively worded items tended to produce a positively-worded item effect, while median and low proportion of negatively worded items indicated no statistical difference between the quality of negatively and positively worded items.

Similar subgroup analysis was performed to examine the effect of the proportion of negated items on the negation effect of personality measures. The sample was categorized into High and Low proportion. When the proportion of negatively worded items was below 0.35, it was coded as Low. When the proportion was between above 0.35 (including 0.35), it was coded as High.

The results suggested that when negated items were in high proportion, there was a statistical negation effect. That is, the quality of negated items was inferior compared with non-negated items. However, low proportion of negated item did not affect the quality of negated and non-negated items (see details in Table 16).

Table 15

Subgroup Analysis of the Proportion of Negatively Worded Items on the Effect

	k	Random Effect (*SE*)	95%CI.l	95%CI.u	*Z*	*p*	*Q*	*df*	*p.h*	I^2
High	20	0.81 (*0.35*)	0.12	1.49	2.30	0.02	48132	19	0	100%
Median	51	-0.31 (*0.22*)	-0.74	0.12	-1.41	0.16	44764	50	0	100%
Low	73	-0.09 (*0.18*)	-0.45	0.27	-0.51	0.61	15067	72	0	100%
Overall	144	-0.05 (*0.13*)	-0.30	0.21	-0.34	0.73	151826	143	0	100%

Q	*Qw*	*Qw.df*	*Qw.p*	*Qb*	*Qb.df*	*Qb.p*
151826	107964	141	0	7.39	2	0.025

Note. Q= Heterogeneity Q statistic, Q_w=Within-study heterogeneity, $Q_w.df$= degree of freedom for within-study heterogeneity, $Q_w.p$= p-value of within-study heterogeneity, Q_b=Between-study heterogeneity, $Q_b.df$= degree of freedom for between-study heterogeneity, $Q_b.p$= p-value of between-study heterogeneity.

Table 16

Subgroup Analysis of the Proportion of Negated Items on the Effect

	k	Random Effect (*SE*)	95%CI.l	95%CI.u	*Z*	*p*	*Q*	*df*	*p.h*	I^2
High	18	-0.46 (*0.15*)	-0.75	-0.18	-3.15	0.002	3480	17	0	100%
Low	101	-0.02 (*0.06*)	-0.14	0.11	-0.26	0.80	9617	100	0	99%
Overall	119	-0.08 (*0.06*)	-0.20	0.03	-1.46	0.14	13586	118	0	99%

Q	*Qw*	*Qw.df*	*Qw.p*	*Qb*	*Qb.df*	*Qb.p*
13586	13097	117	0	7.87	1	0.005

Note. Q= Heterogeneity Q statistic, Q_w=Within-study heterogeneity, $Q_w.df$= degree of freedom for within-study heterogeneity, $Q_w.p$= p-value of within-study heterogeneity, Q_b=Between-study heterogeneity, $Q_b.df$= degree of freedom for between-study heterogeneity, $Q_b.p$= p-value of between-study heterogeneity.

3.6 Publication Bias

Publication bias happens when the studies included in a meta-analysis are not representative. Non-significance research is less likely to be published on journals, and unpublished papers might contain more non-significant findings. This bias could lead to inaccurate estimation of the effect sizes.

In this study, two approaches were adopted to examine publication bias. First, trim-and-fill analysis (Duval & Tweedie, 2000) was conducted for each research question and hypothesis. Only examination of negation types requires 34 imputations on the left side of the distribution in Figure 1, where black dots represents effect sizes for the current study, and white dots represents imputed effect sizes. The trim-and-fill analysis was conducted on the overall effect between negated and non-negated items. It suggested that the corrected summary effect size was equal to -0.29 (*SE*=0.07) and 95% confidence

interval ranged from -0.43 to -0.16. Thus, there was statistically significant difference of item discrimination between negated and non-negated items. Meanwhile, heterogeneity Q statics was equal to 37248.14 (df=173), and τ=0.92, which suggested that a potential moderator existed in the effect sizes.

Empirical research showed that trim-and-fill analysis performed poorly when between-study heterogeneity existed (Peters, Sutton, Jones, Abrams, & Rushton, 2007). As large between-study heterogeneity was observed for negation types, trim-and-fill analysis may not provide an accurate estimation of publication bias under this circumstance.

Meanwhile, funnel plots were drawn for all the effect sizes in the study, which provides information consistent with trim-and-fill analysis. That is, no additional studies are needed for correcting publication bias, except negation type (see Appendix E).

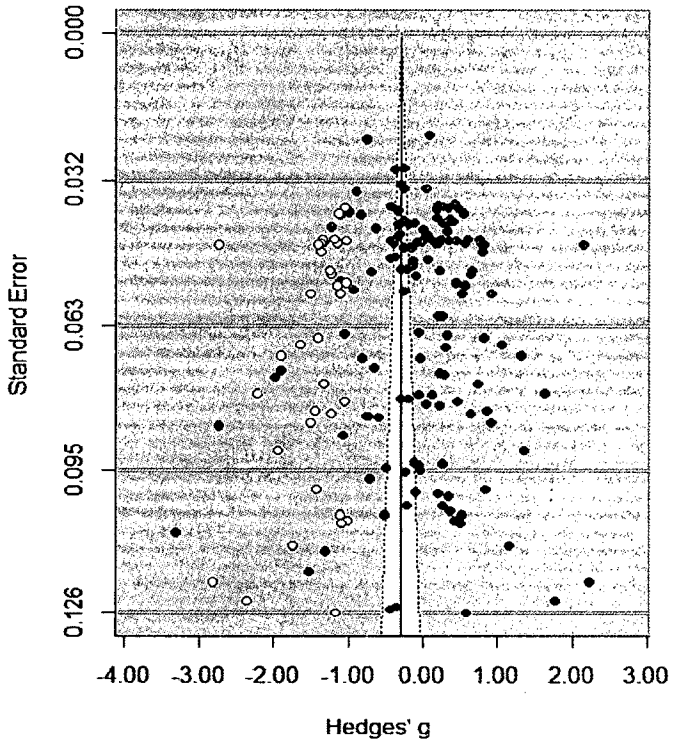


Figure 1. Funnel plot for evaluating publication bias for negation types

CHAPTER 4

DISCUSSION

This section will discuss possible causes of the main findings, how they will be of help in item writing and scale construction, limitations of the current study and ideas for future studies.

4.1 Discussion

4.1.1 The Negatively-Worded Item Effect and Its Operationalization. The three research questions asked about the existence and magnitude of the negatively-worded item effect overall and in specific personality dimensions. The results suggested that the negatively-worded item effect differs across personality scales, and that an overall effect therefore depends upon which scales were included and how the factors were defined. Because personality is bipolar, either *Neuroticism* or *Emotional Stability* could be the fifth factor and this factor had the largest individual effect, which was that items indicating *Emotional Stability* were much better quality than items indicating *Neuroticism*. When *Neuroticism* was included in the overall effect, its strong negative effect combined with the negative effects found for *Extraversion*, *Agreeableness* and *Openness* causing an overall negative effect. However, when *Emotional Stability* was included, cancellation produced a non-significant overall effect.

Regarding the size of the effect, the overall effect size was modest. When *Emotional Stability* was excluded, the overall effect size was an almost trivial -0.05, and it rose to -0.37 when *Neuroticism* was included (or -0.30 when an outlier was removed). The 95% confidence intervals suggest that the largest effect size likely to be observed in

practice would be of medium size (-0.50 to -0.65). Larger effect sizes were observed for individual dimensions (see below) but the overall results suggest that if a negatively-worded item effect occurs, it is likely to be small on average.

These results are different from previous research on the negatively-worded item effect of personality items. Sliter and Zickar (2014) discovered that the negatively-worded item effect happened at an overall level across dimensions. However, even if *Emotional Stability* was re-coded as *Neuroticism*, the conclusions that there was a negatively-worded item effect in Study 2 was unlikely to change due to the small contribution of *Emotional Stability*, as item discrimination difference was equal to 0.08 (Sliter & Zickar, 2014).

However, their findings were based on item discrimination difference between negative and positively worded items without considering standard deviations. For instance, in their Study 2, the item discrimination difference between negatively and positively word items was -0.27 at the overall level. As the pooled standard deviation was unknown, Cohen's *d* and Hedges' *g* could be larger or smaller than -0.27.

Meanwhile, in self-report non-personality measures, the negatively-worded item effect is unlikely to occur. Non-personality items tend to exhibit similar relationships with the psychological construct, no matter what they were negatively or positively worded. As shown in Appendix D, the summary effect of non-personality measures was -0.04, and the summary effect of personality measures was -0.05 when *Emotional Stability* was included. Most of non-personality measures involved in the study was designed to measure one construct, so the usage of positively and negatively worded items is seemingly able to evaluate the underlying concept symmetrically.

This finding contrasts strongly with the literature on the Rosenberg's Self-Esteem Scale (RSES). The current study included six effect sizes from RSES (see Appendix C), where only one study showed that the negatively-worded item effect existed (Hedges' g=-0.28) and other suggested that positively-worded item effect (Hedges' g ranged from 0.16 to 1.04). However, one crucial distinction is that we adopted a different perspective to investigate the negatively-worded item effect. In most RSES study on the effect, they established a structural model, and evaluated the model fit and model dimensionality/structure. That is, whether a one-factor or a two-factor solution (due to item wording) fit the data better. They did not place emphasis on item loadings of negatively and positively worded items. However, our study focuses on item qualities by examining factor loadings and item discrimination, which might lead to totally different conclusions.

Additionally, the proportion of negatively-worded item did not seem to relate to the negatively-worded item effect. On the contrary, a large proportion of negatively-worded items was associated with better quality of negatively-worded items than positively-worded items. However, when negatively-worded items were few, the quality of negatively and positively worded items showed no statistically difference. One possible reason of the counterintuitive finding is that the proportion of negatively worded items affects the estimation of item quality. When negatively worded items are majority, the estimation of the positively worded items is poor. Perhaps, this shows the effect of inattention; respondents seem to respond correctly to negatively worded items when they are in the majority and not when they are in the minority of the items on the scale.

4.1.2 The Negatively-Worded Item Effect and Personality Dimension. This study also found that the negatively-worded item effect was observed in *Agreeableness*, *Extraversion* and *Openness*, but not for *Conscientiousness* and *Emotional Stability*, based on the results of Research Question 3 and additional analysis. The effect was particularly strong for *Emotional Stability/Neuroticism*, where a significant effect size of 0.76 was obtained after deleting an outlier, and for *Agreeableness*, where a statistically significant value of -0.63 was obtained. For conscientiousness, the effect was non-significant as the 95% confidence interval [-0.37, 1.05] included zero, but the mean effect was positive 0.34 and the confidence interval indicates a much greater likelihood of a positive effect. This may suggest that the negatively-worded item effect is more likely to associate with the construct that is most socially desired. *Conscientiousness* and *Emotional Stability* are two most predictive dimensions of job performance (Barrick & Mount, 1991).

The findings shared similarities and differences with Sliter and Zickar study (2014). They found that the negatively-worded item effect occurred in all factors in Study 2, except *Emotional Stability*, where item discrimination difference was equal to 0.08. While we both revealed that *Agreeableness*, *Extraversion,* and *Openness* showed negatively-worded item effects. Similarly, *Emotional Stability* presented a positively-worded item effect in both studies. However, *Conscientiousness* displayed a positively-worded item effect in the current study, but a negatively-worded item effect in Sliter and Zickar study. One possible reason is due to their low-motivation sample, based on what we found in *H2*.

It is also possible that the relationships between the items and underlying constructs are different in *Agreeableness*, *Extraversion* and *Openness* from those in

Conscientiousness and *Emotional Stability*. For *Conscientiousness* and *Emotional Stability*, items that tap into unconscientiousness and emotional turmoil are better indicators of the underlying trait. However, items tapping into *Agreeableness*, *Extraversion* and *Openness* are better items on those scales. Maybe this relates so some difference in the nature of the content on those scales.

Another, more speculative rationale for the effects we observed may be due to the way personality dimensions are defined and personality items are written. Perhaps the differences in these scales indicate nothing more than that item writers are well-accustomed to behaviors, thoughts, and feelings that characterize both poles of *Conscientiousness*, *Extraversion* and *Openness* but that the definition of *Agreeableness* is written in such a way that prevents item-writers from writing effective *Disagreeableness* items. This explanation must be a partial explanation, at best, since it seems unlikely to explain our results for *Neuroticism*.

4.1.3 The Negatively-Worded Item Effects and Sample's Motivation. Hypothesis 2 demonstrated that highly-motivated samples are less likely to produce the negatively-worded item effect than less motivated samples. It is likely that a smaller proportion of careless responses exist in samples with high motivation. The amount of careless responses impacted the relationships between items and the constructs, which resulted in the emergence of the negatively-worded item effect. Low-motivation samples, with more careless responses, might ignore the negative wording and respond to the item as if it was positively-worded, resulting in the effect demonstrated by Schmitt and Stults (1985) and Woods (2006).

Clearly, if the main reason why the negatively-worded item effect occurs is because respondents cannot be bothered to read the items, then any solution that focuses on the items is entirely and hopelessly misguided. It is obvious that researchers must remove respondents who are not actually participating. If that were not the case, then we can dispense with participants altogether and replace them with trivial random number generators, which could produce enormous sample sizes. Alas, this clearly is ridiculous; when we study human behavior using specific stimuli, we must study behavior that results from reasonable attention to those stimuli. Therefore, one implication of these results is that researchers should regard the negatively-worded item effect as an indication of a serious data quality issue that should trigger vigorous attempts to remove inattentive respondents. Another implication is that too much shoddy research has been conducted using samples that are not actually reading the questions to which they are responding and that future research efforts need to find ways to raise the motivation of such samples.

4.1.4 Negated Items. Test of hypothesis 1 did not support that negated items presented larger item discrimination compared with non-negated items. For personality scales, the mean effect size was -0.08, and for all measures the overall effect size was -0.02. These values are in the hypothesized direction but of a trivial size. Analysis by types of negation did not indicate that specific types of negation were more or less problematic; in fact, the use of not/never/n't was associated with a modest and non-significant opposite effect of 0.10. Of course, this doesn't mean that all items with negation are good items or that some items are more confusing with negation. However, these results do suggest that

item characteristics of negated and non-negated items are very similar and that merely including negation in an item is unlikely to create a poor item.

These findings are inconsistent with previous finding by Holden and colleagues (1985; 1990) that negated item showed inferior psychometric properties, compared with non-negated items. They observed negated items showed lower criterion-validity, as well as lower desirability, compared with non-negated items. However, most statistics they presented in the two studies were not statistically significant. Meanwhile, their conclusions depended on external criteria and subjective judgments (e.g., desirability), which could be also impacted by sampling errors. Even though the tendency was clear, what they found might be vulnerable to sampling errors than the meta-analysis.

However, the proportion of negated items seemingly influences the quality of negated and non-negated items. Specifically, when the proportion of negated items was high (even 35% or above items were negated), negated items had inferior item discrimination than non-negated items; when the proportion of negated items was low, negated and non-negated items presented no statistical difference on item discrimination. Considering the finding of $H1$, the effect of negated items may be caused by how many negated items are in the scale, rather than whether the items are negated or not. That is, the proportion of negated items play a more important role in determining the quality of negated and non-negated items, compared with the simple classification of items with or without negation. It is also consistent with the conclusion of Holden and colleagues (1985; 1990).

The publication bias results reported in section 3.6 are curious. Publication bias was not found for any other analysis, and apparent publication bias was found for overall

analysis of negated item effects. The funnel plot in section 3.6 shows that most of our negated item effects have positive values, indicating that negated items functioned better. The trim-and-fill procedure suggested that a large number of negative effects would be needed to balance the plot and thus implies that there are a large number of unpublished studies that have negative negated-items effects. In fact, the imputed overall effect suggests a significant negated item effect after including these imputed values (i.e., contrasting with the lack of differences we observed).

We are extremely puzzled about why there would be a tendency not to report negated item effects. That is, the effect sizes we are calculating and reporting are not actually of interest in the primary studies and their authors (and editors and reviewers) are unlikely to have even noticed these effects. However, it could be the case that poorly functioning negated items have been removed from the studied surveys. After all, the surveys included in this study are all published and many are commercial (e.g., EPQ) or heavily researched (e.g., IPIP, and Rosenberg SES) and poor quality items would have been removed from the scales during pilot testing. The effect of removing poorly functioning items could cause the "publication bias" results that we observed. If so, it would change our interpretation of the results slightly. It might mean that negated items are at least somewhat more likely to be poor (consistent with the prevailing view) but that negated items can be effective (as evidenced by their presence and high quality in in pilot-tested, published surveys).

4.1.5 Theoretical Implication. Our findings expanded the understanding of personality items in several ways. First, the study reviewed the negatively-worded effect in

personality measures by meta-analyzing item-level information, which is more straightforward and intuitive approach than previous studies. Second, the findings suggested that the effect does not occur as consistently and strongly as the field has been lead to believe by the published literature. The negatively-worded item effect depends how the concept was operationalized as well as which personality dimension was considered. Third, the present study also refuted the myth that negation in items makes them harder to read, and thus lowers their quality, or at least suggested that negated items can be of equal quality as non-negated items. However, when the proportion of negated items was high in the scale, the quality of negated item tended to be inferior.

4.1.6 Practice Implication. Several implications can be drawn from the results of the study. First, low respondent motivation is probably a bigger problem than has typically been assumed and is likely to cause the negatively-worded item effect because respondents are not attending to the items. Thus, any solution to the negatively-worded item effect that focuses on the items is misguided. Research participants must be motivated, if possible, or else vigorous data cleaning methods must be used to remove inattentive respondents from the sample. Respondents are likely to be more motivated under high-stake situations, like selection and promotion; however, under research circumstances, participants might be less motivated. Results from low-motivation samples may result in significantly erroneous conclusions.

Our initial analyses in the beginning stages of this project focused on personality surveys completed by samples attracted by a free personality report and we found no negatively-worded item effect in these samples. Thus, we have one example that seems to

illustrate how research may be conducted using volunteers who are attentive. We challenge researchers to find other paradigms that motivate research samples.

Second, item writers and test developers should review item-writing materials and procedures to try to ensure that items written for both poles are of equal quality. Our results could suggest that it is difficult for item-writers to write negatively-worded items for *Agreeableness*, *Extraversion* and *Openness* and positively-worded items for *Emotional Stability* and *Conscientiousness*. It is equally clear that simply not writing such items is a poor solution.

Last, any prohibition on item negation should be relaxed. Our findings indicate little concern that negated items have a detrimental effect on scale quality. Our research does not exclude that negation or some other item-writing practice could be used to create poor items and thus normal precautions should be taken. Probably the best way to view our findings is that when the use of negation is the most natural way to express an idea, it should not be discouraged. However, the proportion of negated items should be relatively low, as we found that the proportion of items with negation affected the quality of negated and non-negated items.

4.2 Limitations and Future Directions

As with any empirical research, this study had several limitations. First, many interesting potential methodological factors (e.g., different types of personality measures, and countries) could not be fully investigated in this study. For example, the EPQ, IPIP or NEO may show different effects but we did not have enough samples to perform this analysis. It would also be interesting to examine the effect of negation and negatively-

worded items in different languages; these effects may present differently in other languages.

Second, only a relatively small number of non-personality measures were included in the study. The conclusion should be interpreted with caution. There are only thirteen non-personality inventories involved, while hundreds of psychological constructs have been evaluated. Even though meta-analysis can be conducted as long as the number of studies involved is greater than one, more studies could contribute to higher power of study, particularly for random-effects models, such as were used here. Potentially every research sample that retained item-level response data could be analyzed and included in this kind of meta-analysis and hopefully in the future larger samples of effects will be available.

The nature of negative-wording is subjective and especially so for bipolar personality scales and tied closely to the Big Five. We observed opposite effects for *Neuroticism* and *Emotional Stability*. If we had included the EPQ *Psychoticism* dimension or data from Cattell's 16PF (which defines *Agreeableness* and *Openness* as *Independence* and *Tough-Mindedness*), we may have found somewhat different results. We made the decision to constrain our exploration in the popular Big Five because if facilitated meta-analysis across studies, but exploration using a wider variety of models may shed additional light on these effects.

Fourth, we calculated our effect sizes using parameter estimates of statistical structural models, but we could not always verify the fit of these models to the data. When we analyzed datasets ourselves, we ensured good model-data fit and it is likely the case that published studies also had reasonable fit. However, misfitting models might

produce misleading results. Hopefully, future research can address this issue by re-analyzing all of the datasets included in the meta-analysis. It is possible that model-fit issues contributed to the sizable variability in effect sizes in our meta-analysis.

Also, this research used only published surveys that had already been pilot-tested and analyzed. It would be highly informative to extend these results to newly-written surveys under-going item analysis. It is possible that large numbers of negatively-worded or negated items are found to be poorly functioning and discarded during the pilot testing process and, if this is the case, would serve as a warning to item writers.

The dataset was analyzed multiple times in the current study, the findings of which are susceptible to inflated type I error. We did not use statistical method (e.g., Bonferroni correction) to control the error. One solution for future research would be to increase sample size.

In this research, standardized mean difference was operationalized as effect size, the calculation of which involved in the pooled standard deviation. However, there is no conclusive answer on whether the pooled standard deviation or the standard deviation of control group. Thus, the interpretation of the pooled standard deviation might be questionable.

Future research should investigate what factors may contribute to the negatively-worded item effect in certain personality dimensions. It has been shown that the negatively-worded item effect depends on what personality dimensions were involved. Meanwhile, *Neuroticism/Emotional Stability* demonstrated asymmetry of the bipolar scales regarding the negatively-worded item effect. Thus, research will be needed to

understand why some dimensions exhibit the effect while others does not, and the symmetry/asymmetry of the bipolar scales.

Finally, future studies should also examine the relationships between modality and negation effect. Even though the research concluded that negation items did not present inferior psychometric properties compared to non-negation items. However, no modality of the negation items was explored. Modality (including epistemic and deontic modality) might be relevant to qualities of negated items.

APPENDIX A

DISCUSSION ON THE FIVE-FACTOR MODEL OF PERSONALITY

Although different theories of personality were proposed over the years, the Five Factor Model (FFM) has been the most popular personality model, as accumulated empirical evidence suggested that almost any personality measurement can be categorized into five factors (Goldberg, 1990). The five factors were driven from factor-analyzing trait adjectives and named by experts (McCrae & John, 1992). Research demonstrated that the five factors are stable across culture and consistent over time (McCrae & Costa, 1997).

FFM, also known as OCEAN, usually refers to five personality dimensions, including *Agreeableness*, *Consciousness*, *Extraversion*, *Neuroticism* and *Openness to experiences*. *Extraversion/Introversion* involves traits, like sociability, assertiveness, ambition and energy; *Neuroticism/Emotional Stability* usually associates with traits, such as anxiety, depression, anger, insecurity, and worry; *Agreeableness/Hostile No-Compliance* includes courtesy, flexibility, trustworthy, cooperation, tolerance, and soft-heart; *Conscientiousness/Unconsciousness* relates to traits, such as dependability, accountability, organization, and hardworking; *Openness to experience*, which is the hardest to define and identify (Barrick & Mount, 1991), refers to traits, like imagination, broad-mind, intelligence and artistic sensitivity.

It is worth noting that FFM also follows a bipolar model, which means that each factor can be measured by two different poles. For instance, *Extraversion* can be measured by being talkative, sociable and energetic, and it can also be measured by being introverted, shy, and lazy. Which polar the items assess totally depends on what test developers intended to score, because when the scale is scored, only one score will be

presented from one particular polar. The interpretation of the scores might involve in two directions.

There is no right or wrong direction of a factor, but the choice of direction closely relates to direction-of-wording. Use "I feel blue." as an example. If the item belongs to Neuroticism, it is a positively worded item, as a high score on the item represents a high neurotic level. However, if it is supposed to measure *Emotional Stability*, it is a negatively worded item, which suggests that a high score on the item stands for a low level of emotional stability. Thus, this item should be reverse coded when it is used for scoring. Even thought the respondents have no knowledge on how the items will scores, test-developers should define the direction of polar for scoring during test development. Therefore, even if personality can be measured by either pole, it is more likely to be scored by only one.

FFM has revolutionized the domain of Industrial and Organizational (I/O) Psychology research. Personality (such as *Conscientiousness*) has shown positive correlations with important organizational and individual outcomes, such as leadership (Judge, Bono, Ilies, & Gerhardt, 2002) and engagement (Macey & Schneider, 2008). Moreover, a small to moderate relationship between personality and job performance has been observed in research and practice (Barrick & Mount, 1991). Thus, understanding of personality and measurement of personality has drawn increasing interest. Studies have been executed by I/O psychologists regarding construction of personality scales, identifying influential factors, understanding the response process and so on (Stark, Chernyshenko, Drawsgow, Williams, 2006; Huang & Mead, 2014).

APPENDIX B

SUMMARY OF ALL STUDIES

Table 17

Summary Information of All Studies in the Meta-Analysis

ID	Author and Year	P	Measure	Item Stats	Sample	Motivation	Published Stats	IRT
1	Marsh & Morin, 2013	1	NEO-FFI	1	Others	Others	1	0
2	Sliter & Zickar, 2014	1	IPIP-100	0	Students	Research	1	1
3	Corwyn, 2000 Adult M1	0	RSES	1	General Population	Research	1	0
4	Corwyn, 2000 Followup	0	RSES	1	General Population	Research	1	0
5	Andrich & Schoubroeck, 1989	0	GHQ-30	1	Teachers	Research	1	0
6	DiStefano & Motl, 2006	0	RSES	1	Students	Research	1	0
7	Dunbar et al., 2000 Mid	0	RSES	1	General Population	Research	1	0
8	Dunbar et al., 2000 Old	0	RSES	1	General Population	Research	1	0
9	Martin, 1999	0	GHQ-12	1	General Population	Research	1	0
10	Martín-Albo et al., 2007	0	RSES	1	Students	Research	1	0
11	Mavor et al., 2010	0	RWA	1	Students	Research	1	0
12	Hevey et al., 2010	0	CFC	1	Students	Research	1	0
13	Apple, 2011	1	IPIP 50	1	Students	Research	0	1
14	Broadfoot, 2008	1	IPIP 50-marker scale	1	Students	Research	0	1
15	Chernyshenko et al., 2007	1	Conscientiousness	1	Students	Research	1	1
16	Chiesi et al., 2013	0	LOT-R	1	Students	Research	1	0
17	Ferrando, 2001	1	Neuroticism	1	Students	Research	1	1
18	Ferrando, & Lorenzo-Seva, 2005	1	E-EPQ-R, N-EPQ-R	1	Students	Research	1	1
19	Stepp et al., 2012	1	NEO PI-R	1	Students	Research	1	1
20	Gignac et al., 2007 S1	1	NEO-FFI	1	Students	Research	1	0
21	Gignac et al., 2007 S2	1	NEO-FFI	1	Students	Research	1	0
22	McAbee BFI	1	BFI	1	General Population	Research	0	0
23	McAbee HEXACO	1	HEXACO	1	General Population	Research	0	0
24	Biderman IPIP	1	IPIP	1	Students	Research	1	0
25	Biderman et al., 2012 SIOP	1	IPIP	1	Students	Research	0	0

ID	Author and Year	P	Measure	Item Stats	Sample	Motivation	Published Stats	IRT
26	Biderman et al., 2011 S1	1	IPIP 50-item	1	Students	Research	1	0
27	Biderman et al., 2011 S2	1	IPIP 50-item	1	Students	Research	1	0
28	Biderman et al., 2011 S3	1	IPIP 50-item	1	Students	Research	1	0
29	Biderman et al., 2011 S4	1	IPIP 50-item	1	Students	Research	1	0
30	Biderman et al., 2011 S5	1	IPIP 50-item	1	Students	Research	1	0
31	Biderman et al., 2011 S5 NEO	1	IPIP 50-item	1	Students	Research	1	0
32	Eysenck, & Barrett, 2013 Au	1	EPQ	1	General Population	Research	0	1
33	Eysenck, & Barrett, 2013 Br	1	EPQ	1	General Population	Research	0	1
34	Eysenck, & Barrett, 2013 Bu	1	EPQ	1	General Population	Research	0	1
35	Eysenck, & Barrett, 2013 Ca	1	EPQ	1	General Population	Research	0	1
36	Eysenck, & Barrett, 2013 Ct	1	EPQ	1	General Population	Research	0	1
37	Eysenck, & Barrett, 2013 Cz	1	EPQ	1	General Population	Research	0	1
38	Eysenck, & Barrett, 2013 Eg	1	EPQ	1	General Population	Research	0	1
39	Eysenck, & Barrett, 2013 UK	1	EPQ	1	General Population	Research	0	1
40	Eysenck, & Barrett, 2013 Fi	1	EPQ	1	General Population	Research	0	1
41	Eysenck, & Barrett, 2013 Fr	1	EPQ	1	General Population	Research	0	1
42	Eysenck, & Barrett, 2013 Ge	1	EPQ	1	General Population	Research	0	1
43	Eysenck, & Barrett, 2013 HK	1	EPQ	1	General Population	Research	0	1
44	Eysenck, & Barrett, 2013 In	1	EPQ	1	General Population	Research	0	1
45	Eysenck, & Barrett, 2013 Is	1	EPQ	1	General Population	Research	0	1
46	Eysenck, & Barrett, 2013 It	1	EPQ	1	General Population	Research	0	1
47	Eysenck, & Barrett, 2013 Jp	1	EPQ	1	General Population	Research	0	1
48	Eysenck, & Barrett, 2013 Le	1	EPQ	1	General Population	Research	0	1
49	Eysenck, & Barrett, 2013 Li	1	EPQ	1	General Population	Research	0	1
50	Eysenck, & Barrett, 2013 Me	1	EPQ	1	General Population	Research	0	1
51	Eysenck, & Barrett, 2013 Ne	1	EPQ	1	General Population	Research	0	1
52	Eysenck, & Barrett, 2013 Ni	1	EPQ	1	General Population	Research	0	1
53	Eysenck, & Barrett, 2013 No	1	EPQ	1	General Population	Research	0	1
54	Eysenck, & Barrett, 2013 Po	1	EPQ	1	General Population	Research	0	1

ID	Author and Year	P	Measure	Item Stats	Sample	Motivation	Published Stats	IRT
55	Eysenck, & Barrett, 2013 Pt	1	EPQ	1	General Population	Research	0	1
56	Eysenck, & Barrett, 2013 Pr	1	EPQ	1	General Population	Research	0	1
57	Eysenck, & Barrett, 2013 Ro	1	EPQ	1	General Population	Research	0	1
58	Eysenck, & Barrett, 2013 Si	1	EPQ	1	General Population	Research	0	1
59	Eysenck, & Barrett, 2013 Sp	1	EPQ	1	General Population	Research	0	1
60	Eysenck, & Barrett, 2013 Sk	1	EPQ	1	General Population	Research	0	1
61	Eysenck, & Barrett, 2013 Sp	1	EPQ	1	General Population	Research	0	1
62	Eysenck, & Barrett, 2013 Sl	1	EPQ	1	General Population	Research	0	1
63	Eysenck, & Barrett, 2013 Ug	1	EPQ	1	General Population	Research	0	1
64	Eysenck, & Barrett, 2013 USA	1	EPQ	1	General Population	Research	0	1
65	Eysenck, & Barrett, 2013 Ru	1	EPQ	1	General Population	Research	0	1
66	Eysenck, & Barrett, 2013 Zi	1	EPQ	1	General Population	Research	0	1
67	Mead, 2006	1	Personality	1	Students	Research	0	1
68	Huang & Mead, 2014	1	IPIP	1	Students	Research	1	1
69	Mead, 2011	1	IPIP	1	Students	Research	0	1
70	Mead, 2012	1	IPIP	1	Students	Research	0	1
71	Mead, 2013	1	IPIP	1	Students	Research	0	1
72	Goldberg, 1995	1	IPIP-100	1	General Population	Research	0	1
73	Leibetseder et al., 2007	1	E-Scale	1	General Population	Research	1	0
74	Maydeu-Olivares et al., 2011	1	EPQ-R	1	General Population	Research	1	1
75	Bollen, & Maydeu-Olivares, 2007	1	LOT	1	General Population	Research	1	1
76	Fleischer, 2014 Mturk	1	IPIP	1	General Population	Personal reports	0	1
77	Fleischer, 2014 Students	1	IPIP	1	Students	Research	0	1

Note. P=Personality measures (1=Personality measures, 0=Non-personality measures), Item Stats=Item-level statistics (1=Item-level statistics, 0=mean/SD), Published stats=Statistics directly from publication (1=Yes, 0=No), IRT=Whether the statistics was from IRT (1=Yes, 0=No).

