Personality and the prediction of work performance : artificial neural networks versus linear regression

Author:
Minbashian, Amirali

Publication Date:
2006

DOI:
https://doi.org/10.26190/unsworks/23615

Personality and the Prediction of Work Performance:

Artificial Neural Networks Versus Linear Regression

Amirali Minbashian

A thesis submitted for the degree of Doctor of Philosophy

School of Psychology, University of New South Wales

February, 2006

Acknowledgments

I am grateful to my supervisor Dr. Jim Bright for encouraging me to pursue my interest, and for providing me with the support to do so. I have appreciated his input. I would also like to thank my co-supervisor Dr. Kevin Bird whose door has always been open to me, and who has taught me much of what I know about data analysis and research methodology.

A number of the datasets employed in this thesis were provided by other institutions and researchers. In this regard I would like to thank the Center for Creative Leadership, the New Zealand Police, Jonathan Black, Douglas Cellar, Jeffrey Conte, Maxine Dalton, and Kelly Hannum.

Special thanks to my fellow student Jo Earl for the many ways in which she contributed to this research, and for all the hours she spent patiently listening to me.

Finally, I am enormously indebted to my family for everything they have done for me over the years. This thesis would not have been completed if it was not for their support.

Abstract

Previous research that has evaluated the effectiveness of personality variables for predicting work performance has predominantly relied on methods designed to detect simple relationships. The research reported in this thesis employed artificial neural networks – a method that is capable of capturing complex nonlinear and configural relationships among variables – and the findings were compared to those obtained by the more traditional method of linear regression.

Six datasets that comprise a range of occupations, personality inventories, and work performance measures were used as the basis of the analyses. A series of studies were conducted to compare the predictive performance of prediction equations that a) were developed using either artificial neural networks or linear regression, and b) differed with respect to the type and number of personality variables that were used as predictors of work performance. Studies 1 and 2 compared the two methods using individual personality variables that assess the broad constructs of the five-factor model of personality. Studies 3 and 4 used combinations of these broad variables as the predictors. Study 5 employed narrow personality variables that assess specific facets of the broad constructs. Additional methodological contributions include the use of a resampling procedure, the use of multiple measures of predictive performance, and the comparison of two procedures for developing neural networks.

Across the studies, it was generally found that the neural networks were rarely able to outperform the simpler linear regression equations, and this was attributed to the lack of reliable nonlinearity and configurality in personality-work performance relationships. However, the neural networks were able to outperform linear regression in the few instances where there was some independent evidence of nonlinear or configural relationships. Consequently, although the findings do not support the usefulness of neural networks for specifically improving the effectiveness of personality variables as predictors of work performance, in a broader sense they provide some grounds for optimism for organisational researchers interested in applying this method to investigate and exploit complex relationships among variables.

Table of Contents

CHAPTER 1: Personality and Work Performance

Introduction

The notion that success at work may be related to an individual's personality has long been intuitively appealing. Over 2300 years ago Plato wrote that people should be assigned to jobs for which they are naturally suited (Cohen, Swerdlik, & Phillips, 1996), and there is ample evidence that the ancient Chinese recognised the importance of personal characteristics in determining the fitness of officials for their jobs (Higgins & Sun, 2002). However, it was not until the early 20th Century that standardised personality inventories were developed and systematically used to make decisions about incumbent and prospective personnel. The first personality inventory, the Woodworth Personal Data Sheet, was developed during World War I with the aim of screening out potentially problematic army recruits (Anastasi, 1988), and since that time there has been a rapid proliferation in the number and use of these instruments in organisations.

Today, personality inventories are one of the most popular methods of employee selection. Survey evidence suggests that they are the most frequently used selection tests for managerial jobs in Canada, France, Greece, Spain, and Sweden, and only come second to medical screening in Australia, Ireland, Italy, the Netherlands, Portugal, Singapore, South Africa, and the United Kingdom (Ryan, McFarland, Baron, & Page, 1999). Their use has also been documented in a number of other countries including Belgium, Croatia, Germany, Hong Kong, Israel, New Zealand, Norway, Portugal, and the United States (Levy-Leboyer, 1994; Ryan et al., 1999; Smith & Abrahamsen, 1992; Taylor, Keelty, & McDonnell, 2002). A recent survey of large Australian businesses found that personality inventories were used to make selection and promotion decisions by 71 of the 85 responding firms (Hartstone & Kirby, 1998). Surveys conducted in the

United Kingdom suggest that they are used at least some of the time by the majority of companies hiring managers (Shackleton & Newell, 1991; Williams, 1994) and graduates (Hodgkinson & Payne, 1998). In the United States the estimates are lower, closer to 10% to 20% usage (American Management Association, 2001; Dipboye, 1992), although even this would represent a substantial number of employees who come into contact with these instruments.

Given the widespread use of personality inventories in organisations, an important issue concerns the effectiveness of personality variables for predicting work performance. A great deal of research has investigated this issue, yet the majority of previous work has relied on methods designed to detect simple (linear and additive) relationships among variables, and there have been few attempts to employ methods capable of capturing more complex relationships. The primary aim of the research reported in this thesis was to apply one such method, *artificial neural networks*, to evaluate the effectiveness of personality variables for predicting work performance, and to compare the findings with those obtained by the more traditional method of linear regression.

This chapter reviews the literature concerning personality and work performance. First, the research designs and methods typically used to evaluate the criterion-related validity of personality variables are outlined, [1] and the importance of a taxonomy of personality constructs within this framework is highlighted. Second, the five-factor model of personality is presented as one such taxonomy, and evidence for its usefulness as a classification scheme is provided. Third, the validity of measures of each of the five factors for predicting work performance is reviewed. Finally, arguments are

[1] The term *criterion-related validity* refers to the effectiveness of a predictor variable in predicting an individual’s standing on some criterion of interest (Anastasi, 1988). In this thesis the term is used specifically to refer to the prediction of work performance.

made for more complex relationships between personality variables and work performance, that if detected would increase the criterion-related validity of personality variables.

A Conceptual Framework

In the approximately 90 years since the first standardised personality inventory was developed there have been hundreds of studies that have examined the validity of personality variables for predicting work performance. Typically, the method of investigation has involved administering a self-report personality inventory that assesses one or more personality variables to a sample of employees, and collecting work performance information for the employees either concurrently or, less often, after a certain period of time has elapsed (Guion, 1991). Validity coefficients are then calculated by correlating the scores on the personality variables (the *predictors*) with the performance scores (the *criterion*) in order to evaluate the extent to which the personality variables predict performance. Primarily, researchers have used supervisor ratings of performance as the criterion, although other popular measures have included objective indices such as production rates, sales volume, or performance on tests (Borman, 1991). Furthermore, researchers have made a distinction between training performance (the amount learned on training and development programs) and performance on-the-job. While there has been interest in investigating the correlates of both types of performance, it has also been recognised that personality variables may be differentially related to the two criteria (e.g., Barrick & Mount, 1991).[2]

A number of intermittent reviews have summarised the findings from the large volume of individual studies that have been conducted (e.g., Ghiselli, 1973; Guion &

[2] In this thesis the term work performance is used to refer to both on-the-job performance and training performance.

Gottier, 1965; Schmitt, Gooding, Noe, & Kirsch, 1984). Early reviews focused on the criterion-related validity of personality variables in general rather than the validity associated with specific personality constructs, and in many cases this resulted in pessimistic conclusions about the validity of personality variables. For example, Ghiselli and Barthol (1953) found that in certain circumstances personality measures were valid predictors of performance, although they suggested caution in the use of personality inventories after noting the number of low and negative results. Similarly, Guion and Gottier (1965) concluded that while there was variation in the validity coefficients obtained in different studies, there was no evidence that personality measures in general were useful tools for personnel selection. The fact that results were summarised across personality constructs, rather than in terms of specific constructs, may have contributed to the pessimistic conclusions of these reviews. This is because scales assessing constructs that would not be expected to predict performance on rational grounds may have been included in the summary process. Such a procedure was justified at the time as there was no well-accepted framework for grouping personality scales. Nevertheless, the results from these early reviews highlight the importance of a classification scheme for organising the myriad of personality variables from different inventories.

As a first step toward a taxonomy of personality constructs, Eysenck (1947) introduced the idea that personality variables can be organised hierarchically, according to the breadth of behaviours represented. In particular, he distinguished between personality variables that assess consistencies in behaviour at three different levels of breadth, namely *habitual responses*, *personality traits*, and *personality types*.[3] Habitual responses refer to specific behaviours that tend to recur under similar circumstances,

[3] Eysenck (1947) used the term *type* to refer to the highest level of the hierarchy, however today the term *higher-level factor* is more commonly used (e.g., Goldberg, 1999).

and thus they represent a relatively narrow range of behaviours. Within personality inventories, habitual responses are often assessed by individual items (Digman, 1990). For example the item “(I) pay my bills on time.” (from the IPIP-NEO inventory; International Personality Item Pool, 2001) represents one such behavioural tendency. At the next level, habitual responses that are related can be grouped to define personality traits that by definition represent a broader set of behaviours. For example, habitual responses such as paying bills promptly, keeping promises, and attending work regularly are all part of the broader trait of dutifulness. At the highest level of breadth, related traits can be grouped to define personality factors that describe broad mental, emotional, motivational, and interpersonal constructs. Within personality inventories, the latter two levels of the hierarchy are typically operationalised in terms of scales composed of multiple items.

Methodologically, the technique of factor analysis has been be used to investigate the number and nature of constructs at the highest levels of the hierarchy. For example, items on personality inventories have frequently been factor analysed to yield a number of first-order factors corresponding to various traits. The first-order factors (or the scales that serve to operationalise them) have in turn been factor analysed to yield second-order factors corresponding to broad personality dimensions. There has been little agreement among researchers about the number or nature of factors at the first-order level (Barrick, Mount & Judge, 2001). Indeed, the failure of first-order factors to emerge with any degree of clarity has prompted doubts about their ability to form an adequate basis for the description of personality (e.g., Kline & Barrett, 1983). In contrast, there has been growing consensus regarding the number of dimensions at the second-order factor level, with much evidence pointing to the existence of five dimensions collectively labelled as the five-factor model of personality. The next section provides a description of the five factors, and evaluates the five-factor model as a taxonomy for classifying personality variables.

The Five-Factor Model

Description of the Five Factors

According to the five-factor model, personality can be adequately described in terms of five major dimensions or factors. Although there has been some disagreement about the specific traits associated with each factor (see Digman, 1990), there is enough consensus to provide broad characterisations of the factors.

One factor has been referred to as *Neuroticism*. Alternatively, it has been conceptualised in terms of its opposite pole and called *Emotional Stability*. Neuroticism is commonly identified with terms representing fear and vulnerability such as anxious, nervous, insecure, tense, easily upset, intolerant of stress, and unstable (e.g., Mount & Barrick, 1995). Conversely, terms such as calm, relaxed, secure, and unflinching load negatively on this factor (Goldberg, 1990). Other traits that have been found to load on Neuroticism include those associated with depression, anger, shame and embarrassment (e.g., Costa & McCrae, 1992). Taken together, this suggests that Neuroticism is broadly indicative of the tendency to experience negative emotions.

A second factor, labelled *Extraversion* or *Surgency*, is characterised by traits that indicate the extent and forcefulness with which an individual interacts with their external environment. This is reflected specifically in the way the individual interacts with others (high scorers tend to be more sociable, talkative, assertive and domineering than low scorers, Goldberg, 1992), and more generally in the way they live their life (high scorers are more energetic, active, daring, adventurous, and enthusiastic than low scorers, Goldberg, 1992). However, as many of the traits that define Extraversion also describe positive emotions, in particular a highly aroused type of positive affect, this factor has also been equated with positive emotionality (e.g., Watson & Tellegen, 1985).

A third factor, *Openness to Experience* (hereafter referred to as Openness), has been the most difficult to define as reflected by a number of alternative labels such as *Intellectance, Culture,* and *Intellect*. John and Srivastava (1999) suggest that it describes the “breadth, depth, originality, and complexity of an individual’s mental and experiential life” (p. 121). Traits such as imagination, creativity, and artistic sensitivity are core aspects of this factor (Johnson, 1994; Saucier, 1994). High scorers also display greater intellectual curiosity and a willingness to consider new ideas, although it has been argued that Openness is distinct from cognitive ability (Costa & McCrae, 1992). More generally Openness manifests in terms of open-mindedness to novelty and unconventionality in many different aspects of life including one’s behavioural experiences, inner feelings, and values (Costa & McCrae, 1992).

A fourth factor, commonly labelled *Agreeableness*, is interpersonal in nature in that it describes the extent to which an individual has a prosocial and communal orientation toward others (John & Srivastava, 1999). Thus, traits that are associated with sympathy and altruism, such as kindness, compassion, generosity, helpfulness, and soft-heartedness load on this factor (Hofstee, De Raad, & Goldberg, 1992). Furthermore, in their interactions with others high scorers tend to be cooperative, trusting, and courteous whereas low scorers are more likely to be uncooperative, sceptical, and impolite. Agreeableness has also been linked to morality and modesty – high scorers are less likely to be described as cruel, dishonest and manipulative, and more likely to be described as humble (Goldberg, 1990).

The essence of the fifth factor, labelled *Conscientiousness*, is self-control (Costa & McCrae, 1992). This factor is directly relevant to the way an individual approaches and completes tasks, and can manifest itself in a number of ways. High scorers approach tasks in a more planful and organised manner, and are more likely to be described as orderly, neat, systematic, meticulous, and efficient (Hofstee et al., 1992). In their interactions with others the highly conscientious show greater levels of dependability, responsibility, and reliability (Goldberg, 1992). Conscientiousness has also been associated with self-discipline and traits representative of the will to achieve, such as being persevering and hardworking (Mount & Barrick, 1995). Furthermore, high scorers are more likely to follow rules and to think before acting (John & Srivastava, 1999).

Evidence for the Five-Factor Model

The five-factor model has its origin in factor-analytic studies of ratings of trait adjectives that were compiled by Allport and Odbert (1936), and subsequently modified and reduced to 35 by Cattell (1947). In an early attempt to study the principal dimensions of personality, Fiske (1949) used 22 of Cattell's trait terms to compare the factorial structures of trait ratings from different sources. The participants were 128 trainees participating in a week-long intensive assessment program. At the end of the program each participant was rated on the trait terms by themselves, by their peers, and by the staff who were supervising them. Separate factor analyses of the self-, peer-, and staff-ratings provided evidence for five factors that recurred across the different sources.

Several other researchers were able to demonstrate the robustness of a five-factor solution. Tupes and Christal (1961/1992) factor analysed ratings on Cattell's 35 trait terms across eight datasets comprising different samples, situations, raters, and lengths and kinds of acquaintanceships. They labelled the resulting five factors Surgency, Agreeableness, Dependability, Emotional Stability, and Culture, and concluded that “(t)here can be no doubt that the five factors found throughout all eight analyses are recurrent” (p.244). Subsequent studies by Norman (1963), Borgatta (1964), and Smith (1967) provided support for five factors that closely resembled those obtained by Tupes and Christal. Moreover, Digman and Takemoto-Chock (1981) reanalysed correlations from six earlier studies, some of which had found solutions with greater than five factors. After correcting for clerical errors that had occurred in some of the studies, and using a common method of factor analysis, they too concluded that the observed relationships were well accounted for in terms of the five-factor structure.

Much of the early work used the set (or a subset) of the 35 trait terms initially selected by Cattell (1947), and therefore did not preclude the possibility that the findings simply indicated the structure in Cattell’s limited set of variables, rather than representing truly substantive findings (see Block, 1995). Goldberg (1981, 1990, 1992) sought to establish the generality of the five-factor structure by analysing more comprehensive sets of trait terms. He demonstrated that the five-factor structure emerged across different sets of trait terms, different methods of factor analysis, and for both peer- and self-ratings. The term *Big Five* was coined to refer to these five factors. Factors corresponding to the Big Five have also been found in studies of trait terms from other languages including German (Hofstee, Kiers, De Raad, Goldberg, & Ostendorf, 1997), Russian (Shmelyov & Pokhil’ko, 1993), Turkish (Goldberg & Somer, 2000), and Filipino (Church, Katigbak, & Reyes, 1996). The consistency of the results across different studies can be taken as evidence for the fundamental nature of the five factors underlying personality trait term ratings.

While studies of trait term ratings provided the initial impetus for the five-factor model, later studies confirmed the usefulness of the model as a framework for organising scales from existing personality inventories. A significant contribution in this regard was the work of Costa and McCrae (1985) who, encouraged by the replicability of the Big Five in analyses of trait ratings, developed a personality questionnaire explicitly designed to measure the five factors – the NEO Personality Inventory (NEO PI). Through a series of studies conducted with other colleagues they were able to demonstrate that the majority of scales from existing personality inventories could be associated with at least one of the five factors. For example, the Myers-Briggs Type Indicator (MBTI; Myers & McCaulley, 1985), a personality inventory that is currently used in many organisations, was developed with the aim of classifying individuals into one of 16 personality categories based on their scores on four variables. McCrae and Costa (1989) found that each of these variables was strongly associated with one of the five NEO PI scales (only Neuroticism was not represented in the MBTI). Similarly, using a combination of rational and empirical analyses, McCrae, Costa, and Piedmont (1993) showed that all but one of the 20 scales of the California Psychological Inventory (CPI; Gough, 1987) could be meaningfully related to one or more of the five factors. Another widely used instrument, Jackson's (1984) Personality Research Form (PRF), contains 20 scales that assess various forms of psychological needs. Factor-analytic studies of the 20 PRF scales have found five factors that replicate well across different samples and cultures (e.g., Paunonen, Jackson, Trzebinski, & Fosterling, 1992; Stumpf, 1993). Costa and McCrae's (1988) joint factor analysis of the NEO PI scales and the PRF scales provided evidence that the five PRF factors correspond to the five dimensions of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness.