APPENDIX C

CODING MANUAL AND CODING SHEET SAMPLE

Coding Manual

Prior to coding, articles must be sampled, obtained, and assigned a study ID; do not code a study unless it has already been assigned a study ID.

Before coding, it is helpful to consider what will be done with these data. Mainly, we will tally the occurrence of various aspects of these studies or compute simple statistics. For example, we might use these data to report that: "Only 1.2% of studies sampled actually reported that the estimation procedure converged." Or we might say: "The mean sample size was 345.4, the median was 471.5, and the standard deviation was 50.1."

As a consequence, it is vitally important to code the individual results of individual IRT/CFA analyses. For example, if a study reported that IRT analysis was conducted for five personality scales. Then, effect sizes should be recoded for each dimension. Another example, if a study described multiple models, please choose model(s) which is similar to IRT. That is, if all items are supposed to measure one construct, they should be loaded on one-factor. If the measure was designed to measure two factors, choose the model with a 2-factor solution. Then, calculate the negatively-worded item effect and difference between negated and non-effect size item effect size on each factor based on item contents.

If a study reported Neuroticism, please swap positively and negatively worded items. Make sure items are positively and negatively worded under "Emotional Stability". Then, calculate effect sizes accordingly.

The coding sheet is a spreadsheet. The study ID and citation were already entered. Create a new row by copying the study ID. If there are multiple analyses or different tests or for different samples, you will need to code a row for each analysis. Please copy all other information regarding article information (See details below). In this case, each analysis should be identified by substudyID or Model. For instance, enter 1 or 2 to distinguish study 1 or 2 from the same article. Another example, an article includes multiple models, enter "Model1", "Model2" or "Model3" for each analysis.

If a study involved more than one categories, please enter 1 in each. For example, a study used both students and applicants samples, please enter 1 under students and 1 under applicants.
Try to fill in all the fields. If you have considered this field and determined that the manuscript does not contain this information, but enter "Cannot tell" under others of the particular category.
If you believe that the data is erroneous, code the reported data along with a note in sheet 2 (in the comment section, if you cannot enter free responses in that field) like "2.3 but I think they mean 3.3".

When coding, you may have questions. You can include short questions in the Notes (in sheet 1) or Comments field (in sheet 2) or you can type up questions. I try to highlight the parts of the study that contain the relevant information; this allows me to quickly find where I found a particular piece of information when discussing the article.

PART1-ARTICLE INFORMATION

ID: To track the articles

Citation: To keep all information of the articles by APA style

Personality: To record whether personality or non-personality measures was involved in the study

If the study involved personality inventory, please enter 1

If the study involved non-personality inventory, please enter 0

If the study involved both personality and non-personality inventories, please enter 2

Inventory Names: To enter the name(s) of measures used in the study

PART2-ANALYSIS, CONSTRUCT AND SAMPLE

SubStudyID: To separate research with multiple studies

Sample Size: To write the sample size of each study

Construct: To write which construct the effect size came from. For example, agreeableness will be entered as "A", conscientiousness will be entered as "C", extraversion will be entered as "E", emotional stability will be entered as "ES", and openness will be entered as "O"; Rosenberg's self-esteem will be entered as "RSES".

Model: To distinguish study with multiple models. For instance, a CFA study reported model 1, model 2, model 3, and all of them will be recorded. Then, create a row for each model, and enter "Model 1", "Model 2" and "Model 3" on each row under Model.

Sample: To code characteristics of samples

If the study used students sample, please enter 1 under students

If the study used applicants sample, please enter 1 under applicants

If the study used employees sample, please enter 1 under employees

If the study used general population, please enter 1 under general population

If the study used sample which was not mentioned before, please write the sample

Motivation: To code motivation and rewards methods in the study

If the purpose of attending the study is for research, please enter 1 under research

If the purpose of attending the study is for selection, please enter 1 under selection

If the purpose of attending the study is for class or credits, please enter 1 under class/credit

If the purpose of attending the study is for personal report (like personality report), please enter 1 under personal report

If there are other purposes, please write down under others.

ItemWording: To record whether the item is negatively or positively worded

Positively worded items refer to the items with positive keys; Negatively worded items refer to the items with negatively keys.

If the item is positively worded, please enter "+"

If the item is negatively worded, please enter "-"

Negation: To record whether the item is with or without negation

Items with negation fit one of the following categories: (1) the items with not (e.g., not, -n't, never), or (2) implicit negative (e.g., negative prefixes as im-, un-, in-), or (3) negative qualifiers (e.g., seldom or rarely). Items that do not fit the categories are without negation.

If the item with negation, please enter 1 under negation

If the item without negation, please enter 0 under negation

IRT Model: To record what IRT model the analysis was used. For example, "1PL" or "2PL"

IRT Software: To identify what IRT software the analysis was conducted with. For instance, "MULTILOG" or "BILOG"

CFA Model: To record what CFA model the analysis was used. For example, "CFA"

CFA Software: To identify what CFA software the analysis was conducted with. For instance, "MPlus", "AMOS", "EQS" or "LISERAL"

The article usually applied either IRT or CFA, thus, only IRT Model and Software or CFA Model and Software will be filled in. For instance, if the article only used IRT, please leave CFA Model and Software alone, and vice versa.

PART3: ITEM STATISTICS AND EFFECT SIZES

Statistics: To record what level statistics were described in the study

If the study reported mean and SD for positively and negatively worded items, please enter 1 in M &SD

If the study reported factor loadings or IRT discrimination on item-level, please enter 1 in item-level

If the study reported other formats, please specify the details in "others"

Negatively Worded Items: To record a effect size of the negative-worded item effect, where d=(Mn-Mp)/SD, Mn stands for the mean of negatively worded items, Mp stands for the mean of positively worded items.

Negated Items: To record a effect size of negated and non-negated items, where d=(Mt-Mnt)/SD, Mt stands for items with negation, Mnt stands for items without negation

Comments: To write your comments, questions, and concerns.

Coding Sheet Sample

ID		58
CITATION		Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. *Psychological reports*, *84*(3), 927-935.
PERSONALITY		
INVENTORY NAME		
SUBSTUDYID		
CONSTRUCT		
ISSCORINGSWAPPED		
MODEL		
SAMPLE SIZE		
SAMPLE	Students	
	Employees	
	Applicants	
	General Population	
	Others	
MOTIVATION	Research	
	Class/Credits	
	Selection	
	Personal Report	
	Others	
STATISTICS	M & SD	
	Item-level	
	others	
IRT	IRT Model	
	IRT Software	
CFA	CFA Model	
	CFA Software	
POSITIVELY WORDED ITEMS	Proportion	
	Mean	
	SD	
NEGATIVELY WORDED ITEMS	Proportion	
	Mean	
	SD	
NEGATED ITEMS	Proportion	
	Mean	
	SD	
NON-NEGATED ITEMS	Proportion	
	Mean	
	SD	
EFFECT SIZES	Negatively Worded Items	
	Negated Items	
COMMENTS		

APPENDIX D

EFFECT SIZES OF RESEARCH QUESTIONS AND HYPOTHESES

Table 18

Effect Sizes of the Negatively-Worded Item Effect for All Personality Studies

ID	D	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
1	A	14021	0.27	0.000	14021	0.71	0.304	0.22	-2.02	0.000	-2.02	0.000	0.015
1	C	14021	0.31	0.000	14021	0.71	0.240	0.17	-2.35	0.000	-2.35	0.000	0.016
1	E	14021	0.24	0.000	14021	0.73	0.134	0.10	-5.11	0.001	-5.11	0.001	0.025
1	ES	14021	0.71	0.021	14021	0.55	0.000	0.02	10.33	0.002	10.33	0.002	0.045
2	A	545	0.39	0.055	545	2.08	0.828	0.59	-2.87	0.007	-2.87	0.007	0.086
2	O	545	0.22	0.031	545	2.98	1.770	1.25	-2.21	0.006	-2.21	0.006	0.077
14	A	439	0.93	0.483	439	1.25	0.571	0.53	-0.62	0.005	-0.61	0.005	0.069
14	C	439	0.72	0.233	439	0.88	0.200	0.22	-0.71	0.005	-0.71	0.005	0.070
14	E	439	1.30	0.480	439	1.12	0.248	0.38	0.45	0.005	0.45	0.005	0.068
14	ES	439	1.09	0.449	439	0.79	0.113	0.33	0.92	0.005	0.92	0.005	0.071
14	O	439	0.84	0.188	439	0.93	0.269	0.23	-0.40	0.005	-0.40	0.005	0.068
15	C	539	0.91	0.366	539	0.75	0.327	0.35	0.46	0.004	0.46	0.004	0.062
18	E	201	0.20	0.000	201	0.92	0.538	0.38	-1.87	0.014	-1.86	0.014	0.120
19	C	1517	1.59	0.523	1517	1.47	0.322	0.43	0.27	0.001	0.27	0.001	0.036
20	A	538	0.47	0.107	538	0.53	0.046	0.08	-0.65	0.004	-0.65	0.004	0.063
20	C	538	0.71	0.141	538	0.55	0.063	0.11	1.47	0.005	1.47	0.005	0.069
20	E	538	0.58	0.000	538	0.52	0.111	0.08	0.82	0.004	0.82	0.004	0.063
20	O	538	0.39	0.032	538	0.50	0.228	0.16	-0.68	0.004	-0.68	0.004	0.063
21	A	539	0.44	0.112	539	0.51	0.099	0.11	-0.59	0.004	-0.59	0.004	0.062
21	C	539	0.71	0.071	539	0.57	0.061	0.07	2.20	0.006	2.20	0.006	0.077
21	E	539	0.54	0.000	539	0.52	0.115	0.08	0.30	0.004	0.30	0.004	0.061
21	O	539	0.36	0.080	539	0.50	0.244	0.18	-0.79	0.004	-0.79	0.004	0.063
22	A	478	0.66	0.092	478	0.51	0.085	0.09	1.67	0.006	1.67	0.006	0.075
22	C	478	0.64	0.096	478	0.61	0.093	0.09	0.31	0.004	0.31	0.004	0.065
22	E	478	0.73	0.059	478	0.74	0.190	0.14	-0.10	0.004	-0.10	0.004	0.065
22	ES	478	0.62	0.102	478	0.67	0.037	0.08	-0.66	0.004	-0.66	0.004	0.066
22	O	478	0.56	0.000	478	0.62	0.133	0.09	-0.65	0.004	-0.65	0.004	0.066
23	C	322	0.48	0.073	322	0.46	0.067	0.07	0.22	0.006	0.22	0.006	0.079
23	E	322	0.51	0.107	322	0.55	0.095	0.10	-0.35	0.006	-0.35	0.006	0.079
23	O	322	0.44	0.124	322	0.41	0.125	0.12	0.24	0.006	0.24	0.006	0.079
24	A	372	0.42	0.222	372	0.56	0.123	0.18	-0.75	0.006	-0.74	0.006	0.076
24	C	372	0.55	0.175	372	0.52	0.119	0.15	0.20	0.005	0.20	0.005	0.074
24	E	372	0.65	0.082	372	0.75	0.007	0.06	-1.68	0.007	-1.68	0.007	0.085
24	ES	372	0.57	0.215	372	0.65	0.099	0.17	-0.50	0.006	-0.50	0.006	0.074
24	O	372	0.52	0.115	372	0.59	0.088	0.10	-0.77	0.006	-0.77	0.006	0.076
25	A	328	0.50	0.185	328	0.58	0.158	0.17	-0.50	0.006	-0.50	0.006	0.079
25	C	328	0.59	0.097	328	0.59	0.109	0.10	-0.02	0.006	-0.02	0.006	0.078
25	E	328	0.64	0.064	328	0.76	0.030	0.05	-2.34	0.010	-2.34	0.010	0.101
25	ES	328	0.60	0.177	328	0.68	0.101	0.14	-0.54	0.006	-0.54	0.006	0.080
25	O	328	0.67	0.047	328	0.54	0.096	0.08	1.60	0.008	1.60	0.008	0.090
26	A	183	0.55	0.131	183	0.59	0.127	0.13	-0.31	0.011	-0.31	0.011	0.105

ID	D	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
26	C	183	0.67	0.232	183	0.45	0.162	0.20	1.12	0.013	1.12	0.013	0.113
26	E	183	0.66	0.074	183	0.77	0.032	0.06	-1.89	0.016	-1.88	0.016	0.126
26	ES	183	0.53	0.102	183	0.60	0.083	0.09	-0.73	0.012	-0.73	0.012	0.108
26	O	183	0.54	0.030	183	0.50	0.131	0.09	0.47	0.011	0.47	0.011	0.106
27	A	202	0.46	0.110	202	0.60	0.084	0.10	-1.38	0.012	-1.38	0.012	0.111
27	C	202	0.54	0.291	202	0.47	0.090	0.22	0.31	0.010	0.31	0.010	0.100
27	E	202	0.59	0.082	202	0.65	0.078	0.08	-0.85	0.011	-0.85	0.011	0.104
27	ES	202	0.54	0.224	202	0.65	0.164	0.20	-0.54	0.010	-0.54	0.010	0.101
27	O	202	0.46	0.118	202	0.51	0.135	0.13	-0.41	0.010	-0.41	0.010	0.101
28	A	311	0.43	0.073	311	0.73	0.062	0.07	-4.43	0.022	-4.42	0.022	0.149
28	C	311	0.51	0.155	311	0.67	0.063	0.12	-1.34	0.008	-1.34	0.008	0.089
28	E	311	0.68	0.051	311	0.77	0.028	0.04	-2.11	0.010	-2.11	0.010	0.100
28	ES	311	0.52	0.279	311	0.69	0.125	0.22	-0.79	0.007	-0.79	0.007	0.083
28	O	311	0.24	0.079	311	0.56	0.123	0.10	-3.12	0.014	-3.11	0.014	0.119
29	A	404	0.41	0.086	404	0.43	0.093	0.09	-0.27	0.005	-0.27	0.005	0.071
29	C	404	0.68	0.163	404	0.36	0.128	0.15	2.21	0.008	2.20	0.008	0.089
29	E	404	0.46	0.195	404	0.49	0.167	0.18	-0.15	0.005	-0.15	0.005	0.070
29	ES	404	0.47	0.086	404	0.56	0.087	0.09	-1.04	0.006	-1.04	0.006	0.075
29	O	404	0.46	0.045	404	0.42	0.112	0.09	0.51	0.005	0.51	0.005	0.072
30	A	189	0.55	0.107	189	0.54	0.159	0.14	0.09	0.011	0.09	0.011	0.103
30	C	189	0.60	0.103	189	0.58	0.099	0.10	0.14	0.011	0.14	0.011	0.103
30	E	189	0.65	0.079	189	0.75	0.003	0.06	-1.83	0.015	-1.83	0.015	0.123
30	ES	189	0.67	0.076	189	0.67	0.082	0.08	0.00	0.011	0.00	0.011	0.103
30	O	189	0.53	0.059	189	0.50	0.144	0.11	0.29	0.011	0.29	0.011	0.103
31	A	189	0.49	0.153	189	0.48	0.122	0.14	0.06	0.011	0.06	0.011	0.103
31	C	189	0.45	0.065	189	0.59	0.112	0.09	-1.58	0.014	-1.57	0.014	0.118
31	E	189	0.47	0.000	189	0.48	0.108	0.08	-0.01	0.011	-0.01	0.011	0.103
31	O	189	0.42	0.232	189	0.48	0.225	0.23	-0.25	0.011	-0.25	0.011	0.103
32	E	654	0.66	0.012	654	0.50	0.136	0.10	1.67	0.004	1.67	0.004	0.064
33	E	1215	0.30	0.129	1215	0.52	0.215	0.18	-1.21	0.002	-1.20	0.002	0.044
34	E	1022	0.52	0.152	1022	0.55	0.175	0.16	-0.20	0.002	-0.20	0.002	0.044
35	E	1212	0.67	0.067	1212	0.57	0.160	0.12	0.77	0.002	0.77	0.002	0.042
36	E	805	0.53	0.059	805	0.62	0.180	0.13	-0.66	0.003	-0.66	0.003	0.051
37	E	1912	0.72	0.133	1912	0.62	0.200	0.17	0.61	0.001	0.61	0.001	0.033
38	E	1792	0.49	0.109	1792	0.52	0.180	0.15	-0.16	0.001	-0.16	0.001	0.033
39	E	4140	0.70	0.041	4140	0.60	0.182	0.13	0.79	0.001	0.79	0.001	0.023
40	E	949	0.67	0.097	949	0.61	0.186	0.15	0.42	0.002	0.42	0.002	0.046
41	E	1449	0.58	0.072	1449	0.47	0.144	0.11	0.95	0.002	0.95	0.002	0.039
42	E	1121	0.67	0.085	1121	0.58	0.179	0.14	0.68	0.002	0.68	0.002	0.043
43	E	729	0.50	0.063	729	0.51	0.119	0.10	-0.06	0.003	-0.06	0.003	0.052
44	E	981	0.31	0.127	981	0.46	0.148	0.14	-1.12	0.002	-1.12	0.002	0.049
45	E	1050	0.55	0.168	1050	0.55	0.184	0.18	0.00	0.002	0.00	0.002	0.044
46	E	781	0.71	0.097	781	0.56	0.164	0.13	1.12	0.003	1.12	0.003	0.054
47	E	1525	0.67	0.072	1525	0.57	0.236	0.17	0.59	0.001	0.59	0.001	0.037
48	E	1239	0.57	0.072	1239	0.51	0.162	0.13	0.49	0.002	0.49	0.002	0.041

ID	D	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
49	E	1404	0.37	0.183	1404	0.54	0.184	0.18	-0.91	0.002	-0.91	0.002	0.040
50	E	988	0.46	0.163	988	0.59	0.180	0.17	-0.72	0.002	-0.72	0.002	0.046
51	E	876	0.69	0.110	876	0.58	0.205	0.16	0.67	0.002	0.67	0.002	0.049
52	E	1280	0.38	0.033	1280	0.46	0.216	0.15	-0.54	0.002	-0.54	0.002	0.040
53	E	802	0.69	0.078	802	0.55	0.204	0.15	0.90	0.003	0.90	0.003	0.052
54	E	1193	0.61	0.177	1193	0.55	0.183	0.18	0.31	0.002	0.31	0.002	0.041
55	E	2378	0.40	0.102	2378	0.55	0.190	0.15	-0.99	0.001	-0.99	0.001	0.031
56	E	1093	0.43	0.150	1093	0.54	0.214	0.18	-0.56	0.002	-0.56	0.002	0.044
57	E	1014	0.54	0.119	1014	0.52	0.188	0.16	0.14	0.002	0.14	0.002	0.044
58	E	775	0.67	0.125	775	0.58	0.165	0.15	0.61	0.003	0.61	0.003	0.052
59	E	994	0.47	0.100	994	0.55	0.135	0.12	-0.64	0.002	-0.64	0.002	0.046
60	E	1200	0.43	0.153	1200	0.52	0.126	0.14	-0.67	0.002	-0.67	0.002	0.042
61	E	1029	0.58	0.153	1029	0.55	0.170	0.16	0.19	0.002	0.19	0.002	0.044
62	E	1030	0.34	0.153	1030	0.44	0.153	0.15	-0.68	0.002	-0.68	0.002	0.045
63	E	1473	0.34	0.077	1473	0.48	0.161	0.13	-1.13	0.002	-1.13	0.002	0.040
64	E	1381	0.66	0.079	1381	0.60	0.168	0.13	0.43	0.001	0.43	0.001	0.038
65	E	1067	0.58	0.122	1067	0.54	0.178	0.15	0.26	0.002	0.26	0.002	0.043
66	E	838	0.37	0.070	838	0.49	0.148	0.12	-0.99	0.003	-0.99	0.003	0.052
67	A	130	0.52	0.000	130	0.91	0.473	0.33	-1.17	0.018	-1.16	0.018	0.134
67	C	130	0.72	0.104	130	0.70	0.104	0.10	0.12	0.015	0.12	0.015	0.124
67	E	130	0.64	0.133	130	1.13	0.433	0.32	-1.52	0.020	-1.52	0.020	0.141
67	ES	130	1.00	0.123	130	0.50	0.118	0.12	4.19	0.049	4.18	0.049	0.222
68	A	301	0.90	0.210	301	1.02	0.178	0.19	-0.63	0.007	-0.63	0.007	0.084
68	C	300	0.90	0.175	300	0.82	0.146	0.16	0.47	0.007	0.47	0.007	0.083
68	E	301	0.91	0.233	301	0.99	0.144	0.19	-0.44	0.007	-0.44	0.007	0.083
68	ES	301	1.12	0.182	301	0.82	0.112	0.15	2.02	0.010	2.01	0.010	0.100
68	O	301	0.84	0.094	301	0.79	0.243	0.18	0.29	0.007	0.29	0.007	0.082
69	A	192	0.89	0.362	192	0.93	0.256	0.31	-0.11	0.010	-0.11	0.010	0.102
69	C	192	0.98	0.113	192	0.75	0.186	0.15	1.52	0.013	1.51	0.013	0.116
69	E	192	1.20	0.282	192	1.11	0.250	0.27	0.33	0.011	0.33	0.011	0.103
69	ES	192	1.00	0.190	192	0.71	0.111	0.16	1.85	0.015	1.85	0.015	0.122
69	O	192	0.90	0.078	192	0.84	0.253	0.19	0.32	0.011	0.32	0.011	0.103
70	A	223	0.86	0.278	223	0.80	0.276	0.28	0.20	0.009	0.20	0.009	0.095
70	C	223	0.63	0.154	223	0.66	0.089	0.13	-0.27	0.009	-0.27	0.009	0.095
70	E	223	1.06	0.256	223	0.98	0.053	0.18	0.44	0.009	0.44	0.009	0.096
70	ES	223	1.02	0.243	223	0.65	0.067	0.18	2.10	0.014	2.10	0.014	0.118
70	O	223	0.87	0.112	223	0.81	0.419	0.31	0.22	0.009	0.22	0.009	0.095
71	A	228	0.87	0.349	228	0.88	0.285	0.32	-0.05	0.009	-0.05	0.009	0.094
71	C	228	0.96	0.332	228	0.73	0.137	0.25	0.92	0.010	0.92	0.010	0.099
71	E	228	1.03	0.254	228	1.18	0.304	0.28	-0.52	0.009	-0.52	0.009	0.095
71	ES	228	1.25	0.224	228	0.71	0.116	0.18	3.03	0.019	3.02	0.019	0.137
71	O	228	0.70	0.208	228	0.79	0.344	0.28	-0.32	0.009	-0.32	0.009	0.094
72	A	501	0.88	0.168	501	0.77	0.165	0.17	0.70	0.004	0.70	0.004	0.065
72	C	501	0.75	0.058	501	0.72	0.105	0.08	0.38	0.004	0.38	0.004	0.064
72	E	501	0.77	0.178	501	0.88	0.163	0.17	-0.62	0.004	-0.62	0.004	0.065

ID	D	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
72	ES	501	0.84	0.149	501	0.83	0.106	0.13	0.03	0.004	0.03	0.004	0.063
72	O	501	0.83	0.118	501	0.83	0.136	0.13	-0.01	0.004	-0.01	0.004	0.063
74	E	1434	0.65	0.144	1434	0.62	0.194	0.17	0.16	0.001	0.16	0.001	0.037
76	A	310	1.02	0.370	310	0.95	0.339	0.35	0.20	0.006	0.20	0.006	0.081
76	C	310	1.10	0.215	310	0.78	0.089	0.16	1.99	0.010	1.99	0.010	0.098
76	E	310	1.21	0.300	310	1.28	0.182	0.25	-0.30	0.007	-0.30	0.007	0.081
76	ES	310	1.14	0.274	310	0.65	0.028	0.19	2.49	0.011	2.49	0.011	0.107
76	O	310	0.98	0.257	310	0.86	0.270	0.26	0.48	0.007	0.48	0.007	0.081
77	A	232	0.78	0.287	232	0.90	0.314	0.30	-0.39	0.009	-0.39	0.009	0.094
77	C	232	0.66	0.142	232	0.63	0.157	0.15	0.19	0.009	0.19	0.009	0.093
77	E	232	0.92	0.372	232	0.95	0.290	0.33	-0.11	0.009	-0.11	0.009	0.093
77	ES	232	0.99	0.205	232	0.81	0.000	0.14	1.22	0.010	1.22	0.010	0.101
77	O	232	0.78	0.085	232	0.80	0.355	0.26	-0.05	0.009	-0.05	0.009	0.093
Mean									-0.04	0.01	-0.04	0.01	0.08
SD									1.51	0.01	1.51	0.01	0.03

Note. ID=UniqueID, D=Dimension, Nn=Sample size of negatively worded items, Mn=Mean of negatively worded items, SDn=Standard deviation of negatively worded items, Np=Sample size of positively worded items, Mp=Mean of positively worded items, SDp=Standard deviation of positively worded items, s.within=Pooled standard deviation , d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

Table 19

Effect Sizes of Negated Effect for All Personality Studies

ID	D	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
2	A	545	0.39	0.06	545	2.08	0.83	0.59	-2.87	0.007	-2.87	0.007	0.086
2	O	545	0.22	0.03	545	2.98	1.77	1.25	-2.21	0.006	-2.21	0.006	0.077
14	A	439	1.10	0.42	439	1.13	0.61	0.52	-0.06	0.005	-0.06	0.005	0.068
14	E	439	1.03	0.21	439	1.28	0.35	0.29	-0.87	0.005	-0.87	0.005	0.071
14	ES	439	1.09	0.53	439	1.02	0.43	0.48	0.14	0.005	0.14	0.005	0.068
14	O	439	0.92	0.17	439	0.90	0.27	0.22	0.10	0.005	0.10	0.005	0.068
15	C	539	0.82	0.39	539	0.83	0.31	0.35	-0.02	0.004	-0.02	0.004	0.061
17	ES	706	0.73	0.27	706	0.56	0.25	0.26	0.65	0.003	0.65	0.003	0.055
19	C	1517	1.59	0.52	1517	1.47	0.32	0.43	0.27	0.001	0.27	0.001	0.036
20	A	538	0.35	0.00	538	0.50	0.08	0.06	-2.58	0.007	-2.58	0.007	0.082
20	ES	538	0.56	0.15	538	0.61	0.08	0.12	-0.38	0.004	-0.38	0.004	0.062
21	A	539	0.26	0.00	539	0.48	0.09	0.06	-3.46	0.009	-3.46	0.009	0.096
21	ES	539	0.56	0.11	539	0.62	0.08	0.10	-0.56	0.004	-0.56	0.004	0.062
22	A	478	0.60	0.00	478	0.57	0.12	0.09	0.37	0.004	0.37	0.004	0.065
22	C	478	0.65	0.11	478	0.61	0.08	0.10	0.42	0.004	0.42	0.004	0.065
22	E	478	0.80	0.00	478	0.73	0.16	0.11	0.60	0.004	0.60	0.004	0.066
22	ES	478	0.69	0.00	478	0.63	0.09	0.06	0.92	0.005	0.92	0.005	0.068
22	O	478	0.60	0.05	478	0.62	0.14	0.10	-0.13	0.004	-0.13	0.004	0.065
23	C	322	0.44	0.09	322	0.48	0.05	0.07	-0.64	0.007	-0.64	0.007	0.081
23	E	322	0.51	0.10	322	0.54	0.10	0.10	-0.31	0.006	-0.31	0.006	0.079
23	O	322	0.43	0.10	322	0.41	0.13	0.12	0.12	0.006	0.12	0.006	0.079
24	A	372	0.51	0.17	372	0.50	0.19	0.18	0.05	0.005	0.05	0.005	0.073
24	E	372	0.70	0.08	372	0.66	0.09	0.08	0.41	0.005	0.41	0.005	0.074
24	ES	372	0.72	0.09	372	0.57	0.13	0.11	1.30	0.007	1.30	0.006	0.081
24	O	372	0.48	0.14	372	0.59	0.08	0.12	-0.94	0.006	-0.93	0.006	0.077
25	A	328	0.55	0.19	328	0.55	0.17	0.18	-0.03	0.006	-0.03	0.006	0.078
25	E	328	0.66	0.10	328	0.66	0.08	0.09	-0.01	0.006	-0.01	0.006	0.078
25	ES	328	0.73	0.08	328	0.60	0.13	0.11	1.14	0.007	1.13	0.007	0.084
25	O	328	0.65	0.06	328	0.56	0.10	0.08	1.05	0.007	1.05	0.007	0.083
26	A	183	0.60	0.10	183	0.56	0.14	0.12	0.32	0.011	0.32	0.011	0.105
26	E	183	0.77	0.03	183	0.66	0.07	0.06	1.93	0.016	1.93	0.016	0.127
26	ES	183	0.63	0.07	183	0.55	0.09	0.08	0.92	0.012	0.92	0.012	0.110
26	O	183	0.54	0.04	183	0.50	0.12	0.09	0.47	0.011	0.47	0.011	0.106
27	A	202	0.48	0.12	202	0.57	0.11	0.12	-0.74	0.011	-0.74	0.011	0.103
27	E	202	0.62	0.13	202	0.59	0.08	0.10	0.26	0.010	0.26	0.010	0.100
27	ES	202	0.69	0.16	202	0.56	0.19	0.17	0.71	0.011	0.71	0.010	0.103
27	O	202	0.47	0.17	202	0.51	0.13	0.15	-0.27	0.010	-0.27	0.010	0.100
28	A	311	0.47	0.03	311	0.68	0.16	0.12	-1.79	0.009	-1.79	0.009	0.095
28	E	311	0.69	0.09	311	0.70	0.06	0.07	-0.19	0.006	-0.19	0.006	0.080
28	ES	311	0.69	0.19	311	0.61	0.19	0.19	0.38	0.007	0.38	0.007	0.081
28	O	311	0.21	0.07	311	0.53	0.14	0.11	-2.93	0.013	-2.93	0.013	0.116

ID	D	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
29	A	404	0.42	0.10	404	0.42	0.09	0.09	-0.03	0.005	-0.03	0.005	0.070
29	E	404	0.25	0.17	404	0.52	0.15	0.16	-1.70	0.007	-1.70	0 007	0.082
29	ES	404	0.49	0.07	404	0.57	0.10	0.08	-1.01	0.006	-1.01	0.006	0.075
29	O	404	0.49	0.04	404	0.42	0.10	0.08	0.84	0.005	0.84	0.005	0.073
30	A	189	0.58	0.11	189	0.52	0.15	0.13	0.43	0.011	0.43	0.011	0.104
30	E	189	0.65	0.15	189	0.68	0.07	0.12	-0.28	0.011	-0.28	0.011	0.103
30	ES	189	0.70	0.06	189	0.65	0.09	0.07	0.64	0.011	0.64	0.011	0.105
30	O	189	0.56	0.05	189	0.50	0.13	0.10	0.62	0.011	0.62	0.011	0.105
31	A	189	0.36	0.00	189	0.50	0.14	0.10	-1.43	0.013	-1.43	0.013	0.115
31	ES	189	0.67	0.09	189	0.62	0.09	0.09	0.46	0.011	0.46	0.011	0.104
32	ES	654	0.58	0.05	654	0.61	0.14	0.10	-0.25	0.003	-0.25	0.003	0.056
33	ES	1215	0.49	0.08	1215	0.48	0.16	0.13	0.11	0.002	0.11	0.002	0.041
34	ES	1022	0.58	0.08	1022	0.55	0.13	0.11	0.23	0.002	0.23	0.002	0.044
35	ES	1212	0.53	0.07	1212	0.58	0.10	0.09	-0.64	0.002	-0.64	0.002	0.042
36	ES	805	0.61	0.16	805	0.57	0.18	0.17	0.22	0.002	0.22	0.002	0.050
37	ES	1912	0.56	0.07	1912	0.56	0.13	0.11	-0.03	0.001	-0.03	0.001	0.032
38	ES	1792	0.51	0.15	1792	0.48	0.19	0.17	0.19	0.001	0.19	0.001	0.033
39	ES	4140	0.58	0.05	4140	0.61	0.11	0.09	-0.35	0.000	-0.35	0.000	0.022
40	ES	949	0.61	0.09	949	0.59	0.15	0.12	0.12	0.002	0.12	0.002	0.046
41	ES	1449	0.60	0.12	1449	0.58	0.10	0.11	0.25	0.001	0.25	0.001	0.037
42	ES	1121	0.61	0.04	1121	0.58	0.12	0.09	0.27	0.002	0.27	0.002	0.042
43	ES	729	0.50	0.12	729	0.56	0.12	0.12	-0.43	0.003	-0.43	0.003	0.053
44	ES	981	0.56	0.14	981	0.52	0.14	0.14	0.23	0.002	0.23	0.002	0.045
45	ES	1050	0.68	0.02	1050	0.61	0.13	0.09	0.74	0.002	0.74	0.002	0.045
46	ES	781	0.52	0.10	781	0.48	0.15	0.12	0.30	0.003	0.30	0.003	0.051
47	ES	1525	0.53	0.11	1525	0.53	0.12	0.12	-0.01	0.001	-0.01	0.001	0.036
48	ES	1239	0.52	0.12	1239	0.48	0.15	0.14	0.28	0.002	0.28	0.002	0.040
49	ES	1404	0.52	0.10	1404	0.52	0.15	0.13	-0.01	0.001	-0.01	0.001	0.038
50	ES	988	0.60	0.07	988	0.57	0.18	0.14	0.25	0.002	0.25	0.002	0.045
51	ES	876	0.59	0.08	876	0.61	0.15	0.12	-0.23	0.002	-0.22	0.002	0.048
52	ES	1280	0.59	0.19	1280	0.53	0.15	0.17	0.32	0.002	0.32	0.002	0.040
53	ES	802	0.54	0.06	802	0.59	0.17	0.13	-0.41	0.003	-0.41	0.003	0.050
54	ES	1193	0.55	0.12	1193	0.52	0.16	0.14	0.16	0.002	0.16	0.002	0.041
55	ES	2378	0.51	0.09	2378	0.50	0.19	0.15	0.06	0.001	0.06	0.001	0.029
56	ES	1093	0.55	0.08	1093	0.53	0.19	0.15	0.15	0.002	0.15	0.002	0.043
57	ES	1014	0.58	0.08	1014	0.54	0.11	0.10	0.38	0.002	0.38	0.002	0.045
58	ES	775	0.64	0.09	775	0.54	0.14	0.12	0.83	0.003	0.83	0.003	0.053
59	ES	994	0.55	0.05	994	0.58	0.13	0.10	-0.27	0.002	-0.27	0.002	0.045
60	ES	1200	0.49	0.11	1200	0.49	0.12	0.11	0.01	0.002	0.01	0.002	0.041
61	ES	1029	0.58	0.15	1029	0.57	0.16	0.15	0.05	0.002	0.05	0.002	0.044
62	ES	1030	0.52	0.12	1030	0.53	0.15	0.14	-0.06	0.002	-0.06	0.002	0.044
63	ES	1473	0.45	0.09	1473	0.45	0.10	0.10	-0.02	0.001	-0.02	0.001	0.037
64	ES	1381	0.58	0.10	1381	0.62	0.11	0.11	-0.39	0.001	-0.39	0.001	0.038
65	ES	1067	0.58	0.07	1067	0.52	0.16	0.12	0.51	0.002	0.51	0.002	0.044
66	ES	838	0.50	0.11	838	0.48	0.09	0.10	0.13	0.002	0.13	0.002	0.049

ID	D	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
67	A	130	0.47	0.07	130	0.96	0.47	0.34	-1.46	0.020	-1.46	0.019	0.140
67	C	130	0.82	0.13	130	0.69	0.09	0.11	1.21	0.018	1.20	0.018	0.135
67	E	130	1.46	0.00	130	0.97	0.43	0.30	1.64	0.021	1.64	0.020	0.143
67	ES	130	0.83	0.28	130	0.84	0.30	0.29	-0.06	0.015	-0.06	0.015	0.124
68	A	301	1.00	0.08	301	0.96	0.23	0.17	0.22	0.007	0.22	0.007	0.082
68	E	301	0.77	0.09	301	1.07	0.12	0.11	-2.82	0.013	-2.82	0.013	0.115
68	ES	301	0.96	0.32	301	1.08	0.20	0.26	-0.46	0.007	-0.46	0.007	0.083
68	O	301	0.80	0.09	301	0.80	0.23	0.17	-0.02	0.007	-0.02	0.007	0.082
69	A	192	1.05	0.19	192	0.85	0.31	0.26	0.79	0.011	0.79	0.011	0.106
69	E	192	1.02	0.34	192	1.24	0.15	0.27	-0.80	0.011	-0.80	0.011	0.106
69	ES	192	0.67	0.05	192	1.01	0.17	0.13	-2.68	0.020	-2.67	0.020	0.141
69	O	192	0.93	0.09	192	0.85	0.23	0.18	0.45	0.011	0.45	0.011	0.103
70	A	223	0.97	0.19	223	0.76	0.27	0.24	0.89	0.010	0.89	0.010	0.099
70	E	223	1.05	0.29	223	1.00	0.07	0.21	0.23	0.009	0.23	0.009	0.095
70	ES	223	0.76	0.09	223	1.00	0.28	0.21	-1.15	0.010	-1.15	0.010	0.102
70	O	223	0.91	0.14	223	0.81	0.39	0.29	0.34	0.009	0.34	0.009	0.095
71	A	228	1.00	0.28	228	0.82	0.30	0.29	0.60	0.009	0.59	0.009	0.096
71	E	228	0.87	0.10	228	1.26	0.24	0.18	-2.16	0.014	-2.15	0.014	0.118
71	ES	228	1.13	0.48	228	1.15	0.29	0.40	-0.04	0.009	-0.04	0.009	0.094
71	O	228	0.76	0.24	228	0.76	0.33	0.29	0.02	0.009	0.02	0.009	0.094
72	A	501	0.97	0.09	501	0.76	0.16	0.13	1.66	0.005	1.66	0.005	0.073
72	E	501	0.90	0.08	501	0.80	0.19	0.15	0.70	0.004	0.70	0.004	0.065
72	ES	501	0.87	0.02	501	0.82	0.16	0.11	0.42	0.004	0.42	0.004	0.064
72	O	501	0.87	0.15	501	0.82	0.13	0.14	0.35	0.004	0.35	0.004	0.064
74	ES	1434	0.57	0.12	1434	0.59	0.11	0.11	-0.22	0.001	-0.22	0.001	0.037
76	A	310	1.18	0.22	310	0.89	0.35	0.29	1.01	0.007	1.01	0.007	0.085
76	E	310	1.19	0.34	310	1.28	0.17	0.27	-0.34	0.007	-0.34	0.007	0.081
76	ES	310	0.90	0.32	310	1.07	0.33	0.32	-0.55	0.007	-0.55	0.007	0.082
76	O	310	1.06	0.32	310	0.86	0.25	0.29	0.70	0.007	0.70	0.007	0.083
77	A	232	0.91	0.13	232	0.82	0.35	0.26	0.35	0.009	0.35	0.009	0.094
77	E	232	0.83	0.37	232	1.01	0.28	0.33	-0.54	0.009	-0.54	0.009	0.095
77	ES	232	0.81	0.00	232	0.99	0.20	0.14	-1.22	0.010	-1.22	0.010	0.101
77	O	232	0.80	0.11	232	0.79	0.33	0.25	0.04	0.009	0.04	0.009	0.093
Mean									-0.09	0.01	-0.09	0.01	0.07
SD									0.95	0.00	0.95	0.00	0.03

Note. ID=UniqueID, D=Dimension, Nt=Sample size of negated items, Mt=Mean of negatied items, SDt=Standard deviation of negated items, Ns=Sample size of non-negated items, Ms=Mean of non-negated items, SDs=Standard deviation of non-negated items, s.within=Pooled standard deviation, d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

Table 20

Effect Sizes of Negation Type for All Personality Studies

ID	T	Nnt	Mnt	SDnt	Nnn	Mnn	SDnn	s.within	d	var.d	g	var.g	se.g
2	2	545	0.46	0.60	545	2.23	1.17	0.93	-1.90	0.005	-1.90	0.005	0.073
14	1	439	1.10	0.24	439	1.00	0.40	0.33	0.29	0.005	0.29	0.005	0.068
14	2	439	1.46	0.00	439	1.00	0.40	0.28	1.63	0.006	1.63	0.006	0.078
14	3	439	0.75	0.16	439	1.00	0.40	0.30	-0.82	0.005	-0.82	0.005	0.070
15	1	539	0.93	0.53	539	0.83	0.31	0.43	0.23	0.004	0.23	0.004	0.061
15	2	539	0.90	0.38	539	0.83	0.31	0.34	0.20	0.004	0.20	0.004	0.061
15	3	539	0.58	0.14	539	0.83	0.31	0.24	-1.05	0.004	-1.05	0.004	0.065
17	1	706	0.74	0.36	706	0.56	0.25	0.31	0.57	0.003	0.57	0.003	0.054
17	2	706	0.65	0.16	706	0.56	0.25	0.21	0.43	0.003	0.43	0.003	0.054
17	3	706	0.73	0.00	706	0.56	0.25	0.18	0.92	0.003	0.92	0.003	0.056
19	1	1517	1.96	0.00	1517	1.47	0.32	0.23	2.14	0.002	2.14	0.002	0.046
19	2	1517	1.22	0.00	1517	1.47	0.32	0.23	-1.11	0.002	-1.11	0.002	0.039
20	1	538	0.35	0.00	538	0.53	0.13	0.09	-1.98	0.006	-1.98	0.006	0.074
20	2	538	0.56	0.15	538	0.53	0.13	0.14	0.20	0.004	0.20	0.004	0.061
21	1	539	0.26	0.00	539	0.53	0.14	0.10	-2.75	0.007	-2.74	0.007	0.085
21	2	539	0.56	0.11	539	0.53	0.14	0.13	0.26	0.004	0.26	0.004	0.061
22	1	478	0.62	0.09	478	0.63	0.13	0.11	-0.05	0.004	-0.05	0.004	0.065
22	2	478	0.67	0.10	478	0.63	0.13	0.11	0.32	0.004	0.32	0.004	0.065
23	1	322	0.45	0.10	322	0.48	0.12	0.11	-0.29	0.006	-0.29	0.006	0.079
23	2	322	0.46	0.10	322	0.48	0.12	0.11	-0.19	0.006	-0.19	0.006	0.079
23	3	322	0.52	0.07	322	0.48	0.12	0.10	0.46	0.006	0.46	0.006	0.080
24	1	372	0.61	0.12	372	0.57	0.13	0.13	0.27	0.005	0.27	0.005	0.074
24	2	372	0.64	0.00	372	0.57	0.13	0.09	0.74	0.006	0.74	0.006	0.076
24	3	372	0.62	0.27	372	0.57	0.13	0.21	0.22	0.005	0.22	0.005	0.074
25	1	328	0.67	0.06	328	0.59	0.12	0.09	0.86	0.007	0.86	0.007	0.082
25	2	328	0.59	0.00	328	0.59	0.12	0.08	-0.06	0.006	-0.06	0.006	0.078
25	3	328	0.62	0.25	328	0.59	0.12	0.20	0.12	0.006	0.12	0.006	0.078
26	1	183	0.61	0.07	183	0.56	0.15	0.12	0.42	0.011	0.42	0.011	0.106
26	2	183	0.75	0.00	183	0.56	0.15	0.10	1.78	0.015	1.77	0.015	0.123
26	3	183	0.64	0.15	183	0.56	0.15	0.15	0.50	0.011	0.50	0.011	0.106
27	1	202	0.57	0.16	202	0.54	0.14	0.15	0.21	0.010	0.21	0.010	0.100
27	2	202	0.53	0.00	202	0.54	0.14	0.10	-0.10	0.010	-0.10	0.010	0.100
27	3	202	0.61	0.23	202	0.54	0.14	0.19	0.35	0.010	0.35	0.010	0.100
28	1	311	0.47	0.24	311	0.62	0.15	0.20	-0.75	0.007	-0.75	0.007	0.083
28	2	311	0.63	0.00	311	0.62	0.15	0.10	0.04	0.006	0.04	0.006	0.080
28	3	311	0.66	0.17	311	0.62	0.15	0.16	0.23	0.006	0.23	0.006	0.080
29	1	404	0.48	0.05	404	0.48	0.15	0.11	-0.03	0.005	-0.03	0.005	0.070
29	2	404	0.13	0.00	404	0.48	0.15	0.11	-3.31	0.012	-3.30	0.012	0.108
29	3	404	0.40	0.10	404	0.48	0.15	0.13	-0.66	0.005	-0.66	0.005	0.072
30	1	189	0.64	0.08	189	0.59	0.12	0.10	0.52	0.011	0.52	0.011	0.105
30	2	189	0.54	0.00	189	0.59	0.12	0.09	-0.53	0.011	-0.52	0.011	0.105

ID	T	Nnt	Mnt	SDnt	Nnn	Mnn	SDnn	s.within	d	var.d	g	var.g	se.g
30	3	189	0.64	0.16	189	0.59	0.12	0.14	0.36	0.011	0.36	0.011	0.104
31	1	189	0.36	0.00	189	0.52	0.15	0.11	-1.53	0.014	-1.53	0.014	0.117
31	2	189	0.67	0.09	189	0.52	0.15	0.12	1.16	0.012	1.16	0.012	0.111
32	1	654	0.61	0.04	654	0.56	0.14	0.10	0.52	0.003	0.52	0.003	0.056
32	2	654	0.54	0.00	654	0.56	0.14	0.10	-0.25	0.003	-0.25	0.003	0.056
33	1	1215	0.52	0.10	1215	0.48	0.19	0.15	0.26	0.002	0.26	0.002	0.041
33	2	1215	0.44	0.01	1215	0.48	0.19	0.13	-0.29	0.002	-0.29	0.002	0.041
34	1	1022	0.59	0.02	1022	0.55	0.15	0.11	0.35	0.002	0.34	0.002	0.045
34	2	1022	0.57	0.16	1022	0.55	0.15	0.15	0.10	0.002	0.10	0.002	0.044
35	1	1212	0.57	0.08	1212	0.58	0.13	0.11	-0.11	0.002	-0.11	0.002	0.041
35	2	1212	0.46	0.03	1212	0.58	0.13	0.09	-1.32	0.002	-1.32	0.002	0.045
36	1	805	0.68	0.08	805	0.59	0.17	0.13	0.66	0.003	0.66	0.003	0.051
36	2	805	0.45	0.06	805	0.59	0.17	0.13	-1.11	0.003	-1.11	0.003	0.054
37	1	1912	0.56	0.02	1912	0.60	0.17	0.12	-0.29	0.001	-0.29	0.001	0.033
37	2	1912	0.49	0.01	1912	0.60	0.17	0.12	-0.89	0.001	-0.89	0.001	0.034
38	1	1792	0.51	0.16	1792	0.50	0.18	0.17	0.05	0.001	0.05	0.001	0.033
38	2	1792	0.45	0.19	1792	0.50	0.18	0.18	-0.27	0.001	-0.27	0.001	0.034
39	1	4140	0.62	0.01	4140	0.61	0.15	0.10	0.08	0.000	0.08	0.000	0.022
39	2	4140	0.53	0.04	4140	0.61	0.15	0.11	-0.74	0.001	-0.74	0.001	0.023
40	1	949	0.58	0.05	949	0.61	0.16	0.12	-0.25	0.002	-0.25	0.002	0.046
40	2	949	0.58	0.12	949	0.61	0.16	0.14	-0.16	0.002	-0.16	0.002	0.046
41	1	1449	0.60	0.19	1449	0.53	0.13	0.16	0.46	0.001	0.46	0.001	0.038
41	2	1449	0.56	0.06	1449	0.53	0.13	0.10	0.26	0.001	0.26	0.001	0.037
42	1	1121	0.62	0.08	1121	0.59	0.15	0.12	0.31	0.002	0.31	0.002	0.042
42	2	1121	0.59	0.01	1121	0.59	0.15	0.10	0.00	0.002	0.00	0.002	0.042
43	1	729	0.52	0.16	729	0.53	0.12	0.14	-0.10	0.003	-0.10	0.003	0.052
43	2	729	0.43	0.10	729	0.53	0.12	0.11	-0.93	0.003	-0.93	0.003	0.055
44	1	981	0.52	0.26	981	0.48	0.15	0.22	0.19	0.002	0.19	0.002	0.045
44	2	981	0.56	0.02	981	0.48	0.15	0.11	0.79	0.002	0.79	0.002	0.047
45	1	1050	0.66	0.01	1050	0.58	0.16	0.11	0.77	0.002	0.77	0.002	0.045
45	2	1050	0.67	0.00	1050	0.58	0.16	0.11	0.81	0.002	0.81	0.002	0.045
46	1	781	0.50	0.11	781	0.53	0.16	0.14	-0.22	0.003	-0.22	0.003	0.051
46	2	781	0.49	0.13	781	0.53	0.16	0.15	-0.30	0.003	-0.30	0.003	0.051
47	1	1525	0.62	0.09	1525	0.56	0.18	0.14	0.41	0.001	0.41	0.001	0.037
47	2	1525	0.43	0.06	1525	0.56	0.18	0.13	-0.99	0.001	-0.99	0.001	0.038
48	1	1239	0.55	0.07	1239	0.50	0.15	0.12	0.39	0.002	0.39	0.002	0.041
48	2	1239	0.46	0.21	1239	0.50	0.15	0.18	-0.26	0.002	-0.26	0.002	0.040
49	1	1404	0.46	0.16	1404	0.52	0.17	0.16	-0.34	0.001	-0.34	0.001	0.038
49	2	1404	0.54	0.03	1404	0.52	0.17	0.12	0.17	0.001	0.17	0.001	0.038
50	1	988	0.59	0.07	988	0.57	0.18	0.13	0.16	0.002	0.16	0.002	0.045
50	2	988	0.57	0.03	988	0.57	0.18	0.13	-0.02	0.002	-0.02	0.002	0.045
51	1	876	0.55	0.03	876	0.61	0.17	0.12	-0.44	0.002	-0.44	0.002	0.048
51	2	876	0.56	0.04	876	0.61	0.17	0.13	-0.38	0.002	-0.38	0.002	0.048
52	1	1280	0.53	0.28	1280	0.49	0.18	0.24	0.18	0.002	0.18	0.002	0.040
52	2	1280	0.54	0.04	1280	0.49	0.18	0.13	0.36	0.002	0.36	0.002	0.040

ID	T	Nnt	Mnt	SDnt	Nnn	Mnn	SDnn	s.within	d	var.d	g	var.g	se.g
53	1	802	0.56	0.02	802	0.58	0.18	0.13	-0.14	0.002	-0.14	0.002	0.050
53	2	802	0.49	0.05	802	0.58	0.18	0.14	-0.68	0.003	-0.68	0.003	0.051
54	1	1193	0.50	0.06	1193	0.54	0.17	0.13	-0.33	0.002	-0.33	0.002	0.041
54	2	1193	0.51	0.15	1193	0.54	0.17	0.16	-0.20	0.002	-0.20	0.002	0.041
55	1	2378	0.47	0.00	2378	0.52	0.19	0.13	-0.37	0.001	-0.37	0.001	0.029
55	2	2378	0.48	0.04	2378	0.52	0.19	0.13	-0.26	0.001	-0.26	0.001	0.029
56	1	1093	0.63	0.01	1093	0.53	0.20	0.14	0.76	0.002	0.76	0.002	0.044
56	2	1093	0.48	0.04	1093	0.53	0.20	0.14	-0.32	0.002	-0.32	0.002	0.043
57	1	1014	0.61	0.12	1014	0.53	0.15	0.14	0.56	0.002	0.56	0.002	0.045
57	2	1014	0.53	0.07	1014	0.53	0.15	0.12	0.00	0.002	0.00	0.002	0.044
58	1	775	0.65	0.09	775	0.57	0.15	0.12	0.63	0.003	0.63	0.003	0.052
58	2	775	0.60	0.12	775	0.57	0.15	0.14	0.23	0.003	0.23	0.003	0.051
59	1	994	0.55	0.03	994	0.56	0.13	0.10	-0.08	0.002	-0.08	0.002	0.045
59	2	994	0.52	0.00	994	0.56	0.13	0.09	-0.39	0.002	-0.39	0.002	0.045
60	1	1200	0.54	0.14	1200	0.50	0.13	0.13	0.30	0.002	0.30	0.002	0.041
60	2	1200	0.43	0.10	1200	0.50	0.13	0.12	-0.63	0.002	-0.63	0.002	0.042
61	1	1029	0.59	0.11	1029	0.56	0.16	0.14	0.25	0.002	0.25	0.002	0.044
61	2	1029	0.48	0.19	1029	0.56	0.16	0.17	-0.44	0.002	-0.44	0.002	0.045
62	1	1030	0.54	0.14	1030	0.48	0.16	0.15	0.44	0.002	0.44	0.002	0.045
62	2	1030	0.43	0.03	1030	0.48	0.16	0.12	-0.37	0.002	-0.37	0.002	0.044
63	1	1473	0.39	0.16	1473	0.45	0.13	0.15	-0.43	0.001	-0.43	0.001	0.037
63	2	1473	0.47	0.01	1473	0.45	0.13	0.10	0.21	0.001	0.21	0.001	0.037
64	1	1381	0.67	0.02	1381	0.62	0.14	0.10	0.55	0.002	0.55	0.002	0.039
64	2	1381	0.48	0.06	1381	0.62	0.14	0.11	-1.24	0.002	-1.24	0.002	0.042
65	1	1067	0.54	0.08	1067	0.53	0.16	0.13	0.07	0.002	0.07	0.002	0.043
65	2	1067	0.61	0.09	1067	0.53	0.16	0.13	0.60	0.002	0.60	0.002	0.044
66	1	838	0.46	0.19	838	0.48	0.12	0.16	-0.14	0.002	-0.14	0.002	0.049
66	2	838	0.48	0.00	838	0.48	0.12	0.09	0.07	0.002	0.07	0.002	0.049
67	1	130	0.75	0.38	130	0.91	0.37	0.37	-0.44	0.016	-0.44	0.016	0.126
67	2	130	0.81	0.11	130	0.91	0.37	0.27	-0.36	0.016	-0.36	0.016	0.125
67	3	130	1.07	0.13	130	0.91	0.37	0.27	0.58	0.016	0.57	0.016	0.127
68	1	301	0.86	0.14	301	0.97	0.22	0.19	-0.60	0.007	-0.60	0.007	0.083
68	2	301	1.19	0.00	301	0.97	0.22	0.16	1.35	0.008	1.35	0.008	0.090
68	3	301	0.79	0.10	301	0.97	0.22	0.17	-1.06	0.008	-1.06	0.008	0.087
69	1	192	1.01	0.28	192	0.94	0.25	0.26	0.26	0.011	0.26	0.010	0.103
69	2	192	0.71	0.00	192	0.94	0.25	0.18	-1.32	0.013	-1.32	0.013	0.113
69	3	192	0.89	0.22	192	0.94	0.25	0.24	-0.21	0.010	-0.21	0.010	0.102
70	1	223	1.04	0.23	223	0.83	0.28	0.25	0.85	0.010	0.84	0.010	0.099
70	2	223	0.82	0.00	223	0.83	0.28	0.20	-0.03	0.009	-0.03	0.009	0.095
70	3	223	0.78	0.08	223	0.83	0.28	0.20	-0.24	0.009	-0.24	0.009	0.095
71	1	228	0.93	0.21	228	0.94	0.33	0.28	-0.06	0.009	-0.06	0.009	0.094
71	2	228	1.47	0.00	228	0.94	0.33	0.24	2.23	0.014	2.22	0.014	0.119
71	3	228	0.77	0.07	228	0.94	0.33	0.24	-0.72	0.009	-0.71	0.009	0.097
72	1	501	0.89	0.10	501	0.78	0.15	0.12	0.82	0.004	0.82	0.004	0.066
72	2	501	0.90	0.06	501	0.78	0.15	0.11	1.05	0.005	1.05	0.005	0.067

ID	T	Nnt	Mnt	SDnt	Nnn	Mnn	SDnn	s.within	d	var.d	g	var.g	se.g
72	3	501	0.94	0.09	501	0.78	0.15	0.12	1.31	0.005	1.31	0.005	0.070
74	1	1434	0.65	0.01	1434	0.61	0.15	0.11	0.34	0.001	0.34	0.001	0.038
74	2	1434	0.48	0.17	1434	0.61	0.15	0.16	-0.82	0.002	-0.82	0.002	0.039
76	1	310	1.24	0.25	310	0.98	0.30	0.27	0.92	0.007	0.92	0.007	0.084
76	2	310	1.12	0.00	310	0.98	0.30	0.21	0.64	0.007	0.64	0.007	0.082
76	3	310	0.81	0.14	310	0.98	0.30	0.23	-0.74	0.007	-0.74	0.007	0.083
77	1	232	0.90	0.24	232	0.83	0.29	0.26	0.27	0.009	0.27	0.009	0.093
77	2	232	0.81	0.00	232	0.83	0.29	0.20	-0.11	0.009	-0.11	0.009	0.093
77	3	232	0.71	0.18	232	0.83	0.29	0.24	-0.50	0.009	-0.50	0.009	0.094
Mean									-0.01	0.005	-0.01	0.005	0.063
SD									0.78	0.004	0.78	0.004	0.026

Note. ID=UniqueID, T=Negation Type (1=Not, 2=Negative prefix, 3=Negative qualifier), Nnt=Sample size of negated items, Mnt=Mean of negatied items, SDnt=Standard deviation of negated items, Nnn=Sample size of non-negated items, Mnn=Mean of non-negated items, SDnn=Standard deviation of non-negated items, s.within=Pooled standard deviation, d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

Table 21

Effect Sizes of the Negatively-Worded Item Effect for All Studies

ID	D	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
1	A	1	14021	0.27	0.00	14021	0.71	0.30	0.22	-2.02	0.000	-2.02	0.000	0.015
1	C	1	14021	0.31	0.00	14021	0.71	0.24	0.17	-2.35	0.000	-2.35	0.000	0.016
1	E	1	14021	0.24	0.00	14021	0.73	0.13	0.10	-5.11	0.001	-5.11	0.001	0.025
1	ES	1	14021	0.71	0.02	14021	0.55	0.00	0.02	10.33	0.002	10.33	0.002	0.045
2	A	1	545	0.39	0.06	545	2.08	0.83	0.59	-2.87	0.007	-2.87	0.007	0.086
2	O	1	545	0.22	0.03	545	2.98	1.77	1.25	-2.21	0.006	-2.21	0.006	0.077
3	R	0	939	0.76	0.12	939	0.66	0.11	0.12	0.83	0.002	0.83	0.002	0.048
4	R	0	343	0.76	0.15	343	0.65	0.14	0.14	0.72	0.006	0.72	0.006	0.079
5	G	0	1967	0.26	0.97	1967	-0.26	0.55	0.79	0.65	0.001	0.65	0.001	0.033
6	R	0	757	0.65	0.06	757	0.64	0.04	0.05	0.16	0.003	0.16	0.003	0.051
7	R	0	852	0.67	0.10	852	0.57	0.09	0.09	1.04	0.003	1.04	0.003	0.052
8	R	0	858	0.65	0.10	858	0.55	0.08	0.09	0.96	0.003	0.96	0.003	0.051
9	G12	0	169	-0.65	0.13	169	0.52	0.08	0.11	-11.07	0.193	-11.04	0.192	0.439
10	R	0	420	-0.38	0.52	420	-0.24	0.48	0.50	-0.28	0.005	-0.28	0.005	0.069
11	RW	0	545	0.53	0.07	545	0.63	0.09	0.08	-1.16	0.004	-1.16	0.004	0.065
12	CF	0	590	0.53	0.15	590	0.47	0.16	0.16	0.41	0.003	0.41	0.003	0.059
14	A	1	439	0.93	0.48	439	1.25	0.57	0.53	-0.62	0.005	-0.61	0.005	0.069
14	C	1	439	0.72	0.23	439	0.88	0.20	0.22	-0.71	0.005	-0.71	0.005	0.070
14	E	1	439	1.30	0.48	439	1.12	0.25	0.38	0.45	0.005	0.45	0.005	0.068
14	ES	1	439	1.09	0.45	439	0.79	0.11	0.33	0.92	0.005	0.92	0.005	0.071
14	O	1	439	0.84	0.19	439	0.93	0.27	0.23	-0.40	0.005	-0.40	0.005	0.068
15	C	1	539	0.91	0.37	539	0.75	0.33	0.35	0.46	0.004	0.46	0.004	0.062
16	LR	0	469	1.65	0.23	469	1.78	0.42	0.34	-0.39	0.004	-0.39	0.004	0.066
18	E	1	201	0.20	0.00	201	0.92	0.54	0.38	-1.87	0.014	-1.86	0.014	0.120
19	C	1	1517	1.59	0.52	1517	1.47	0.32	0.43	0.27	0.001	0.27	0.001	0.036
20	A	1	538	0.47	0.11	538	0.53	0.05	0.08	-0.65	0.004	-0.65	0.004	0.063
20	C	1	538	0.71	0.14	538	0.55	0.06	0.11	1.47	0.005	1.47	0.005	0.069
20	E	1	538	0.58	0.00	538	0.52	0.11	0.08	0.82	0.004	0.82	0.004	0.063
20	O	1	538	0.39	0.03	538	0.50	0.23	0.16	-0.68	0.004	-0.68	0.004	0.063
21	A	1	539	0.44	0.11	539	0.51	0.10	0.11	-0.59	0.004	-0.59	0.004	0.062
21	C	1	539	0.71	0.07	539	0.57	0.06	0.07	2.20	0.006	2.20	0.006	0.077
21	E	1	539	0.54	0.00	539	0.52	0.12	0.08	0.30	0.004	0.30	0.004	0.061
21	O	1	539	0.36	0.08	539	0.50	0.24	0.18	-0.79	0.004	-0.79	0.004	0.063
22	A	1	478	0.66	0.09	478	0.51	0.08	0.09	1.67	0.006	1.67	0.006	0.075
22	C	1	478	0.64	0.10	478	0.61	0.09	0.09	0.31	0.004	0.31	0.004	0.065
22	E	1	478	0.73	0.06	478	0.74	0.19	0.14	-0.10	0.004	-0.10	0.004	0.065
22	ES	1	478	0.62	0.10	478	0.67	0.04	0.08	-0.66	0.004	-0.66	0.004	0.066
22	O	1	478	0.56	0.00	478	0.62	0.13	0.09	-0.65	0.004	-0.65	0.004	0.066
23	C	1	322	0.48	0.07	322	0.46	0.07	0.07	0.22	0.006	0.22	0.006	0.079
23	E	1	322	0.51	0.11	322	0.55	0.10	0.10	-0.35	0.006	-0.35	0.006	0.079
23	O	1	322	0.44	0.12	322	0.41	0.12	0.12	0.24	0.006	0.24	0.006	0.079