One of the most extensively researched personality systems is that of Cattell who proposed that personality be described in terms of 16 first-order factors, and who developed the popular 16 PF inventory (Cattell, Eber, & Tatsuoka, 1970) as a means of operationalising these factors. While factor analysis of the 16 PF items has sometimes failed to support the proposed first-order factor structure (e.g., Kline & Barrett, 1983), a large scale study of the second-order factor structure of the 16 PF found that the major portion of variance in the first-order scales was captured by five dimensions (Krug & John, 1986). Three of these factors can be clearly identified with Neuroticism, Extraversion, and Conscientiousness, while the other two show some resemblance to the Agreeableness and Openness factors.

The five-factor model has also been shown to be useful for classifying variables from personality instruments that were designed within clinical contexts. The Minnesota Multiphasic Personality Inventory (MMPI) contains 550 items that were initially developed by referring to psychiatric reports, textbooks, and previously published personality items, and later retained based on their ability to differentiate between psychiatric and non-psychiatric subjects (Cohen et al., 1996). Although factor analysis of the MMPI items has yielded more than five factors (e.g., Costa, Zonderman, McCrae, & Williams, 1985; Johnson, Null, Butcher, & Johnson, 1984), rational judgments and correlations with the NEO PI variables suggest that most of the MMPI factors reflect aspects of Neuroticism, Extraversion, Openness, or Agreeableness (Costa, Busch, Zonderman, & McCrae, 1986). Similarly, the 100 items of the California Q-Set (CQS; Block, 1961) were developed based on the judgments of panels of psychodynamically-oriented clinicians. McCrae, Costa, and Busch's (1986) factor analysis of these items suggested six replicable factors. Five of the factors could be identified with the dimensions of the five-factor model based on their correlations with the NEO PI variables and the items loading on each factor. A sixth factor labelled as Attractive-Narcissistic represented physical attractiveness.

At the second-order factor level, the main competitor to the five-factor model is Eysenck’s model comprising the three factors Neuroticism, Extraversion, and Psychoticism (Eysenck & Eysenck, 1985). The first two factors of Eysenck’s model are consistent with the first two factors of the five-factor model. Furthermore, it has been proposed that Eysenck’s Psychoticism factor represents a blend of the Agreeableness and Conscientiousness factors (e.g., Digman, 1990). Support for these assertions has come from a study conducted by McCrae and Costa (1985) in which the NEO PI measures were jointly factor analysed with personality instruments developed by Eysenck to operationalise his theory. The results suggested the presence of five factors that could be clearly identified with the dimensions of the five-factor model. Eysenck’s measures of Neuroticism and Extraversion loaded on the same factors as the NEO PI measures of Neuroticism and Extraversion, and the Psychoticism measure loaded negatively on the factors corresponding to Agreeableness and Conscientiousness. Finally, none of the Eysenck measures loaded on the Openness factor. The absence of the Openness factor from Eysenck’s system may be attributed to the fact that he considered intellect-related variables as being separate from those aspects of personality that are non-intellectual (Digman, 1990). Thus, the main difference between the two systems concerns whether Agreeableness and Conscientiousness should be viewed as distinct second-order factors, or whether they are more appropriately categorised as constituents of the second-order factor Psychoticism when conceptualised in terms of its opposite pole.

In summary, the majority of variables measured by existing personality inventories can be meaningfully organised within the five-factor taxonomy. An implication of this finding has been that many of the reviews conducted since 1990 have used the five-factor model as a framework for developing and testing hypotheses about the validity of personality variables for predicting work performance. The next section discusses the findings from these reviews.

The Five Personality Factors and Work Performance

In one of the initial applications of the five-factor model to the prediction of work performance, Barrick and Mount (1991) proposed a set of hypotheses regarding the expected relationships between measures of the five factors and work performance. It was hypothesised that measures of Conscientiousness and Neuroticism would be valid predictors of performance across occupations because traits associated with Conscientiousness (such as being responsible, hard-working, and persistent) would facilitate performance in all jobs, whereas traits associated with Neuroticism (such as being nervous, temperamental, and high-strung) would hamper performance. Openness was hypothesised to predict performance on training programs because this factor is associated with characteristics that indicate a positive attitude to learning, such as curiosity and broad-mindedness. Extraversion and Agreeableness were hypothesised to be important for dealing with others, and therefore measures of these two factors were expected to predict performance for occupations that require interpersonal interaction.

A number of meta-analyses conducted since the early 1990's used the five-factor model as a basis for investigating relationships between personality variables and work performance (see Barrick & Mount, 2003), and therefore provided evidence relevant to these hypotheses. Two main conclusions could be drawn from the findings of these reviews. First, it emerged that there was some empirical support for the various hypothesised personality-performance linkages, although the findings were not always consistent across meta-analyses. It was generally found that Conscientiousness was a valid predictor of performance across occupational groups and types of performance criteria (e.g., Barrick & Mount, 1991; Mount & Barrick, 1995; Salgado, 1997). Neuroticism was found to predict performance in some reviews (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Salgado, 1997; Tett, Jackson, & Rothstein, 1991), although in others the correlation coefficients could not be distinguished from zero or were in the opposite direction to that hypothesised (e.g. Barrick & Mount, 1991; Vinchur, Schippmann, Switzer, & Roth, 1998). Openness predicted training performance, but not performance on the job (e.g., Barrick & Mount, 1991; Hough et al., 1990; Salgado, 1997). There was also some support for the hypotheses linking Extraversion and Agreeableness to jobs involving interpersonal interaction. Extraversion was a valid predictor of performance for sales and managerial occupations (Barrick & Mount, 1991; Vinchur et al., 1998). Additionally, Extraversion was found to predict performance on training programs that were highly interactive in nature, such as police academy, sales, and flight attendant training programs (Barrick & Mount, 1991). Agreeableness was related to performance for jobs involving teamwork or the provision of customer service (Hough, 1992; Mount, Barrick, & Stewart, 1998).

A second finding was that the magnitude of the validity coefficients rarely exceeded .30.[4] This was initially attributed (at least partly) to the post hoc classification of scales into one of the five-factor categories: the possibility of misclassification, and the fact that the individual scales typically only represent facets of the five constructs rather than the constructs themselves, could serve to attenuate the validity estimates. Hurtz and Donovan (2000) addressed this issue by only including studies that had used

[4] Unless otherwise stated, the validity coefficients reported in this section are those that were corrected for statistical artefacts (such as sampling error, range restriction, and measurement error) that would otherwise underestimate the true validity of the personality variables.

a personality inventory explicitly designed to measure the five-factor constructs. Validity coefficients were summarised across four occupational categories (sales, customer service, management, and skilled/semi-skilled workers). The estimated true validity coefficients for Conscientiousness and Neuroticism (which was conceptualised in terms of its opposing pole, Emotional Stability) were .20 and .13 averaged across occupations, and did not exceed .26 for any given occupation. The validity of Openness for predicting training performance was estimated to be .13. Extraversion exhibited validity coefficients of .15 and .12 for sales and managerial work, and the validity of Agreeableness in customer service jobs was estimated to be .17. Based on these results Hurtz and Donovan (2000) concluded that the benefits of the five factor measures for personnel selection purposes are likely to be small.

Similar validity coefficients were obtained by Barrick et al. (2001), who conducted a meta-analysis of previous meta-analyses in order to provide a summary of what had been learnt from the previous reviews. Results were summarised across five occupational groups (sales, management, professionals, police, and skilled/semi-skilled workers). Conscientiousness was the best predictor of performance across different occupations. Its corrected correlation coefficient ranged from .23 to .26 for different occupations. The corrected correlation between measures of Emotional Stability and performance across occupations was .13. When each occupational group was considered separately, Emotional Stability was related to performance for police (corrected $r = .12$) and skilled/semi-skilled workers (corrected $r = .15$), but not for any other occupation. Openness predicted training performance (corrected $r = .33$), and also had a low (though non-zero) correlation with the job performance of professionals (corrected $r = .11$). Extraversion was related to performance in some jobs involving interpersonal interaction, such as management (corrected $r = .21$) and police work (corrected $r = .12$).

Additionally, Extraversion was positively associated with performance on training programs (corrected r = .28). Agreeableness was not related to performance in any occupational group, although it predicted the specific criterion of teamwork (corrected r = .34).

The magnitude of the validity coefficients for personality variables can be evaluated with respect to a number of different benchmarks. For example, based on his observations of effect sizes typically obtained in social science research, Cohen (1977) proposed several benchmarks that could be used as conventions in the absence of prior research in the area. Specifically, Cohen suggested that an *uncorrected* correlation coefficient of approximately .50 to .60 represents an upper bound on what could be expected in the social sciences, an uncorrected correlation of .30 is moderate in size, and that an uncorrected correlation of .10 borders on the trivially small. The uncorrected (i.e., observed) correlations obtained in previous meta-analyses have not been reported here, although they rarely exceed .20 even in the most optimistic cases, and are usually less than .10 (for examples see Barrick et al., 2001; Hurtz & Donovan, 2000). Thus using Cohen’s benchmarks, one could conclude that the criterion-related validity of measures of the five factors are at best low to moderate.

A more appropriate frame of reference for interpreting the magnitude of personality-performance correlation coefficients may be the findings from the literature concerning alternate predictors of work performance. Schmidt and Hunter (1998) summarised the results of meta-analyses concerning 19 different selection methods. The validity coefficients, corrected for range restriction and measurement error in the criterion, ranged from in excess of .50 for the most valid methods to negligible values for predictors such as age and graphology. The most accurate predictors were work sample tests (r = .54), general mental ability tests (r = .51), and structured interviews (r

= .51). If one were to take the value of .30 as an estimate of the corrected validity coefficient associated with measures of the most valid five-factor dimension, Conscientiousness, this predictor would clearly have lower predictive validity than 11 of the predictors examined by Schmidt and Hunter (1998). Thus, relative to the best performing selection methods, the magnitude of the validity coefficients associated with measures of Conscientiousness and the other five-factor dimensions tend to be low.

In summary, therefore, the validity coefficients associated with measures of the five factors are low when compared to the benchmarks proposed by Cohen (1977), and relative to what has been achieved by some other predictors of work performance. As recently noted by Murphy and Dzieweczynski (2005), although the application of the five-factor model has helped to resolve some of the inconsistencies in earlier reviews, the pessimistic conclusions about the validity of personality variables compared to other selection methods are still relevant today. This is perhaps surprising given the proposal that the individual differences captured by the five factors evolved as a result of their importance for forecasting the utility of an individual to the economy of their group (Hogan, 1996), and for selecting allies who are likely to facilitate the achievement of adaptively significant goals (Buss, 1996). The low validity coefficients may reflect limits on the ability of personality variables to predict performance. That is, the information provided by the five dimensions (or more precisely by the scores on personality scales that measure these dimensions) may be inherently less useful for predicting work performance than the information provided by other predictors. An alternative possibility, however, is that the low validity is not due to limits in the informativeness of personality scores, but rather due to the methods used to examine the ability of personality variables to predict performance. The majority of studies have used methods designed to detect simple relationships. For example, the commonly used procedure of calculating correlations assesses the degree of linear association between one personality variable and performance. In other words, it is assumed that the personality variable is uniformly related to performance over the entire range of scores on the variable. Similarly, regression analysis has typically been employed to study linear and additive relationships between personality variables and work performance. However, the relationship between personality variables and performance may be nonlinear or configural in nature, in which case the low criterion-related validity of personality scales may be the result of examining complex relationships with simple methods. Such complex relationships are discussed in the next section.

Complexity in Personality-Performance Relationships

Nonlinear Relationships

Nonlinear relationships occur when a predictor is not uniformly related to the criterion across the predictor's entire range. There are many ways in which two variables can be nonlinearly related, and not all of these functions are likely descriptors of the relationship between personality and work performance. Existing theories of work performance that incorporate personality characteristics as explanatory variables provide little guidance as to the types of nonlinearity (if any) to expect. Within these theories, personality variables are conceptualised as indices of basic underlying dispositions that influence work performance indirectly through their effects on various mediating variables such as goals and goal-setting (Barrick, Mitchell, & Stewart, 2003), or work-related knowledge, skills, and habits (Motowidlo, Borman, & Schmit, 1997). The nature of the functional relationships between variables is not specified. Indeed, the methods used to test the theories suggest the implicit assumption of linearity.

Nevertheless, a number of researchers have highlighted the importance of investigating nonlinear relationships (e.g., Guion, 1991), and in some instances conceptually plausible ways in which personality may be nonlinearly related to performance have been outlined (e.g., Murphy, 1996; Robie & Ryan, 1999). One can consider cases where the *direction* of the relationship between a personality variable and performance changes at different values of the personality variable. For example, Murphy (1996) proposed that too much or too little of a given personality characteristic may have a negative effect on performance. This can be described by an inverted-U function where an increase in the personality variable is associated with an increase in work performance up to a certain optimal point after which a decrease in performance occurs. Murphy argued that Conscientiousness and Extraversion may be related to job performance in this way. The extremely conscientious individual may be so conventional and rule-bound that their performance is impaired relative to their more moderately conscientious colleagues; and while a certain level of Extraversion is likely to be useful in jobs requiring interpersonal interaction, the extremely extraverted individual may spend all their time interacting with others at the expense of completing their own tasks. Similarly, Mount et al. (1998) hypothesised that although Agreeableness is likely to be a desirable characteristic in service-oriented jobs, too much of this characteristic may hamper performance: The highly agreeable employee may be too cooperative with customers to take actions that are in the company’s best interests. The relationship between work performance and the two other factors – Neuroticism and Openness – may also show an inverted-U pattern. Neuroticism has been posited to adversely effect performance by reducing the motivation to exert and maintain effort on work tasks (Barrick et al., 2003). However, to the extent that negative affect can be a motivator, a certain amount of fear of failure may also be required for optimal performance. Therefore, although the relationship between Neuroticism and performance is expected to be negative across most of the predictor's range, at low levels of Neuroticism the relationship may be positive. Similarly, Openness is likely to be a desirable attribute for performance on training programs, however the high levels of fantasy and intellectual curiosity that are characteristic of the upper extremes of this factor may distract the trainee away from the relevant material in the training program.

Other possible forms of nonlinearity may involve changes in the *strength* of the relationship between a personality variable and performance at different values of the personality variable. For example, performance may increase by large amounts as a function of increases in a personality variable up to given point after which the relationship ceases to exist or becomes weaker. Murphy (1996) argued that this is likely to occur in situations where only a certain level of an attribute or skill is required to facilitate performance. He cited the example of interpersonal sensitivity – an individual with no interpersonal sensitivity is likely to function poorly in social situations, although the difference in performance between normal sensitivity and being finely attuned to others may be small. A similar type of reasoning could be extended to the relationship between other types of skills (social and technical) and performance. Given that personality variables have been hypothesised to partly influence performance via their effects on skills – for example Extraversion and Agreeableness have been associated with the acquisition of social skills (Motowidlo et al., 1997), and Openness with the ability to solve technical problems (Hogan, 1996) – it may well be the case that personality variables are also nonlinearly related to performance in this way. It is also interesting to note that the way in which many practitioners use personality (and other psychological) tests for selection purposes implicitly assumes this type of nonlinear relationship between the predictor and work performance. Specifically, it is often the case that cut-off scores are set on the predictor and those candidates who score above the cut-off are eligible for selection whereas those scoring below the cut-off are rejected. This procedure is adopted in favour of one where candidates are selected sequentially starting with the highest scorers (Guion, 1991). The assumption that is being made is that the cut-off score represents the minimum level of the predictor attribute that is desirable, and that beyond this minimum level increases in the predictor attribute confer little benefit.

Alternatively, it has been suggested that personality may be more strongly related to work performance at the extremes than in the mid-range (see Sinclair, Banas, & Radwinsky, 1999). According to the concept of *traitedness*, the relevance of any given personality trait to an individual’s behaviour is greater for some individuals than others (Siem, 1998). Hence, individuals who possess high traitedness for a particular trait are more consistent in their behaviour with respect to that trait and therefore more predictable than low traited individuals (Siem, 1998). Furthermore, there is some evidence that traitedness is related to trait level such that individuals who score either at the high or low end of a particular personality trait scale display higher traitedness than individuals who score in the middle of the scale (Paunonen, 1988; Siem, 1998). Taken together, the above findings suggest that personality traits are not uniformly related to work criteria over their entire ranges and therefore the relationship between personality and work criteria is nonlinear. Specifically, the relationship between a given personality trait and a criterion will be stronger (and therefore exhibit a steeper slope) for low and high levels of the trait than for intermediate levels (where the slope will be flatter).

Despite the rationale for nonlinearity, there have only been a few attempts to empirically investigate nonlinear relationships between personality variables and work performance. Day and Silverman (1989) hypothesised that extreme scores on two personality scales, labelled *orientation towards direction from others* and *impulse expression*, were likely to signal poor performance in an accounting job. The former scale assessed the tendency to be dependent on the direction of others versus the tendency to be non-conforming and rebellious. The latter scale assessed the tendency to act without deliberation and avoid routine at one extreme, versus the tendency to be fearful and apprehensive, neat and systematic, and rigid and exacting at the other extreme. On description, therefore, impulse expression seems to capture elements of the Conscientiousness and Neuroticism factors. In partial support of their hypotheses, Day and Silverman found an inverted-U relationship between impulse expression and supervisor ratings of performance in a sample of 40 accountants.

Sinclair and his colleagues (Sinclair et al., 1999; Sinclair & Lyne, 1997) presented two conference papers in which they used polynomial regression to investigate quadratic and cubic relationships between the seven scales of the Hogan Personality Inventory (HPI; Hogan & Hogan, 1995) and supervisor ratings on various measures of performance.[5] In total they analysed samples from six occupations: (a) inspectors/operators in a manufacturing consortium, (b) adjustors in a manufacturing consortium, (c) retail service clerks, (d) bank employees, (e) marketers, and (f) customer service representatives. A number of significant nonlinear relationships were detected, although these were scattered across the personality measures. A limitation of this research was that the sample sizes were small (typically less than 100), a factor that would decrease the likelihood of detecting nonlinear effects.

Subsequent work on nonlinearity between personality and work performance has predominantly focused on the Conscientiousness factor. Robie and Ryan (1999)

---

[5] A quadratic relationship exists when the relationship between two variables can be described as a curve with one bend (for example, an inverted-U function). A cubic relationship exists when the relationship between two variables can be described as a curve with two bends.

examined five diverse samples that were substantially larger than those used by Sinclair et al. (1999). They failed to find any evidence for either quadratic or cubic relationships between measures of this construct and supervisor ratings of overall job performance across the five samples. In contrast, La Huis, Martin, and Davis (2005) obtained a significant quadratic relationship between Conscientiousness and supervisor ratings of job performance in two separate clerical samples; and Cucina and Vasilopoulos (2005) found that Conscientiousness was quadratically related to the academic performance of 262 undergraduate university students. Consequently, although there is some evidence in favour of nonlinear personality-performance relationships, the results are far from clear-cut.

Configural Relationships

Configural relationships occur when the strength and/or direction of the relationship between a predictor and criterion depends on an individual’s standing on some other variable (Ganzach, 1997). Within organisational psychology such relationships have been commonly investigated as moderator (or interaction) effects, in which the analysis proceeds by examining whether the predictor and the moderator have a significant multiplicative effect on the criterion after controlling for the additive effects of the two variables (see Baron & Kenny, 1986).

Most of the work on moderators of the relationship between personality variables and performance has focused on situational moderators such as the cooperative and competitive demands placed on employees by the job (Barrick et al., 2003), the level of autonomy associated with the job (Barrick & Mount, 1993), reward structures (Stewart, 1996), and organisational politics (Hochwater, Witt, & Kacmar, 2000). There has also been some focus on the interactive effect of cognitive ability and personality on performance (e.g., Mount, Barrick & Strauss, 1999; Sackett, Gruys, & Ellington, 1998; Wright, Kacmar, McMahan, & Deleeuw; 1995). Far less attention has been devoted to examining the extent to which the effects of a particular personality variable is moderated by other personality variables. Nevertheless, a rationale for moderator effects among personality variables can be found in the writings of several researchers, especially the type of moderator effect in which the *strength* of the relationship between a relevant personality variable and work performance differs at various levels of one or more other relevant personality variables. For example, Buss (1996) has argued that the five factors represent (in part) alternative strategies that can be drawn upon to solve problems. Implicit in this theorising is the idea that different personality characteristics can be useful for facilitating performance within a particular domain, and that being low on one relevant characteristic is not necessarily disadvantageous if the individual is high on some other useful characteristic. For example, an individual high on Openness may solve a problem by thinking of creative solutions, whereas the individual low on this characteristic but high on Conscientiousness may do so through persistence and hard work. A third individual who is low on both these characteristics but high on Extraversion may solve the same problem by eliciting the cooperation of others. This implies the type of moderator effect where performance is poor only if the individual is low on all characteristics that are useful for solving the problem. In other words, the relationship between a personality variable and performance is stronger when alternative personality variables that facilitate performance are low.

Conversely, one might expect situations in which the relationship between a personality variable and performance is stronger when other relevant personality characteristics are high. That is, a high score on one personality variable enhances the effects of a high score on another personality variable. For example, in order to have successful client relationships, a salesperson will need to display high levels of sociability to establish such relationships but will need to combine this with high levels of likeability to maintain the relationships (Hogan and Hogan, 1995); and in jobs involving interpersonal interaction being conscientious may add little to performance if the employee is not agreeable (Witt, Burke, Barrick & Mount, 2002). In both these examples the relationship between the personality variables and performance would be better represented as a multiplicative effect, where the benefits of being high on both personality variables is greater than the additive effects of the two variables.

Three recent studies by Witt and his colleagues have empirically investigated the multiplicative effects of the five factors on performance using moderated regression (Burke & Witt, 2002; Witt, 2002; Witt et al., 2002). Burke and Witt (2002) hypothesised that low levels of Openness would have a destructive effect on performance when combined with either high Extraversion, low Emotional Stability, or low Agreeableness (but not otherwise), and therefore that the relationship between Openness and performance would be stronger at higher levels of Extraversion, and lower levels of Emotional Stability and Agreeableness. This was tested using a sample of 114 employees of a financial services firm who completed measures of the five factors and who were subsequently rated on 13 performance items by their supervisors. The results supported the hypotheses concerning the moderating effects of Extraversion and Emotional Stability, but not Agreeableness. However, as only one sample was examined it was not possible to establish the generality of the findings.

Other studies utilised multiple samples and focused particularly on moderators of the Conscientiousness-performance relationship. Witt (2002) examined the interactive effect of Extraversion and Conscientiousness on job performance ratings across three samples. In all three datasets there was a significant moderation effect: Conscientiousness was more strongly related to performance at high levels of Extraversion than at low levels. Similarly, when Witt et al. (2002) investigated the interactive effect of Conscientiousness and Agreeableness on performance, they obtained a significant effect in five of the seven samples examined. Specifically, Conscientiousness was more strongly related to performance at high levels of Agreeableness than at low levels.

To summarise, there are various conceptual reasons to believe that the relationship between personality variables and performance may be nonlinear and/or configural in nature. Furthermore, there is some support (albeit mixed) for such relationships from the few empirical studies that have addressed this issue, although most of the research to date has focused on the Conscientiousness factor. In addition to the effects discussed, the situation may be more complex further still. For example, there is the possibility of higher order interactions where a moderation effect is itself moderated by other variables, or where the effect of the predictor on the outcome may vary nonlinearly as a function of the moderator. Unfortunately, theories of job performance are not precise enough to specify the exact nature of such relationships. Nevertheless, a method that is flexible enough to empirically detect such complex trends may well improve the effectiveness of personality variables for predicting work performance, as well as contribute to theory development by providing a better understanding of the nature of the relationships between personality variables and performance. Artificial neural networks represent one group of techniques that offer this flexibility, and it is to this method that we next turn our attention.

CHAPTER 2: Artificial Neural Networks

Introduction

Artificial neural networks, broadly defined, are networks of many simple processing elements (called *units*) that are interconnected via communication channels (called *weights*) carrying numeric data (Sarle, 2001a).[1] While any given unit only performs relatively simple computations, the network as a whole is typically capable of representing a variety of complex relationships (Reed & Marks, 1999).

There are many kinds of artificial neural networks, and many different tasks for which they are used. For example, Sarle (2001a) notes that they are used by physicists to model phenomena in statistical mechanics, by biologists to interpret nucleotide sequences, by cognitive scientists to describe high-level brain functions, by computer scientists to investigate the properties of non-symbolic information processing, and by engineers for signal processing and automatic control. The focus of this thesis is on artificial neural networks as a statistical procedure for mapping relationships between sets of input and output variables. In particular, the aim is to develop prediction equations that capture empirical relationships between scores on personality variables (the inputs) and work performance criteria (the output), and that can then be used to make predictions about an individual's work performance given their personality scores. This procedure provides an alternative way of evaluating the effectiveness of personality variables for predicting work performance, and may well result in prediction equations that produce more accurate predictions than equations developed using linear regression.

[1] Artificial neural networks are often simply referred to as neural networks. The two terms are used interchangeably in this thesis.

This chapter provides much of the rationale for the specific procedures that are used to implement neural networks in the subsequent studies. In particular, I focus on the most widely used type of neural network, the *multilayer perceptron*. The first section of this chapter provides a brief introduction to such neural networks and discusses their capability to represent different functional relationships between predictors and a criterion. The second section addresses issues related to developing neural networks. The third section is concerned with issues related to testing the predictive performance of networks. Techniques related to assessing the nature of the relationships detected by the network and the importance of individual predictors are discussed in the fourth section. The fifth section provides a review of previous studies that have used artificial neural networks to evaluate the effectiveness of psychological variables for predicting work-related criteria. In all sections the emphasis will be on comparing artificial neural networks to traditional regression-based methods of data analysis. At the end of this chapter the aims of the research conducted for this thesis are summarised, and a preview of subsequent chapters is provided.

Representational Capability

Multilayer perceptron neural networks provide a general framework for representing nonlinear and configural relationships between sets of input and output variables (Bishop, 1995). This can be best demonstrated by a network diagram, as in Figure 2.1, which illustrates some of the main characteristics of such neural networks.

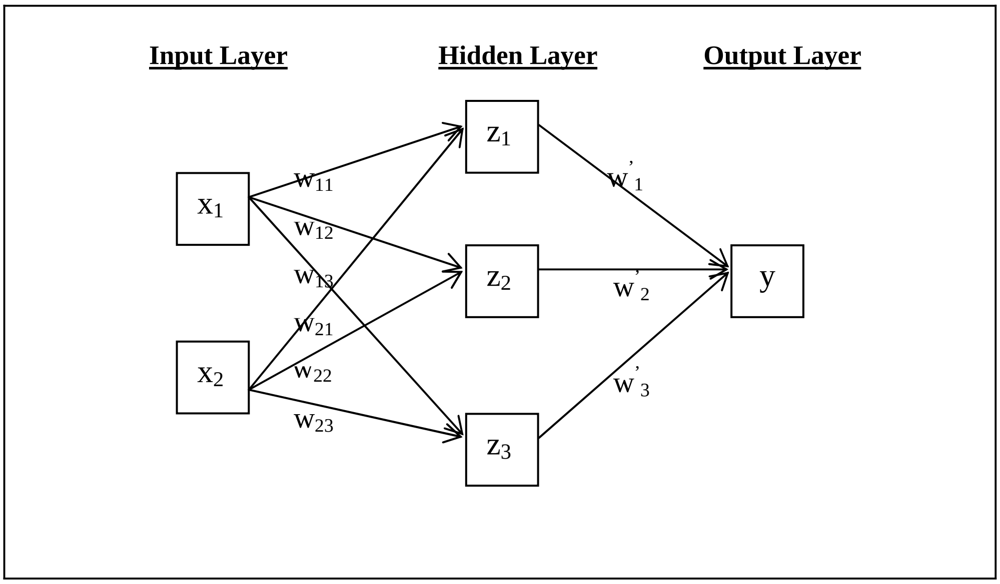


Figure 2.1

A diagrammatic representation of a multilayer perceptron neural network with two input units, three hidden units, and one output unit.

Each network is organised in terms of layers of units that are interconnected via weights. The weights are simply numerical values that represent the strength of the connections between units, much in the same way that regression coefficients represent the strength of the relationships between predictors and the criterion in linear regression. The network in Figure 2.1 has three layers: an *input* layer, an *output* layer, and a *hidden* layer that links the inputs to the outputs. It is possible to have neural networks with more than one hidden layer, although for practical applications one hidden layer is sufficient to represent virtually all relationships of interest (Masters, 1993). The layers are connected according to a *feedforward* topology where the units in one layer have one-way connections running to units in the next layer.

The input layer is used to represent the predictor or input variables. For the purposes of this thesis the inputs are the personality variables. When the input variables are continuous, as is the case with the majority of personality variables, then they can each be represented by one unit in the input layer. Unlike the other units in the network, the input units do not conduct any processing of the data (Masters, 1993).

The units in the hidden layer conduct some intermediate processing of the input values. Specifically, at each hidden unit the values of the inputs are weighted and summed. Each hidden unit also contains an adjustable *bias* term that is added to the sum of the weighted inputs.[2] The total sum is passed through an *activation function* to obtain the value of the hidden unit. This process can be expressed mathematically. Specifically,

$$z_j = f\left(\sum_i w_{ij}\, x_i + w_{0j}\right) \qquad (1)$$

where z_j is the value of the jth hidden unit, x_i is the value of the ith input unit, w_{ij} is the value of the weight linking input unit i to hidden unit j, w_{0j} is the value of the bias term of hidden unit j, and f(.) denotes the activation function that transforms the weighted linear combination of inputs and the bias term ($\sum_i w_{ij}\, x_i + w_{0j}$) into the output z_j. The role of the activation function is to introduce nonlinearity into the network. In practice, the most commonly used activation functions are sigmoidal (S-shaped) functions such as the logistic function or the hyperbolic tangent function (Sarle, 2001b). The hyperbolic tangent function is plotted in Figure 2.2 and is described by the equation $f(a) = (e^a - e^{-a}) / (e^a + e^{-a})$, where $a = (\sum_i w_{ij}\, x_i + w_{0j})$ as discussed above. Thus, each hidden unit is capable of representing a relatively simple type of nonlinear function.

[2] For simplicity the bias term is not depicted in Figure 2.1. The bias term plays a similar role to the intercept term (a) in a regression equation (Y = a + bX). A regression equation would be constrained to pass through the origin if the intercept term were omitted; equivalently, a neural network without bias terms would be constrained in terms of the functions that could be approximated (Sarle, 2001b).

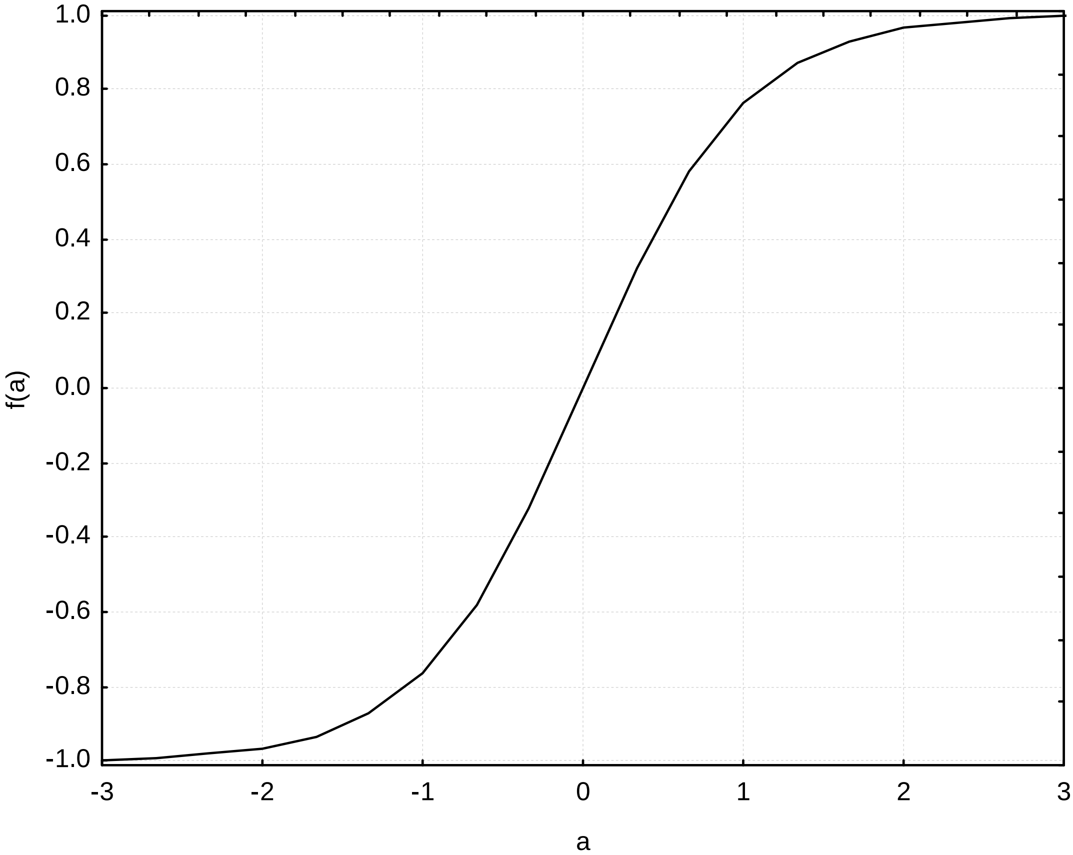


Figure 2.2

The hyperbolic tangent function: $f(a) = (e^a - e^{-a}) / (e^a + e^{-a})$

The output layer represents the criterion to be predicted. Only one unit in the output layer is needed if the criterion is continuous, as is the case with most measures of work performance, including the ones used in this thesis.[3] The output unit weighs and sums the values of the hidden units and adds a bias term to produce a single output value corresponding to a predicted score on the criterion. Typically, an identity activation function is used in the output layer (Hastie, Tibshirani, & Friedman, 2001), which is synonymous with using no activation function, and therefore the output unit

[3] If the criterion is an unordered categorical variable with more than two categories then multiple output units are required (Sarle, 2001b).

can be expressed as a weighted linear combination of the hidden units plus the output unit bias term:

$$y = \sum_j w'_j z_j + w'_0 \qquad (2)$$

where y is the value of the output unit (the predicted criterion score), z_j is the value of the jth hidden unit, w'_j is the weight linking hidden unit j to the output unit, and w'_0 is the value of the output unit bias term.

Thus, from the above it can be seen that the multilayer perceptron neural network can be thought of as a complex equation that represents the relationship between the input units (the predictors) and the output unit (the criterion) by:

1. Taking nonlinear functions of the linear combination of inputs at each hidden unit (Equation 1).
2. Linearly combining the nonlinear functions at the output unit (Equation 2).