ID	D	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
24	A	1	372	0.42	0.22	372	0.56	0.12	0.18	-0.75	0.006	-0.74	0.006	0.076
24	C	1	372	0.55	0.18	372	0.52	0.12	0.15	0.20	0.005	0.20	0.005	0.074
24	E	1	372	0.65	0.08	372	0.75	0.01	0.06	-1.68	0.007	-1.68	0.007	0.085
24	ES	1	372	0.57	0.22	372	0.65	0.10	0.17	-0.50	0.006	-0.50	0.006	0.074
24	O	1	372	0.52	0.11	372	0.59	0.09	0.10	-0.77	0.006	-0.77	0.006	0.076
25	A	1	328	0.50	0.19	328	0.58	0.16	0.17	-0.50	0.006	-0.50	0.006	0.079
25	C	1	328	0.59	0.10	328	0.59	0.11	0.10	-0.02	0.006	-0.02	0.006	0.078
25	E	1	328	0.64	0.06	328	0.76	0.03	0.05	-2.34	0.010	-2.34	0.010	0.101
25	ES	1	328	0.60	0.18	328	0.68	0.10	0.14	-0.54	0.006	-0.54	0.006	0.080
25	O	1	328	0.67	0.05	328	0.54	0.10	0.08	1.60	0.008	1.60	0.008	0.090
26	A	1	183	0.55	0.13	183	0.59	0.13	0.13	-0.31	0.011	-0.31	0.011	0.105
26	C	1	183	0.67	0.23	183	0.45	0.16	0.20	1.12	0.013	1.12	0.013	0.113
26	E	1	183	0.66	0.07	183	0.77	0.03	0.06	-1.89	0.016	-1.88	0.016	0.126
26	ES	1	183	0.53	0.10	183	0.60	0.08	0.09	-0.73	0.012	-0.73	0.012	0.108
26	O	1	183	0.54	0.03	183	0.50	0.13	0.09	0.47	0.011	0.47	0.011	0.106
27	A	1	202	0.46	0.11	202	0.60	0.08	0.10	-1.38	0.012	-1.38	0.012	0.111
27	C	1	202	0.54	0.29	202	0.47	0.09	0.22	0.31	0.010	0.31	0.010	0.100
27	E	1	202	0.59	0.08	202	0.65	0.08	0.08	-0.85	0.011	-0.85	0.011	0.104
27	ES	1	202	0.54	0.22	202	0.65	0.16	0.20	-0.54	0.010	-0.54	0.010	0.101
27	O	1	202	0.46	0.12	202	0.51	0.14	0.13	-0.41	0.010	-0.41	0.010	0.101
28	A	1	311	0.43	0.07	311	0.73	0.06	0.07	-4.43	0.022	-4.42	0.022	0.149
28	C	1	311	0.51	0.15	311	0.67	0.06	0.12	-1.34	0.008	-1.34	0.008	0.089
28	E	1	311	0.68	0.05	311	0.77	0.03	0.04	-2.11	0.010	-2.11	0.010	0.100
28	ES	1	311	0.52	0.28	311	0.69	0.13	0.22	-0.79	0.007	-0.79	0.007	0.083
28	O	1	311	0.24	0.08	311	0.56	0.12	0.10	-3.12	0.014	-3.11	0.014	0.119
29	A	1	404	0.41	0.09	404	0.43	0.09	0.09	-0.27	0.005	-0.27	0.005	0.071
29	C	1	404	0.68	0.16	404	0.36	0.13	0.15	2.21	0.008	2.20	0.008	0.089
29	E	1	404	0.46	0.19	404	0.49	0.17	0.18	-0.15	0.005	-0.15	0.005	0.070
29	ES	1	404	0.47	0.09	404	0.56	0.09	0.09	-1.04	0.006	-1.04	0.006	0.075
29	O	1	404	0.46	0.05	404	0.42	0.11	0.09	0.51	0.005	0.51	0.005	0.072
30	A	1	189	0.55	0.11	189	0.54	0.16	0.14	0.09	0.011	0.09	0.011	0.103
30	C	1	189	0.60	0.10	189	0.58	0.10	0.10	0.14	0.011	0.14	0.011	0.103
30	E	1	189	0.65	0.08	189	0.75	0.00	0.06	-1.83	0.015	-1.83	0.015	0.123
30	ES	1	189	0.67	0.08	189	0.67	0.08	0.08	0.00	0.011	0.00	0.011	0.103
30	O	1	189	0.53	0.06	189	0.50	0.14	0.11	0.29	0.011	0.29	0.011	0.103
31	A	1	189	0.49	0.15	189	0.48	0.12	0.14	0.06	0.011	0.06	0.011	0.103
31	C	1	189	0.45	0.07	189	0.59	0.11	0.09	-1.58	0.014	-1.57	0.014	0.118
31	E	1	189	0.47	0.00	189	0.48	0.11	0.08	-0.01	0.011	-0.01	0.011	0.103
31	O	1	189	0.42	0.23	189	0.48	0.23	0.23	-0.25	0.011	-0.25	0.011	0.103
32	E	1	654	0.66	0.01	654	0.50	0.14	0.10	1.67	0.004	1.67	0.004	0.064
33	E	1	1215	0.30	0.13	1215	0.52	0.21	0.18	-1.21	0.002	-1.20	0.002	0.044
34	E	1	1022	0.52	0.15	1022	0.55	0.17	0.16	-0.20	0.002	-0.20	0.002	0.044
35	E	1	1212	0.67	0.07	1212	0.57	0.16	0.12	0.77	0.002	0.77	0.002	0.042
36	E	1	805	0.53	0.06	805	0.62	0.18	0.13	-0.66	0.003	-0.66	0.003	0.051
37	E	1	1912	0.72	0.13	1912	0.62	0.20	0.17	0.61	0.001	0.61	0.001	0.033

ID	D	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
38	E	1	1792	0.49	0.11	1792	0.52	0.18	0.15	-0.16	0.001	-0.16	0.001	0.033
39	E	1	4140	0.70	0.04	4140	0.60	0.18	0.13	0.79	0.001	0.79	0.001	0.023
40	E	1	949	0.67	0.10	949	0.61	0.19	0.15	0.42	0.002	0.42	0.002	0.046
41	E	1	1449	0.58	0.07	1449	0.47	0.14	0.11	0.95	0.002	0.95	0.002	0.039
42	E	1	1121	0.67	0.08	1121	0.58	0.18	0.14	0.68	0.002	0.68	0.002	0.043
43	E	1	729	0.50	0.06	729	0.51	0.12	0.10	-0.06	0.003	-0.06	0.003	0.052
44	E	1	981	0.31	0.13	981	0.46	0.15	0.14	-1.12	0.002	-1.12	0.002	0.049
45	E	1	1050	0.55	0.17	1050	0.55	0.18	0.18	0.00	0.002	0.00	0.002	0.044
46	E	1	781	0.71	0.10	781	0.56	0.16	0.13	1.12	0.003	1.12	0.003	0.054
47	E	1	1525	0.67	0.07	1525	0.57	0.24	0.17	0.59	0.001	0.59	0.001	0.037
48	E	1	1239	0.57	0.07	1239	0.51	0.16	0.13	0.49	0.002	0.49	0.002	0.041
49	E	1	1404	0.37	0.18	1404	0.54	0.18	0.18	-0.91	0.002	-0.91	0.002	0.040
50	E	1	988	0.46	0.16	988	0.59	0.18	0.17	-0.72	0.002	-0.72	0.002	0.046
51	E	1	876	0.69	0.11	876	0.58	0.21	0.16	0.67	0.002	0.67	0.002	0.049
52	E	1	1280	0.38	0.03	1280	0.46	0.22	0.15	-0.54	0.002	-0.54	0.002	0.040
53	E	1	802	0.69	0.08	802	0.55	0.20	0.15	0.90	0.003	0.90	0.003	0.052
54	E	1	1193	0.61	0.18	1193	0.55	0.18	0.18	0.31	0.002	0.31	0.002	0.041
55	E	1	2378	0.40	0.10	2378	0.55	0.19	0.15	-0.99	0.001	-0.99	0.001	0.031
56	E	1	1093	0.43	0.15	1093	0.54	0.21	0.18	-0.56	0.002	-0.56	0.002	0.044
57	E	1	1014	0.54	0.12	1014	0.52	0.19	0.16	0.14	0.002	0.14	0.002	0.044
58	E	1	775	0.67	0.12	775	0.58	0.16	0.15	0.61	0.003	0.61	0.003	0.052
59	E	1	994	0.47	0.10	994	0.55	0.13	0.12	-0.64	0.002	-0.64	0.002	0.046
60	E	1	1200	0.43	0.15	1200	0.52	0.13	0.14	-0.67	0.002	-0.67	0.002	0.042
61	E	1	1029	0.58	0.15	1029	0.55	0.17	0.16	0.19	0.002	0.19	0.002	0.044
62	E	1	1030	0.34	0.15	1030	0.44	0.15	0.15	-0.68	0.002	-0.68	0.002	0.045
63	E	1	1473	0.34	0.08	1473	0.48	0.16	0.13	-1.13	0.002	-1.13	0.002	0.040
64	E	1	1381	0.66	0.08	1381	0.60	0.17	0.13	0.43	0.001	0.43	0.001	0.038
65	E	1	1067	0.58	0.12	1067	0.54	0.18	0.15	0.26	0.002	0.26	0.002	0.043
66	E	1	838	0.37	0.07	838	0.49	0.15	0.12	-0.99	0.003	-0.99	0.003	0.052
67	A	1	130	0.52	0.00	130	0.91	0.47	0.33	-1.17	0.018	-1.16	0.018	0.134
67	C	1	130	0.72	0.10	130	0.70	0.10	0.10	0.12	0.015	0.12	0.015	0.124
67	E	1	130	0.64	0.13	130	1.13	0.43	0.32	-1.52	0.020	-1.52	0.020	0.141
67	ES	1	130	1.00	0.12	130	0.50	0.12	0.12	4.19	0.049	4.18	0.049	0.222
68	A	1	301	0.90	0.21	301	1.02	0.18	0.19	-0.63	0.007	-0.63	0.007	0.084
68	C	1	300	0.90	0.18	300	0.82	0.15	0.16	0.47	0.007	0.47	0.007	0.083
68	E	1	301	0.91	0.23	301	0.99	0.14	0.19	-0.44	0.007	-0.44	0.007	0.083
68	ES	1	301	1.12	0.18	301	0.82	0.11	0.15	2.02	0.010	2.01	0.010	0.100
68	O	1	301	0.84	0.09	301	0.79	0.24	0.18	0.29	0.007	0.29	0.007	0.082
69	A	1	192	0.89	0.36	192	0.93	0.26	0.31	-0.11	0.010	-0.11	0.010	0.102
69	C	1	192	0.98	0.11	192	0.75	0.19	0.15	1.52	0.013	1.51	0.013	0.116
69	E	1	192	1.20	0.28	192	1.11	0.25	0.27	0.33	0.011	0.33	0.011	0.103
69	ES	1	192	1.00	0.19	192	0.71	0.11	0.16	1.85	0.015	1.85	0.015	0.122
69	O	1	192	0.90	0.08	192	0.84	0.25	0.19	0.32	0.011	0.32	0.011	0.103
70	A	1	223	0.86	0.28	223	0.80	0.28	0.28	0.20	0.009	0.20	0.009	0.095
70	C	1	223	0.63	0.15	223	0.66	0.09	0.13	-0.27	0.009	-0.27	0.009	0.095

ID	D	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
70	E	1	223	1.06	0.26	223	0.98	0.05	0.18	0.44	0.009	0.44	0.009	0.096
70	ES	1	223	1.02	0.24	223	0.65	0.07	0.18	2.10	0.014	2.10	0.014	0.118
70	O	1	223	0.87	0.11	223	0.81	0.42	0.31	0.22	0.009	0.22	0.009	0.095
71	A	1	228	0.87	0.35	228	0.88	0.28	0.32	-0.05	0.009	-0.05	0.009	0.094
71	C	1	228	0.96	0.33	228	0.73	0.14	0.25	0.92	0.010	0.92	0.010	0.099
71	E	1	228	1.03	0.25	228	1.18	0.30	0.28	-0.52	0.009	-0.52	0.009	0.095
71	ES	1	228	1.25	0.22	228	0.71	0.12	0.18	3.03	0.019	3.02	0.019	0.137
71	O	1	228	0.70	0.21	228	0.79	0.34	0.28	-0.32	0.009	-0.32	0.009	0.094
72	A	1	501	0.88	0.17	501	0.77	0.16	0.17	0.70	0.004	0.70	0.004	0.065
72	C	1	501	0.75	0.06	501	0.72	0.10	0.08	0.38	0.004	0.38	0.004	0.064
72	E	1	501	0.77	0.18	501	0.88	0.16	0.17	-0.62	0.004	-0.62	0.004	0.065
72	ES	1	501	0.84	0.15	501	0.83	0.11	0.13	0.03	0.004	0.03	0.004	0.063
72	O	1	501	0.83	0.12	501	0.83	0.14	0.13	-0.01	0.004	-0.01	0.004	0.063
73	Ep	0	972	0.41	0.08	972	0.56	0.11	0.10	-1.50	0.003	-1.50	0.003	0.051
74	E	1	1434	0.65	0.14	1434	0.62	0.19	0.17	0.16	0.001	0.16	0.001	0.037
75	L	0	389	1.01	0.06	389	0.35	0.32	0.23	2.87	0.010	2.87	0.010	0.102
76	A	1	310	1.02	0.37	310	0.95	0.34	0.35	0.20	0.006	0.20	0.006	0.081
76	C	1	310	1.10	0.21	310	0.78	0.09	0.16	1.99	0.010	1.99	0.010	0.098
76	E	1	310	1.21	0.30	310	1.28	0.18	0.25	-0.30	0.007	-0.30	0.007	0.081
76	ES	1	310	1.14	0.27	310	0.65	0.03	0.19	2.49	0.011	2.49	0.011	0.107
76	O	1	310	0.98	0.26	310	0.86	0.27	0.26	0.48	0.007	0.48	0.007	0.081
77	A	1	232	0.78	0.29	232	0.90	0.31	0.30	-0.39	0.009	-0.39	0.009	0.094
77	C	1	232	0.66	0.14	232	0.63	0.16	0.15	0.19	0.009	0.19	0.009	0.093
77	E	1	232	0.92	0.37	232	0.95	0.29	0.33	-0.11	0.009	-0.11	0.009	0.093
77	ES	1	232	0.99	0.20	232	0.81	0.00	0.14	1.22	0.010	1.22	0.010	0.101
77	O	1	232	0.78	0.09	232	0.80	0.36	0.26	-0.05	0.009	-0.05	0.009	0.093
Mean										-0.08	0.008	-0.08	0.008	0.078
SD										1.72	0.016	1.72	0.016	0.042

Note. ID=UniqueID, D=Dimension, P=Personality measures (1=Personality measures, 0=non-personality measures), Nn=Sample size of negatively worded items, Mn=Mean of negatively worded items, SDn=Standard deviation of negatively worded items, Np=Sample size of positively worded items, Mp=Mean of positively worded items, SDp=Standard deviation of positively worded items, s.within=Pooled standard deviation , d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g, R=RSES, RW=RWA, CF=CFC, G12=GHQ-12, L=LOT, LR=LOT-R, Ep=Empathy.

Table 22

Effect Sizes of the Negatively-Worded Item Effect for High and Low Motivation

ID	D	M	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
1	A	L	14021	0.27	0.00	14021	0.71	0.30	0.22	-2.02	0.000	-2.02	0.000	0.015
1	C	L	14021	0.31	0.00	14021	0.71	0.24	0.17	-2.35	0.000	-2.35	0.000	0.016
1	E	L	14021	0.24	0.00	14021	0.73	0.13	0.10	-5.11	0.001	-5.11	0.001	0.025
1	ES	L	14021	0.71	0.02	14021	0.55	0.00	0.02	10.33	0.002	10.33	0.002	0.045
2	A	L	545	0.39	0.06	545	2.08	0.83	0.59	-2.87	0.007	-2.87	0.007	0.086
2	O	L	545	0.22	0.03	545	2.98	1.77	1.25	-2.21	0.006	-2.21	0.006	0.077
14	A	L	439	0.93	0.48	439	1.25	0.57	0.53	-0.62	0.005	-0.61	0.005	0.069
14	C	L	439	0.72	0.23	439	0.88	0.20	0.22	-0.71	0.005	-0.71	0.005	0.070
14	E	L	439	1.30	0.48	439	1.12	0.25	0.38	0.45	0.005	0.45	0.005	0.068
14	ES	L	439	1.09	0.45	439	0.79	0.11	0.33	0.92	0.005	0.92	0.005	0.071
14	O	L	439	0.84	0.19	439	0.93	0.27	0.23	-0.40	0.005	-0.40	0.005	0.068
15	C	L	539	0.91	0.37	539	0.75	0.33	0.35	0.46	0.004	0.46	0.004	0.062
18	E	L	201	0.20	0.00	201	0.92	0.54	0.38	-1.87	0.014	-1.86	0.014	0.120
19	C	L	1517	1.59	0.52	1517	1.47	0.32	0.43	0.27	0.001	0.27	0.001	0.036
20	A	L	538	0.47	0.11	538	0.53	0.05	0.08	-0.65	0.004	-0.65	0.004	0.063
20	C	L	538	0.71	0.14	538	0.55	0.06	0.11	1.47	0.005	1.47	0.005	0.069
20	E	L	538	0.58	0.00	538	0.52	0.11	0.08	0.82	0.004	0.82	0.004	0.063
20	O	L	538	0.39	0.03	538	0.50	0.23	0.16	-0.68	0.004	-0.68	0.004	0.063
21	A	L	539	0.44	0.11	539	0.51	0.10	0.11	-0.59	0.004	-0.59	0.004	0.062
21	C	L	539	0.71	0.07	539	0.57	0.06	0.07	2.20	0.006	2.20	0.006	0.077
21	E	L	539	0.54	0.00	539	0.52	0.12	0.08	0.30	0.004	0.30	0.004	0.061
21	O	L	539	0.36	0.08	539	0.50	0.24	0.18	-0.79	0.004	-0.79	0.004	0.063
22	A	L	478	0.66	0.09	478	0.51	0.08	0.09	1.67	0.006	1.67	0.006	0.075
22	C	L	478	0.64	0.10	478	0.61	0.09	0.09	0.31	0.004	0.31	0.004	0.065
22	E	L	478	0.73	0.06	478	0.74	0.19	0.14	-0.10	0.004	-0.10	0.004	0.065
22	ES	L	478	0.62	0.10	478	0.67	0.04	0.08	-0.66	0.004	-0.66	0.004	0.066
22	O	L	478	0.56	0.00	478	0.62	0.13	0.09	-0.65	0.004	-0.65	0.004	0.066
23	C	L	322	0.48	0.07	322	0.46	0.07	0.07	0.22	0.006	0.22	0.006	0.079
23	E	L	322	0.51	0.11	322	0.55	0.10	0.10	-0.35	0.006	-0.35	0.006	0.079
23	O	L	322	0.44	0.12	322	0.41	0.12	0.12	0.24	0.006	0.24	0.006	0.079
24	A	L	372	0.42	0.22	372	0.56	0.12	0.18	-0.75	0.006	-0.74	0.006	0.076
24	C	L	372	0.55	0.18	372	0.52	0.12	0.15	0.20	0.005	0.20	0.005	0.074
24	E	L	372	0.65	0.08	372	0.75	0.01	0.06	-1.68	0.007	-1.68	0.007	0.085
24	ES	L	372	0.57	0.22	372	0.65	0.10	0.17	-0.50	0.006	-0.50	0.006	0.074
24	O	L	372	0.52	0.11	372	0.59	0.09	0.10	-0.77	0.006	-0.77	0.006	0.076
25	A	L	328	0.50	0.19	328	0.58	0.16	0.17	-0.50	0.006	-0.50	0.006	0.079
25	C	L	328	0.59	0.10	328	0.59	0.11	0.10	-0.02	0.006	-0.02	0.006	0.078
25	E	L	328	0.64	0.06	328	0.76	0.03	0.05	-2.34	0.010	-2.34	0.010	0.101
25	ES	L	328	0.60	0.18	328	0.68	0.10	0.14	-0.54	0.006	-0.54	0.006	0.080
25	O	L	328	0.67	0.05	328	0.54	0.10	0.08	1.60	0.008	1.60	0.008	0.090
26	A	L	183	0.55	0.13	183	0.59	0.13	0.13	-0.31	0.011	-0.31	0.011	0.105

ID	D	M	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
26	C	L	183	0.67	0.23	183	0.45	0.16	0.20	1.12	0.013	1.12	0.013	0.113
26	E	L	183	0.66	0.07	183	0.77	0.03	0.06	-1.89	0.016	-1.88	0.016	0.126
26	ES	L	183	0.53	0.10	183	0.60	0.08	0.09	-0.73	0.012	-0.73	0.012	0.108
26	O	L	183	0.54	0.03	183	0.50	0.13	0.09	0.47	0.011	0.47	0.011	0.106
27	A	L	202	0.46	0.11	202	0.60	0.08	0.10	-1.38	0.012	-1.38	0.012	0.111
27	C	L	202	0.54	0.29	202	0.47	0.09	0.22	0.31	0.010	0.31	0.010	0.100
27	E	L	202	0.59	0.08	202	0.65	0.08	0.08	-0.85	0.011	-0.85	0.011	0.104
27	ES	L	202	0.54	0.22	202	0.65	0.16	0.20	-0.54	0.010	-0.54	0.010	0.101
27	O	L	202	0.46	0.12	202	0.51	0.14	0.13	-0.41	0.010	-0.41	0.010	0.101
28	A	L	311	0.43	0.07	311	0.73	0.06	0.07	-4.43	0.022	-4.42	0.022	0.149
28	C	L	311	0.51	0.15	311	0.67	0.06	0.12	-1.34	0.008	-1.34	0.008	0.089
28	E	L	311	0.68	0.05	311	0.77	0.03	0.04	-2.11	0.010	-2.11	0.010	0.100
28	ES	L	311	0.52	0.28	311	0.69	0.13	0.22	-0.79	0.007	-0.79	0.007	0.083
28	O	L	311	0.24	0.08	311	0.56	0.12	0.10	-3.12	0.014	-3.11	0.014	0.119
29	A	L	404	0.41	0.09	404	0.43	0.09	0.09	-0.27	0.005	-0.27	0.005	0.071
29	C	L	404	0.68	0.16	404	0.36	0.13	0.15	2.21	0.008	2.20	0.008	0.089
29	E	L	404	0.46	0.19	404	0.49	0.17	0.18	-0.15	0.005	-0.15	0.005	0.070
29	ES	L	404	0.47	0.09	404	0.56	0.09	0.09	-1.04	0.006	-1.04	0.006	0.075
29	O	L	404	0.46	0.05	404	0.42	0.11	0.09	0.51	0.005	0.51	0.005	0.072
30	A	L	189	0.55	0.11	189	0.54	0.16	0.14	0.09	0.011	0.09	0.011	0.103
30	C	L	189	0.60	0.10	189	0.58	0.10	0.10	0.14	0.011	0.14	0.011	0.103
30	E	L	189	0.65	0.08	189	0.75	0.00	0.06	-1.83	0.015	-1.83	0.015	0.123
30	ES	L	189	0.67	0.08	189	0.67	0.08	0.08	0.00	0.011	0.00	0.011	0.103
30	O	L	189	0.53	0.06	189	0.50	0.14	0.11	0.29	0.011	0.29	0.011	0.103
31	A	L	189	0.49	0.15	189	0.48	0.12	0.14	0.06	0.011	0.06	0.011	0.103
31	C	L	189	0.45	0.07	189	0.59	0.11	0.09	-1.58	0.014	-1.57	0.014	0.118
31	E	L	189	0.47	0.00	189	0.48	0.11	0.08	-0.01	0.011	-0.01	0.011	0.103
31	O	L	189	0.42	0.23	189	0.48	0.23	0.23	-0.25	0.011	-0.25	0.011	0.103
32	E	L	654	0.66	0.01	654	0.50	0.14	0.10	1.67	0.004	1.67	0.004	0.064
33	E	L	1215	0.30	0.13	1215	0.52	0.21	0.18	-1.21	0.002	-1.20	0.002	0.044
34	E	L	1022	0.52	0.15	1022	0.55	0.17	0.16	-0.20	0.002	-0.20	0.002	0.044
35	E	L	1212	0.67	0.07	1212	0.57	0.16	0.12	0.77	0.002	0.77	0.002	0.042
36	E	L	805	0.53	0.06	805	0.62	0.18	0.13	-0.66	0.003	-0.66	0.003	0.051
37	E	L	1912	0.72	0.13	1912	0.62	0.20	0.17	0.61	0.001	0.61	0.001	0.033
38	E	L	1792	0.49	0.11	1792	0.52	0.18	0.15	-0.16	0.001	-0.16	0.001	0.033
39	E	L	4140	0.70	0.04	4140	0.60	0.18	0.13	0.79	0.001	0.79	0.001	0.023
40	E	L	949	0.67	0.10	949	0.61	0.19	0.15	0.42	0.002	0.42	0.002	0.046
41	E	L	1449	0.58	0.07	1449	0.47	0.14	0.11	0.95	0.002	0.95	0.002	0.039
42	E	L	1121	0.67	0.08	1121	0.58	0.18	0.14	0.68	0.002	0.68	0.002	0.043
43	E	L	729	0.50	0.06	729	0.51	0.12	0.10	-0.06	0.003	-0.06	0.003	0.052
44	E	L	981	0.31	0.13	981	0.46	0.15	0.14	-1.12	0.002	-1.12	0.002	0.049
45	E	L	1050	0.55	0.17	1050	0.55	0.18	0.18	0.00	0.002	0.00	0.002	0.044
46	E	L	781	0.71	0.10	781	0.56	0.16	0.13	1.12	0.003	1.12	0.003	0.054
47	E	L	1525	0.67	0.07	1525	0.57	0.24	0.17	0.59	0.001	0.59	0.001	0.037
48	E	L	1239	0.57	0.07	1239	0.51	0.16	0.13	0.49	0.002	0.49	0.002	0.041

ID	D	M	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
49	E	L	1404	0.37	0.18	1404	0.54	0.18	0.18	-0.91	0.002	-0.91	0.002	0.040
50	E	L	988	0.46	0.16	988	0.59	0.18	0.17	-0.72	0.002	-0.72	0.002	0.046
51	E	L	876	0.69	0.11	876	0.58	0.21	0.16	0.67	0.002	0.67	0.002	0.049
52	E	L	1280	0.38	0.03	1280	0.46	0.22	0.15	-0.54	0.002	-0.54	0.002	0.040
53	E	L	802	0.69	0.08	802	0.55	0.20	0.15	0.90	0.003	0.90	0.003	0.052
54	E	L	1193	0.61	0.18	1193	0.55	0.18	0.18	0.31	0.002	0.31	0.002	0.041
55	E	L	2378	0.40	0.10	2378	0.55	0.19	0.15	-0.99	0.001	-0.99	0.001	0.031
56	E	L	1093	0.43	0.15	1093	0.54	0.21	0.18	-0.56	0.002	-0.56	0.002	0.044
57	E	L	1014	0.54	0.12	1014	0.52	0.19	0.16	0.14	0.002	0.14	0.002	0.044
58	E	L	775	0.67	0.12	775	0.58	0.16	0.15	0.61	0.003	0.61	0.003	0.052
59	E	L	994	0.47	0.10	994	0.55	0.13	0.12	-0.64	0.002	-0.64	0.002	0.046
60	E	L	1200	0.43	0.15	1200	0.52	0.13	0.14	-0.67	0.002	-0.67	0.002	0.042
61	E	L	1029	0.58	0.15	1029	0.55	0.17	0.16	0.19	0.002	0.19	0.002	0.044
62	E	L	1030	0.34	0.15	1030	0.44	0.15	0.15	-0.68	0.002	-0.68	0.002	0.045
63	E	L	1473	0.34	0.08	1473	0.48	0.16	0.13	-1.13	0.002	-1.13	0.002	0.040
64	E	L	1381	0.66	0.08	1381	0.60	0.17	0.13	0.43	0.001	0.43	0.001	0.038
65	E	L	1067	0.58	0.12	1067	0.54	0.18	0.15	0.26	0.002	0.26	0.002	0.043
66	E	L	838	0.37	0.07	838	0.49	0.15	0.12	-0.99	0.003	-0.99	0.003	0.052
67	A	H	130	0.52	0.00	130	0.91	0.47	0.33	-1.17	0.018	-1.16	0.018	0.134
67	C	H	130	0.72	0.10	130	0.70	0.10	0.10	0.12	0.015	0.12	0.015	0.124
67	E	H	130	0.64	0.13	130	1.13	0.43	0.32	-1.52	0.020	-1.52	0.020	0.141
67	ES	H	130	1.00	0.12	130	0.50	0.12	0.12	4.19	0.049	4.18	0.049	0.222
68	A	H	301	0.90	0.21	301	1.02	0.18	0.19	-0.63	0.007	-0.63	0.007	0.084
68	C	H	300	0.90	0.18	300	0.82	0.15	0.16	0.47	0.007	0.47	0.007	0.083
68	E	H	301	0.91	0.23	301	0.99	0.14	0.19	-0.44	0.007	-0.44	0.007	0.083
68	ES	H	301	1.12	0.18	301	0.82	0.11	0.15	2.02	0.010	2.01	0.010	0.100
68	O	H	301	0.84	0.09	301	0.79	0.24	0.18	0.29	0.007	0.29	0.007	0.082
69	A	H	192	0.89	0.36	192	0.93	0.26	0.31	-0.11	0.010	-0.11	0.010	0.102
69	C	H	192	0.98	0.11	192	0.75	0.19	0.15	1.52	0.013	1.51	0.013	0.116
69	E	H	192	1.20	0.28	192	1.11	0.25	0.27	0.33	0.011	0.33	0.011	0.103
69	ES	H	192	1.00	0.19	192	0.71	0.11	0.16	1.85	0.015	1.85	0.015	0.122
69	O	H	192	0.90	0.08	192	0.84	0.25	0.19	0.32	0.011	0.32	0.011	0.103
70	A	H	223	0.86	0.28	223	0.80	0.28	0.28	0.20	0.009	0.20	0.009	0.095
70	C	H	223	0.63	0.15	223	0.66	0.09	0.13	-0.27	0.009	-0.27	0.009	0.095
70	E	H	223	1.06	0.26	223	0.98	0.05	0.18	0.44	0.009	0.44	0.009	0.096
70	ES	H	223	1.02	0.24	223	0.65	0.07	0.18	2.10	0.014	2.10	0.014	0.118
70	O	H	223	0.87	0.11	223	0.81	0.42	0.31	0.22	0.009	0.22	0.009	0.095
71	A	H	228	0.87	0.35	228	0.88	0.28	0.32	-0.05	0.009	-0.05	0.009	0.094
71	C	H	228	0.96	0.33	228	0.73	0.14	0.25	0.92	0.010	0.92	0.010	0.099
71	E	H	228	1.03	0.25	228	1.18	0.30	0.28	-0.52	0.009	-0.52	0.009	0.095
71	ES	H	228	1.25	0.22	228	0.71	0.12	0.18	3.03	0.019	3.02	0.019	0.137
71	O	H	228	0.70	0.21	228	0.79	0.34	0.28	-0.32	0.009	-0.32	0.009	0.094
72	A	L	501	0.88	0.17	501	0.77	0.16	0.17	0.70	0.004	0.70	0.004	0.065
72	C	L	501	0.75	0.06	501	0.72	0.10	0.08	0.38	0.004	0.38	0.004	0.064
72	E	L	501	0.77	0.18	501	0.88	0.16	0.17	-0.62	0.004	-0.62	0.004	0.065

ID	D	M	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
72	ES	L	501	0.84	0.15	501	0.83	0.11	0.13	0.03	0.004	0.03	0.004	0.063
72	O	L	501	0.83	0.12	501	0.83	0.14	0.13	-0.01	0.004	-0.01	0.004	0.063
74	E	L	1434	0.65	0.14	1434	0.62	0.19	0.17	0.16	0.001	0.16	0.001	0.037
76	A	L	310	1.02	0.37	310	0.95	0.34	0.35	0.20	0.006	0.20	0.006	0.081
76	C	L	310	1.10	0.21	310	0.78	0.09	0.16	1.99	0.010	1.99	0.010	0.098
76	E	L	310	1.21	0.30	310	1.28	0.18	0.25	-0.30	0.007	-0.30	0.007	0.081
76	ES	L	310	1.14	0.27	310	0.65	0.03	0.19	2.49	0.011	2.49	0.011	0.107
76	O	L	310	0.98	0.26	310	0.86	0.27	0.26	0.48	0.007	0.48	0.007	0.081
77	A	H	232	0.78	0.29	232	0.90	0.31	0.30	-0.39	0.009	-0.39	0.009	0.094
77	C	H	232	0.66	0.14	232	0.63	0.16	0.15	0.19	0.009	0.19	0.009	0.093
77	E	H	232	0.92	0.37	232	0.95	0.29	0.33	-0.11	0.009	-0.11	0.009	0.093
77	ES	H	232	0.99	0.20	232	0.81	0.00	0.14	1.22	0.010	1.22	0.010	0.101
77	O	H	232	0.78	0.09	232	0.80	0.36	0.26	-0.05	0.009	-0.05	0.009	0.093
Mean										-0.04	0.01	-0.04	0.01	0.08
SD										1.51	0.01	1.51	0.01	0.03

Note. ID=UniqueID, D=Dimension, M=Sample motivation (L=Low motivation, H=High motivation), Nn=Sample size of negatively worded items, Mn=Mean of negatively worded items, SDn=Standard deviation of negatively worded items, Np=Sample size of positively worded items, Mp=Mean of positively worded items, SDp=Standard deviation of positively worded items, s.within=Pooled standard deviation , d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

Table 23

Effect Sizes of the Negatively-Worded Item Effect by Wording Proportion

ID	D	PG	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
1	A	M	0.33	14021	0.27	0.00	14021	0.71	0.30	0.22	-2.02	0.00	-2.02	0.00	0.01
1	C	M	0.33	14021	0.31	0.00	14021	0.71	0.24	0.17	-2.35	0.00	-2.35	0.00	0.02
1	E	M	0.33	14021	0.24	0.00	14021	0.73	0.13	0.10	-5.11	0.00	-5.11	0.00	0.02
1	ES	H	0.67	14021	0.71	0.02	14021	0.55	0.00	0.02	10.33	0.00	10.33	0.00	0.05
2	A	M	0.50	545	0.39	0.06	545	2.08	0.83	0.59	-2.87	0.01	-2.87	0.01	0.09
2	O	M	0.50	545	0.22	0.03	545	2.98	1.77	1.25	-2.21	0.01	-2.21	0.01	0.08
14	A	M	0.40	439	0.93	0.48	439	1.25	0.57	0.53	-0.62	0.00	-0.61	0.00	0.07
14	C	M	0.40	439	0.72	0.23	439	0.88	0.20	0.22	-0.71	0.00	-0.71	0.00	0.07
14	E	L	0.30	439	1.30	0.48	439	1.12	0.25	0.38	0.45	0.00	0.45	0.00	0.07
14	ES	H	0.80	439	1.09	0.45	439	0.79	0.11	0.33	0.92	0.01	0.92	0.01	0.07
14	O	L	0.30	439	0.84	0.19	439	0.93	0.27	0.23	-0.40	0.00	-0.40	0.00	0.07
15	C	M	0.50	539	0.91	0.37	539	0.75	0.33	0.35	0.46	0.00	0.46	0.00	0.06
18	E	L	0.09	201	0.20	0.00	201	0.92	0.54	0.38	-1.87	0.01	-1.86	0.01	0.12
19	C	L	0.29	1517	1.59	0.52	1517	1.47	0.32	0.43	0.27	0.00	0.27	0.00	0.04
20	A	H	0.67	538	0.47	0.11	538	0.53	0.05	0.08	-0.65	0.00	-0.65	0.00	0.06
20	C	L	0.17	538	0.71	0.14	538	0.55	0.06	0.11	1.47	0.00	1.47	0.00	0.07
20	E	L	0.08	538	0.58	0.00	538	0.52	0.11	0.08	0.82	0.00	0.82	0.00	0.06
20	O	L	0.25	538	0.39	0.03	538	0.50	0.23	0.16	-0.68	0.00	-0.68	0.00	0.06
21	A	H	0.67	539	0.44	0.11	539	0.51	0.10	0.11	-0.59	0.00	-0.59	0.00	0.06
21	C	L	0.17	539	0.71	0.07	539	0.57	0.06	0.07	2.20	0.01	2.20	0.01	0.08
21	E	L	0.08	539	0.54	0.00	539	0.52	0.12	0.08	0.30	0.00	0.30	0.00	0.06
21	O	L	0.25	539	0.36	0.08	539	0.50	0.24	0.18	-0.79	0.00	-0.79	0.00	0.06
22	A	M	0.44	478	0.66	0.09	478	0.51	0.08	0.09	1.67	0.01	1.67	0.01	0.08
22	C	M	0.44	478	0.64	0.10	478	0.61	0.09	0.09	0.31	0.00	0.31	0.00	0.07
22	E	M	0.38	478	0.73	0.06	478	0.74	0.19	0.14	-0.10	0.00	-0.10	0.00	0.06
22	ES	M	0.63	478	0.62	0.10	478	0.67	0.04	0.08	-0.66	0.00	-0.66	0.00	0.07
22	O	L	0.20	478	0.56	0.00	478	0.62	0.13	0.09	-0.65	0.00	-0.65	0.00	0.07
23	C	M	0.47	322	0.48	0.07	322	0.46	0.07	0.07	0.22	0.01	0.22	0.01	0.08
23	E	M	0.47	322	0.51	0.11	322	0.55	0.10	0.10	-0.35	0.01	-0.35	0.01	0.08
23	O	M	0.41	322	0.44	0.12	322	0.41	0.12	0.12	0.24	0.01	0.24	0.01	0.08
24	A	M	0.40	372	0.42	0.22	372	0.56	0.12	0.18	-0.75	0.01	-0.74	0.01	0.08
24	C	M	0.40	372	0.55	0.18	372	0.52	0.12	0.15	0.20	0.01	0.20	0.01	0.07
24	E	H	0.80	372	0.65	0.08	372	0.75	0.01	0.06	-1.68	0.01	-1.68	0.01	0.09
24	ES	L	0.30	372	0.57	0.22	372	0.65	0.10	0.17	-0.50	0.01	-0.50	0.01	0.07
24	O	L	0.30	372	0.52	0.11	372	0.59	0.09	0.10	-0.77	0.01	-0.77	0.01	0.08
25	A	M	0.40	328	0.50	0.19	328	0.58	0.16	0.17	-0.50	0.01	-0.50	0.01	0.08
25	C	M	0.40	328	0.59	0.10	328	0.59	0.11	0.10	-0.02	0.01	-0.02	0.01	0.08

ID	D	PG	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
25	E	H	0.80	328	0.64	0.06	328	0.76	0.03	0.05	-2.34	0.01	-2.34	0.01	0.10
25	ES	L	0.30	328	0.60	0.18	328	0.68	0.10	0.14	-0.54	0.01	-0.54	0.01	0.08
25	O	L	0.30	328	0.67	0.05	328	0.54	0.10	0.08	1.60	0.01	1.60	0.01	0.09
26	A	M	0.40	183	0.55	0.13	183	0.59	0.13	0.13	-0.31	0.01	-0.31	0.01	0.11
26	C	M	0.40	183	0.67	0.23	183	0.45	0.16	0.20	1.12	0.01	1.12	0.01	0.11
26	E	H	0.80	183	0.66	0.07	183	0.77	0.03	0.06	-1.89	0.02	-1.88	0.02	0.13
26	ES	L	0.30	183	0.53	0.10	183	0.60	0.08	0.09	-0.73	0.01	-0.73	0.01	0.11
26	O	L	0.30	183	0.54	0.03	183	0.50	0.13	0.09	0.47	0.01	0.47	0.01	0.11
27	A	M	0.40	202	0.46	0.11	202	0.60	0.08	0.10	-1.38	0.01	-1.38	0.01	0.11
27	C	M	0.40	202	0.54	0.29	202	0.47	0.09	0.22	0.31	0.01	0.31	0.01	0.10
27	E	H	0.80	202	0.59	0.08	202	0.65	0.08	0.08	-0.85	0.01	-0.85	0.01	0.10
27	ES	L	0.30	202	0.54	0.22	202	0.65	0.16	0.20	-0.54	0.01	-0.54	0.01	0.10
27	O	L	0.30	202	0.46	0.12	202	0.51	0.14	0.13	-0.41	0.01	-0.41	0.01	0.10
28	A	M	0.40	311	0.43	0.07	311	0.73	0.06	0.07	-4.43	0.02	-4.42	0.02	0.15
28	C	M	0.40	311	0.51	0.15	311	0.67	0.06	0.12	-1.34	0.01	-1.34	0.01	0.09
28	E	H	0.80	311	0.68	0.05	311	0.77	0.03	0.04	-2.11	0.01	-2.11	0.01	0.10
28	ES	L	0.30	311	0.52	0.28	311	0.69	0.13	0.22	-0.79	0.01	-0.79	0.01	0.08
28	O	L	0.30	311	0.24	0.08	311	0.56	0.12	0.10	-3.12	0.01	-3.11	0.01	0.12
29	A	M	0.40	404	0.41	0.09	404	0.43	0.09	0.09	-0.27	0.00	-0.27	0.00	0.07
29	C	M	0.40	404	0.68	0.16	404	0.36	0.13	0.15	2.21	0.01	2.20	0.01	0.09
29	E	H	0.80	404	0.46	0.19	404	0.49	0.17	0.18	-0.15	0.00	-0.15	0.00	0.07
29	ES	L	0.30	404	0.47	0.09	404	0.56	0.09	0.09	-1.04	0.01	-1.04	0.01	0.08
29	O	L	0.30	404	0.46	0.05	404	0.42	0.11	0.09	0.51	0.01	0.51	0.01	0.07
30	A	M	0.40	189	0.55	0.11	189	0.54	0.16	0.14	0.09	0.01	0.09	0.01	0.10
30	C	M	0.40	189	0.60	0.10	189	0.58	0.10	0.10	0.14	0.01	0.14	0.01	0.10
30	E	H	0.80	189	0.65	0.08	189	0.75	0.00	0.06	-1.83	0.02	-1.83	0.01	0.12
30	ES	L	0.30	189	0.67	0.08	189	0.67	0.08	0.08	0.00	0.01	0.00	0.01	0.10
30	O	L	0.30	189	0.53	0.06	189	0.50	0.14	0.11	0.29	0.01	0.29	0.01	0.10
31	A	H	0.67	189	0.49	0.15	189	0.48	0.12	0.14	0.06	0.01	0.06	0.01	0.10
31	C	L	0.17	189	0.45	0.07	189	0.59	0.11	0.09	-1.58	0.01	-1.57	0.01	0.12
31	E	L	0.08	189	0.47	0.00	189	0.48	0.11	0.08	-0.01	0.01	-0.01	0.01	0.10
31	O	L	0.25	189	0.42	0.23	189	0.48	0.23	0.23	-0.25	0.01	-0.25	0.01	0.10
32	E	L	0.14	654	0.66	0.01	654	0.50	0.14	0.10	1.67	0.00	1.67	0.00	0.06
33	E	L	0.14	1215	0.30	0.13	1215	0.52	0.21	0.18	-1.21	0.00	-1.20	0.00	0.04
34	E	L	0.14	1022	0.52	0.15	1022	0.55	0.17	0.16	-0.20	0.00	-0.20	0.00	0.04
35	E	L	0.14	1212	0.67	0.07	1212	0.57	0.16	0.12	0.77	0.00	0.77	0.00	0.04
36	E	L	0.14	805	0.53	0.06	805	0.62	0.18	0.13	-0.66	0.00	-0.66	0.00	0.05
37	E	L	0.14	1912	0.72	0.13	1912	0.62	0.20	0.17	0.61	0.00	0.61	0.00	0.03
38	E	L	0.14	1792	0.49	0.11	1792	0.52	0.18	0.15	-0.16	0.00	-0.16	0.00	0.03
39	E	L	0.14	4140	0.70	0.04	4140	0.60	0.18	0.13	0.79	0.00	0.79	0.00	0.02
40	E	L	0.14	949	0.67	0.10	949	0.61	0.19	0.15	0.42	0.00	0.42	0.00	0.05
41	E	L	0.14	1449	0.58	0.07	1449	0.47	0.14	0.11	0.95	0.00	0.95	0.00	0.04

ID	D	PG	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
42	E	L	0.14	1121	0.67	0.08	1121	0.58	0.18	0.14	0.68	0.00	0.68	0.00	0.04
43	E	L	0.14	729	0.50	0.06	729	0.51	0.12	0.10	-0.06	0.00	-0.06	0.00	0.05
44	E	L	0.14	981	0.31	0.13	981	0.46	0.15	0.14	-1.12	0.00	-1.12	0.00	0.05
45	E	L	0.14	1050	0.55	0.17	1050	0.55	0.18	0.18	0.00	0.00	0.00	0.00	0.04
46	E	L	0.14	781	0.71	0.10	781	0.56	0.16	0.13	1.12	0.00	1.12	0.00	0.05
47	E	L	0.14	1525	0.67	0.07	1525	0.57	0.24	0.17	0.59	0.00	0.59	0.00	0.04
48	E	L	0.14	1239	0.57	0.07	1239	0.51	0.16	0.13	0.49	0.00	0.49	0.00	0.04
49	E	L	0.14	1404	0.37	0.18	1404	0.54	0.18	0.18	-0.91	0.00	-0.91	0.00	0.04
50	E	L	0.14	988	0.46	0.16	988	0.59	0.18	0.17	-0.72	0.00	-0.72	0.00	0.05
51	E	L	0.14	876	0.69	0.11	876	0.58	0.21	0.16	0.67	0.00	0.67	0.00	0.05
52	E	L	0.14	1280	0.38	0.03	1280	0.46	0.22	0.15	-0.54	0.00	-0.54	0.00	0.04
53	E	L	0.14	802	0.69	0.08	802	0.55	0.20	0.15	0.90	0.00	0.90	0.00	0.05
54	E	L	0.14	1193	0.61	0.18	1193	0.55	0.18	0.18	0.31	0.00	0.31	0.00	0.04
55	E	L	0.14	2378	0.40	0.10	2378	0.55	0.19	0.15	-0.99	0.00	-0.99	0.00	0.03
56	E	L	0.14	1093	0.43	0.15	1093	0.54	0.21	0.18	-0.56	0.00	-0.56	0.00	0.04
57	E	L	0.14	1014	0.54	0.12	1014	0.52	0.19	0.16	0.14	0.00	0.14	0.00	0.04
58	E	L	0.14	775	0.67	0.12	775	0.58	0.16	0.15	0.61	0.00	0.61	0.00	0.05
59	E	L	0.14	994	0.47	0.10	994	0.55	0.13	0.12	-0.64	0.00	-0.64	0.00	0.05
60	E	L	0.14	1200	0.43	0.15	1200	0.52	0.13	0.14	-0.67	0.00	-0.67	0.00	0.04
61	E	L	0.14	1029	0.58	0.15	1029	0.55	0.17	0.16	0.19	0.00	0.19	0.00	0.04
62	E	L	0.14	1030	0.34	0.15	1030	0.44	0.15	0.15	-0.68	0.00	-0.68	0.00	0.05
63	E	L	0.14	1473	0.34	0.08	1473	0.48	0.16	0.13	-1.13	0.00	-1.13	0.00	0.04
64	E	L	0.14	1381	0.66	0.08	1381	0.60	0.17	0.13	0.43	0.00	0.43	0.00	0.04
65	E	L	0.14	1067	0.58	0.12	1067	0.54	0.18	0.15	0.26	0.00	0.26	0.00	0.04
66	E	L	0.14	838	0.37	0.07	838	0.49	0.15	0.12	-0.99	0.00	-0.99	0.00	0.05
67	A	L	0.08	130	0.52	0.00	130	0.91	0.47	0.33	-1.17	0.02	-1.16	0.02	0.13
67	C	M	0.62	130	0.72	0.10	130	0.70	0.10	0.10	0.12	0.02	0.12	0.02	0.12
67	E	L	0.25	130	0.64	0.13	130	1.13	0.43	0.32	-1.52	0.02	-1.52	0.02	0.14
67	ES	H	0.67	130	1.00	0.12	130	0.50	0.12	0.12	4.19	0.05	4.18	0.05	0.22
68	A	M	0.40	301	0.90	0.21	301	1.02	0.18	0.19	-0.63	0.01	-0.63	0.01	0.08
68	C	M	0.40	300	0.90	0.18	300	0.82	0.15	0.16	0.47	0.01	0.47	0.01	0.08
68	E	M	0.50	301	0.91	0.23	301	0.99	0.14	0.19	-0.44	0.01	-0.44	0.01	0.08
68	ES	H	0.80	301	1.12	0.18	301	0.82	0.11	0.15	2.02	0.01	2.01	0.01	0.10
68	O	L	0.30	301	0.84	0.09	301	0.79	0.24	0.18	0.29	0.01	0.29	0.01	0.08
69	A	M	0.40	192	0.89	0.36	192	0.93	0.26	0.31	-0.11	0.01	-0.11	0.01	0.10
69	C	M	0.40	192	0.98	0.11	192	0.75	0.19	0.15	1.52	0.01	1.51	0.01	0.12
69	E	M	0.50	192	1.20	0.28	192	1.11	0.25	0.27	0.33	0.01	0.33	0.01	0.10
69	ES	H	0.80	192	1.00	0.19	192	0.71	0.11	0.16	1.85	0.01	1.85	0.01	0.12
69	O	L	0.30	192	0.90	0.08	192	0.84	0.25	0.19	0.32	0.01	0.32	0.01	0.10
70	A	M	0.40	223	0.86	0.28	223	0.80	0.28	0.28	0.20	0.01	0.20	0.01	0.09
70	C	M	0.40	223	0.63	0.15	223	0.66	0.09	0.13	-0.27	0.01	-0.27	0.01	0.10
70	E	M	0.50	223	1.06	0.26	223	0.98	0.05	0.18	0.44	0.01	0.44	0.01	0.10

ID	D	PG	P	Nn	Mn	SDn	Np	Mp	SDp	s.within	d	var.d	g	var.g	se.g
70	ES	H	0.80	223	1.02	0.24	223	0.65	0.07	0.18	2.10	0.01	2.10	0.01	0.12
70	O	L	0.30	223	0.87	0.11	223	0.81	0.42	0.31	0.22	0.01	0.22	0.01	0.09
71	A	M	0.40	228	0.87	0.35	228	0.88	0.28	0.32	-0.05	0.01	-0.05	0.01	0.09
71	C	M	0.40	228	0.96	0.33	228	0.73	0.14	0.25	0.92	0.01	0.92	0.01	0.10
71	E	M	0.50	228	1.03	0.25	228	1.18	0.30	0.28	-0.52	0.01	-0.52	0.01	0.10
71	ES	H	0.80	228	1.25	0.22	228	0.71	0.12	0.18	3.03	0.02	3.02	0.02	0.14
71	O	L	0.30	228	0.70	0.21	228	0.79	0.34	0.28	-0.32	0.01	-0.32	0.01	0.09
72	A	L	0.30	501	0.88	0.17	501	0.77	0.16	0.17	0.70	0.00	0.70	0.00	0.07
72	C	M	0.45	501	0.75	0.06	501	0.72	0.10	0.08	0.38	0.00	0.38	0.00	0.06
72	E	M	0.50	501	0.77	0.18	501	0.88	0.16	0.17	-0.62	0.00	-0.62	0.00	0.06
72	ES	H	0.75	501	0.84	0.15	501	0.83	0.11	0.13	0.03	0.00	0.03	0.00	0.06
72	O	M	0.35	501	0.83	0.12	501	0.83	0.14	0.13	-0.01	0.00	-0.01	0.00	0.06
74	E	L	0.13	1434	0.65	0.14	1434	0.62	0.19	0.17	0.16	0.00	0.16	0.00	0.04
76	A	M	0.40	310	1.02	0.37	310	0.95	0.34	0.35	0.20	0.01	0.20	0.01	0.08
76	C	M	0.40	310	1.10	0.21	310	0.78	0.09	0.16	1.99	0.01	1.99	0.01	0.10
76	E	M	0.50	310	1.21	0.30	310	1.28	0.18	0.25	-0.30	0.01	-0.30	0.01	0.08
76	ES	H	0.80	310	1.14	0.27	310	0.65	0.03	0.19	2.49	0.01	2.49	0.01	0.11
76	O	L	0.30	310	0.98	0.26	310	0.86	0.27	0.26	0.48	0.01	0.48	0.01	0.08
77	A	M	0.40	232	0.78	0.29	232	0.90	0.31	0.30	-0.39	0.01	-0.39	0.01	0.09
77	C	M	0.40	232	0.66	0.14	232	0.63	0.16	0.15	0.19	0.01	0.19	0.01	0.09
77	E	M	0.50	232	0.92	0.37	232	0.95	0.29	0.33	-0.11	0.01	-0.11	0.01	0.09
77	ES	H	0.80	232	0.99	0.20	232	0.81	0.00	0.14	1.22	0.01	1.22	0.01	0.10
77	O	L	0.30	232	0.78	0.09	232	0.80	0.36	0.26	-0.05	0.01	-0.05	0.01	0.09

Note. ID=UniqueID, D=Dimension, PG=Proportion group of negatively worded items (L=Low, M=Medium, H=High), P=Proportion of negatively worded items, Nn=Sample size of negatively worded items, Mn=Mean of negatively worded items, SDn=Standard deviation of negatively worded items, Np=Sample size of positively worded items, Mp=Mean of positively worded items, SDp=Standard deviation of positively worded items, s.within=Pooled standard deviation , d=Cohen'd, var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

Table 24

Effect Sizes of the Negated Item Effect by Negation Proportion

ID	D	PG	NP	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
2	A	H	0.50	545	0.39	0.06	545	2.08	0.83	0.59	-2.87	0.01	-2.87	0.01	0.09
2	O	H	0.50	545	0.22	0.03	545	2.98	1.77	1.25	-2.21	0.01	-2.21	0.01	0.08
14	A	L	0.30	439	1.10	0.42	439	1.13	0.61	0.52	-0.06	0.00	-0.06	0.00	0.07
14	E	H	0.40	439	1.03	0.21	439	1.28	0.35	0.29	-0.87	0.00	-0.87	0.00	0.07
14	ES	L	0.20	439	1.09	0.53	439	1.02	0.43	0.48	0.14	0.00	0.14	0.00	0.07
14	O	L	0.20	439	0.92	0.17	439	0.90	0.27	0.22	0.10	0.00	0.10	0.00	0.07
15	C	H	0.55	539	0.82	0.39	539	0.83	0.31	0.35	-0.02	0.00	-0.02	0.00	0.06
17	ES	L	0.19	706	0.73	0.27	706	0.56	0.25	0.26	0.65	0.00	0.65	0.00	0.05
19	C	L	0.29	1517	1.59	0.52	1517	1.47	0.32	0.43	0.27	0.00	0.27	0.00	0.04
20	A	L	0.08	538	0.35	0.00	538	0.50	0.08	0.06	-2.58	0.01	-2.58	0.01	0.08
20	ES	L	0.25	538	0.56	0.15	538	0.61	0.08	0.12	-0.38	0.00	-0.38	0.00	0.06
21	A	L	0.08	539	0.26	0.00	539	0.48	0.09	0.06	-3.46	0.01	-3.46	0.01	0.10
21	ES	L	0.25	539	0.56	0.11	539	0.62	0.08	0.10	-0.56	0.00	-0.56	0.00	0.06
22	A	L	0.11	478	0.60	0.00	478	0.57	0.12	0.09	0.37	0.00	0.37	0.00	0.07
22	C	L	0.33	478	0.65	0.11	478	0.61	0.08	0.10	0.42	0.00	0.42	0.00	0.07
22	E	L	0.13	478	0.80	0.00	478	0.73	0.16	0.11	0.60	0.00	0.60	0.00	0.07
22	ES	L	0.13	478	0.69	0.00	478	0.63	0.09	0.06	0.92	0.00	0.92	0.00	0.07
22	O	L	0.20	478	0.60	0.05	478	0.62	0.14	0.10	-0.13	0.00	-0.13	0.00	0.06
23	C	L	0.34	322	0.44	0.09	322	0.48	0.05	0.07	-0.64	0.01	-0.64	0.01	0.08
23	E	L	0.31	322	0.51	0.10	322	0.54	0.10	0.10	-0.31	0.01	-0.31	0.01	0.08
23	O	L	0.28	322	0.43	0.10	322	0.41	0.13	0.12	0.12	0.01	0.12	0.01	0.08
24	A	L	0.30	372	0.51	0.17	372	0.50	0.19	0.18	0.05	0.01	0.05	0.01	0.07
24	E	L	0.20	372	0.70	0.08	372	0.66	0.09	0.08	0.41	0.01	0.41	0.01	0.07
24	ES	H	0.40	372	0.72	0.09	372	0.57	0.13	0.11	1.30	0.01	1.30	0.01	0.08
24	O	L	0.20	372	0.48	0.14	372	0.59	0.08	0.12	-0.94	0.01	-0.93	0.01	0.08
25	A	L	0.30	328	0.55	0.19	328	0.55	0.17	0.18	-0.03	0.01	-0.03	0.01	0.08
25	E	L	0.20	328	0.66	0.10	328	0.66	0.08	0.09	-0.01	0.01	-0.01	0.01	0.08
25	ES	H	0.40	328	0.73	0.08	328	0.60	0.13	0.11	1.14	0.01	1.13	0.01	0.08
25	O	L	0.20	328	0.65	0.06	328	0.56	0.10	0.08	1.05	0.01	1.05	0.01	0.08
26	A	L	0.30	183	0.60	0.10	183	0.56	0.14	0.12	0.32	0.01	0.32	0.01	0.11
26	E	L	0.20	183	0.77	0.03	183	0.66	0.07	0.06	1.93	0.02	1.93	0.02	0.13
26	ES	H	0.40	183	0.63	0.07	183	0.55	0.09	0.08	0.92	0.01	0.92	0.01	0.11
26	O	L	0.20	183	0.54	0.04	183	0.50	0.12	0.09	0.47	0.01	0.47	0.01	0.11
27	A	L	0.30	202	0.48	0.12	202	0.57	0.11	0.12	-0.74	0.01	-0.74	0.01	0.10
27	E	L	0.20	202	0.62	0.13	202	0.59	0.08	0.10	0.26	0.01	0.26	0.01	0.10
27	ES	H	0.40	202	0.69	0.16	202	0.56	0.19	0.17	0.71	0.01	0.71	0.01	0.10
27	O	L	0.20	202	0.47	0.17	202	0.51	0.13	0.15	-0.27	0.01	-0.27	0.01	0.10
28	A	L	0.30	311	0.47	0.03	311	0.68	0.16	0.12	-1.79	0.01	-1.79	0.01	0.09