This is a very powerful and general approach to mapping relationships as it allows the same network to represent a wide variety of nonlinear and configural relationships between the predictors and the criterion simply by varying the values of the weights (Hastie et al., 2001). To illustrate, one can consider networks with different numbers of hidden units. If a network has no hidden units, and therefore no hidden layer, then the inputs link directly to the output unit and the network is equivalent to a linear regression equation, $y = \sum_i w_i x_i + w_0$. Thus, linear regression can be conceptualised as a network with no hidden units. Clearly such a network is only capable of representing linear and additive relationships between the inputs and the output regardless of the values of the weights.

The introduction of a hidden layer with one hidden unit increases the range of functions that can be approximated. First, the sigmoidal activation function in the hidden unit allows the network to represent S-shaped functions. Furthermore, such a network is also capable of representing linear and step functions (Bishop, 1995). The former is achieved by setting the weights feeding into the hidden unit to very small values so that the summed input lies close to zero, where the sigmoidal function is approximately linear. Conversely, a step function can be represented by setting the weights into the hidden unit to very large values. Any rescaling that needs to be performed is then incorporated into the weights linking the hidden units to the output unit (Bishop, 1995).

When there is more than one hidden unit, the process of combining multiple sigmoidal functions at the output unit can generate many other types of nonlinear and configural functions, some of which are far more complex than the original sigmoidal functions. For example, depending on the weights chosen, the outputs from two hidden units can be combined to produce functions that look like ridges; and combining multiple ridges can produce functions that contain many bumps, curves, bends and other similar structures (Bishop, 1995).

The actual range of functions that can be approximated by a given network when its weights and bias terms are varied is referred to as the network's *representational capability* (Reed & Marks, 1999). Generally, as the number of hidden units is increased the network's representational capability also increases. One of the properties of multilayer perceptrons that makes them appealing is their capability as *universal approximators.* That is, given enough hidden units the network can in theory approximate all bounded continuous functions (Reed & Marks, 1999, cite a number of studies that have established this property through mathematical proof). It is not yet clear exactly how many hidden units are required to approximate any given function (Sarle, 2001c), although investigations with artificial data sets have demonstrated that even networks with only a small number of hidden units can approximate a wide variety of relationships. For example, Bishop (1995) generated artificial datasets containing quadratic, sine, absolute value, and step functions. He found that the same network, containing five hidden units, was able to represent each function by altering the values of the weights and bias terms.

The capability of representing many types of functions makes the neural network approach to detecting relationships in data very flexible in that few a priori assumptions need to be made about the nature of the relationships between the input and output variables.[4] This capability is likely to be particularly useful in situations where there are general expectations of nonlinear or configural relationships among variables but the existing theories are not precise enough to specify the exact nature of such relationships. Nevertheless, that a neural network is capable of representing various relationships does not necessarily guarantee that it will detect the true relationships underlying the data or that it will produce more accurate predictions than prediction equations developed using simpler methods such as linear regression. The act of developing a network – that is of selecting the appropriate weights – is one of the factors that plays an important role in successfully implementing neural network applications, and this is discussed next.

---

[4] It is possible to develop regression equations that have the same representational capabilities as neural networks simply by adding a sufficiently large number of power and product terms (Bishop, 1995). However, the number of weights that would need to be estimated as a function of the number of inputs increases at a much faster rate for this method than for neural networks (Paik, 2000), and therefore neural networks offer a more practical way of performing flexible function approximation.

Network Development

Developing a neural network requires a dataset, called the *training set,* that contains predictor and criterion scores for the training cases. The goal of network development is to use the training data to develop a prediction equation that accurately predicts the criterion scores of unseen cases (cases not in the training set) given the predictor scores (see Reed & Marks, 1999). It is assumed that the training data reflects an unknown underlying systematic component – namely the population level functional relationship between the predictors and the criterion – but is corrupted by random noise (Bishop, 1995). In order to make the most accurate predictions for unseen cases the network needs to capture the underlying functional relationship while ignoring the noise in the training set (Geman, Bienenstock, & Doursat, 1992). Hence, any differences between the form of the prediction equation derived by the network and the true underlying function will decrease the predictive performance of the network for unseen cases (Smith, 1993).

The extent to which a network captures the underlying relationship while ignoring training set noise is closely related to the representational capability of the network. If the network has little representational capability then it may not be able to represent the underlying function and this will result in a prediction equation that is too simple. This would occur, for example, if a linear equation were used when the underlying function is nonlinear. The equation would not fully reflect the underlying function and hence would yield predictions that are on average different to what would be predicted by the underlying function. In this case, the network would be said to *underfit* the data (Sarle, 2001c). On the other hand, if the network has a high level of representational capability then it is more likely that the underlying function will be contained within the range of relationships that the network can represent. However, this too comes at a price as high representational capability of the network will make it very sensitive to the noise in the particular training set, and hence it may *overfit* the data by capturing too much of the idiosyncrasies of the training data (Geman, Bienenstock et al., 1992). Consequently, the network will do a good job of fitting the training set but will yield poor predictions for cases not used to train the network. Therefore, there is a trade off between accurately fitting the training data and controlling the complexity of the prediction equation that needs to be balanced if the network is to display good predictions for unseen cases (Bishop, 1995). Procedures for accurately fitting the training data are presented immediately below, followed by a discussion of various approaches to controlling the complexity of the network.

Fitting the Training Data

A network with a given number of hidden units is fitted to the training data by assigning values to its weights such that some error function is minimised. This is referred to as *training* and is similar to parameter estimation in linear regression models. In linear regression the aim is to select the set of weights (regression coefficients) that minimises the sum of the squared errors (the error function), where the error for each case is defined as the difference between the criterion value predicted by the regression equation and the observed criterion value (Pedhazur, 1997). As the regression equation is a linear function of the weights, the error function is a quadratic function of the weights, and hence the optimal set of weights (optimal in the sense that it minimises the overall error on the training set) can be determined exactly, and relatively easily, by differentiation (see Bishop, 1995).

Neural networks are also commonly trained with the aim of minimising the sum of the squared errors (Hastie et al., 2001), although the process of estimating optimal weights is a lot more complex and arduous compared to linear regression. The error function to be minimised is highly nonlinear in the weights (Masters, 1993), and therefore the optimal weights need to be estimated iteratively. This process can be described as follows:

1. Small random values are assigned as weights.
2. The network is presented with the training set.
3. The predictor scores for each training case are weighted and propagated through the network in order to produce predicted output values.
4. Error is calculated by comparing predicted outputs to the observed output values.
5. The derivatives of the error with respect to the weights are computed by progressively working backwards through the network.
6. The weights are adjusted so as to decrease error.

The process is iterated by repeating steps 2 to 6 many times until error is minimised (Reed & Marks, 1999).

Backpropagation refers to the procedure for calculating the derivatives of the error with respect to the weights (step 5), and is also the name of the most commonly used training algorithm for adjusting the weights (step 6). In short, the derivatives are calculated through an application of the chain rule of calculus (see Hinton, 1992 for a description of this procedure). The derivatives indicate the direction in which the error function increases the most. The backpropagation training algorithm adjusts the weights by moving a small distance in the opposite direction, where error decreases most steeply (Smith, 1993). By iterating the above process many times the error gradually decreases and converges towards a minimum. The backpropagation algorithm has been shown in many practical applications to be a successful method of training artificial neural networks (e.g., Bounds, Lloyd, & Mathew, 1990; Gorman & Sejnowski, 1988; Rajavelu, Musavi, Shirvaikar, 1989; Silverman & Noetzel, 1990), however it can be very slow and sometimes requires many thousands of iterations before error is minimised. A more complex algorithm, called *conjugate gradients*, reduces the training time required by assuming a quadratic error surface (Reed & Marks, 1999). As the quadratic error assumption tends to only hold in the vicinity of a minimum it is useful to initially train a network with a few iterations of backpropagation in order to obtain the approximate position of a minimum, followed by conjugate gradients training to obtain a final solution (Reed & Marks, 1999). This procedure for training neural networks was adopted in the present research.

Controlling the Complexity of the Network

Given the large representational capability of neural networks, the process outlined above for fitting the network to the training data may overfit the data by capturing some of the noise in the training set, and consequently produce poor predictions for unseen cases. The extent to which overfitting occurs also depends on characteristics of the data within the domain of interest, such as the amount of noise in the data and the number of training cases available (Reed & Marks, 1999). If the data is very noisy or only small samples are available for training purposes then a simpler linear model may outperform a neural network even if the underlying function is nonlinear (Sarle 2000c). Several approaches to reducing the extent to which neural networks overfit the data have been developed based on controlling the complexity of the network. Three of the more common methods are discussed below.

The simplest approach involves experimenting with different numbers of hidden units, including networks with only a small number of hidden units (e.g., Zhang, Hu, Patuwo, & Indro, 1999). As previously discussed, the range of functions that can be approximated by a neural network increases as the number of hidden units is increased. By including networks with only a small number of hidden units one limits the representational capability of these networks and hence their ability to overfit. However, even small networks can overfit the data (e.g., Caruana, Lawrence, & Giles, 2000), and therefore this approach is not always effective.

A second approach, referred to as *weight regularisation*, is based on constraining the size of the weights in the network by adding a penalty term to the error function. For example, a common penalty term referred to as the *weight decay* term involves squaring the weights, summing them, and multiplying the sum by a regularisation constant that influences the extent to which the weights are constrained (Reed & Marks, 1999). When a network is trained with this term added to its error function it is penalised for having large weights (in the form of an increase in the overall error function). The training process is therefore motivated to reduce the size of weights that are not essential to the solution (Reed & Marks, 1999). Large weights cause a lot of curvature in the prediction equation, and the use of a penalty term has the effect of smoothing the equation. Weight regularisation techniques have been successfully used to reduce overfitting and improve predictive accuracy for unseen cases in a number of practical applications of neural networks (e.g., Ennett & Frize, 2003; Weigend, Huberman, & Rumelhart, 1990).

A third approach is the procedure known as *early stopping* (e.g., Hagiwara, 2002; Sarle, 1995). It involves using only a subset of the total training cases to adjust the weights, and using the remaining cases to monitor prediction error *during the training process*. Specifically, the available training cases are partitioned into a *training set* and a *validation set*. A network with a large number of hidden units is trained using the training set and prediction error for the validation cases is calculated periodically during training, ideally after every iteration. Generally, as training progresses, error on the validation set initially decreases but then starts to increase as the network starts to overfit the training set (e.g., LeBaron & Weigend, 1998; Wang, Venkatesh, & Judd, 1994; cf. Prechelt, 1998). By monitoring error on the validation set one can stop the training process at the point where error starts to increase and select the set of weights corresponding to this point. In other words, this procedure uses the validation set as an indicator of when the network starts to overfit the training data.

Like weight regularisation, early stopping has been successfully applied in a number of studies that have used artificial neural networks to solve practical problems (e.g., Gencay & Qi, 2001; Edwards & Murray, 2000; Tetko, Livingstone, & Luik, 1995). Finoff, Hergert, and Zimmermann (1993) compared the performance of several techniques for controlling complexity including variants of the three approaches presented here. They tested each technique on a number of artificially generated datasets that varied in terms of the amount of noise in the data, the degree of nonlinearity, and the number of input and output variables. There was no clear difference in predictive accuracy for unseen cases between the weight regularisation and early stopping procedures, although both approaches were uniformly superior to simply experimenting with the number of hidden units. In the present research combinations of all three procedures were employed.

Testing Predictive Performance

Regardless of the approach used to develop artificial neural networks, a network's predictive performance should be evaluated using data that is independent of that used during the training process. Error on the training set provides an overoptimistic estimate of predictive accuracy. A network with a large number of free parameters (or weights) can produce very low training error by fitting the noise in the training set yet perform poorly when generalising to unseen cases. Similarly, when using the early stopping procedure error on the validation set is also a biased estimate of the predictive performance of the chosen network (Sarle, 2001c). This is because the early stopping procedure involves a choice among a potentially large number of networks (each iteration of the training process representing one possible network), and therefore provides a good deal of opportunity for overfitting the validation set. Consequently, in addition to the training set (and validation set if early stopping is employed), a separate set of data that is not used in any way during network development (referred to as the *test set*) is required to assess the predictive performance of the network. This is in contrast to the way in which predictive performance is typically assessed in the context of linear methods. For example, with a single predictor it is common to compute a validity coefficient over the entire data available. Similarly, when there is more than one predictor it is often the case that a multiple correlation coefficient (R) is computed over the entire data and adjusted for upward bias using one of the formulae that have been developed for this purpose (see Pedhazur, 1997).

The simplest approach to obtaining subsets of data, referred to as the *hold out* method, involves randomly partitioning the total data into a training set and a test set. Prediction equations are then developed using the former set and tested on the latter. A common heuristic is to allocate approximately two-thirds of the cases to the training set, and the remaining cases to the test set (Weiss & Kulikowski, 1991). If early stopping is being used then the training data needs to be further partitioned into a training set and a validation set.

The hold out method provides an unbiased estimate of the predictive performance associated with a particular prediction equation. It can also be used to compare the predictive performance of two prediction equations. However, it is less appropriate when the researcher is interested in the predictive performance associated with a particular method of generating prediction equations, or with comparing the performance of two such methods (Dietterich, 1998). This is because it does not take into account variability that occurs as a result of the random choice of the training set. The variability in the estimate of predictive performance across different random partitions of the data tends to be large (Martin & Hirschberg, 1996; Reed & Marks, 1999; also see LeBaron & Weigend, 1998). Therefore, misleading results may be obtained if one relies on a single partition of the data. A more appropriate approach is the use of a *resampling* procedure. For example, Nadeau and Bengio (2003) outline a simple form of resampling that involves repeating the random partitioning of the data into a training set and a test set multiple times. For each partition, prediction equations are developed using the training set and tested on the test set. Predictive performance is estimated by averaging over the multiple test sets. In this way a more stable estimate of predictive performance is obtained.

A second issue concerns the choice of the measures used to assess the predictive performance of prediction equations. Artificial neural networks, like linear regression equations, are usually trained to minimise squared errors.[5] It follows that one potential measure of predictive performance is the *mean square error* on the test set. However, squared errors are purely mathematical constructs that are not easily interpreted (Masters, 1993). Furthermore, the act of squaring errors magnifies the effects of large

[5] Least-squares error functions have certain properties that are very useful for training prediction equations. For example, their derivatives are more easily computed than other error functions, and they correspond to maximum likelihood under certain realistic assumptions (see Bishop, 1995).

errors, which is not necessarily desirable. Therefore, for the purposes of testing a network it is often convenient to use a different error function to that used to train the networks (Bishop, 1995), or to at least include other measures of predictive accuracy in addition to mean square error. For example, the *mean absolute error* takes the absolute value of the difference between predicted and observed criterion scores for each test case and averages across the entire test set. Thus, it indicates the average magnitude of the error. The mean absolute error places equal emphasis on large and small errors. Furthermore, given that the mean absolute error is expressed in the original scale of the criterion, it is more easily interpreted than mean square error.

Both the mean square error and the mean absolute error assess predictive performance as a function of the discrepancy between predicted and observed criterion scores. Hence, predictions are evaluated in terms of the *absolute agreement* between predicted and observed scores. Absolute agreement takes into account, among other things, the extent to which the prediction equation correctly estimates the mean and standard deviation of the observed scores (Kirlik & Strauss, 2003). As a result it is useful when one is interested in predicting the specific value of an individual's criterion score. Measures of absolute agreement can be distinguished from *relational* indices of predictive performance such as the test set correlation between predicted and observed criterion scores, also referred to as the *cross-validity coefficient*. The cross-validity coefficient is not sensitive to differences in either the magnitude or scale of predicted and observed criterion scores (see Kirlik & Strauss, 2003), and is therefore more relevant when one aims to predict the relative standing of cases on the criterion rather than the actual value of the criterion. Furthermore, although both absolute and relational measures are relevant for the assessment of predictive performance, the cross-validity coefficient is more closely related to the validity coefficients that have typically been used to evaluate the validity of predictor variables within organisational contexts. [6, 7]

Analysing the Role of Predictors

Once a network has been tested and shown to exhibit satisfactory predictive performance, one may be interested in investigating the nature of the relationships between the predictors and the criterion that are detected by the network. Furthermore, the relative importance of the different predictors for predicting criterion scores may also be of interest.

It is useful to first consider the manner in which the role of predictors is assessed within regression-based methods. In linear regression the weight or coefficient associated with each predictor indicates the linear relationship between the predictor and the criterion, holding constant the effects of the other predictors (Pedhazur, 1997). There are various inferential tests that determine whether each weight is significantly different from zero, and the sign of the weight can be used to indicate the direction of the relationship. Furthermore, the relative importance of predictors is often (possibly inappropriately) determined by standardising the weights and then comparing their magnitudes (see Pedhazur, 1997). More complicated forms of regression, such as polynomial regression or moderated multiple regression, represent each type of relationship as a separate variable (see Cohen, 1978). For example, a quadratic relationship is represented by a variable that is the squared value of the predictor

---

[6] Guion (1998) distinguishes between the concepts of validity and accuracy, and argues that relational measures are more closely aligned to the definition of validity whereas absolute measures are more appropriately a measure of accuracy. Nevertheless, the two types of measures are related in that accuracy is a function of validity.

[7] It should be kept in mind that a negative cross-validity coefficient indicates that the predicted values are negatively related to the observed values and therefore that the predictive performance of the prediction equation is poor. This contrasts with the interpretation of validity coefficients, for which it is the magnitude of the coefficient (rather than both sign and magnitude) that indicates predictive performance.

hypothesised to be quadratically related to the criterion. Similarly, a multiplicative relationship of two predictors with the criterion is represented by a variable that is the product of the two predictors. The weight associated with the variable can then be tested to determine whether the hypothesised quadratic or multiplicative relationship is statistically significant.[8] The actual nature of the quadratic or multiplicative relationship is subsequently inferred by referring to the sign of the weight or else by graphical methods.