ID	D	PG	NP	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
28	E	L	0.20	311	0.69	0.09	311	0.70	0.06	0.07	-0.19	0.01	-0.19	0.01	0.08
28	ES	H	0.40	311	0.69	0.19	311	0.61	0.19	0.19	0.38	0.01	0.38	0.01	0.08
28	O	L	0.20	311	0.21	0.07	311	0.53	0.14	0.11	-2.93	0.01	-2.93	0.01	0.12
29	A	L	0.30	404	0.42	0.10	404	0.42	0.09	0.09	-0.03	0.00	-0.03	0.00	0.07
29	E	L	0.20	404	0.25	0.17	404	0.52	0.15	0.16	-1.70	0.01	-1.70	0.01	0.08
29	ES	H	0.40	404	0.49	0.07	404	0.57	0.10	0.08	-1.01	0.01	-1.01	0.01	0.07
29	O	L	0.20	404	0.49	0.04	404	0.42	0.10	0.08	0.84	0.01	0.84	0.01	0.07
30	A	L	0.30	189	0.58	0.11	189	0.52	0.15	0.13	0.43	0.01	0.43	0.01	0.10
30	E	L	0.20	189	0.65	0.15	189	0.68	0.07	0.12	-0.28	0.01	-0.28	0.01	0.10
30	ES	H	0.40	189	0.70	0.06	189	0.65	0.09	0.07	0.64	0.01	0.64	0.01	0.11
30	O	L	0.20	189	0.56	0.05	189	0.50	0.13	0.10	0.62	0.01	0.62	0.01	0.11
31	A	L	0.08	189	0.36	0.00	189	0.50	0.14	0.10	-1.43	0.01	-1.43	0.01	0.12
31	ES	L	0.25	189	0.67	0.09	189	0.62	0.09	0.09	0.46	0.01	0.46	0.01	0.10
32	ES	L	0.22	654	0.58	0.05	654	0.61	0.14	0.10	-0.25	0.00	-0.25	0.00	0.06
33	ES	L	0.22	1215	0.49	0.08	1215	0.48	0.16	0.13	0.11	0.00	0.11	0.00	0.04
34	ES	L	0.22	1022	0.58	0.08	1022	0.55	0.13	0.11	0.23	0.00	0.23	0.00	0.04
35	ES	L	0.22	1212	0.53	0.07	1212	0.58	0.10	0.09	-0.64	0.00	-0.64	0.00	0.04
36	ES	L	0.22	805	0.61	0.16	805	0.57	0.18	0.17	0.22	0.00	0.22	0.00	0.05
37	ES	L	0.22	1912	0.56	0.07	1912	0.56	0.13	0.11	-0.03	0.00	-0.03	0.00	0.03
38	ES	L	0.22	1792	0.51	0.15	1792	0.48	0.19	0.17	0.19	0.00	0.19	0.00	0.03
39	ES	L	0.22	4140	0.58	0.05	4140	0.61	0.11	0.09	-0.35	0.00	-0.35	0.00	0.02
40	ES	L	0.22	949	0.61	0.09	949	0.59	0.15	0.12	0.12	0.00	0.12	0.00	0.05
41	ES	L	0.22	1449	0.60	0.12	1449	0.58	0.10	0.11	0.25	0.00	0.25	0.00	0.04
42	ES	L	0.22	1121	0.61	0.04	1121	0.58	0.12	0.09	0.27	0.00	0.27	0.00	0.04
43	ES	L	0.22	729	0.50	0.12	729	0.56	0.12	0.12	-0.43	0.00	-0.43	0.00	0.05
44	ES	L	0.22	981	0.56	0.14	981	0.52	0.14	0.14	0.23	0.00	0.23	0.00	0.05
45	ES	L	0.22	1050	0.68	0.02	1050	0.61	0.13	0.09	0.74	0.00	0.74	0.00	0.05
46	ES	L	0.22	781	0.52	0.10	781	0.48	0.15	0.12	0.30	0.00	0.30	0.00	0.05
47	ES	L	0.22	1525	0.53	0.11	1525	0.53	0.12	0.12	-0.01	0.00	-0.01	0.00	0.04
48	ES	L	0.22	1239	0.52	0.12	1239	0.48	0.15	0.14	0.28	0.00	0.28	0.00	0.04
49	ES	L	0.22	1404	0.52	0.10	1404	0.52	0.15	0.13	-0.01	0.00	-0.01	0.00	0.04
50	ES	L	0.22	988	0.60	0.07	988	0.57	0.18	0.14	0.25	0.00	0.25	0.00	0.05
51	ES	L	0.22	876	0.59	0.08	876	0.61	0.15	0.12	-0.23	0.00	-0.22	0.00	0.05
52	ES	L	0.22	1280	0.59	0.19	1280	0.53	0.15	0.17	0.32	0.00	0.32	0.00	0.04
53	ES	L	0.22	802	0.54	0.06	802	0.59	0.17	0.13	-0.41	0.00	-0.41	0.00	0.05
54	ES	L	0.22	1193	0.55	0.12	1193	0.52	0.16	0.14	0.16	0.00	0.16	0.00	0.04
55	ES	L	0.22	2378	0.51	0.09	2378	0.50	0.19	0.15	0.06	0.00	0.06	0.00	0.03
56	ES	L	0.22	1093	0.55	0.08	1093	0.53	0.19	0.15	0.15	0.00	0.15	0.00	0.04
57	ES	L	0.22	1014	0.58	0.08	1014	0.54	0.11	0.10	0.38	0.00	0.38	0.00	0.04
58	ES	L	0.22	775	0.64	0.09	775	0.54	0.14	0.12	0.83	0.00	0.83	0.00	0.05
59	ES	L	0.22	994	0.55	0.05	994	0.58	0.13	0.10	-0.27	0.00	-0.27	0.00	0.05
60	ES	L	0.22	1200	0.49	0.11	1200	0.49	0.12	0.11	0.01	0.00	0.01	0.00	0.04

ID	D	PG	NP	Nt	Mt	SDt	Ns	Ms	SDs	s.within	d	var.d	g	var.g	se.g
61	ES	L	0.22	1029	0.58	0.15	1029	0.57	0.16	0.15	0.05	0.00	0.05	0.00	0.04
62	ES	L	0.22	1030	0.52	0.12	1030	0.53	0.15	0.14	-0.06	0.00	-0.06	0.00	0.04
63	ES	L	0.22	1473	0.45	0.09	1473	0.45	0.10	0.10	-0.02	0.00	-0.02	0.00	0.04
64	ES	L	0.22	1381	0.58	0.10	1381	0.62	0.11	0.11	-0.39	0.00	-0.39	0.00	0.04
65	ES	L	0.22	1067	0.58	0.07	1067	0.52	0.16	0.12	0.51	0.00	0.51	0.00	0.04
66	ES	L	0.22	838	0.50	0.11	838	0.48	0.09	0.10	0.13	0.00	0.13	0.00	0.05
67	A	L	0.17	130	0.47	0.07	130	0.96	0.47	0.34	-1.46	0.02	-1.46	0.02	0.14
67	C	L	0.15	130	0.82	0.13	130	0.69	0.09	0.11	1.21	0.02	1.20	0.02	0.13
67	E	L	0.08	130	1.46	0.00	130	0.97	0 43	0.30	1.64	0.02	1.64	0.02	0.14
67	ES	H	0.50	130	0.83	0.28	130	0.84	0.30	0.29	-0.06	0.02	-0.06	0.02	0.12
68	A	L	0.30	301	1.00	0.08	301	0.96	0.23	0.17	0.22	0.01	0.22	0.01	0.08
68	E	H	0.40	301	0.77	0.09	301	1.07	0.12	0.11	-2.82	0.01	-2.82	0.01	0.12
68	ES	L	0.20	301	0.96	0.32	301	1.08	0.20	0.26	-0.46	0.01	-0.46	0.01	0.08
68	O	L	0.20	301	0.80	0.09	301	0.80	0.23	0.17	-0.02	0.01	-0.02	0.01	0.08
69	A	L	0.30	192	1.05	0.19	192	0.85	0.31	0.26	0.79	0.01	0.79	0.01	0.11
69	E	H	0.40	192	1.02	0.34	192	1.24	0.15	0.27	-0.80	0.01	-0.80	0.01	0.11
69	ES	L	0.20	192	0.67	0.05	192	1.01	0.17	0.13	-2.68	0.02	-2.67	0.02	0.14
69	O	L	0.20	192	0.93	0.09	192	0.85	0.23	0.18	0.45	0.01	0.45	0.01	0.10
70	A	L	0.30	223	0.97	0.19	223	0.76	0.27	0.24	0.89	0.01	0.89	0.01	0.10
70	E	H	0.40	223	1.05	0.29	223	1.00	0.07	0.21	0.23	0.01	0.23	0.01	0.10
70	ES	L	0.20	223	0.76	0.09	223	1.00	0.28	0.21	-1.15	0.01	-1.15	0.01	0.10
70	O	L	0.20	223	0.91	0.14	223	0.81	0.39	0.29	0.34	0.01	0.34	0.01	0.10
71	A	L	0.30	228	1.00	0.28	228	0.82	0.30	0.29	0.60	0.01	0.59	0.01	0.10
71	E	H	0.40	228	0.87	0.10	228	1.26	0.24	0.18	-2.16	0.01	-2.15	0.01	0.12
71	ES	L	0.20	228	1.13	0.48	228	1.15	0.29	0.40	-0.04	0.01	-0.04	0.01	0.09
71	O	L	0.20	228	0.76	0.24	228	0.76	0.33	0.29	0.02	0.01	0.02	0.01	0.09
72	A	L	0.20	501	0.97	0.09	501	0.76	0.16	0.13	1.66	0.01	1.66	0.01	0.07
72	E	L	0.25	501	0.90	0.08	501	0.80	0.19	0.15	0.70	0.00	0.70	0.00	0.07
72	ES	L	0.25	501	0.87	0.02	501	0.82	0.16	0.11	0.42	0.00	0.42	0.00	0.06
72	O	L	0.15	501	0.87	0.15	501	0.82	0.13	0.14	0.35	0.00	0.35	0.00	0.06
74	ES	L	0.21	1434	0.57	0.12	1434	0.59	0.11	0.11	-0.22	0.00	-0.22	0.00	0.04
76	A	L	0.30	310	1.18	0.22	310	0.89	0.35	0.29	1.01	0.01	1.01	0.01	0.09
76	E	H	0.40	310	1.19	0.34	310	1.28	0.17	0.27	-0.34	0.01	-0.34	0.01	0.08
76	ES	L	0.20	310	0.90	0.32	310	1.07	0.33	0.32	-0.55	0.01	-0.55	0.01	0.08
76	O	L	0.20	310	1.06	0.32	310	0.86	0.25	0.29	0.70	0.01	0.70	0.01	0.08
77	A	L	0.30	232	0.91	0.13	232	0.82	0.35	0.26	0.35	0.01	0.35	0.01	0.09
77	E	H	0.40	232	0.83	0.37	232	1.01	0.28	0.33	-0.54	0.01	-0.54	0.01	0.09
77	ES	L	0.20	232	0.81	0.00	232	0.99	0.20	0.14	-1.22	0.01	-1.22	0.01	0.10
77	O	L	0.20	232	0.80	0.11	232	0.79	0.33	0.25	0.04	0.01	0.04	0.01	0.09

Note. ID=UniqueID, D=Personality dimension, PG=Proportion group of negated items (L=Low, H=High), NP=Proportion of negated items, Nt=Sample size of negated items, Mt=Mean of negatied items, SDt=Standard deviation of negated items, Ns=Sample size of non-negated items, Ms=Mean of non-negated items, SDs=Standard deviation of non-negated items, s.within=Pooled standard deviation, d=Cohen'd,

var.d=Variance of Cohen's d, g=Hedges' g, var.g=Variance of Hedges' g, se.g=Standard error of Hedges' g.

APPENDIX E

FOREST PLOTS OF RESEARCH QUESTIONS AND HYPOTHESES

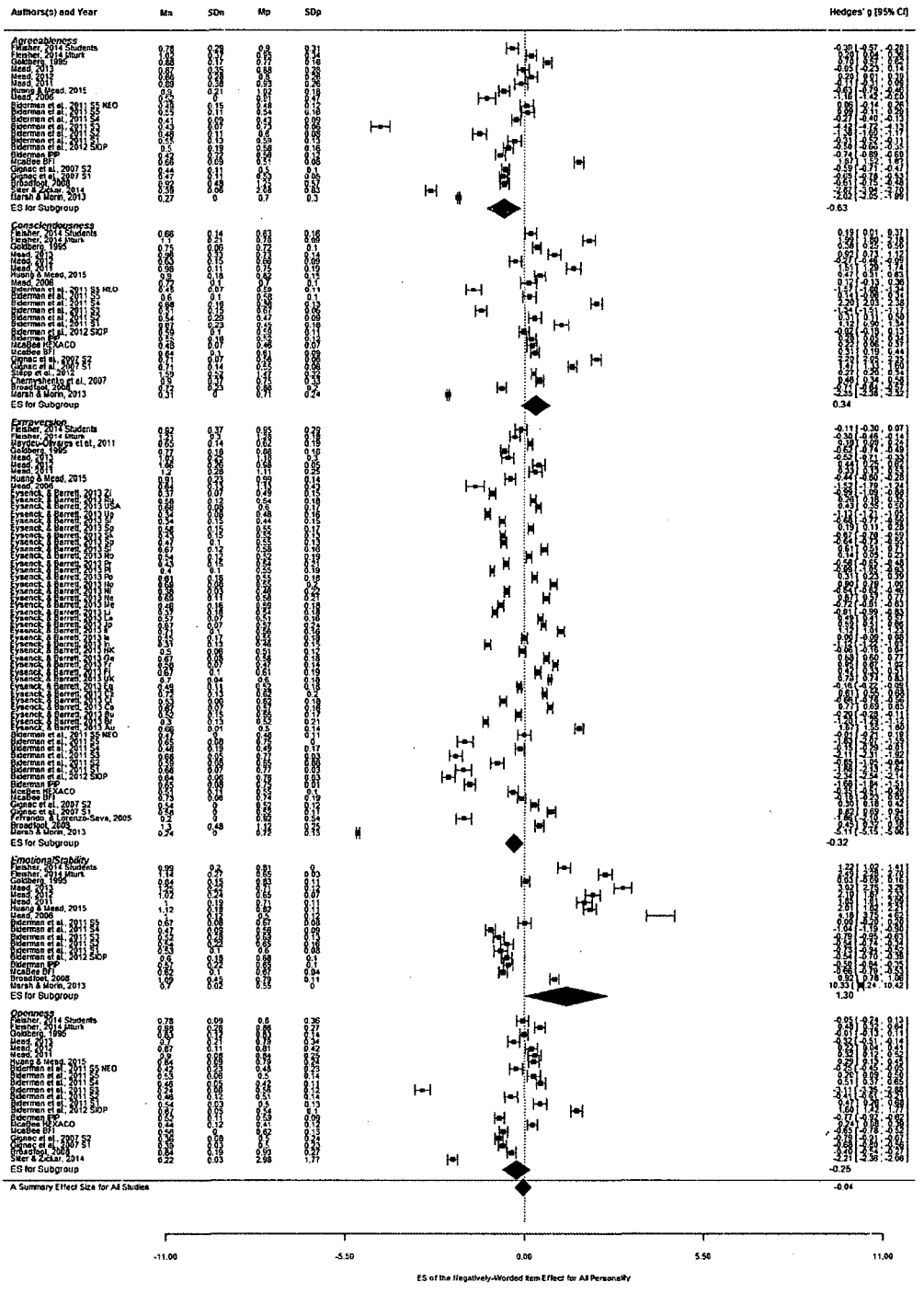


Figure 2. Forest plot of the negatively-worded item effect for all personality studies

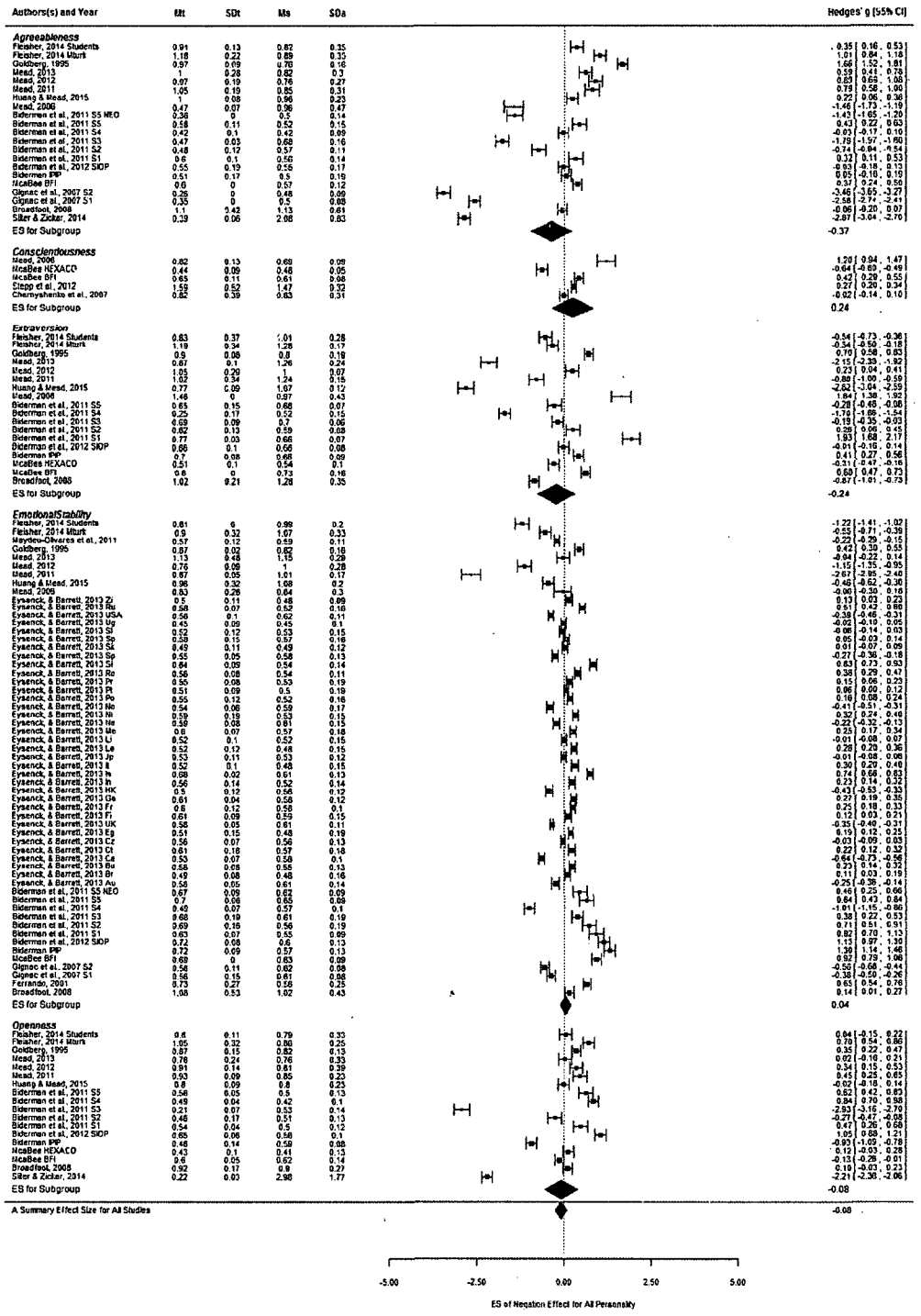


Figure 3. Forest plot of negation effect for all personality studies

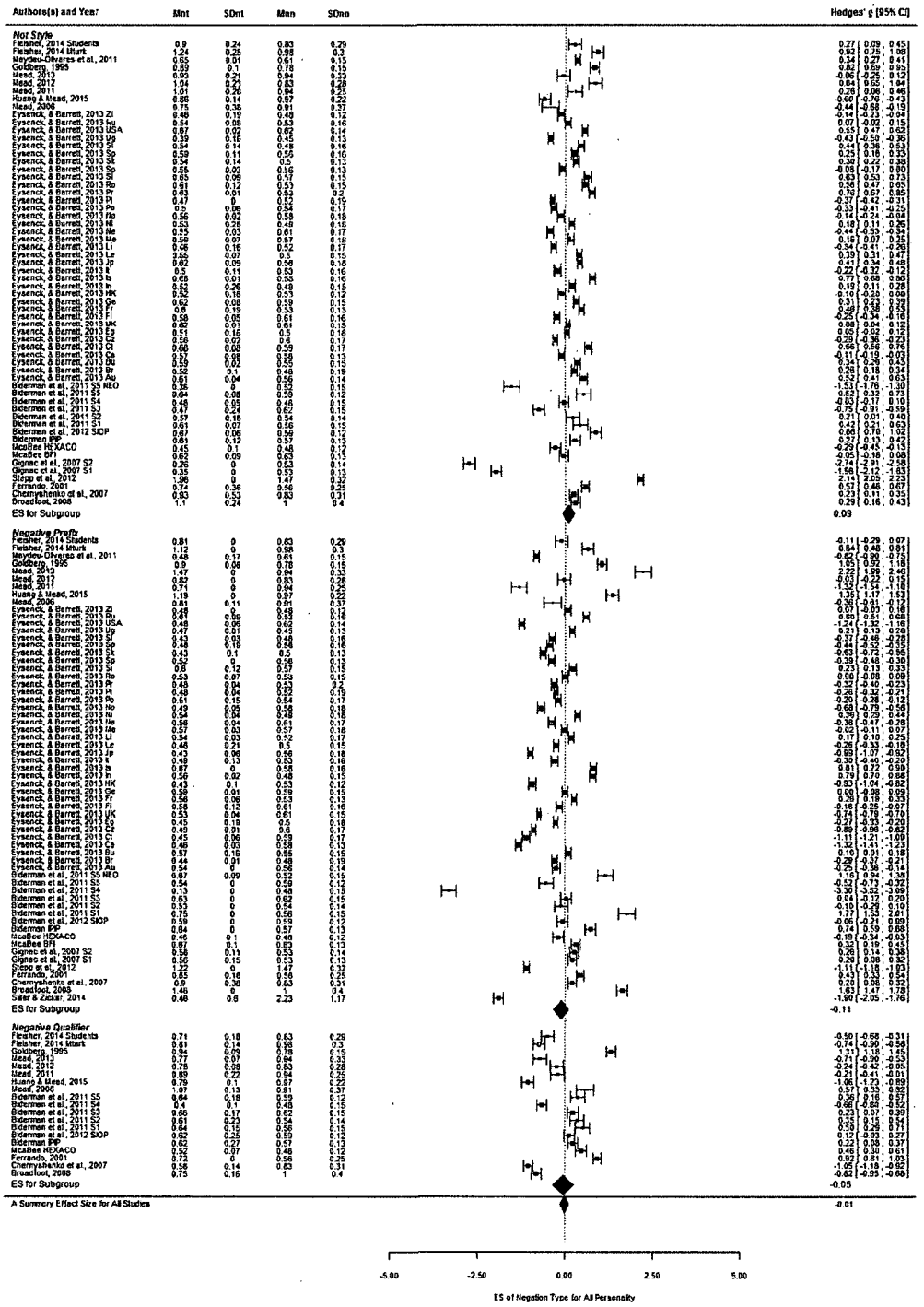


Figure 4. Forest plot of negation type for all personality studies

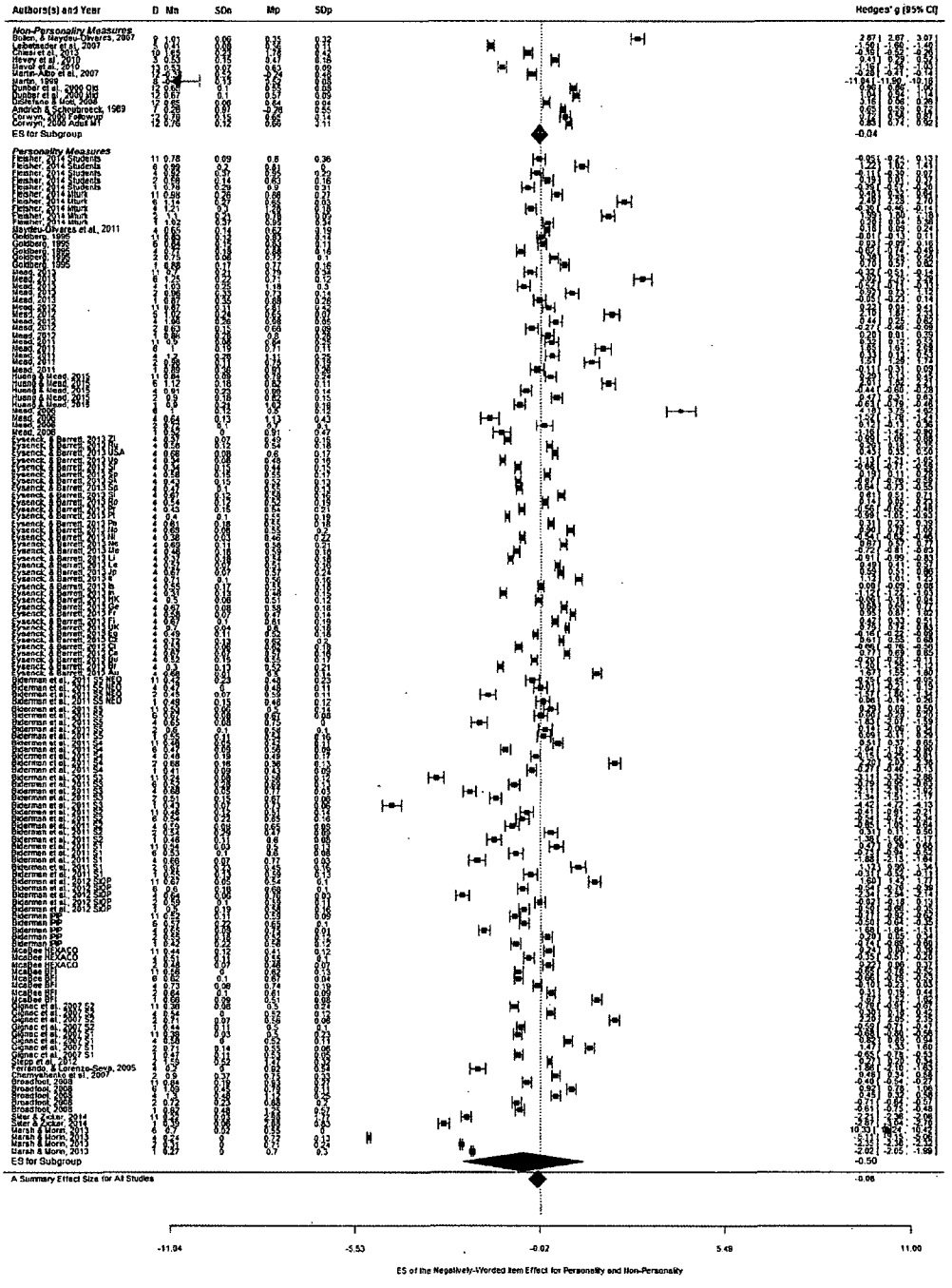


Figure 5. Forest plot of the negatively-worded item effect for personality and non-personality measures

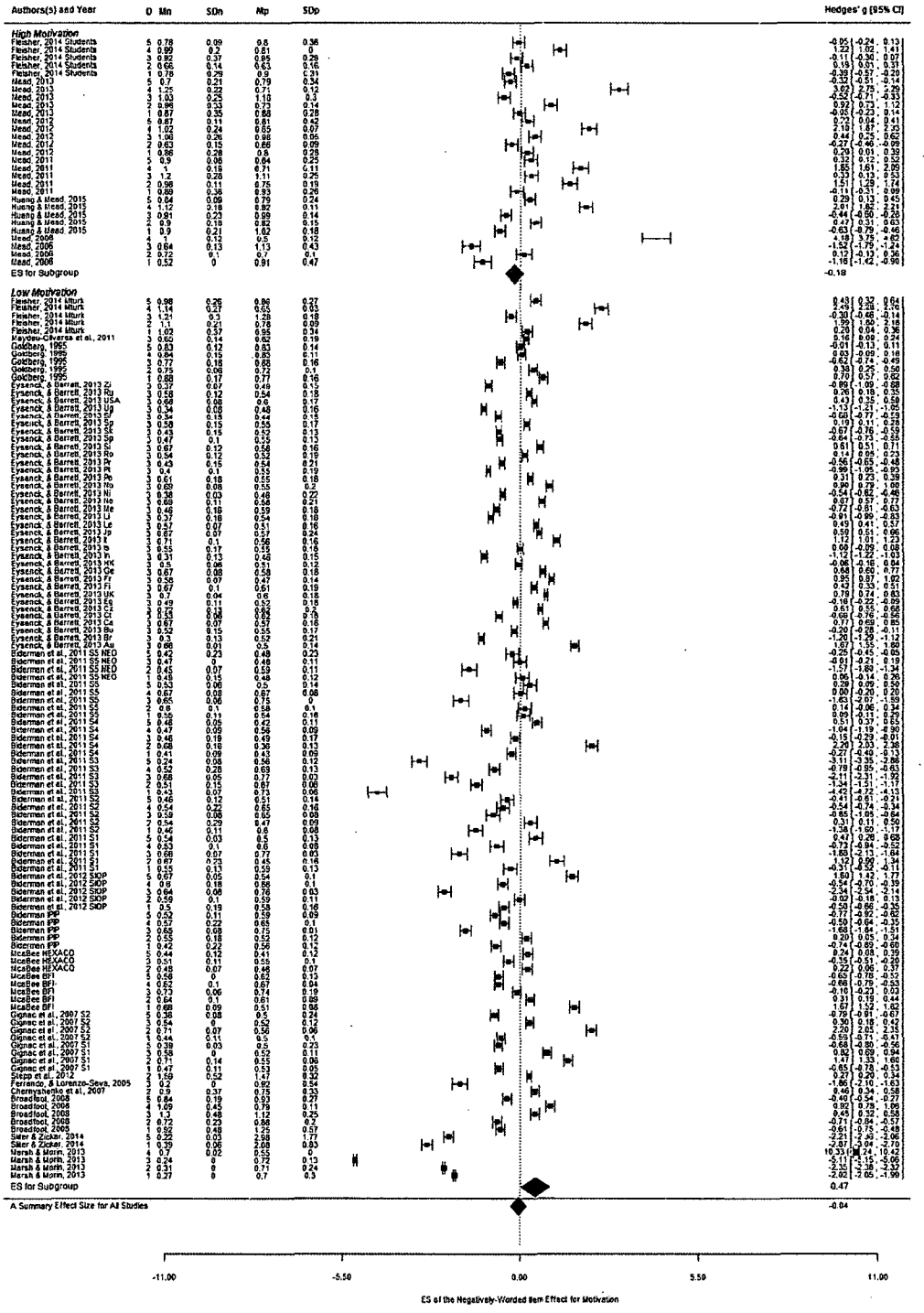


Figure 6. Forest plot for the negatively-worded item effect for motivation

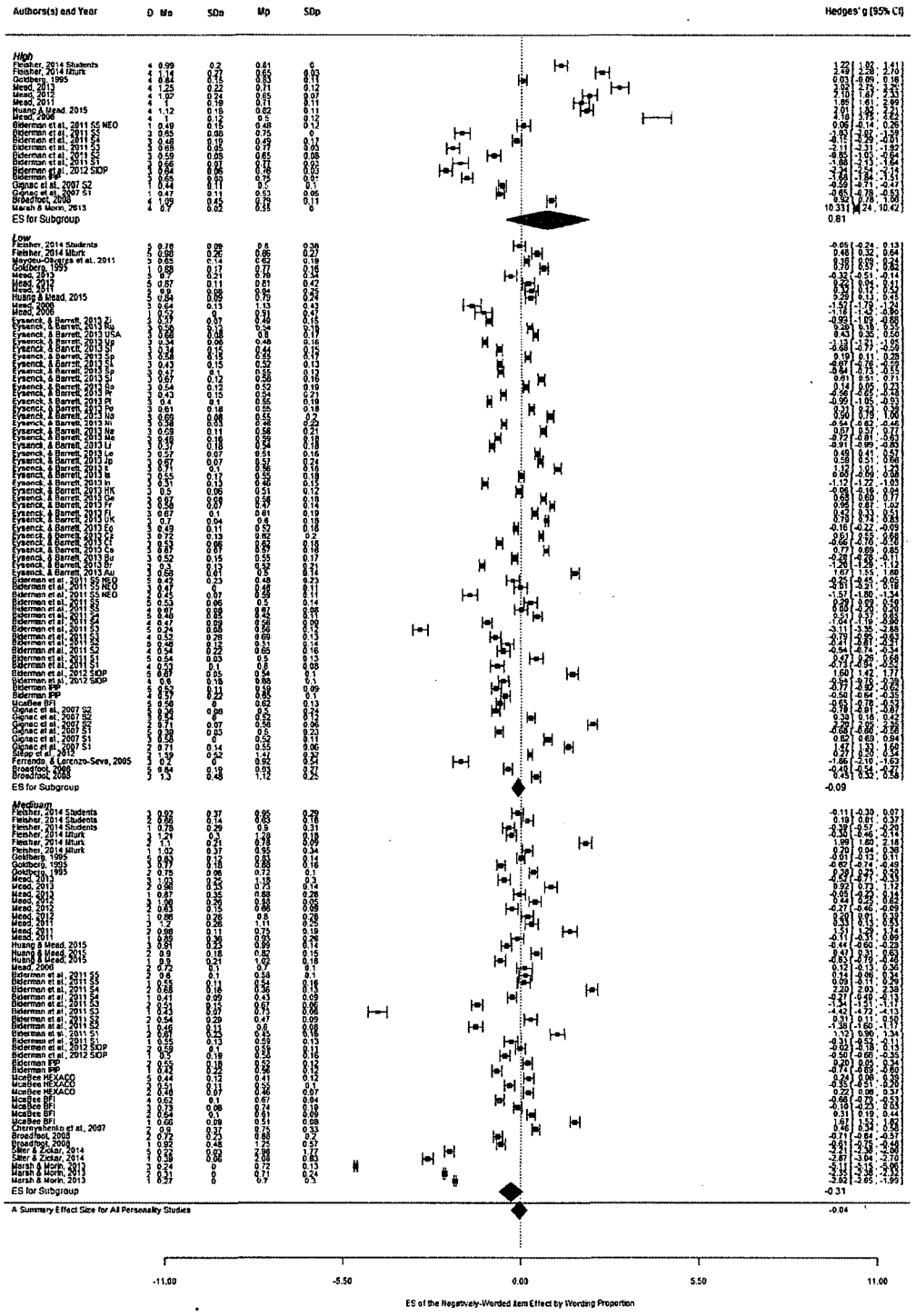


Figure 7. Forest plot for the negatively-worded item effect by wording proportion