In contrast, it is a lot more difficult to extract information about the role of the predictors in an artificial neural network. The prediction equation represented by a network usually takes a very complicated form; and the relationship of each predictor to the criterion is not represented by a single weight, but rather is distributed across many weights that are transformed in complex ways before linking to the output unit. As a result, it is normally not useful to try to interpret the weights in a network (Masters, 1993). Nevertheless, there are several approaches that can be used to provide insight into the role of different predictors in a neural network. Below I discuss graphical methods and sensitivity analysis.

One can use graphs to depict the nature of the relationship between predictors and the criterion (e.g., Somers, 1999). For example, a series of two-dimensional graphs can be used to visualise the relationship between individual predictors (plotted on the horizontal axis) and the criterion (plotted on the vertical axis). The graphs are generated by varying the values of the predictor, and determining the predicted criterion score at each value of the predictor. When there are multiple predictors, the value of all predictors other than the one being examined can be held constant, for example by fixing them at their mean values. Configural relationships involving two predictors can

---

[8] The proper way to conduct this analysis is via hierarchical regression in which the linear effects of predictors are held constant when testing the quadratic or multiplicative effects (See Cohen, 1978).

be considered using a three-dimensional graph in which the two predictors are plotted as the width and depth of the graph and the criterion is plotted as the height. Again, all predictors not being examined can be held constant at their mean values. Of course graphs derived from neural network analyses cannot be used to assess the statistical significance of any particular relationship. However, they are potentially useful exploratory tools for suggesting possible nonlinear or configural relationships between predictors and the criterion that can then be more precisely examined using other techniques, such as regression-based hypothesis testing.

A direct measure of the relative importance of predictors in a neural network is the change in predictive performance that occurs when each predictor is omitted from the network (Sarle, 2000). This can be implemented via *sensitivity analysis* (StatSoft Inc., 1999), in which the information provided by a predictor is made unavailable by clamping its value to a typical value (for example the mean of the predictor) and predictive performance is computed using scores on the other predictors. This is repeated for each predictor. A predictor whose omission results in a large deterioration in predictions is considered more important for predictive purposes than a predictor whose omission has little impact on predictions or else improves predictions.

Applications to Organisational Psychology

Artificial neural networks, and in particular multilayer perceptrons, have been successfully applied to many real-world problems similar to the problem addressed in the present research. For example, neural networks have been used to predict outcomes relevant to clinical psychology (Price et al., 2000) and social work (Marshall & English, 2000). Within organisational contexts there have been successful applications to many of the tasks encountered in the manufacturing, marketing, and finance departments of businesses (e.g., Edwards & Murray, 2000; Kuo, Wu, & Wang, 2002; Zhang et al., 1999). Yet, despite the widespread popularity of artificial neural networks in related domains, there have been few attempts to extend this methodology to organisational psychology in particular. A review of the literature yielded only a small number of published studies that had used artificial neural network methods to address issues related to behaviour in organisations. Nevertheless, the few investigations that have been conducted hold promise for the usefulness of artificial neural networks in organisational psychology research (Hanges, Lord, Godfrey, & Raver, 2002).

Collins and Clark (1993) were possibly the first researchers to apply artificial neural networks to the prediction of workplace behaviour. Using a sample of 81 managers working in project teams, they sought to predict managers' perceptions of team performance (dichotomised as high versus low) from their perceptions of team satisfaction, team cohesion, team task-orientation and work pressure. They developed one multilayer perceptron neural network using two-thirds of the data, and assessed the predictive performance of the network by calculating a cross-validity coefficient using the remaining data. The cross-validity coefficient of the neural network was identical to that of an equation developed using the simpler technique of discriminant analysis. However, the number of hidden units in the neural network was chosen in an ad-hoc manner, and was large relative to the number of training cases. This probably resulted in an overfitted prediction equation, as indicated by the large discrepancy between the network's performance on the training set and the test set. In a second analysis only indirectly related to work performance, Collins and Clarke investigated whether artificial neural networks could be used to distinguish between managers who either were or were not incarcerated for white-collar crimes, using scales of the California Psychological Inventory (Gough, 1987) as predictors. In this case the sample size was much larger (N = 649, training n = 435, test n = 214) and fewer hidden units were used to develop the neural network than in the first study. The cross-validity coefficient of the neural network (r=.70) was higher than that of the discriminant model (r=.66). Collins and Clark concluded that their results warranted further research on artificial neural networks as a statistical tool for personnel psychologists.

As part of his doctoral dissertation Scarborough (1996) examined the feasibility of artificial neural networks as a methodology for criterion validation of employee selection testing. The participants were 1085 telephone sales agents working within a large service organisation. Thirty-five multilayer perceptrons were developed to predict the revenue generated by the sales agents using information from a test battery consisting of biographical and personality predictors. The predictive performance of the networks was significantly greater than that of a linear regression model and roughly equivalent to that of a proprietary nonlinear equation.

Griffin (1998) used artificial neural networks to assess the validity of psychomotor and other aptitude tests for predicting the flight grades earned by student naval aviators during the first 6-months of training. Half of the 434 cases were used to develop a linear regression equation along with many different types of feedforward networks including multilayer perceptrons. Cross-validity coefficients computed on the remaining cases showed that the multilayer perceptrons outperformed all the other artificial neural networks but were no more accurate than the linear regression equation. Griffin speculated that this may have been due to the linear nature of the data. In concordance with this claim, a review of 174 studies found that nonlinear relationships between aptitude measures and job performance – as opposed to linear aptitude-performance relationships – do not occur at levels substantially greater than chance (Coward & Sackett, 1990).

Somers (1999) used artificial neural networks to study the relationship between work attitudes (organisational commitment, job satisfaction and job withdrawal intentions) and turnover (whether or not the employee left the organisation within the subsequent 12 months) among a sample of 577 nurses. A multilayer perceptron with three hidden units was developed using 462 training cases, and used to predict turnover for each of the 115 nurses assigned to the test set. The network achieved a correct classification rate of 88%, which was higher than the 76% achieved by the benchmark logistic regression equation. Furthermore, graphical representations were used to demonstrate a number of nonlinear relationships between work attitudes and turnover that may have accounted for the superior predictive performance of the neural network, and these relationships were interpreted in terms of existing theories of turnover.[9]

More recently, Somers (2001) extended his investigation to attitudinal predictors of job performance ratings. He developed a neural network and a linear regression equation using a training sample of 185 nurses, and tested the equations on a test set consisting of 47 nurses. The cross-validity coefficient of the network (r=.27) was higher than that of the regression equation (r=.17). Somers then graphically identified a number of nonlinear and multiplicative relationships that may have accounted for the findings, and he highlighted the need to incorporate nonlinear effects into existing theories of attitude-performance relations.

In summary, the findings described above suggest that the application of artificial neural network methodology to issues encountered in organisational psychology is potentially useful. In particular they show that neural networks are generally able to deal with the level of noise present in datasets collected within

[9] Somers (1999) also trained and tested a Learning Vector Quantization neural network and found that it performed favourably compared to the logistic regression equation. Learning Vector Quantization networks are only appropriate for classification tasks (in which the criterion is categorical) and are therefore not relevant for the present research, which used continuous critera.

organisational contexts, and yield more accurate predictions (if only slightly) than traditional methods when large samples are available and there is some expectation of complex relationships between the predictors and the criterion.

Aims and Preview

The overriding aim of the present thesis was to compare artificial neural networks and linear regression as two methods of evaluating the effectiveness of personality variables for predicting work performance. Specifically, the aim was to develop prediction equations that relate scores on personality variables to measures of work performance, using either neural networks or linear regression to produce the equations, and to then compare the predictive performance of the two types of equations. A second related aim was to explore the nature and extent of any potential nonlinear and configural relationships between personality variables and work performance. The findings were anticipated to be relevant to two broad areas of research as described below.

First, the present work represents a comparison of two different methodological approaches to evaluating the effectiveness of psychological variables as predictors of behaviour in organisations, and is therefore relevant to the literature on organisational research methods. Specifically, the flexible representational capability of the neural network approach was compared to the simpler and more traditional regression-based method that assumes linear and additive relationships among variables. This is in line with recent calls for the application of more complex methods to the analysis of organisational data (e.g., Hanges et al., 2002; Mount, Barrick, & Ryan, 2003). The choice of the best method is contingent on the nature of the data within the domain of interest (e.g., the noisiness of the data, the presence of complex relationships, and the number of cases available), and is therefore an empirical question. As described in this chapter, the findings from some of the few studies that have applied artificial neural networks to organisational data suggest optimism about the usefulness of this method for issues encountered in organisational psychology. The present study investigated whether this optimism could be extended to the specific case where personality variables are used to predict work performance. Furthermore, much of the previous work has relied on a single measure of predictive performance, a single partition of the data into training and test subsets, or a single approach to developing neural networks. Methodological contributions of the present research include the use of multiple measures of predictive performance, the use of a resampling procedure, and the comparison of two alternate procedures for developing neural networks.

Second, the present research is relevant to a number of issues within the literature on personality and work performance. The research investigated the specific nature of the relationships between personality variables and work performance. As outlined in chapter 1, current theories of work performance implicitly assume linear and additive relationships between personality variables and performance (e.g., Barrick et al., 2003); by examining more complex relationships the present research addresses the tenability of the linearity and additivity assumptions. Furthermore, the present research addresses the effectiveness of different personality variables for predicting work performance. Unlike previous research, however, this issue was evaluated in the context of a method that is capable of capturing highly nonlinear and configural relationships, and therefore the obtained results provide an assessment of the effectiveness of personality variables in light of the possibility of such complex relationships.

The above issues were investigated by conducting a comparison of artificial neural networks and linear regression across six datasets that comprised a broad range of occupations, personality inventories, and work performance measures. The details of these datasets, along with some preliminary analyses, are presented in chapter 3. Chapters 4 and 5 report the results of studies that compared artificial neural networks and linear regression using measures of the dimensions of the five-factor model as the predictors. Chapter 4 presents analyses for each personality measure *individually*, thereby facilitating the exploration of nonlinear relationships between personality and work performance. Chapter 5 presents analyses for *combinations* of the five factor measures, thus facilitating the exploration of configural relationships. Chapter 6 reports the results of a study that compared the predictive performance of the broad personality variables represented by the five-factor model to that of narrower personality variables that assess specific facets of these broader constructs. This study also compared artificial neural networks and linear regression in the context of narrow personality measures. Finally, Chapter 7 presents a general discussion of the findings.

CHAPTER 3: The Datasets

Introduction

Previous studies that have applied artificial neural networks to organisational data have relied on one or two datasets from the domain of interest. In such cases one cannot easily conclude whether the success of an application is specifically limited to characteristics of the particular dataset (for example the occupational group under examination or the inventory used), or whether it can be generalised to other datasets collected within the domain. The present research compared artificial neural networks and linear regression using six datasets that comprise a range of occupations, personality inventories, and work performance measures. In this way it was possible to assess the extent to which the results generalise across different dataset characteristics.

This chapter introduces and describes the six datasets that were used as the basis of the analysis. The chapter is divided into two sections. The first section describes the method by which each dataset was obtained. Detailed descriptions of the participants, the measures, and the procedure are provided for each dataset. The second section presents the findings from preliminary analyses conducted on each dataset. The analyses were carried out to evaluate the validity with which the intended constructs were being measured, to assess the reliability of the predictor and criterion scales, and to provide some descriptive statistics for the datasets.

Description of the Datasets

The present research used both existing and newly collected datasets as part of the comparison of neural networks and linear regression. A review of the recent literature was used to draw up a list of researchers who had previously conducted organisational-based field studies in which personality and performance measures were assessed. Each researcher was contacted by an email that outlined the purpose of the research, and access was requested to any relevant datasets that they had collected. For a dataset to be suitable for the present research three criteria had to be met. First, the dataset had to include a personality instrument that had been explicitly based on the five-factor model and all five factors needed to have been assessed. Second, the dataset had to include a measure of on-the-job performance or training performance. Third, given that large sample sizes are usually required to develop and test artificial neural networks, the dataset had to include complete data for at least 100 cases.

Through the process described above four existing datasets were acquired that met the criteria for selection (Datasets 2 to 5 in Table 3.1). In addition, I collected two new datasets specifically for the purposes of the present research (Datasets 1 and 6 in Table 3.1). This yielded a total of six datasets.

Table 3.1

Characteristics of the six datasets used in the present thesis.

Dataset	Sample size	Personality inventory	Performance measure
1. University students	227	IPIP-NEO	Academic marks
2. Police recruits	286	NEO PI-R	Test marks
3. Flight attendants	305	NEO PI	Instructor ratings
4. Managers	179	NEO PI-R	Supervisor ratings
5. Bus drivers	486	HPI	Supervisor ratings
6. Professionals	120	CPS-2	Supervisor ratings

Note: IPIP-NEO = International Personality Item Pool NEO, NEO PI-R = Revised NEO Personality Inventory, NEO PI = NEO Personality Inventory, HPI = Hogan Personality Inventory, CPS-2 = Congruence Personality Scale form 2.

As can be seen in Table 3.1, the six samples represent a wide range of occupations, and encompass both on-the-job performance and performance on training programs. The first sample consists of university students, and is therefore not strictly an occupational sample. It is included here given that performance on training programs often contains an academic component (e.g., Driskell, Hogan, Salas, & Hoskin, 1994). The other five samples comprise five of the six major occupational categories included in previous meta-analyses (e.g., Barrick et al., 2001; Hurtz & Donovan, 2000), namely: police, customer service workers (flight attendants), managers, skilled/semi-skilled workers (bus drivers), and professionals; only the sales category is not represented here. Datasets 2 and 3 were collected during training programs, and therefore relate to training performance. Datasets 4, 5, and 6 relate to on-the-job performance.

Three of the datasets used the NEO Personality Inventory (NEO PI; Costa & McCrae, 1985) or its revised version (NEO PI-R; Costa & McCrae, 1992) to measure the five factors. A fourth dataset employed the widely used Hogan Personality Inventory (HPI; Hogan & Hogan, 1995). Two five-factor inventories that have been less extensively used in organisational research, the International Personality Item Pool NEO (IPIP-NEO; International Personality Item Pool, 2001) and the Congruence Personality Scale form 2 (CPS-2; Pryor & Taylor, 2000), were used in the remaining two datasets. Performance was assessed using scores obtained on objective performance tests (Datasets 1 and 2) or ratings provided by supervisors/instructors (Datasets 3, 4, 5, and 6). Each dataset is described in more detail below. Given that I was not involved in the collection of Datasets 2 to 5 descriptions of these datasets is largely based on notes provided by the original researchers and the documentation in published papers. Therefore, these descriptions are sometimes less detailed than those for the datasets that I collected, namely Datasets 1 and 6.

Dataset 1: University students

Participants. The participants were 234 university students (181 females and 53 males, mean age = 20) who were enrolled in a 14-week introductory psychology course at the University of New South Wales. Seven of the participants were subsequently omitted from the analysis due to incomplete or missing performance scores resulting in a final sample of 227 students.

Measures. Personality was assessed using the IPIP-NEO (International Personality Item Pool, 2001). This instrument consists of 300 items that are responded to on a five-point scale with the labels *very inaccurate*, *moderately inaccurate*, *neither accurate nor inaccurate*, *moderately accurate*, and *very accurate*.[1] The items are scored from 1 to 5 (or 5 to 1 for items scored in reverse direction), and are summed to obtain 30 lower-level scales (10 items per scale). The lower-level scales were designed to measure the same constructs as those assessed by the 30 facet scales of the NEO PI-R (Table 3.2 lists the facet labels of these two inventories), although the two inventories differ in terms of the content of the items and the number of items per scale. The lower-level scales of the IPIP-NEO are summed to obtain five higher-level scales assessing Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (each higher-level scale is the sum of six lower-level scales or 60 items). Higher scores indicate greater levels of the personality attribute.

Performance was measured by the student's overall mark (out of 100) for the course. The overall mark was composed of results obtained in a final examination, as well as a research report, a methodology assignment, a tutorial field study, and exams completed during the semester. Higher scores indicate better performance.

---

[1] The items of this inventory are available in the public domain (http://ipip.ori.org/newNEOKey.htm).

Table 3.2

The 30 lower-level facet scales of the IPIP-NEO and NEO PI-R.

Facet	IPIP-NEO label	NEO PI-R label
N1	Anxiety	Anxiety
N2	Anger	Angry Hostility
N3	Depression	Depression
N4	Self-Consciousness	Self-Consciousness
N5	Immoderation	Impulsiveness
N6	Vulnerability	Vulnerability
E1	Friendliness	Warmth
E2	Gregariousness	Gregariousness
E3	Assertiveness	Assertiveness
E4	Activity Level	Activity
E5	Excitement-Seeking	Excitement-Seeking
E6	Cheerfulness	Positive Emotions
O1	Imagination	Fantasy
O2	Artistic Interests	Aesthetics
O3	Emotionality	Feelings
O4	Adventurousness	Actions
O5	Intellect	Ideas
O6	Liberalism	Values
A1	Trust	Trust
A2	Morality	Straightforwardness
A3	Altruism	Altruism
A4	Cooperation	Compliance
A5	Modesty	Modesty
A6	Sympathy	Tender-Mindedness
C1	Self-Efficacy	Competence
C2	Orderliness	Order
C3	Dutifulness	Dutifulness
C4	Achievement-Striving	Achievement-Striving
C5	Self-Discipline	Self-Discipline
C6	Cautiousness	Deliberation

Note: The letters N, E, O, A, C in the first column indicate the higher-level scale that the lower-level facet relates to.

Procedure. Students were recruited through a web site set up by the School of Psychology for this purpose, and participated in the study in return for course credit. They completed a paper and pencil version of the personality inventory during the semester in groups of up to ten students. At the beginning of each testing session each participant was provided with a copy of the personality inventory and a participant information statement and consent form. The latter outlined the purpose of the study, requested permission to access the student’s psychology grades, and assured the student that any information obtained in connection with the study would remain confidential. Upon signing the consent form and completing the questionnaire each participant was given a debriefing form and provided with the opportunity to ask questions. At the end of the semester each student’s mark for the different assessment components of the course was accessed.

Dataset 2: Police recruits

Access to this dataset was provided by Jonathan Black from the New Zealand Police. For further details see Black (2000).

Participants. The participants were 286 police recruits (190 males and 96 females, mean age = 27) attending a 22-week basic training program at the Royal New Zealand Police College. The program was designed to develop the skills and knowledge required for police work, and included firearms training, physical training and self-defence, driving, computer studies, social science skills training, and police law and procedures.

Measures. Personality was assessed using the NEO PI-R (Costa & McCrae, 1992). This instrument consists of 240 self-report items that are responded to on a five-point scale ranging from *strongly disagree* to *strongly agree*. Each item is scored from 0 to 4 (or 4 to 0 for reverse scored items). The five higher-level *domain* scales (Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness) are defined by groups of lower-level *facet* scales. Eight items are summed to derive each of the thirty facet scales (refer to Table 3.2 for a list of the facet labels). Each of the five domain scales is in turn derived by summing the six facet scales that relate to that domain. This yields a score ranging from 0 to 192 for each domain scale. Higher scores indicate greater levels of the attribute.

A composite index of overall course performance (out of 1000) was used to operationalise the training performance criterion. The index was the aggregate of scores on 17 academic (e.g., assignments, exams) and practical (e.g., firearms, driving) tests completed as part of the training program. Higher scores indicate better performance.

Procedure. The recruits completed the personality inventory within a month of starting training. They were assured that their results would remain confidential and would be used for research purposes only. Performance tests relating to the different components of training were completed during the program, and at the end of the training course the test scores achieved by each participant were obtained.

Dataset 3: Flight attendants

Access to this dataset was provided by Douglas Cellar from DePaul University. For further details see Cellar, Miller, Doverspike, and Klawsky (1996).

Participants and procedure. This dataset was collected as part of a 6-week training program for 423 flight attendant trainees (361 females, 50 males, and 12 unspecified) who had recently been recruited by a large international airline. The training program involved role-playing exercises, participative exercises, discussions, lectures, and take-home study. During the last two weeks of training the trainees completed the personality inventory, and performance ratings were collected from instructors. Performance ratings were not available for 98 of the trainees, and an additional 20 trainees had missing personality information. The final sample of 305 trainees consists of those with complete personality and performance data.

Measures. Personality was assessed using the NEO PI (Costa & McCrae, 1985). The NEO PI differs from the revised version in that only three of the five dimensions (Neuroticism, Extraversion, and Openness) are defined in terms of lower-level facet scales. The inventory does not contain facet scales for Conscientiousness and Agreeableness, but rather assesses these dimensions using two 18-item global scales. Furthermore, ten of the items that were used in the original version to measure facets of the first three dimensions were subsequently replaced in the revised version.

Performance was assessed using ratings provided by instructors on eight behaviourally anchored rating scales. Seven of the eight scales measured performance on dimensions determined by job analysis to be important for training and flight attendant success, namely a) learning and applying knowledge, b) demonstrating responsible work habits, c) work-related communications, d) interpersonal skills, e) customer interaction, f) teamwork, and g) problem solving. The final scale was a global rating of overall training performance. Each scale was rated from 1 to 9, where higher scores indicate better performance. Performance was operationalised as the average score over the eight items.

Dataset 4: Managers

Access to this dataset was provided by the Center for Creative Leadership. For further details see Dalton, Ernst, Leslie, and Deal (2002).

Participants and procedure. The participants were 211 managers (184 males and 27 females, mean age = 45) from four organisations, who were at approximately the same level of management. Ninety-eight managers were from a Swiss-based pharmaceutical company, 48 worked for a Swiss-based hospitality and service company, 40 were from a Swedish-based truck manufacturing and construction company, and 25 were from a U.S.-based technology company. The managers were asked to complete the personality inventory along with other questionnaires that formed part of the original study. Concurrently, performance ratings were collected from the supervisor of each manager. All respondents were assured that their responses would remain confidential. One participant was excluded due to 54 missing responses on the personality inventory,[2] and 31 participants were excluded due to missing performance data. The final sample therefore consisted of data for 179 cases.

Measures. Personality was assessed using the NEO PI-R (Costa & McCrae, 1992). This instrument was discussed in connection with Dataset 2. Supervisor ratings on a 10-item scale designed to assess achievement on the job were used as the performance criterion (See Table 3.4 in this chapter for a list of the ten items). The items were part of a broader set of scales that were developed and revised in consultation with the Director of Executive Development from one of the participating companies. The items are responded to on a five-point scale ranging from *strongly disagree* to *strongly agree*. Each item is scored from 1 to 5 where higher scores indicate better performance. Performance was operationalised as the average score over the ten items.

[2] The NEO PI-R manual (Costa & McCrae, 1992) recommends that the inventory should not be scored if 41 or more responses are missing.

Dataset 5: Bus drivers

Access to this dataset was provided by Jeffrey Conte from San Diego State University. For further details see Jacobs, Conte, Day, Silva, and Harris (1996).

Participants and procedure. The participants were 864 bus drivers from nine bus properties across North America who had volunteered to participate in the study in return for their hourly wage rate. The participants, who represented a wide range of experience levels (from recently hired drivers to those with over 20 years experience), completed a predictor battery that included the personality inventory used here. They were assured that there responses would be confidential, and that any reports generated would be at the aggregate summary level. Subsequently, performance data was obtained for 486 of the bus drivers (391 males and 95 females) from six of the properties. This group constituted the sample used in this thesis.

Measures. Personality was assessed using the HPI (Hogan & Hogan, 1995). This inventory is based on the five-factor model of personality, and is specifically designed for use in organisational settings. It contains 206 items that together define seven scales. Four of the scales (labelled *Adjustment*, *Likeability*, *Prudence*, and *Intellectance*) roughly correspond to the dimensions Neuroticism (reverse scored), Agreeableness, Conscientiousness, and Openness in the five-factor model. Extraversion is operationalised as two distinct scales in the HPI, namely *Ambition* and *Sociability*. The former captures the potency aspects of Extraversion (e.g., competitiveness, energy and leadership), whereas the latter relates to the desire to interact with others. Finally, the HPI contains a seventh scale, *School Success*, that assesses the extent to which an individual is interested in and has aptitude for academic activities (Hogan & Hogan, 1995). This scale is most closely related to the Openness dimension. However, given that many of the items of School Success assess characteristics associated with cognitive ability, such as the extent of the respondent's vocabulary and their ability to multiply large numbers quickly, this scale was not included in the present research. For the six scales included here, the number of items per scale are 37 for the Adjustment scale, 29 for Ambition, 24 for Sociability, 22 for Likeability, 31 for Prudence, and 25 for Intellectance. Each item is responded to as *true* or *false*, and scored as 1 or 0. Thus the possible scores for each scale range from 0 to the number of items per scale, where higher scores indicate greater levels of the personality attribute.

Performance was assessed using ratings provided by supervisors on nine items. As part of the original study a comprehensive job analysis had been conducted. Written material (e.g., formal job descriptions, training manuals and research articles), observations of bus drivers performing their job, interviews, and a job analysis survey had been used to identify the key responsibilities and tasks performed by bus drivers. This information had then been used to develop the nine behaviourally anchored rating scales that assessed performance on the key elements of the job. The scales assessed a) dependability, b) schedule adherence, c) safety, d) drive quality, e) attention to details, f) passenger interactions, g) service orientation, h) interactions with supervisors, and i) interactions with co-workers. The scores for each scale had been standardised within each of the six properties to a mean of 50 and a standard deviation of 10, where higher scores indicate better performance. In the present study the average of the standardised scores over the nine scales was used to operationalise performance.

Dataset 6: Professionals

Participants. This dataset was collected as part of a longitudinal study in a large Australian-based professional services company that had hired 228 recent university graduates. The graduates had started work between December 3, 2001 and July 1, 2002.

Upon joining the company they were provided with training relating to the firm's systems and procedures and the jobs that they would perform. They then worked with the company's clients in order to provide solutions to complex business issues related to corporate finance and accounting. Of the 228 graduates, 131 agreed to participate in the present study (response rate = 57%). An additional 11 participants were subsequently omitted due to missing performance data (eight had left the firm and no records were available for the other three), resulting in a final sample of 120 participants (64 females, 55 males, and 1 unspecified) with a mean age of 23.

Measures. Personality was assessed using the CPS-2 (Pryor & Taylor, 2000). This instrument was designed to measure five dimensions (labelled *Emotional Orientation*, *Social Orientation*, *Cognitive Orientation*, *Interpersonal Orientation*, and *Task Orientation*) that correspond to Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. It has the advantage of being brief (most test-takers require no more than 10 to 15 minutes to complete the inventory), and is specifically designed for use within organisational and vocational contexts. The instrument consists of 75 behavioural and attitudinal statements that are responded to on a 7-point scale with the labels *Never*, *Almost Never*, *Seldom*, *Sometimes*, *Often*, *Almost Always*, and *Always*. The items are scored from 1 to 7 (or 7 to 1 for reverse scored items). Measures of the five dimensions are obtained by summing the 15 items that correspond to each dimension. These scales are in the same direction as the NEO instruments. For example, higher levels of Emotional Orientation correspond to higher Neuroticism, and higher levels of Task Orientation correspond to higher Conscientiousness.

Information obtained as part of the company's formal performance appraisal system was used to operationalise work performance for this dataset. Specifically, each employee’s overall performance was rated by their supervisor on a five-point scale ranging from 1 (labelled *Exceptional*) to 5 (labelled *Below expectations*). For the purposes of the present research I reverse-scored the item so that higher scores indicated better performance.

Procedure. During July 2002 the graduates were mailed a packet consisting of a participant information statement and consent form, a reply-paid envelope, the personality questionnaire used in the present study, and other work-related questionnaires that were used as part of an unrelated study. The graduates were asked to participate in the research by completing the questionnaires and returning them to the researchers in the reply-paid envelope. They were assured that their responses would remain confidential. All graduates were identified only by code, and any information provided to the company was as a group summary. Consistent with the longitudinal nature of the study, supervisor ratings of performance were collected approximately twelve months after the personality questionnaire had been administered.

Preliminary Analyses

Prior to applying artificial neural networks to the data a number of preliminary analyses were carried out to evaluate the validity with which the intended constructs were being measured, to assess the reliability of the predictor and criterion scales, and to provide some descriptive statistics for the datasets.

Construct Validity

Personality variables. All the personality inventories used in this thesis were originally constructed with the aim of achieving content validity in relation to the five-factor model. That is, items were developed so as to provide an adequate sampling of the range of characteristics and traits that underlie the five factors. Furthermore, the construct validity of these inventories is supported by a good deal of previous research. For example ample evidence exists for the validity of the HPI and the original and revised NEO personality inventories; much of this information can be found in the test manuals for these instruments (see Hogan & Hogan, 1995; McCrae & Costa, 1992). Similarly, the test manual for the CPS-2 (Pryor & Taylor, 2000) and the associated bulletins (see www.congruence.com.au) provide some evidence for the validity of this instrument. This includes theoretically expected correlations with other tests that measure similar constructs, theoretically expected differences between a sample of rehabilitation clients and a sample of students, and an analysis of the factor-structure of the items.

There is also support for the construct validity of the IPIP-NEO scales. Correlations (corrected for unreliability) between the 30 lower-level scales of the IPIP-NEO and the corresponding NEO PI-R facet scales have been shown to range from .86 to .99 (Goldberg, 1999), thus suggesting that the same constructs are being assessed by the two inventories. Furthermore, Johnson (2000) factor-analysed the lower-level scales of a web-based version of this instrument and concluded that the scales generally loaded on the appropriate factors, although three of the scales (E4, O3, and C6) had primary loadings on factors other than the intended ones. In order to further evaluate the factor-structure of this inventory a principal components analysis with varimax rotation was conducted.[3] The scree plot associated with the analysis of the IPIP- NEO facet scales provided support for the retention of five components (see Figure 3.1), which together accounted for 64.9% of the variance in the data. Table 3.3 presents the loadings of each scale on the five varimax-rotated components. Taking into account the loadings, it is

---

[3] Principal components analysis with varimax rotation was used because it is consistent with the procedures used in the test manuals of the NEO PI-R and HPI, and Johnson's (2000) factor analysis of the IPIP NEO. However, the analysis was also reconducted using the principal axis method of factoring, and the direct oblimin method of rotation, and the results were essentially the same.

clear that the components correspond to Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness. All but two of the facets had their primary loading on the intended component. The exceptions were Activity Level (E4) and Emotionality (O3), which loaded primarily on the Conscientiousness and Neuroticism components, with secondary loadings of .37 and .41 on the intended components. Thus, the findings are very similar to those obtained by Johnson's (2000) analysis of this inventory, and support the validity of the IPIP-NEO as an instrument for assessing the five factors.

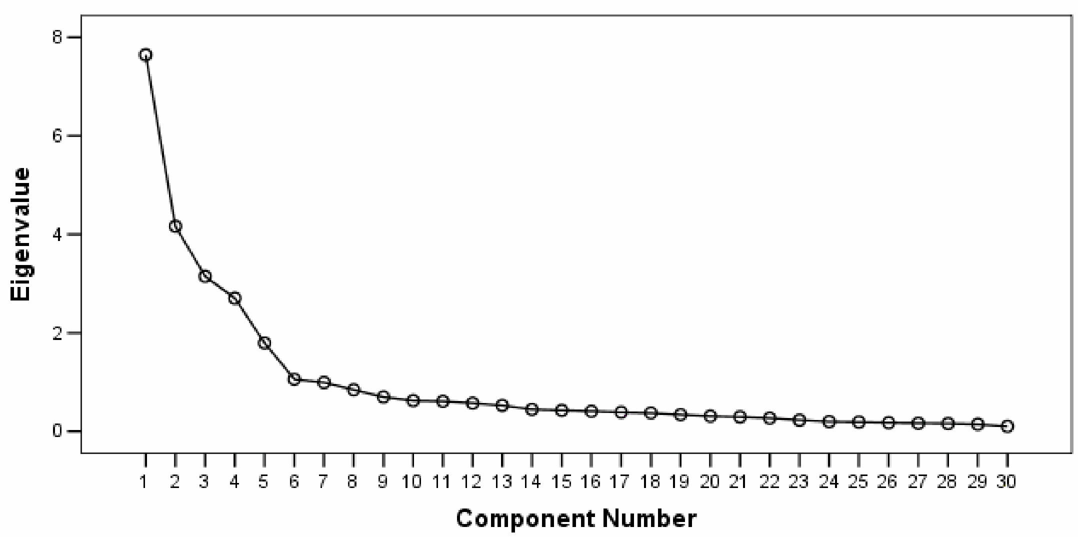


Figure 3.1

Scree plot of the eigenvalues associated with the principal components analysis of the IPIP-NEO subscales.

Table 3.3

Rotated component (RC) loadings of the IPIP-NEO scales for a five-factor solution.

Facet scale	RC1	RC2	RC3	RC4	RC5
N1: Anxiety	**<u>.81</u>**	-.30	-.02	-.01	-.20
N2: Anger	**<u>.81</u>**	-.07	.06	-.21	-.06
N3: Depression	**<u>.79</u>**	-.24	-.30	-.08	.02
N4: Self-Consciousness	**<u>.56</u>**	<u>-.55</u>	-.20	.18	-.20
N5: Immoderation	**<u>.44</u>**	.33	-.30	-.17	.22
N6: Vulnerability	**<u>.87</u>**	-.15	-.23	.05	-.17
E1: Friendliness	-.19	**<u>.79</u>**	.09	.29	.00
E2: Gregariousness	-.08	**<u>.88</u>**	.02	.02	.04
E3: Assertiveness	-.15	**<u>.64</u>**	<u>.44</u>	-.28	.25
E4: Activity Level	.00	.37	**<u>.62</u>**	-.03	.05
E5: Excitement-Seeking	-.10	**<u>.66</u>**	-.09	-.21	<u>.43</u>
E6: Cheerfulness	-.34	**<u>.65</u>**	-.01	.17	.26
O1: Imagination	.22	.07	-.03	-.05	**<u>.66</u>**
O2: Artistic Interests	.01	.17	.07	.22	**<u>.70</u>**
O3: Emotionality	**<u>.57</u>**	.09	.25	.28	<u>.41</u>
O4: Adventurousness	-.40	.33	.03	.00	**<u>.57</u>**
O5: Intellect	-.25	.01	.22	-.06	**<u>.78</u>**
O6: Liberalism	-.16	.08	-.25	.08	**<u>.48</u>**
A1: Trust	-.32	<u>.40</u>	-.01	**<u>.49</u>**	.04
A2: Morality	-.03	-.12	.26	**<u>.73</u>**	.00
A3: Altruism	-.06	.38	.20	**<u>.70</u>**	.22
A4: Cooperation	-.34	-.09	.01	**<u>.69</u>**	-.12
A5: Modesty	.25	-.32	-.24	**<u>.55</u>**	-.03
A6: Sympathy	.14	.13	-.04	**<u>.73</u>**	.16
C1: Self-Efficacy	<u>-.52</u>	.20	**<u>.63</u>**	.04	.27
C2: Orderliness	.21	-.25	**<u>.57</u>**	.00	-.26
C3: Dutifulness	-.11	-.10	**<u>.60</u>**	<u>.52</u>	-.08
C4: Achievement-Striving	-.06	.13	**<u>.84</u>**	.02	.15
C5: Self-Discipline	-.22	.04	**<u>.78</u>**	.13	.00
C6: Cautiousness	-.30	<u>-.51</u>	**<u>.55</u>**	.12	-.11

Note: Loadings over .40 in absolute magnitude are underlined. Boldface indicates the primary loading of each facet.