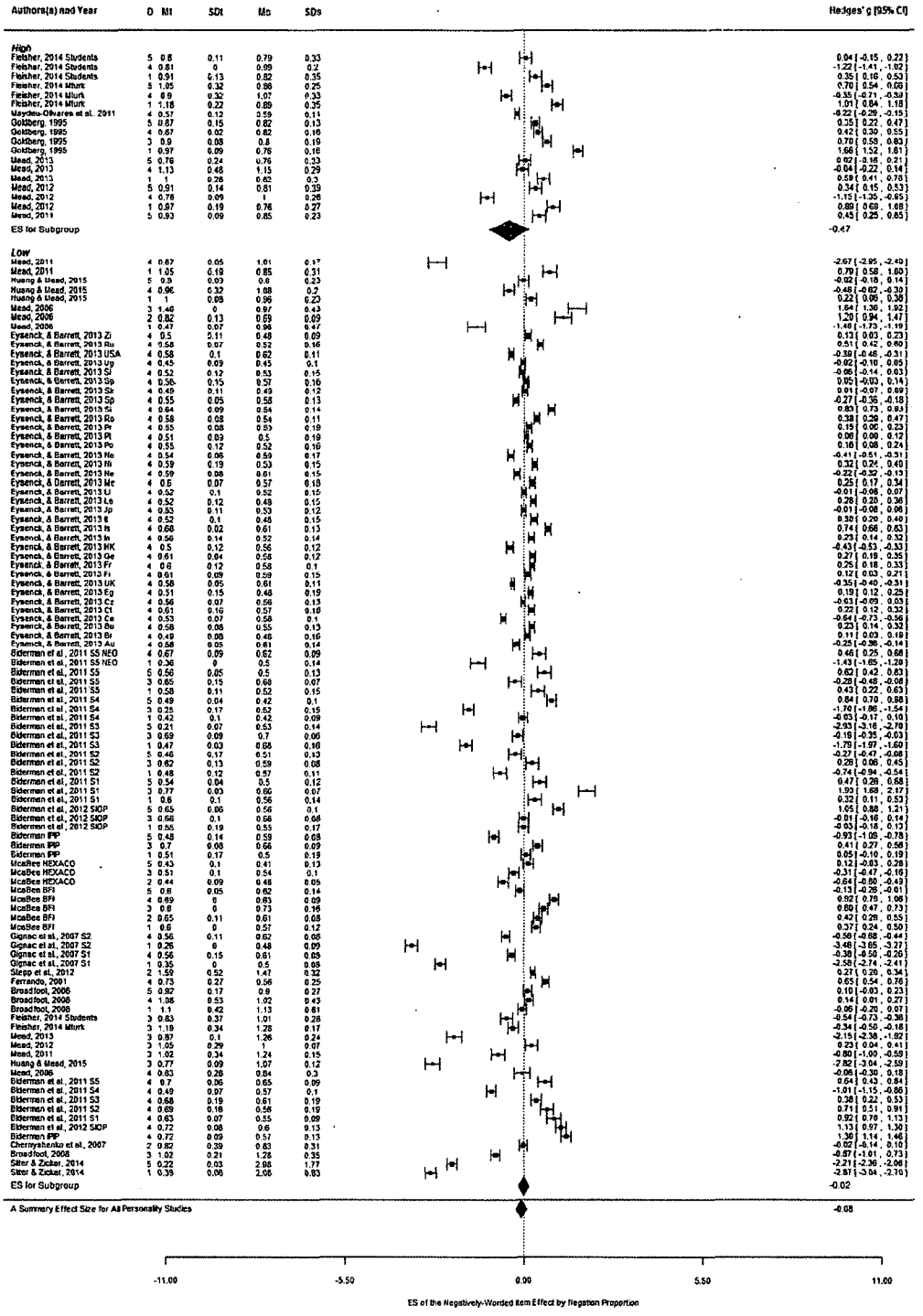


Figure 8. Forest plot for the negated item effect by negation proportion

APPENDIX F

FUNNEL PLOTS OF RESEARCH QUESTIONS AND HYPOTHESES

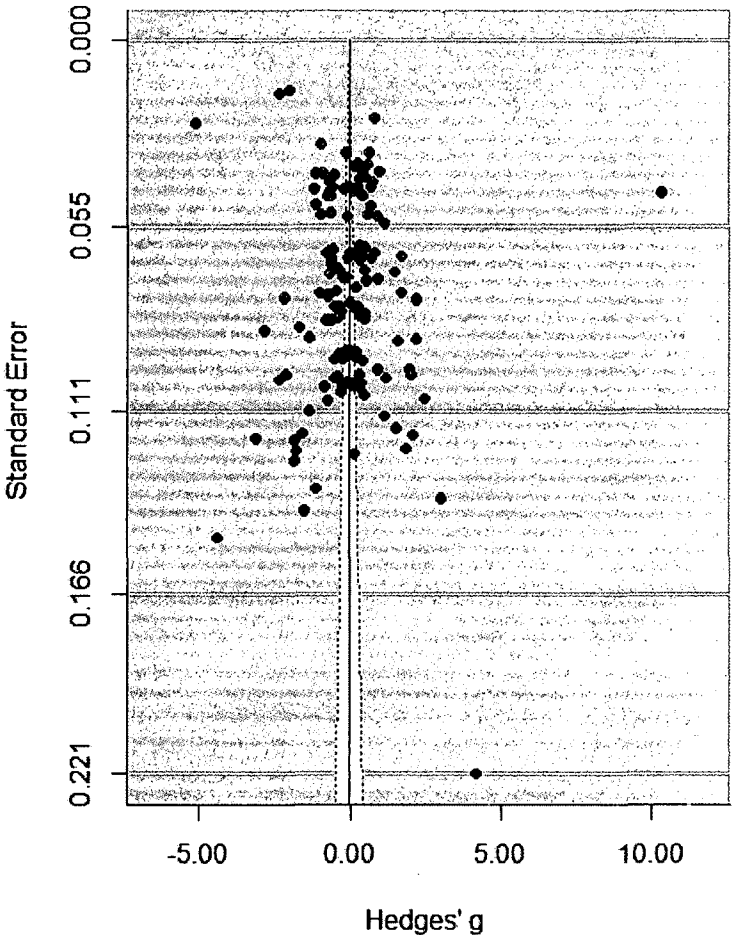


Figure 9. Funnel plot for evaluating publication bias for all personality studies

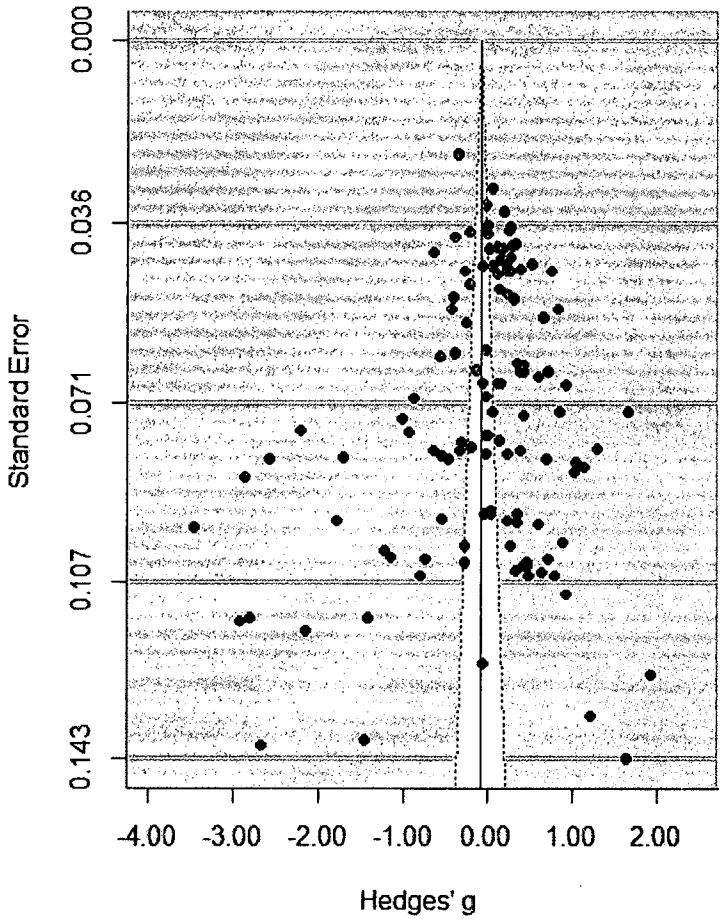


Figure 10. Funnel plot for evaluating publication bias for all negation studies

APPENDIX G

R CODE FOR ANALYSIS AND PLOTS

R code for analysis

```
# install.packages("MAd", dependencies=T)
# install.packages("metafor",dependencies=T)
# install.packages("sqldf",dependencies=T)
# install.packages("Hmisc", dependencies=T)
# install.packages("robumeta", dependencies=T)


##This program used all POOLED.SD for computing ES
##Calculate summary ES in Md packages
##Use RVE and meta-regression to estimate moderate analysis

#read in data file
RawData <- read.csv("CodingSheetData.csv",header=T, sep=",")
#show column names
colnames(RawData)
#select data with valid fields
RawData <- RawData [,c(1:21)]


#ITEMWORINDG SECTION--Data Manipulation
#calculate mean, sd, and frequency for positively and negatively
 worded items
library(sqldf)
ItemWording <- sqldf("select UniqueID, AuthorAndYear, Construct,
 IsSwapped, SampleSize, ItemWording, AVG(ItemStatistics) AS Mean,
 STDEV(ItemStatistics) AS SD, COUNT(*) AS Frequency from RawData
      GROUP BY UniqueID, AuthorAndYear, Construct, IsSwapped,
 SampleSize, ItemWording
      ORDER BY UniqueID, AuthorAndYear, Construct, IsSwapped,
 SampleSize, ItemWording")

#denormalize dataset into SPSS-like layout
WordingEffect <- sqldf(" SELECT n.*, p.Mean AS Mp, p.SD AS SDp,
 p.Frequency AS Fp
                        FROM
                          (select * from ItemWording WHERE
 ItemWording='-' ) AS n
                        INNER JOIN
                          (select * from ItemWording WHERE
 ItemWording='+' ) AS p
                        ON n.UniqueID=p.UniqueID And
 n.AuthorAndYear=p.AuthorAndYear AND n.Construct=p.Construct AND
 n.IsSwapped=p.IsSwapped")

#rename colnames
colnames(WordingEffect)[c(7:9)] <- c("Mn","SDn","Fn")

#create new variable for Personality (0=Non-Personality;
 1=Personality)
```

```
WordingEffect$Personality <- rep(0,nrow(WordingEffect))
for (nR in 1:nrow(WordingEffect) )
{
  if (WordingEffect$Construct[nR] %in% c("A","C","E","ES","O") )
 {
    WordingEffect$Personality[nR] <- 1  }
}


#calculate the proportion of negatively worded items
WordingEffect$WordingProportion <-
 WordingEffect$Fn/(WordingEffect$Fn+WordingEffect$Fp)


#sample-dependent test as OCEAN are from the same sample
#calculate weighted mean and weighted sd for variance of item
 wording
Var.Wording <- sqldf("SELECT UniqueID,AuthorAndYear, Construct,
 SampleSize, SDn, SDn*SDn AS Varn, SDp, SDp*SDp AS Varp FROM
 WordingEffect
                     WHERE Personality=1 AND (SDn>0 OR SDp>0)")

library(Hmisc)
WM.Negative <- wtd.mean(Var.Wording$Varn, Var.Wording$SampleSize)
WM.Positive <- wtd.mean(Var.Wording$Varp, Var.Wording$SampleSize)

WSD.Negative <- sqrt(wtd.var(Var.Wording$Varn,
 Var.Wording$SampleSize) )
WSD.Positive <- sqrt(wtd.var(Var.Wording$Varp,
 Var.Wording$SampleSize) )

print(c(WM.Negative,WM.Positive, WSD.Negative, WSD.Positive))


#####################################################################
 ###############################

WordingEffectInput <- sqldf("SELECT DISTINCT UniqueID,
 AuthorAndYear, Construct, SampleSize AS Nn, Mn, SDn, SampleSize
 AS Np, Mp,SDp
                             FROM WordingEffect WHERE
 Personality=1 AND (SDn>0 OR SDp>0)")


#RQ1: calculate a summary effect size from hedges' g under random
 effects
#calculate a summary effect size on each personality dimension
library(MAd)
Dimensions <- as.character(unique(WordingEffectInput$Construct))
for (iD in 1:length(Dimensions))
  {
    WordingEffectInput.temp <-
```

```
WordingEffectInput[which(WordingEffectInput$Construct==Dimension
s[iD]),]
    WordingD <- compute_ds(WordingEffectInput.temp$Nn,
WordingEffectInput.temp$Mn,WordingEffectInput.temp$SDn,
WordingEffectInput.temp$Np, WordingEffectInput.temp$Mp,
WordingEffectInput.temp$SDp, WordingEffectInput.temp, denom =
"pooled.sd")
    WordingG <- compute_gs(WordingD$d, WordingD$var.d ,
WordingD$Nn, WordingD$Np, WordingD)

    SummaryESWording <- omni(WordingG$g, WordingG$var.g, WordingG,
type="weighted", method = "random")

    print(Dimensions[iD])
    print(SummaryESWording)
  }

#calculate a summary effect size across all dimensions
WordingD <- compute_ds(WordingEffectInput$Nn,
 WordingEffectInput$Mn,WordingEffectInput$SDn,
 WordingEffectInput$Np, WordingEffectInput$Mp,
 WordingEffectInput$SDp, WordingEffectInput, denom = "pooled.sd")
WordingG <- compute_gs(WordingD$d, WordingD$var.d , WordingD$Nn,
 WordingD$Np, WordingD)

SummaryESWording2 <- omni(WordingG$g, WordingG$var.g, WordingG,
 type="weighted", method = "random")
print(SummaryESWording2)

#save the effect sizes into a table
write.table(WordingG[,-2], "WordingG.csv", quote=F, row.names=F,
 sep=",")


#remove outlines and calculate ES
WordingGNO <- sqldf("SELECT * FROM WordingG WHERE g>-10 AND g<10")

omni(WordingGNO$g, WordingGNO$var.g, WordingGNO, type="weighted",
 method = "random")

#RQ1: Redo the analysis by reversing ES into N
#Reverse ES into N
WordingEffectInput.four <- sqldf("SELECT UniqueID, Construct,
 SampleSize AS Nn, Mn, SDn, SampleSize AS Np, Mp,SDp
                            FROM WordingEffect WHERE
 Personality=1  AND (SDn>0 OR SDp>0)
                            AND CONSTRUCT <> 'ES' ")

WordingEffectInput.Neuro <- sqldf("SELECT UniqueID, Construct,
 SampleSize AS Nn, Mp AS Mn, SDp AS SDn, SampleSize AS Np, Mn AS
 Mp,SDn AS SDp
                            FROM WordingEffect WHERE
```

```
Personality=1  AND (SDn>0 OR SDp>0)
                              AND CONSTRUCT='ES' ")

#rename the construct into N
WordingEffectInput.Neuro$Construct <- rep("N",
 nrow(WordingEffectInput.Neuro))
#reconstruct the dataset for input
WordingEffectInputN <-
 rbind(WordingEffectInput.four,WordingEffectInput.Neuro)


#calculate a summary effect size across all dimensions
WordingND <- compute_ds(WordingEffectInputN$Nn,
 WordingEffectInputN$Mn,WordingEffectInputN$SDn,
 WordingEffectInputN$Np, WordingEffectInputN$Mp,
 WordingEffectInputN$SDp, WordingEffectInputN, denom =
 "pooled.sd")
WordingNG <- compute_gs(WordingND$d, WordingND$var.d ,
 WordingND$Nn, WordingND$Np, WordingND)

SummaryESWordingN2 <- omni(WordingNG$g, WordingNG$var.g,
 WordingNG, type="weighted", method = "random")
print(SummaryESWordingN2)


#remove outlines and calculate ES
WordingGNNO <- sqldf("SELECT * FROM WordingNG WHERE g>-10 AND
 g<10")

omni(g,var.g, WordingGNNO, type="weighted", method = "random")


# ES without one outlier
WordingGES <- sqldf("SELECT * FROM WordingNG WHERE g>-10 AND g<10
 AND Construct='N' ")

omni(g,var.g, WordingGES, type="weighted", method = "random")


##########################################################################
 ########################

#RQ3: Does personality moderate the wording effect?
#ANOVA for hedges' g
fit.dimension <- aov(g ~ UniqueID, data=WordingG)
s.d <- summary(fit.dimension)
#average number of ES per study
k.d <- nrow(WordingG)/length(unique(WordingG$UniqueID))

#calculate ICC=(MSB-MSW)/[MSB+(k-1)*MSW]
rho.d <- ( s.d[[1]] [[3]] [[1]]- s.d[[1]] [[3]] [[2]])/( s.d[[1]]
 [[3]] [[1]]+(k.d-1)* s.d[[1]] [[3]] [[2]])
```

```
#meta-regression by RVE
library(robumeta)
r.d <- robu(formula=g ~ Construct, data=WordingG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.d, small=F,
 modelweights = "HIER")

print(r.d)

#remove outlier--retest moderator
fit.dimension <- aov(g ~ UniqueID, data=WordingGNO)
s.d <- summary(fit.dimension)
#average number of ES per study
k.d <- nrow(WordingGNO)/length(unique(WordingGNO$UniqueID))

#calculate ICC=(MSB-MSW)/[MSB+(k-1)*MSW]
rho.d <- ( s.d[[1]] [[3]] [[1]]- s.d[[1]] [[3]] [[2]])/( s.d[[1]]
 [[3]] [[1]]+(k.d-1)* s.d[[1]] [[3]] [[2]])

#meta-regression by RVE
library(robumeta)
r.d <- robu(formula=g ~ Construct, data=WordingGNO,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.d, small=F,
 modelweights = "HIER")

print(r.d)


#REVERSE ES INTO N
WordingG.Neuro <- WordingG[which(WordingG$Construct=="ES"),c(1:3,
 13:14)]
WordingG.Neuro[,4] <- WordingG.Neuro[,4] *(-1)
WordingG.Neuro[,3] <- rep("N", nrow(WordingG.Neuro[,-3]))
WordingDimensionNG <- rbind(WordingG[WordingG$Construct %in%
 c("A","E","C","O"), c(1:3, 13:14)], WordingG.Neuro)

#use RVE on meta-regression
fit.dimensionN <- aov(g ~ UniqueID, data=WordingDimensionNG)
s.dN <- summary(fit.dimensionN)
#average number of ES per study
k.dN <-
 nrow(WordingDimensionNG)/length(unique(WordingDimensionNG$Unique
 ID))

#calculate ICC=(MSB-MSW)/[MSB+(k-1)*MSW]
rho.dN <- ( s.dN[[1]] [[3]] [[1]]- s.dN[[1]] [[3]]
 [[2]])/( s.dN[[1]] [[3]] [[1]]+(k.d-1)* s.dN[[1]] [[3]] [[2]])

#meta-regression by RVE
library(robumeta)
r.dN <- robu(formula=g ~ Construct, data=WordingDimensionNG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.dN, small=F,
```

```
modelweights = "HIER")

print(r.dN)

#remove one outlier
WordingDimensionNG <- sqldf("SELECT * FROM WordingDimensionNG
 WHERE g<10 AND g >-10")
#use RVE on meta-regression
fit.dimensionN <- aov(g ~ UniqueID, data=WordingDimensionNG)
s.dN <- summary(fit.dimensionN)
#average number of ES per study
k.dN <-
 nrow(WordingDimensionNG)/length(unique(WordingDimensionNG$Unique
 ID))

#calculate ICC=(MSB-MSW)/[MSB+(k-1)*MSW]
rho.dN <- ( s.dN[[1]] [[3]] [[1]]- s.dN[[1]] [[3]]
 [[2]])/( s.dN[[1]] [[3]] [[1]]+(k.d-1)* s.dN[[1]] [[3]] [[2]])

#meta-regression by RVE
library(robumeta)
r.dN <- robu(formula=g ~ Construct, data=WordingDimensionNG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.dN, small=F,
 modelweights = "HIER")

print(r.dN)


#####################################################################
 ########################

#H1: NEGATION EFFECT exists in personality
#calculate mean, sd and frequency for negation and non-negation
 items
#t for negated items, and s for non-negated items
Negation <- sqldf("select UniqueID, AuthorAndYear, Construct,
 SampleSize, Negation, AVG(ItemStatistics) AS Mean,
 STDEV(ItemStatistics) AS SD, COUNT(*) AS Frequency from RawData
      GROUP BY UniqueID, AuthorAndYear, Construct, SampleSize,
 Negation
      ORDER BY UniqueID, AuthorAndYear, Construct, SampleSize,
 Negation")

#denormalize data into SPSS format
NegationEffect <- sqldf(" SELECT t.*, s.Mean AS Ms, s.SD AS SDs,
 s.Frequency AS Fs
                        FROM
                          (select * from Negation WHERE
 Negation=1 ) AS t
                        INNER JOIN
                          (select * from Negation WHERE
 Negation=0 ) AS s
```

```
                        ON t.UniqueID=s.UniqueID ANd
 t.AuthorAndYear=s.AuthorAndYear AND t.Construct=s.Construct")

#column names
colnames(NegationEffect)
#rename columns
colnames(NegationEffect)[c(6:8)] <- c("Mt","SDt","Ft")
#create a new variable for personality
NegationEffect$Personality <- rep(0, nrow(NegationEffect))
for (nR in 1:nrow(NegationEffect) )
{
  if (NegationEffect$Construct[nR] %in% c("A","C","E","ES","O") )
 {
    NegationEffect$Personality[nR] <- 1  }
}

#calculate weighted mean and weighted sd for variance
Var.Negation <- sqldf("SELECT UniqueID, AuthorAndYear, Construct,
 SampleSize, SDt, SDt*SDt AS Vart, SDs, SDs*SDs AS Vars FROM
 NegationEffect
                     WHERE Personality=1 AND (SDt>0 OR SDs>0)")

library(Hmisc)
WM.Negation <- wtd.mean(Var.Negation$Vart,
 Var.Negation$SampleSize)
WM.NonNegation <- wtd.mean(Var.Negation$Vars,
 Var.Negation$SampleSize)

WSD.Negation <- sqrt(wtd.var(Var.Negation$Vart,
 Var.Negation$SampleSize) )
WSD.NonNegation <- sqrt(wtd.var(Var.Negation$Vars,
 Var.Negation$SampleSize) )

print(c(WM.Negation,WM.NonNegation, WSD.Negation,
 WSD.NonNegation))

#calculate a summary effect size for all personality dimensions
NegationEffectInput <- sqldf("SELECT UniqueID,AuthorAndYear,
 Construct, SampleSize AS Nt, Mt, SDt, SampleSize AS Ns, Ms,SDs
 FROM NegationEffect
                              WHERE Personality=1 AND (SDt>0 OR
 SDs>0) ")

NegationD <- compute_ds(NegationEffectInput$Nt,
 NegationEffectInput$Mt, NegationEffectInput$SDt,
 NegationEffectInput$Ns, NegationEffectInput$Ms,
 NegationEffectInput$SDs, NegationEffectInput, denom =
 "pooled.sd")
NegationG <- compute_gs(NegationD$d, NegationD$var.d ,
 NegationD$Nt, NegationD$Ns, NegationD)

SummaryESNegation <- omni(NegationG$g, NegationG$var.g, NegationG,
```

```
type="weighted", method = "random")

print(SummaryESNegation)

#save NegationG
write.table(NegationG [-2],"NegationG.csv", quote=F, row.names=F,
sep=",")

#calculate a summary effect size for each personality dimension
Dimensions.Negation <-
as.character(unique(NegationEffectInput$Construct))
for (iND in 1:length(Dimensions.Negation))
{
  NegationEffectInput.temp <-
NegationEffectInput[which(NegationEffectInput$Construct==Dimensi
ons.Negation[iND]),]
  NegationD <- compute_ds(NegationEffectInput.temp$Nt,
NegationEffectInput.temp$Mt, NegationEffectInput.temp$SDt,
NegationEffectInput.temp$Ns, NegationEffectInput.temp$Ms,
NegationEffectInput.temp$SDs, NegationEffectInput.temp, denom =
"pooled.sd")
  NegationG <- compute_gs(NegationD$d, NegationD$var.d ,
NegationD$Nt, NegationD$Ns, NegationD)

  SummaryESNegation <- omni(NegationG$g, NegationG$var.g,
NegationG, type="weighted", method = "random")

  print(Dimensions.Negation[iND])
  print(SummaryESNegation)
}


#Whether Different Types of Negation Work Differently
NegationType <- sqldf("select UniqueID, AuthorAndYear, SampleSize,
NegType, AVG(ItemStatistics) AS Mean, STDEV(ItemStatistics) AS
SD, COUNT(*) AS Frequency from RawData
      WHERE NegType in (1,2,3) AND Construct in ('A', 'C', 'E',
'ES', 'O')
      GROUP BY UniqueID, AuthorAndYear, SampleSize, NegType
      ORDER BY UniqueID, AuthorAndYear, SampleSize, NegType")

NonNegation <- sqldf("select UniqueID, AuthorAndYear, SampleSize,
Negation, AVG(ItemStatistics) AS Mean, STDEV(ItemStatistics) AS
SD, COUNT(*) AS Frequency from RawData
      WHERE Negation=0 AND Construct in ('A', 'C', 'E', 'ES', 'O')
      GROUP BY UniqueID, AuthorAndYear, SampleSize, Negation
      ORDER BY UniqueID, AuthorAndYear, SampleSize, Negation")

NegationTypeInput <- sqldf("SELECT nt.*, nn.SampleSize AS Nnn,
nn.Mean AS Mnn, nn.SD AS SDnn, nn.Frequency AS Fnn
                           FROM NegationType AS nt
                           LEFT JOIN NonNegation As nn
```

```
                                ON nt. UniqueID=nn.UniqueID AND
 nt.AuthorAndYear=nn.AuthorAndYear AND
 nt.SampleSize=nn.SampleSize
                                WHERE nt.SD>0 OR nn.SD>0")

colnames(NegationTypeInput) [c(3:7)] <- c("Nnt","NegType", "Mnt",
 "SDnt", "Fnt")

NegationTypeD <- compute_ds(Nnt, Mnt, SDnt, Nnn, Mnn, SDnn,
 NegationTypeInput, denom = "pooled.sd")
NegationTypeG <- compute_gs(NegationTypeD$d, NegationTypeD$var.d ,
 NegationTypeD$Nnt, NegationTypeD$Nnn, NegationTypeD)

SummaryESNegationType <- omni(g, var.g, NegationTypeG,
 type="weighted", method = "random")

print(SummaryESNegationType)

#save NegationTypeG
write.table(NegationTypeG[,-2], "NegationTypeG.csv", quote=F,
 row.names=F, sep=",")

macat(g,var.g, NegType, NegationTypeG, method="random")

##################################################################
 ########################
#H2: Non-Personality and Personality
#calculate heges' g under random effects
WordingEffectPersonality <- sqldf("SELECT UniqueID, Construct,
 SampleSize AS Nn, Mn, SDn, SampleSize AS Np, Mp,SDp,Personality
 FROM WordingEffect
                                    WHERE SDn>0 OR SDp>0")

WordingPersonalityD <- compute_ds(WordingEffectPersonality$Nn,
 WordingEffectPersonality$Mn, WordingEffectPersonality$SDn,
 WordingEffectPersonality$Np, WordingEffectPersonality$Mp,
 WordingEffectPersonality$SDp, WordingEffectPersonality, denom =
 "pooled.sd")
WordingPersonalityG <- compute_gs(WordingPersonalityD$d,
 WordingPersonalityD$var.d , WordingPersonalityD$Nn,
 WordingPersonalityD$Np, WordingPersonalityD)

SummaryESWordingPersonality <- omni(WordingPersonalityG$g,
 WordingPersonalityG$var.g, WordingPersonalityG, type="weighted",
 method = "random")
print(SummaryESWordingPersonality)

#save WordingPersonalityG
write.table(WordingPersonalityG,"WordingPersonalityG.csv",
 quote=F, row.names=F, sep=",")
```

```
##USE RVE TO do subgroup analysis
#conduct an ANOVA
fit.personality <- aov(g ~ UniqueID, data=WordingPersonalityG)
s.p <- summary(fit.personality)
k.p <-
 nrow(WordingPersonalityG)/length(unique(WordingPersonalityG$Uniq
 ueID))

rho.p <- ( s.p[[1]] [[3]] [[1]]- s.p[[1]] [[3]] [[2]])/( s.p[[1]]
 [[3]] [[1]]+(k.p-1)* s.p[[1]] [[3]] [[2]])

r.p<- robu(formula=g ~ Personality, data=WordingPersonalityG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.p, small=F,
 modelweights = "HIER")

print(r.p)

#remove outliers
WordingPersonalityG <- sqldf("SELECT * FROM WordingPersonalityG
 WHERE g<10 AND g> -10")
#conduct an ANOVA
fit.personality <- aov(g ~ UniqueID, data=WordingPersonalityG)
s.p <- summary(fit.personality)
k.p <-
 nrow(WordingPersonalityG)/length(unique(WordingPersonalityG$Uniq
 ueID))

rho.p <- ( s.p[[1]] [[3]] [[1]]- s.p[[1]] [[3]] [[2]])/( s.p[[1]]
 [[3]] [[1]]+(k.p-1)* s.p[[1]] [[3]] [[2]])

r.p<- robu(formula=g ~ Personality, data=WordingPersonalityG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.p, small=F,
 modelweights = "HIER")

print(r.p)


#####################################################################
 ###################
#H3: Sample Motivation
WordingEffectMotivation <- sqldf("SELECT UniqueID, AuthorAndYear,
 Construct, SampleSize AS Nn, Mn, SDn, SampleSize AS Np, Mp,SDp
 FROM WordingEffect
                                WHERE Personality=1 AND (SDn>0 OR
 SDp>0 )")

#assign High and Low Motivation Groups
WordingEffectMotivation$Motivation <- rep("L",
 nrow(WordingEffectMotivation))
WordingEffectMotivation$Motivation[WordingEffectMotivation$Unique
 ID %in% c(67,68,69,70,71,77)] <-
 rep("H",length(WordingEffectMotivation$Motivation[WordingEffectM
```

```
otivation$UniqueID %in% c(67,68,69,70,71,77)]))

#calculate heges' g
WordingMotivationD <- compute_ds(WordingEffectMotivation$Nn,
 WordingEffectMotivation$Mn, WordingEffectMotivation$SDn,
 WordingEffectMotivation$Np, WordingEffectMotivation$Mp,
 WordingEffectMotivation$SDp, WordingEffectMotivation, denom =
 "pooled.sd")
WordingMotivationG <- compute_gs(WordingMotivationD$d,
 WordingMotivationD$var.d, WordingMotivationD$Nn,
 WordingMotivationD$Np, WordingMotivationD)

SummaryESWordingMotivation <- omni(WordingMotivationG$g,
 WordingMotivationG$var.g, WordingMotivationG, type="weighted",
 method = "random")

SummaryESWordingMotivation

#save WordingMotivationG
write.table(WordingMotivationG[,-2], "WordingMotivationG.csv",
 quote=F, row.names=F, sep=",")

##USE RVE TO do subgroup analysis
#conduct an ANOVA
fit.motivation <- aov(g ~ UniqueID, data=WordingMotivationG)
s.m <- summary(fit.motivation)
k.m <-
 nrow(WordingMotivationG)/length(unique(WordingMotivationG$Unique
 ID))

rho.m <- ( s.m[[1]] [[3]] [[1]]- s.m[[1]] [[3]] [[2]])/( s.m[[1]]
 [[3]] [[1]]+(k.m-1)* s.m[[1]] [[3]] [[2]])

r.m<- robu(formula=g ~ Motivation, data=WordingMotivationG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.m, small=F,
 modelweights = "HIER")

print(r.m)

#remove one outlier
WordingMotivationG <- sqldf("SELECT * FROM WordingMotivationG
 WHERE g<10 AND g>-10")
#conduct an ANOVA
fit.motivation <- aov(g ~ UniqueID, data=WordingMotivationG)
s.m <- summary(fit.motivation)
k.m <-
 nrow(WordingMotivationG)/length(unique(WordingMotivationG$Unique
 ID))

rho.m <- ( s.m[[1]] [[3]] [[1]]- s.m[[1]] [[3]] [[2]])/( s.m[[1]]
 [[3]] [[1]]+(k.m-1)* s.m[[1]] [[3]] [[2]])
```

```
r.m<- robu(formula=g ~ Motivation, data=WordingMotivationG,
 studynum=UniqueID, var.eff.size=var.g, rho=rho.m, small=F,
 modelweights = "HIER")

print(r.m)

#################################################################

#test the effect of proportion of negatively worded items
WordingProportionInput <- sqldf("SELECT DISTINCT UniqueID,
 AuthorAndYear, Construct, WordingProportion, SampleSize AS Nn,
 Mn, SDn, SampleSize AS Np, Mp,SDp
                          FROM WordingEffect WHERE
 Personality=1 AND (SDn>0 OR SDp>0)")

unique(WordingProportionInput$WordingProportion)

WordingProportionInput$Proportion[WordingProportionInput$WordingP
 roportion < 0.32]<- "Low"
WordingProportionInput$Proportion[WordingProportionInput$WordingP
 roportion > 0.66]<- "High"
WordingProportionInput$Proportion[WordingProportionInput$WordingP
 roportion > 0.33 & WordingProportionInput$WordingProportion <
 0.66]<- "Median"

WordingProportionD <- compute_ds(WordingProportionInput$Nn,
 WordingProportionInput$Mn, WordingProportionInput$SDn,
 WordingProportionInput$Np, WordingProportionInput$Mp,
 WordingProportionInput$SDp, WordingProportionInput, denom =
 "pooled.sd")
WordingProportionG <- compute_gs(WordingProportionD$d,
 WordingProportionD$var.d, WordingProportionD$Nn,
 WordingProportionD$Np, WordingProportionD)

SummaryESWordingProportion <- omni(WordingProportionG$g,
 WordingProportionG$var.g, WordingProportionG, type="weighted",
 method = "random")

SummaryESWordingProportion

macat(g, var.g, Proportion, WordingProportionG, method= "random")

#################################################################
 ############

NegationEffect$NegationProportion <-
 NegationEffect$Ft/(NegationEffect$Ft+NegationEffect$Fs)

unique(NegationEffect$NegationProportion)

NegationProportionInput <- sqldf("SELECT DISTINCT UniqueID,
 AuthorAndYear, Construct, NegationProportion, SampleSize AS Nt,
```

```
Mt, SDt, SampleSize AS Ns, Ms,SDs
                            FROM NegationEffect WHERE
Personality=1 AND (SDt>0 OR SDs>0)")


NegationProportionInput$Proportion[NegationProportionInput$Negati
onProportion < 0.349]<- "Low"
NegationProportionInput$Proportion[NegationProportionInput$Negati
onProportion > 0.349]<- "High"

NegationProportionD <- compute_ds(NegationProportionInput$Nt,
NegationProportionInput$Mt, NegationProportionInput$SDt,
NegationProportionInput$Ns, NegationProportionInput$Ms,
NegationProportionInput$SDs, NegationProportionInput, denom =
"pooled.sd")
NegationProportionG <- compute_gs(NegationProportionD$d,
NegationProportionD$var.d, NegationProportionD$Nt,
NegationProportionD$Ns, NegationProportionD)

SummaryESNegationProportion <- omni(NegationProportionG$g,
NegationProportionG$var.g, NegationProportionG, type="weighted",
method = "random")

SummaryESNegationProportion

macat(g, var.g, Proportion, NegationProportionG, method=
"random")
```

R code for plotting

```
#Produce forest plots and funnel plots
library(sqldf)
library(MAd)
library(metafor)

#RQ1
WordingG <- read.csv("WordingG.csv", sep=",", header=T)
AuthorAndYear <-
 read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
 header=T, sep=",")
WordingG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM WordingG w
 INNER JOIN AuthorAndYear a
                  ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_allpersonality.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(WordingG$g~1, var=WordingG$var.g, data = WordingG)

par("usr")

forest(WE,
       xlim=c(-16,12), ilab=cbind(round(WordingG$Mn,2),
 round(WordingG$SDn,2), round(WordingG$Mp,2),
 round(WordingG$SDp,2)),
       ilab.xpos=c(-11,-9.5,-8,-6.5), #probably need to change
       order=order(WordingG$Construct, decreasing=T),
       alim=c(-11,11),
       ylim=c(-1,166.5),
       rows=c(3:23, 28:45, 50:108, 113:136, 141:162 ),
       xlab="ES of the Negatively-Worded Item Effect for All
 Personality", mlab="A Summary Effect Size for All Studies",
       slab=WordingG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
 match or not
text(-16, c (24, 46, 109, 137, 163), pos= 4, c("Openness",
 "EmotionalStability","Extraversion", "Conscientiousness",
 "Agreeableness"))

par(font=2)

text(c(-11,-9.5,-8,-6.5), 167, c ("Mn", "SDn", "Mp", "SDp"))
```

```
text(-16, 167, "Authors(s) and Year", pos=4)
text(12, 167, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
 data=WordingG[which(WordingG$Construct=="A"),])
res.c <- mareg(g~1, var.g,
 data=WordingG[which(WordingG$Construct=="C"),])
res.e <- mareg(g~1, var.g,
 data=WordingG[which(WordingG$Construct=="E"),])
res.n <- mareg(g~1, var.g,
 data=WordingG[which(WordingG$Construct=="ES"),])
res.o <- mareg(g~1, var.g,
 data=WordingG[which(WordingG$Construct=="O"),])


addpoly(res.a, row=139.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=111.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.e, row=48.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.n, row=26.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.o, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()


#trim and fill, funnel plot
WE0 <- rma(g, var.g, data=WordingG)
TWE <- trimfill(WE0)
TWE
funnel(TWE, xlab="Hedges' g ")


#####################################################################
 #########
#H1
NegationG <- read.csv("NegationG.csv", sep=",", header=T)
AuthorAndYear <-
 read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
 header=T, sep=",")
NegationG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM NegationG w
 INNER JOIN AuthorAndYear a
                  ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_allnegation.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(NegationG$g~1, var=NegationG$var.g, data = NegationG)
```

```
par("usr")

forest(WE,
       xlim=c(-16,10), ilab=cbind(round(NegationG$Mt,2),
 round(NegationG$SDt,2), round(NegationG$Ms,2),
 round(NegationG$SDs,2)),
       ilab.xpos=c(-11,-9.5,-8,-6.5), #probably need to change
       order=order(NegationG$Construct, decreasing=T),
       alim=c(-5,5),
       ylim=c(-1,141.5),
       rows=c( 3:20, 25:81, 86:103, 108:112, 117:137),
       xlab="ES of Negation Effect for All Personality", mlab="A
 Summary Effect Size for All Studies",
       slab=NegationG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
 match or not
text(-16, c (21, 82, 104, 113, 138), pos= 4, c("Openness",
 "EmotionalStability","Extraversion", "Conscientiousness",
 "Agreeableness"))

par(font=2)

text(c(-11,-9.5,-8,-6.5), 141, c ("Mt", "SDt", "Ms", "SDs"))
text(-16, 141, "Authors(s) and Year", pos=4)
text(10, 141, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
 data=NegationG[which(NegationG$Construct=="A"),])
res.c <- mareg(g~1, var.g,
 data=NegationG[which(NegationG$Construct=="C"),])
res.e <- mareg(g~1, var.g,
 data=NegationG[which(NegationG$Construct=="E"),])
res.n <- mareg(g~1, var.g,
 data=NegationG[which(NegationG$Construct=="ES"),])
res.o <- mareg(g~1, var.g,
 data=NegationG[which(NegationG$Construct=="O"),])


addpoly(res.a, row=115.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=106.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.e, row=84.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.n, row=23.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.o, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()
```

```
#trim and fill, funnel plot
WN0 <- rma(g, var.g, data=NegationG)
TNE <- trimfill(WN0)
TNE
funnel(TNE, xlab="Hedges' g ")

#####################################################################
################
#H1
NegationTypeG <- read.csv("NegationTypeG.csv", sep=",", header=T)
AuthorAndYear <-
read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
header=T, sep=",")
NegationTypeG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM
NegationTypeG w INNER JOIN AuthorAndYear a
                  ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_negationtype.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(NegationTypeG$g~1, var=NegationTypeG$var.g, data =
NegationTypeG)

par("usr")

forest(WE,
       xlim=c(-16,10), ilab=cbind( round(NegationTypeG$Mnt,2),
round(NegationTypeG$SDnt,2), round(NegationTypeG$Mnn,2),
round(NegationTypeG$SDnn,2)),
       ilab.xpos=c( -11,-9.5,-8,-6.5), #probably need to change
       order=order(NegationTypeG$NegType, decreasing=T),
       alim=c(-5,5),
       ylim=c(-1,154.5),
       rows=c( 3:21, 26:86, 91:150),
       xlab="ES of Negation Type for All Personality", mlab="A
Summary Effect Size for All Studies",
       slab=NegationTypeG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
match or not
text(-16, c (22, 87, 151), pos= 4, c("Negative Qualifier",
"Negative Prefix","Not Style"))

par(font=2)

text(c( -11,-9.5,-8,-6.5), 154, c ("Mnt", "SDnt", "Mnn", "SDnn"))
```

```
text(-16, 154, "Authors(s) and Year", pos=4)
text(10, 154, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
 data=NegationTypeG[which(NegationTypeG$NegType==1),])
res.c <- mareg(g~1, var.g,
 data=NegationTypeG[which(NegationTypeG$NegType==2),])
res.e <- mareg(g~1, var.g,
 data=NegationTypeG[which(NegationTypeG$NegType==3),])


addpoly(res.a, row=89.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=24.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.e, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()

#trim and fill, funnel plot
WT0 <- rma(g, var.g, data=NegationTypeG)
TTE <- trimfill(WT0)
TTE
funnel(TTE, xlab="Hedges' g ")

#####################################################################
 ##############################

#H2

MotivationG <- read.csv("WordingMotivationG.csv", sep=",",
 header=T)
AuthorAndYear <-
 read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
 header=T, sep=",")
MotivationG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM
 MotivationG w INNER JOIN AuthorAndYear a
                      ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_Motivation.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(MotivationG$g~1, var=MotivationG$var.g, data =
 MotivationG)

par("usr")

forest(WE,
       xlim=c(-16,12),
```

```
ilab=cbind(MotivationG$Construct,round(MotivationG$Mn,2),
round(MotivationG$SDn,2), round(MotivationG$Mp,2),
round(MotivationG$SDp,2)),
      ilab.xpos=c(-11.5,-11,-9.5,-8,-6.5), #probably need to
change
      order=order(MotivationG$Motivation, decreasing=T),
      alim=c(-11,11),
      ylim=c(-1,154.5),
      rows=c( 3:117, 122:150),
      xlab="ES of the Negatively-Worded Item Effect for
Motivation", mlab="A Summary Effect Size for All Studies",
      slab=MotivationG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
 match or not
text(-16, c (118, 151), pos= 4, c("Low Motivation", "High
 Motivation"))

par(font=2)

text(c(-11.5,-11,-9.5,-8,-6.5), 154, c ("D","Mn", "SDn", "Mp",
 "SDp"))
text(-16, 154, "Authors(s) and Year", pos=4)
text(12, 154, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
 data=MotivationG[which(MotivationG$Motivation=="L"),])
res.c <- mareg(g~1, var.g,
 data=MotivationG[which(MotivationG$Motivation=="H"),])


addpoly(res.a, row=120.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()

#trim and fill, funnel plot
WM0 <- rma(g, var.g, data=MotivationG)
TME <- trimfill(WM0)
TME
funnel(TME, xlab="Hedges' g ")

#################################################################


PersonalityG <- read.csv("WordingPersonalityG.csv", sep=",",
```

```
 header=T)
AuthorAndYear <-
 read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
 header=T, sep=",")
PersonalityG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM
 PersonalityG w INNER JOIN AuthorAndYear a
                       ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_PersonalityNot.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(PersonalityG$g~1, var=PersonalityG$var.g, data =
 PersonalityG)

par("usr")

forest(WE,
       xlim=c(-16,12),
 ilab=cbind(PersonalityG$Construct,round(PersonalityG$Mn,2),
 round(PersonalityG$SDn,2), round(PersonalityG$Mp,2),
 round(PersonalityG$SDp,2)),
       ilab.xpos=c(-11.5,-11,-9.5,-8,-6.5), #probably need to
 change
       order=order(PersonalityG$Personality, decreasing=T),
       alim=c(-11,11),
       ylim=c(-1,167.5),
       rows=c( 3:146, 151:163),
       xlab="ES of the Negatively-Worded Item Effect for
 Personality and Non-Personality", mlab="A Summary Effect Size
 for All Studies",
       slab=PersonalityG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
 match or not
text(-16, c (147, 164), pos= 4, c("Personality Measures", "Non-
 Personality Measures"))

par(font=2)

text(c(-11.5, -11,-9.5,-8,-6.5), 167, c ("D","Mn", "SDn", "Mp",
 "SDp"))
text(-16, 167, "Authors(s) and Year", pos=4)
text(12, 167, "Hedges' g [95% CI]", pos=2)

par(op)
```

```
res.a <- mareg(g~1, var.g,
 data=PersonalityG[which(PersonalityG$Personality==1),])
res.c <- mareg(g~1, var.g,
 data=PersonalityG[which(PersonalityG$Personality==0),])


addpoly(res.a, row=149.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()

#trim and fill, funnel plot
WP0 <- rma(g, var.g, data=PersonalityG)
TPE <- trimfill(WP0)
TPE
funnel(TPE, xlab="Hedges' g ")


#####################################################################
 ##########################
#WordingProportion addtional analysis plots
WordingProportionG <- read.csv("WordingProportionG.csv", sep=",",
 header=T)
AuthorAndYear <-
 read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
 header=T, sep=",")
WordingProportionG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM
 WordingProportionG w INNER JOIN AuthorAndYear a
                   ON w.UniqueID=a.UniqueID")

#levels(WordingProportionG$Proportion) <- c(3,1,2)

#create a forest plot
png(filename="forest_plot_with_wordingproportion.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(WordingProportionG$g~1, var=WordingProportionG$var.g,
 data = WordingProportionG)

par("usr")

forest(WE,
       xlim=c(-16,12), ilab=cbind(WordingProportionG$Construct,
 round(WordingProportionG$Mn,2), round(WordingProportionG$SDn,2),
 round(WordingProportionG$Mp,2),
 round(WordingProportionG$SDp,2)),
       ilab.xpos=c(-11.5, -11,-9.5,-8,-6.5), #probably need to
 change
       order=order(WordingProportionG$Proportion, decreasing=T),
       alim=c(-11,11),
```

```
       ylim=c(-1,160),
       rows=c(3:53, 58:130, 135:154), #5 units in between
       xlab="ES of the Negatively-Worded Item Effect by Wording
Proportion", mlab="A Summary Effect Size for All Personality
Studies",
       slab=WordingProportionG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
 match or not
text(-16, c (54, 131, 155), pos= 4, c("Mediuam", "Low", "High"))

par(font=2)

text(c(-11.5, -11,-9.5,-8,-6.5), 161, c ("D","Mn", "SDn", "Mp",
"SDp"))
text(-16, 161, "Authors(s) and Year", pos=4)
text(12, 161, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
data=WordingProportionG[which(WordingProportionG$Proportion=="Hi
gh"),])
res.c <- mareg(g~1, var.g,
data=WordingProportionG[which(WordingProportionG$Proportion=="Lo
w"),])
res.e <- mareg(g~1, var.g,
data=WordingProportionG[which(WordingProportionG$Proportion=="Me
dian"),])


addpoly(res.a, row=133.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=56.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.e, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()

#trim and fill, funnel plot --#dimensiion is 480*640
WP0 <- rma(g, var.g, data=WordingProportionG)
TPE <- trimfill(WP0)
TPE
funnel(TPE, xlab="Hedges' g ")

#####################################################################
 ##################################
#NegationProportion addtional analysis plots
NegationProportionG <- read.csv("NegationProportionG.csv",
sep=",", header=T)
AuthorAndYear <-
```

```
read.csv("141121UniqueIDCodingSheetDataAuthorsYear.csv",
header=T, sep=",")
NegationProportionG <- sqldf("SELECT a.AuthorAndYear,w.*  FROM
NegationProportionG w INNER JOIN AuthorAndYear a
                            ON w.UniqueID=a.UniqueID")

#create a forest plot
png(filename="forest_plot_with_Negationproportion.png",
    res=95, width=1240, height=1754, type="cairo")

par(mar=c(4,4,1,2))

WE <- mareg(NegationProportionG$g~1,
var=NegationProportionG$var.g, data = NegationProportionG)

par("usr")

forest(WE,
       xlim=c(-16,12), ilab=cbind(NegationProportionG$Construct,
round(NegationProportionG$Mt,2),
round(NegationProportionG$SDt,2),
round(NegationProportionG$Ms,2),
round(NegationProportionG$SDs,2)),
       ilab.xpos=c(-11.5, -11,-9.5,-8,-6.5), #probably need to
change
       order=order(NegationProportionG$Proportion, decreasing=F),
       alim=c(-11,11),
       ylim=c(-1,130),
       rows=c(3:103, 108:125), #5 units in between
       xlab="ES of the Negatively-Worded Item Effect by Negation
Proportion", mlab="A Summary Effect Size for All Personality
Studies",
       slab=NegationProportionG$AuthorAndYear, cex=0.7
)

op <- par(cex=0.75, font=4)

#need to check the order of block and the name of each block
match or not
text(-16, c (104, 126), pos= 4, c("Low", "High"))

par(font=2)

text(c(-11.5, -11,-9.5,-8,-6.5), 130, c ("D","Mt", "SDt", "Ms",
"SDs"))
text(-16, 130, "Authors(s) and Year", pos=4)
text(12, 130, "Hedges' g [95% CI]", pos=2)

par(op)

res.a <- mareg(g~1, var.g,
data=NegationProportionG[which(NegationProportionG$Proportion=="
```

```
 High"),])
res.c <- mareg(g~1, var.g,
 data=NegationProportionG[which(NegationProportionG$Proportion=="
 Low"),])


addpoly(res.a, row=106.5, cex=.75, mlab="ES for Subgroup")
addpoly(res.c, row=1.5, cex=.75, mlab="ES for Subgroup")

dev.off()

#trim and fill, funnel plot --#dimensiion is 480*640
WP0 <- rma(g, var.g, data=NegationProportionG)
TPE <- trimfill(WP0)
TPE
funnel(TPE, xlab="Hedges' g ")
```

BIBLIOGRAPHY

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

*Andrich, D., & Schoubroeck, L. V. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, *19*(02), 469-485.

*Apple, M. T. (2011). The Big Five personality traits and foreign language speaking confidence among Japanese EFL students. Doctoral Dissertation: Temple University.

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Prospect Heights, IL: Waveland Press.

Barnette, J. J. (1996). Responses that may indicate nonattending behaviors in three self-administered educational attitude surveys. *Research in the Schools*, *3*(2), 49-59.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361-370.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta□analysis. *Personnel Psychology*, *44*(1), 1-26.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, *22*, 231-240.

*Biderman, M. D., Nguyen, N. T., Cunningham, C. J., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality*, *45*(5), 417-429.

*Biderman, M. D., Worthy, R., & Nguyen, N. T. (2012). Manipulating method variance. Paper accepted for presentation at the 27th annual conference of *The Society for Industrial and Organizational Psychology*, San Diego, CA, April.

*Bollen, K. A., & Maydeu-Olivares, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, *72*(3), 309-326.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

*Broadfoot, A. A. (2008). Comparing the Dominance Approach to the Ideal-Point Approach in the Measurement and Predictability of Personality. Doctoral dissertation, Bowling Green State University.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Buss, M. D., & Penke, L. (2012). Evolutionary Personality Psychology. In L. Cooper & R. Larsen (Eds.): *Handbook of Personality Processes and Individual Differences*. American Psychological Association.

Campbell, D. T., Siegman, C. R., & Rees, M. B. (1967). Direction-of-wording effects in the relationships between scales. *Psychological Bulletin, 68*(5), 293.

Chamberlain, V. M., & Cummings, M. N. (1984). Development of an instructor/course evaluation instrument. *College Student Journal, 18*, 246-250.

Carmines, E. G., & Zeller. R. A, (1979). *Reliability and validity assessment*: Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-017. Beverly Hills, CA: Sage.

*Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. Psychological Assessment, 19(1), 88.

*Chiesi, F., Galli, S., Primi, C., Innocenti Borgi, P., & Bonacchi, A. (2013). The Accuracy of the Life Orientation Test–Revised (LOT–R) in Measuring Dispositional Optimism: Evidence From Item Response Theory Analyses. *Journal of personality assessment, 95*(5), 523-529.

Cochran, W. G. The combination of estimates from different experiments. *Biometrics, 10*, 101-129.

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

Cooper, M. H. (2010). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. SAGE Publications.

Costa, Jr, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*(6), 653-665.

*Corwyn, R. F. (2000). The factor structure of global self-esteem among adolescents and adults. Journal of Research in Personality, 34(4), 357-379.

Cronbach, L. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*, 3-31.

Del Re, A. C., & Hoyt, W. T. (2012). MAd: meta-analysis with mean differences. *R package version 0.8*.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual differences, 33*(4), 533-552.

*DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464.

DiStefano, C., & Motl, R. W. (2009a). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences, 46*(3), 309-313.

DiStefano, C., & Motl, R. W. (2009b). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling, 16*(1), 134-146.

*Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). Question wording effects in the assessment of global self-esteem. *European Journal of Psychological Assessment, 16*(1), 13.

Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*(2), 155.

*Eysenck, S., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences, 54*(4), 485-489.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.

*Ferrando, P. J. (2001). The measurement of neuroticism using MMQ, MPI, EPI and EPQ items: a psychometric analysis based on item response theory. *Personality and Individual Differences, 30*(4), 641-656.

*Ferrando, P. J., & Lorenzo-Seva, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and Internet-administered questionnaires. *Psychological Methods*, *10*(2), 193.

Fisher, Z., & Tipton, E. (2014). Robumeta: An R-package for robust variance estimation in meta-analysis. *R package version*.

*Fleischer, A., & Mead, A. D. (2014). Paper accepted for presentation at the 29th annual conference of *The Society for Industrial and Organizational Psychology*, Honolulu, HI, May.

Fleishman, J., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliabilities. *Educational and Psychological Measurement*, *47*, 925-939.

Gana, K., Alaphilippe, D., & Bailly, N. (2005). Factorial structure of the French version of the Rosenberg Self-Esteem Scale among the elderly. *International Journal of Testing*, *5*(2), 169-176.

*Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit, reliability, and the five-factor model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*(5), 1051-1062.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216.

Goldberg, L. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26-42.

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

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter? *Personality and Individual Differences*, *35*(6), 1241-1254.

Grothendieck, G. (2012). sqldf: Perform SQL Selects on R Data Frames. *R package version*.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10.

Harrell Jr, F. E. (2008). Hmisc: harrell miscellaneous. *R package version*, 3-5.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, *93*(2), 388.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta□regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

*Hevey, D., Pertl, M., Thomas, K., Maher, L., Craig, A., & Ni Chuinneagain, S. (2010). Consideration of future consequences scale: Confirmatory factor analysis. *Personality and Individual Differences*, *48*(5), 654-657.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta□analysis. *Statistics in Medicine*, *21*(11), 1539-1558.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967-988.

Holden, R. R., Fekken, G. C., & Jackson, D. N. (1985). Structured personality test item characteristics and validity. *Journal of Research in Personality*, *19*(4), 386-394.

Holden, R. R., & Fekken, G. C. (1990). Structured psychopathological test item characteristics and validity. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*(1), 35-40.

*Huang, J. & Mead, A. D. (2014). Effect of personality item writing on psychometric properties of ideal-point and likert scales. *Psychological Assessment*, *26*(4), 1162-1172.

Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193-206.

Jackson, D. N., & Lay, C. H. (1968). Homogeneous dimensions of personality scale content. *Multivariate Behavioral Research*, *3*(3), 321-337.

James, W. (1890). *The Principles of Psychology*. New York: Holt.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765-780.

Kaufman, P., Rasinski, K. A., Lee, R., & West, J. (1991). *National Education Longitudinal Study of 1988*. Quality of the responses of eighth-grade students in NELS88. Washington, DC: U.S. Department of Education.

Knight, R. G., Chisholm, B. J., Marsh, N. V., & Godfrey, H. P. (1988). Some normative, reliability, and factor analytic data for the revised UCLA Loneliness Scale. *Journal of Clinical Psychology*, *44*, 203-206.

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, *96*(1), 202.

Lai, J. C. (1994). Differential predictive power of the positively versus the negatively worded items of the Life Orientation Test. *Psychological Reports*, *75*, 1507-1515.

*Leibetseder, M., Laireiter, A. R., & Köller, T. (2007). Structural analysis of the E-scale. *Personality and Individual Differences*, *42*(3), 547-561.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 44–53.

Lyrakos, G. N., Damigos, D., Mavreas, V., Georgia, K., & Dimoliatis, I. D. (2010). A translation and validation study of the Life Orientation Test Revised in the Greek speaking population of nurses among three hospitals in Athens and Ioannina. *Social indicators research*, *95*(1), 129-142.

Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and organizational Psychology*, *1*(1), 3-30.

Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810–819.

*Marsh, H. W., Nagengast, B., & Morin, A. J. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194.

*Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. Psychological reports, 84(3), 927-935.

*Martín-Albo, J., Núñez, J. L., Navarro, J. G., & Grijalvo, F. (2007). The Rosenberg Self-Esteem Scale: translation and validation in university students. *The Spanish Journal of Psychology*, *10*(02), 458-467.

*Mavor, K. I., Louis, W. R., & Sibley, C. G. (2010). A bias-corrected exploratory and confirmatory factor analysis of right-wing authoritarianism: Support for a three-factor structure. *Personality and Individual Differences*, *48*(1), 28-33.

*Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 333-356.

McCrae, R. R., & Costa, Jr, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509-516.

McCrae, R. R., & John, O. P. (1992). An introduction to the five□factor model and its applications. *Journal of Personality*, *60*(2), 175-215.

McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, *59*(4), 927-953.

McPherson, J., & Mohr, P. (2005). The role of item extremity in the emergence of keying-related factors: an exploration with the life orientation test. *Psychological Methods*, *10*(1), 120-131.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437-455.

Melnick, S. A., & Gable, R. K. (1990). The use of negative item stems: A cautionary note. *Educational Research Quarterly*, *14*(3), 31-36.

Miller, T. R., & Cleary, T. A. (1993). Direction of wording effects in balanced scales. *Educational and Psychological Measurement*, *53*(1), 51-60.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*(3), 683-729.

Motl, R. W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, *9*(4), 562-578.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: the red herring. *Journal of Applied Psychology*, *81*(6), 660-679.

Owen, T. J. (1993). Accentuate the positive-and the negative: Rethinking the use of self-esteem, self-deprecation, and self-confidence. *Social Psychology Quarterly*, *56*, 288-99.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between□study heterogeneity. *Statistics in medicine*, *26*(25), 4544-4562.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, M. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903.

Pilotte, W. J., & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*, 603-610.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling*, *13*(1), 99-117.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517-529.

Roberts, J. S., Laughlin, J. E., & Wedell, D. H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, *59*(2), 211-233.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review*, 141-156.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641.

Rosenthal, R., & DiMatteo, M. R. (2002). Meta□Analysis. *Stevens' handbook of experimental psychology*.

Sánchez-Meca, J., & Marín-Martínez, F. (2010.) Meta-analysis in Psychological Research. *International Journal of Psychological Research*, *3*(1), 151-163.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. *Journal of personality and social psychology*, *89*(4), 623-642.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*(4), 367-373.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *Journal of Management*, *21*(6), 1177-1193.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, *41*(4), 1101-1114.

*Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement*, *74*(2), 214-226.

Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, *91*(1), 25.

*Stepp, S. D., Yu, L., Miller, J. D., Hallquist, M. N., Trull, T. J., & Pilkonis, P. A. (2012). Integrating competing dimensional models of personality: Linking the SNAP, TCI, and NEO using Item Response Theory. *Personality Disorders: Theory, Research, and Treatment*, *3*(2), 107.

Sutton, A. J., & Higgins, J. (2008). Recent developments in meta□analysis. *Statistics in Medicine*, *27*(5), 625-650.

Tamir, P. (1993). Positive and negative multiple choice items: How different are they? *Studies in Educational Evaluation*, *19*, 311-325.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.

Tomas, J. M., & Oliver, A. (1999). Rosenberg's self□esteem scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 84-98.

Trott, D. M., & Jackson, D. N. (1967). An experimental analysis of acquiescence. *Journal of Experimental Research in Personality*, *2*(4), 278-288..

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247.

Viechtbauer, W. (2010). Metafor: meta-analysis package for R. *R package version*, 1-0.

Wang, J., Siegal, H. A., Falck, R. S., & Carlson, R. G. (2001). Factorial structure of Rosenberg's Self-Esteem Scale among crack-cocaine drug users. *Structural Equation Modeling*, *8*(2), 275-286.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186-191.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-200.