Performance measures. Principal components analysis was also applied to some of the performance measures used in this thesis. The information required to perform this analysis was not available for Dataset 2. Furthermore, Dataset 6 consisted of only one performance item and therefore principal components analysis was not applicable. In Datasets 1, 3, and 5 only a single component had an eigenvalue greater than 1 and the scree plots all supported the retention of one component. In Dataset 1 the first principal component accounted for 51% of the variance in the five assessment tasks. The loadings were .80 for the during-semester exam, .75 for the assignment, .72 for the research report, .72 for the final exam, and .56 for the field study. In Dataset 3 the first principal component accounted for 74% of the variance in the eight performance items. The loadings were .86 for the learning and applying knowledge item, .85 for the demonstrating responsible work habits item, .88 for the work-related communications item, .85 for the interpersonal skills item, .84 for the customer interaction item, .83 for the teamwork item, .84 for the problem-solving item, and .93 for the overall performance item. In Dataset 5 the first principal component accounted for 60% of the variance in the nine performance items. The loadings were .56 for the dependability item, .66 for the safety item, .82 for the drive quality item, .77 for the interactions with coworkers item, .80 for the interactions with supervisors item, .82 for the interactions with passengers item, .79 for the attention to details item, .83 for the schedule adherence item, and .87 for the service orientation item. These results suggest that within each dataset the performance items are tapping into one underlying dimension; this provides a rationale for combining the individual performance elements within these datasets in order to obtain measures of overall performance. As previously noted, for Dataset 1 each student’s overall mark for the course was used, and for Datasets 3 and 5 the average supervisor rating over all performance items was used.

The scree plot relating to the principal components analysis of the performance items in Dataset 4 also supported the retention of one component (see Figure 3.2), however there were two principal components with eigenvalues greater than 1. Table 3.4 compares the loadings for the one- and two-component solutions. For the latter case the results of an oblique rotation (the direct oblimin method) are presented here as the two components were correlated. The one-component solution (which accounted for 46% of the variance) produced loadings greater than .40 for all ten items on the component. The two-component solution (which accounted for 57% of the variance) contained one component that was defined by items concerned with job-related knowledge (items 1 and 6), and another component that concerned job-related achievement in general. Given the strong correlation between the two components (r=.44), and the desire to avoid performance scales consisting of only two items, all ten items were averaged to obtain the performance criterion for this dataset.

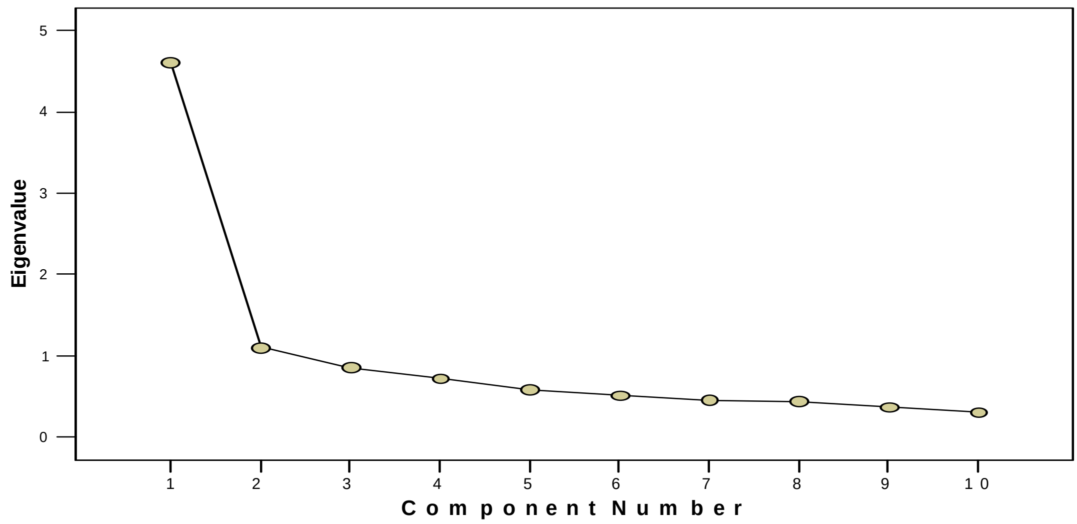


Figure 3.2

Scree plot of the eigenvalues associated with the principal components analysis of the performance items for Dataset 4.

Table 3.4

Loadings of the Dataset 4 performance items for one- and two-component solutions.

Item	One-component solution	Two-component solution	
	I	I	II
1. Has superior knowledge of the business.	**.62**	.18	**.66**
2. Meets company goals and expectations for the position.	**.77**	**.82**	-.03
3. Uses the complexity of the job to produce innovative outcomes.	**.66**	**.56**	.19
4. Takes calculated entrepreneurial risk.	**.73**	**.66**	.14
5. Consistently drives for better outcomes.	**.76**	**.89**	-.15
6. Has broad knowledge of political, economic, and technological issues.	**.48**	-.14	**.93**
7. Demonstrates independence and initiative.	**.73**	**.84**	-.11
8. Demonstrates confidence in the face of ambiguity.	**.72**	**.49**	.38
9. Is professionally competent.	**.67**	**.68**	.02
10. Could effectively handle the most senior position in the company.	**.61**	.36	**.39**

Note: Loadings over .40 in absolute magnitude are underlined. Boldface indicates the primary loading of each subscale.

Reliability

Reliability refers to the consistency of measurement (Cohen et al., 1996). It is important in the present context because measurement error in the predictors or the criterion limits the predictive performance that can be achieved by prediction equations (Sarle, 2001c). Two of the most commonly used indices of reliability are internal consistency and test-retest reliability. The former is an index of the homogeneity of the items in a scale, whereas the latter assesses the consistency of measurement from two administrations of the scale (Anastasi, 1988).

Personality variables. The reliability of the original and revised NEO PI is well established. A summary of the studies that have found evidence for the test-retest reliability of these instruments is presented in the test manual (Costa & McCrae, 1992). The internal consistency of the Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness scales, calculated as coefficient alpha, has been shown to be .92, .89, .87, .86, and .90 for the NEO PI-R (Costa & McCrae, 1992). Of the present datasets that used the NEO PI or NEO PI-R, the corresponding coefficient alphas were .84, .72, .71, .78, and .85 for Dataset 3, and .88, .87, .88, .85, and .90 for Dataset 4.[4] Taking the conventionally used value of .70 as a rule of thumb for minimum levels of adequate reliability (e.g., Nunnally, 1978), each of the above scales exceeds this value.

Similarly, the scales of the HPI have also been shown to be reliable. The test manual (Hogan & Hogan, 1995) reports internal consistency (coefficient alpha) and four-week test-retest reliabilities of .89 and .86 for the Adjustment scale, .86 and .83 for

---

[4] Dataset 2 also employed the NEO PI-R, however scores at the item level were not available and therefore it was not possible to calculate alpha coefficients for this dataset. Furthermore, for the same reason, alpha coefficients could not be calculated for the HPI scales employed in Dataset 5.

the Ambition scale, .83 and .79 for the Sociability scale, .78 and .83 for the Intellectance scale, .71 and .80 for the Likeability scale, and .78 and .74 for the Prudence scale.

Goldberg (1999) reported alpha coefficients for the 30 lower-level scales of the IPIP-NEO that ranged from .71 to .88. In the present study, the alpha coefficients for the higher-level scales measuring Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (calculated using Dataset 1) were .95, .93, .89, .88, and .93. These results can be taken as evidence for the reliability of the IPIP-NEO scales.

The test manual of the CPS-2 (Pryor & Taylor, 2000) reports coefficient alphas of .92, .88, .88, .75, and .83 for the Emotional Orientation, Social Orientation, Cognitive Orientation, Interpersonal Orientation, and Task Orientation scales. The corresponding figures, calculated using the sample of professionals (Dataset 6) in the present research, were .85, .78, .82, .67, and .80. Based on these results it seems that the coefficient alpha of the Interpersonal Orientation scale is lower than that of the other scales, and borders on what is conventionally considered adequate reliability. The other scales, however, display adequate reliability.

Performance measures. Alpha coefficients were also obtained for the performance measure in each dataset. The alpha coefficients were .58 for the performance measure in Dataset 1, .72 for Dataset 2, .95 for Dataset 3, .86 for Dataset 4, and .91 for Dataset 5.[5] The relatively low alpha coefficients associated with the performance measure of Dataset 1 and to a lesser extent of Dataset 2 are not surprising. The elements of these analyses consist of objective performance tests that were collected under different formats (e.g., multiple-choice tests, written reports etc.) and at different points in time, and would therefore be expected to be more heterogeneous than the self-report items that comprised the performance measures for the other datasets.

[5] The performance measure in Dataset 6 consisted of a score on a single item. Therefore an alpha coefficient could not be calculated.

Overall the results provide some support for the reliability of the work performance criteria used in the present research.

Descriptive Statistics

Table 3.5 presents the means, standard deviations, and minimum and maximum observed values for the variables in the six datasets. The following points are noted. First, mean scores for the variables measuring Neuroticism tended to be lower than mean scores for measures of the other five dimensions. This pattern is consistent with the pattern of means reported in the relevant test manuals. Second, there were a number of instances where the means of the personality variables in the present datasets were substantially different to the corresponding means in the normative distributions (as reported in the relevant test manuals), although for the most part the differences were consistent with what would be expected given the nature of the occupations involved. For example, the police recruits, flight attendants, and managers scored substantially higher on Extraversion and lower on Neuroticism relative to the respective normative samples, which is consistent with the idea that these jobs are more likely to attract individuals who enjoy interacting with others and who are emotionally stable. Third, the standard deviations of the personality variables in the current datasets were often slightly smaller than the corresponding standard deviations in the normative samples. This is not surprising given that individuals within an occupation are likely to be more homogeneous than normative samples, which contain individuals from various occupational backgrounds. Nevertheless, even within the current samples there was a good deal of variability in personality scores, as can be seen from the broad range of observed scores represented within each variable. Similarly, there was a substantial range of individual differences on the performance criteria.[6]

Table 3.5

Means, standard deviations, and minimum and maximum observed values for the predictor and criterion variables in each dataset set.

Dataset and variable	Mean	SD	Min	Max
1. University students				
Neuroticism	179.2	34.5	104	269
Extraversion	202.6	29.0	115	273
Openness	215.7	23.6	156	278
Agreeableness	217.2	21.2	157	276
Conscientiousness	207.4	27.6	133	270
Performance	62.8	9.1	37	93
2. Police recruits				
Neuroticism	74.6	18.5	5	121
Extraversion	123.5	15.4	81	165
Openness	112.0	15.8	65	166
Agreeableness	124.0	14.7	77	169
Conscientiousness	125.1	18.7	65	187
Performance	832.9	41.8	711	941
3. Flight attendants				
Neuroticism	65.0	19.2	19	129
Extraversion	129.2	15.3	65	166
Openness	123.9	15.5	76	178
Agreeableness	53.8	7.1	31	71
Conscientiousness	52.7	8.1	16	70
Performance	7.1	1.3	2.1	9.0
4. Managers				
Neuroticism	65.9	15.9	20	119
Extraversion	120.8	16.1	66	160
Openness	117.4	16.1	78	163
Agreeableness	118.0	14.7	68	163
Conscientiousness	130.7	16.5	76	171
Performance	3.87	0.6	1.5	5.0
5. Bus drivers				
Adjustment	25.2	6.6	5	37
Ambition	22.6	4.8	5	29
Sociability	11.9	4.3	1	23
Intellectance	14.1	4.6	2	25
Likeability	19.0	2.6	6	22

[6] In addition to the analyses presented here, Appendix A provides correlation matrices that summarise the correlations between the variables in each dataset.

Prudence	19.5	4.3	2	30
Performance	50.0	7.7	26.6	70.5
<u>6. Professionals</u>				
Emotional Orientation	46.8	10.8	27	75
Social Orientation	73.0	9.0	41	99
Cognitive Orientation	73.2	10.3	47	95
Interpersonal Orientation	73.4	7.5	45	90
Task Orientation	77.8	9.2	53	99
Performance	2.5	0.7	1	4

Summary

The present chapter provided a description of the six datasets that were used in this thesis. Within each dataset a personality inventory that assessed personality variables within the five-factor framework was identified, and a criterion variable assessing work performance was defined. Across the six datasets there were a total of 31 personality variables that could be grouped according to the dimension of the five-factor model that was being assessed.[7] The construct validity and reliability of the personality and performance variables were evaluated, and descriptive statistics were provided for these variables. The chapters that follow report the results of studies that employed these datasets to compare artificial neural networks and linear regression.

[7] The personality inventory for dataset five contained six personality variables and the other five datasets contained five variables each.

CHAPTER 4: Analyses Using Narrow Personality Variables

Introduction

As an introduction to the studies reported in this chapter it is useful to recapitulate the main arguments from previous chapters. In chapter 1 it was argued that the five-factor model has provided researchers with a useful framework for developing hypotheses about the validity of personality variables for predicting work performance. However, the results of empirical studies have not always supported these hypotheses, or have yielded validity coefficients that are low in magnitude (e.g., Hurtz & Donovan, 2000). The low or zero validity coefficients for personality variables that are theoretically expected to predict performance may be due to the presence of nonlinear or configural relationships that cannot be detected by the linear methods traditionally used in this domain. Several researchers have provided conceptual arguments for different forms of complexity in personality-performance relationships (e.g., Murphy, 1996; Witt, 2002), however the existing theories of work performance provide little guidance on the exact functional relationships that might be expected.

Chapter 2 introduced artificial neural networks as a method that has the flexibility to detect many different forms of complex relationships between input (predictor) and output (criterion) variables, and may therefore be useful in situations where the researcher makes few assumptions about the nature of the underlying relationships. However, implementing the method is more complex than conducting the traditional linear analyses, and requires various choices to be made regarding the training and testing of networks. Furthermore, the greater representational capability of neural networks does not guarantee that they will detect the systematic relationships that underlie a dataset, or that they will produce more accurate predictions than prediction equations developed using the simpler linear regression method. The choice of the best method is contingent on factors related to the nature of the data within the domain of interest – such as the presence of complex relationships, the noisiness of the data, and the number of cases available – and is therefore an empirical question.

The studies reported in this chapter and the next compared the predictive performance of artificial neural network and linear regression prediction equations using measures of the five factors as predictors of work performance. The six datasets presented in chapter 3 were used in order to provide a comparison of the two methods across a range of occupations, personality inventories, and work performance measures. The present chapter considers the specific case where each personality variable is included in the prediction equation separately.[1] First, a set of hypotheses about the expected differences between the predictive performance of the neural network and linear regression methods is generated. Following this, I report the findings from a study that was designed to provide a test of the hypotheses. Nonlinear relationships between personality variables and performance were also examined as part of the analyses. A second study is then reported in which linear regression is compared to an alternative procedure for developing neural networks. The chapter concludes with a summary of the main findings.

---

[1] The more general case of prediction equations that contain multiple predictors is considered in chapters 5 and 6.

Hypotheses

Based on the arguments for nonlinearity in personality-performance relationships presented in chapter 1, the present research tested the hypothesis that neural networks would produce more accurate predictions of work performance than linear regression equations when personality variables are used as the predictors. However it was not expected that neural networks would outperform linear regression equations for all personality variables. Specifically, it was not expected that neural networks would confer an advantage over linear regression when the personality variable was not theoretically relevant to performance in the task under consideration.[2] This is because if no relationship exists between the personality variable and work performance then all the variability in the training set represents noise and any attempt to fit a prediction equation (linear or otherwise) to the training data will result in overfitting. In this situation a neural network is likely to be less accurate than linear regression as it has greater capacity for representing the noise in the data.

Consequently, as a first step towards hypothesis development it was important to specify whether each personality variable was theoretically relevant to performance in the job under consideration. It has been suggested that Neuroticism (when conceptualised in terms of it opposite pole Emotional Stability) and Conscientiousness are important factors for performance in all jobs as they tap into motivational variables that are required for accomplishing all work tasks (Barrick et al., 2003). Thus, it was expected that measures of these two factors would be relevant in all six datasets. Openness assesses attributes that are associated with

[2] For the present purposes when the term *theoretically relevant* is applied to a predictor it is used to indicate the expectation that the predictor contains information pertaining to the criterion, and consequently that there exists some mathematical function relating scores on the predictor to scores on the criterion.

positive attitudes towards learning experiences (Barrick & Mount, 1991), and is therefore important in situations where performance is contingent on learning new skills or acquiring knowledge. For the present datasets Openness was expected to be relevant for performance in Datasets 1, 2, and 3 (which consisted of participants at university or on training programs), and in Dataset 6 (which consisted of recent graduates who had only just started their jobs). Extraversion is likely to be important for individuals in jobs with a strong social component, especially those involved with mentoring, leading or persuading (Barrick et al., 2001), such as the managers in Dataset 4. Extraversion is also relevant on training programs that are highly interactive, such as police academy training (Dataset 2) and flight attendant training (Dataset 3) among others (Barrick & Mount, 1991). Finally, Agreeableness is relevant for jobs with strong cooperative demands (Barrick et al., 2003), such as jobs that involve high levels of teamwork or the provision of customer service. Among the present datasets, customer interaction was an explicit component of the performance measure for the flight attendants (Dataset 3) and service orientation was explicitly assessed as part of the performance ratings for the bus drivers (Dataset 5). Following directly from the above arguments, Table 4.1 provides a summary of the personality variables that were specified as theoretically relevant for performance in each dataset. A cross (×) indicates that the predictor was not specified as theoretically relevant for that dataset. A tick (✓) indicates that the predictor was specified as theoretically relevant for the dataset.

Table 4.1

The personality variables specified as theoretically relevant in each dataset.

Dataset	N	E	O	A	C
1. University students	✓	×	✓	×	✓
2. Police recruits	✓	✓	✓	×	✓
3. Flight attendants	✓	✓	✓	✓	✓
4. Managers	✓	✓	×	×	✓
5. Bus drivers	✓	×	×	✓	✓
6. Professionals	✓	×	✓	×	✓

✓ = Specified as theoretically relevant.
× = Specified as not theoretically relevant.

Note: For simplicity the labels N, E, O, A, and C were used to denote the predictors in each dataset even though the personality inventories in Datasets 5 and 6 use alternative labels for their measures of the five-factor personality variables. Furthermore, as the specification did not differ for the two scales that operationalise Extraversion in Dataset 5, namely Ambition and Sociability, no distinction is made between these two predictors in the table.

Theoretical relevance by itself does not justify the hypothesis that neural networks are likely to outperform linear regression equations for a particular predictor; it is also important that there are expectations of a nonlinear relationship between the predictor and the criterion. In chapter 1, arguments for possible nonlinear relationships between personality and performance were presented for all five of the personality factors for performance contexts in which they are relevant. For example it was suggested that being extreme on any personality variable that is relevant for the job (either too low or too high) can impair performance. Similarly, the arguments for nonlinearity that derive from theories of traitedness (see chapter 1) apply to personality traits in general (assuming that the trait is relevant to the behaviour being predicted) rather than to specific traits. Consequently I did not develop hypotheses about the specific theoretically relevant predictors for which neural networks are most likely to outpredict linear regression.

Study 1

The primary aim of Study 1 was to compare prediction equations that: a) were developed using either artificial neural networks or linear regression, and b) contained a measure of one of the five personality factors as the predictor. This facilitated an examination of the hypothesis that neural networks would outpredict linear regression equations when a theoretically relevant personality variable is used as the predictor. The findings can also be used to compare the predictive performance of the different personality variables. A second aim of the study was to explore the extent and nature of nonlinear relationships between personality variables and work performance.

Method

All six datasets described in chapter 3 were used in this study. There were three major steps involved in conducting the study, namely:

1. Partitioning the datasets into training and test sets.
2. Developing prediction equations.
3. Testing the prediction equations.

Partitioning the Datasets

A resampling procedure was used (see chapter 2 for a discussion of the merits of this approach). The total data in each dataset was randomly divided into a training set consisting of approximately two-thirds of the total data and a test set consisting of the remaining third of the data. This random partitioning of the data was repeated twenty times in order to yield twenty training/test set partitions for each dataset.

Developing the Prediction Equations

The training cases in each partition were used to develop linear regression and neural network prediction equations for each of the 31 predictors (six predictors in Dataset 5 and five predictors in each of the other datasets). The *STATISTICA Neural Networks* software package (StatSoft Inc., 1998) was used to implement both types of equations. Linear equations were developed using ordinary least squares regression. Neural networks were developed according to the following specifications and procedures: [3]

- *Architecture.* All neural networks were multilayer perceptrons with one input layer, one hidden layer, and one output layer. The input layer contained one unit that represented the relevant personality variable and the output layer contained one unit that represented the performance criterion. The number of hidden units in the hidden layer varied, as described below. The hyperbolic tangent activation function was used at each hidden unit, and the identity activation function was used at the output unit.
- *Number of hidden units.* I experimented with four different levels of hidden units. The actual number of hidden units at each level was determined as a function of the number of free parameters (weights and bias terms) relative to training cases, and therefore varied between datasets (see Table 4.2). Sarle (2001c) suggests that for the types of analyses conducted here one should choose the number of hidden units such that there are at least two training cases per free parameter. This ratio was used to determine the largest hidden unit level for each dataset. I also experimented with four, ten, and twenty times as many

[3] Much of the rationale for implementing neural networks according to these specifications and procedures was provided in chapter 2.

training cases as free parameters. The latter ratio yields approximately two or three hidden units for the majority of the present datasets, which is close to the minimum of one hidden unit required for a multilayer perceptron neural network. For simplicity, the four levels of hidden units are hereafter referred to as *H1* (the smallest networks – 20:1 ratio), *H2* (10:1 ratio), *H3* (4:1 ratio), and *H4* (2:1 ratio).

Table 4.2

The number of training and test cases per partition, and the number of hidden units at each hidden unit level, by dataset.

Dataset	Training cases	Test cases	Hidden units at each level (H1, H2, H3, H4)
1. University students	151	76	2, 5, 12, 25
2. Police recruits	191	95	3, 6, 16, 31
3. Flight attendants	203	102	3, 6, 16, 33
4. Managers	119	60	2, 4, 10, 20
5. Bus drivers	324	162	5, 10, 26, 53
6. Professionals	80	40	1, 2, 6, 13

- *Weight regularisation.* The STATISTICA Neural Networks software package uses the Weigend weight elimination method of weight regularisation (see Weigend et al., 1990). The Weigend penalty term, as implemented by the software, is given by $\lambda\sum_i w_i^2/(1+ w_i^2)$, where λ refers to the regularisation coefficient, and w_i refers to the value of each weight in the network. This term is added to the sum-of-squares error function during training. By penalising networks with larger weights, this method encourages networks with smaller weights and hence less curvature. Consequently this procedure provides some protection against overfitting the training data.

- *Training algorithm.* The backpropagation and conjugate gradient algorithms were used to train networks. Backpropagation was used during the first 100 iterations, and this was followed by iterations of conjugate gradient training. The backpropagation algorithm requires the specification of a learning rate parameter and a momentum parameter that together control the rate at which the weights are updated. The software default values (learning rate = 0.1, momentum = 0.3) were used. Training was stopped once the error function failed to decrease for 50 consecutive iterations, and the weights that resulted in the lowest error function were used to define the neural network.
- *Multiple random restarts.* As described in chapter 2, the process of training a neural network begins with the assignment of random weights. A characteristic of local search processes such as backpropagation and conjugate gradients is that for certain sets of initial weights the training process can get trapped in local minima in the error function and consequently produce a final set of weights that does not fit the training data well.[4] A common strategy to avoid this problem is to repeat the training process a number of times using different sets of randomly selected initial weights, and to select the network that produces the lowest training error (Sarle, 2001b). Following Zhang et al. (1999), the training process was repeated 15 times at each hidden unit level, and the set of weights that provided the best fit to the training data was selected as the operational network for that hidden unit level.

[4] Networks that contain a large number of hidden units relative to training cases are less likely to experience this problem as they contain few if any minima (see Reed & Marks, 1999).

In summary, prediction equations were developed for 31 predictors across 20 partitions. Within each partition the training data was used to develop five prediction equations: one linear regression equation and four neural networks (one network at each of four hidden unit levels referred to as H1, H2, H3, and H4). Thus, a total of 3,100 equations (31 predictors x 20 partitions x 5 equations per partition) were obtained that formed the basis of the subsequent analyses.

Testing the Prediction Equations

Each prediction equation was presented with the predictor scores of the relevant test cases, and these scores were used to generate predicted criterion scores for the test set. The predicted and observed criterion scores for the test set were used to derive two measures of predictive performance for each equation:

1. *Mean absolute error (MAE).* The difference between the predicted criterion score and the observed criterion score for each test case was used to define the prediction error of the equation in relation to that case. The absolute value of the prediction errors were averaged over the entire test set to obtain an MAE value for each equation. The higher the MAE, the lower the predictive performance.[5]
2. *Cross-validity coefficient.* A cross-validity coefficient was calculated for each equation by correlating the predicted and observed criterion scores on the test set. This provided a relational index of predictive performance (see chapter 2). Higher coefficients equate to greater predictive performance. Negative coefficients indicate poor predictive performance.

---

[5] Mean square error (MSE) was also derived for each equation, and the correlations between MAE and MSE were calculated for each predictor x prediction equation combination. All 155 correlations exceeded .68, and the mean of the correlation coefficients was .90. As a result of the strong relation between these two measures, and due to the advantages of MAE outlined in chapter 2, only MAE values were included in the subsequent analyses.

Data analysis

To formally test for statistically significant differences in predictive performance between neural networks and linear regression, two planned contrasts were tested for each of the 31 predictors. The first contrast tested the difference in MAE between the linear regression equations and the neural networks, where the results for the neural networks were averaged across the four hidden unit levels. The second contrast was identical to the first with the exception that the cross-validity coefficients were substituted as the measure of predictive performance. [6] Thus both contrasts compared the predictive performance of the twenty linear equations that were developed for the predictor in question to the predictive performance of the corresponding eighty neural networks. Each contrast was tested for significance at Type I error rates of $\alpha = .05$ and $\alpha = .01$. As there are five predictors in all but one of the datasets, the latter figure corresponds roughly to carrying out a Bonferroni adjustment for the number of predictors within each dataset.

Given that within each partition the linear equations and neural networks are developed using the same training set and evaluated using the same test, it is possible to test the above contrasts by computing a contrast score (linear - the average of the hidden unit levels) within each partition and to then conduct a t-test on the contrast scores from the twenty partitions (see Nadeau & Bengio, 2003). This is similar to conducting a paired samples t-test where each partition represents one observation. However, a difficulty in applying this test to the resampling procedure adopted above is that the overlap of the training sets (and usually also the test sets) across partitions creates

---

[6] The distribution of an observed correlation coefficient r is skewed when the population correlation ρ is non-zero, and consequently some researchers advocate the use of Fisher's Z transformation when averaging across r (e.g., Corey, Dunlap, & Burke, 1998; cf. Hunter & Schmidt, 1990). In the present research, the arithmetic operations and statistical tests involving cross-validity coefficients were conducted with both transformed and untransformed values. The results were essentially the same and therefore for simplicity only the untransformed results are reported.

unknowable dependencies between the partitions, and consequently violates the independence of observations assumption associated with traditional hypothesis testing procedures (Neal, 1998). Specifically, the standard error of the contrast will be underestimated, and the risk of a type I error will be increased. To remedy this, Nadeau and Bengio (2003) propose a corrected t-test procedure in which the standard error is estimated by multiplying the standard deviation of the contrast scores by $\sqrt{\{(1/J) + (n_2/n_1)\}}$, where √ refers to the square root function, J refers to the number of partitions, n_1 refers to the number of training cases per partition, and n_2 refers to the number of test cases per partition. They show that under the assumption that the predictive performance estimates of the methods under consideration depend on the number of training cases but not on the actual training cases themselves then the adjustment proposed above results in an unbiased estimate of the standard error of the contrast. When this assumption is not met (as is likely to occur with neural networks) then the standard error may be either underestimated, in which case the procedure will result in liberal inferences, or overestimated, in which case inferences will be conservative. However, Nadeau and Bengio conducted a simulation with artificial data that found any departures from the nominal type I error rate (either on the conservative or liberal side) tended to be small, and that good statistical power was achieved with J = 15 partitions (in the present study J =20). Therefore, in the present research this corrected t-test was used to conduct the significance tests when comparing neural networks and linear equations.[7]

[7] An Excel spreadsheet was developed to conduct the corrected t-test. A template of this spreadsheet can be found in Appendix B.

Results and Discussion

Artificial Neural Networks Versus Linear Regression

The comparison of the neural network and linear regression equations are discussed separately below for each of the two measures of predictive performance.

MAE. Table 4.3 presents the MAE values for the linear regression and neural network equations by predictor and dataset. The results are averaged across the twenty partitions within each dataset.[8] The values in boldface are averages across the predictors within each dataset. The first two columns of numbers provide the MAE values for the linear equations and the average of the four neural network hidden unit levels. The final four columns provide a breakdown of the neural network results by hidden unit level. Underlined values indicate that the neural network MAE was lower than that of the corresponding linear regression equations.[9]

It is clear from Table 4.3 that the linear regression equations generally outperformed the neural networks with respect to MAE. Comparing the bold values in the first two columns of Table 4.3, it can be seen that the linear equations produced lower MAE than the neural networks for all six datasets when the results were averaged across predictors. When each predictor was considered separately the linear equations outperformed the neural networks for 28 of the 31 predictors. Furthermore, the final four columns of Table 4.3 indicate that for the majority of predictors the linear equations outperformed all four hidden unit levels, including networks with only a small number of hidden units. Thus, the typically high MAE values of the neural network equations relative to the linear equations in the present datasets seems to generalise

[8] Appendix C provides the MAE value and cross-validity coefficient for every prediction equation that was developed as part of the studies reported in this thesis, as well as summary statistics (means and standard deviations) for the predictive performance measures across the twenty partitions in each dataset. The results for Study 1 and 2 can be found in Table C1 to Table C31 of Appendix C.

[9] In this and subsequent tables there were occasions where more decimal places than shown in the table were required to determine whether the neural networks outperformed the linear equations.

across hidden unit levels. Moreover, for most of the predictors MAE increased as the number of hidden units increased. For example, of the four hidden unit levels, the H1 networks produced the lowest MAE (averaged across predictors) in four of the datasets, and came second to the H2 networks in the remaining two datasets. Therefore, the increased representational capability associated with more hidden units was a disadvantage rather than an advantage for the present datasets.

In the few instances where the neural networks produced lower MAE than the linear regression equations the magnitude of the differences are best described as small. For example, in Dataset 1 the largest difference in favour of the neural networks occurred for the H1 networks that used Neuroticism as the predictor. The average MAE for these networks (MAE = 7.382) was 0.013 of a unit less than that obtained by the corresponding linear equations (MAE = 7.395). This represents a decrease of less than 0.2%, and is trivially small when one considers that the criterion in this dataset was a mark out of 100. Across all datasets the largest percentage decrease in MAE as a result of adopting neural networks occurred for the Intellectance measure in Dataset 5. In this case the H4 networks obtained an MAE that was approximately 1% smaller than that obtained by the corresponding linear equations. It should also be noted that the contrasts that tested for differences between the two types of equations using the corrected t-test procedure outlined in the Data Analysis section were all statistically nonsignificant.[10]

[10] The results of the corrected t-tests for all comparisons between neural networks and linear regression that were conducted in this thesis are presented in Appendix D. Refer to Tables D1 and D2 for the results pertaining to Study 1.

Table 4.3

MAE values for the linear regression (LR) and artificial neural network (ANN) equations, by dataset and predictor.

Dataset and predictor	LR	ANN	H1	H2	H3	H4
<u>1. University students</u>						
Neuroticism	7.395	<u>7.394</u>	<u>7.382</u>	<u>7.384</u>	7.403	7.406
Extraversion	7.289	7.292	7.291	7.297	<u>7.281</u>	7.300
Openness	7.329	7.362	7.344	7.358	7.370	7.378
Agreeableness	7.401	7.444	7.410	7.446	7.447	7.473
Conscientiousness	7.226	7.257	7.273	7.260	7.253	7.243
	7.328	**7.350**	**7.340**	**7.349**	**7.351**	**7.360**
<u>2. Police recruits</u>						
Neuroticism	32.39	32.61	32.57	32.62	32.61	32.63
Extraversion	32.53	32.57	32.62	32.63	<u>32.53</u>	<u>32.51</u>
Openness	32.97	33.35	33.28	33.36	33.39	33.35
Agreeableness	32.53	32.89	32.84	32.90	32.88	32.93
Conscientiousness	31.55	31.68	31.66	31.65	31.71	31.71
	32.40	**32.62**	**32.59**	**32.63**	**32.62**	**32.62**
<u>3. Flight attendants</u>						
Neuroticism	1.063	1.069	1.068	1.068	1.069	1.069
Extraversion	1.063	1.068	1.067	1.068	1.067	1.068
Openness	1.070	1.075	1.075	1.075	1.075	1.076
Agreeableness	1.061	1.075	1.072	1.075	1.076	1.077
Conscientiousness	1.063	1.064	1.064	1.065	1.064	1.065
	1.064	**1.070**	**1.069**	**1.070**	**1.070**	**1.071**
<u>4. Managers</u>						
Neuroticism	0.488	0.494	0.490	0.493	0.497	0.498
Extraversion	0.505	0.506	0.506	0.506	0.506	0.507
Openness	0.507	0.514	0.511	0.513	0.516	0.516
Agreeableness	0.509	0.514	0.509	0.515	0.516	0.516
Conscientiousness	0.480	0.483	0.482	0.483	0.483	0.484
	0.498	**0.502**	**0.500**	**0.502**	**0.504**	**0.504**
<u>5. Bus drivers</u>						
Adjustment	6.155	6.164	6.162	6.162	6.164	6.169
Ambition	6.140	6.152	6.151	6.150	6.152	6.154
Sociability	6.165	6.195	6.192	6.195	6.197	6.197
Intellectance	6.160	<u>6.105</u>	<u>6.113</u>	<u>6.104</u>	<u>6.106</u>	<u>6.097</u>
Likeability	6.146	6.162	6.161	6.159	6.166	6.161
Prudence	6.081	6.094	6.094	6.090	6.096	6.094
	6.141	**6.145**	**6.145**	**6.143**	**6.147**	**6.145**
<u>6. Professionals</u>						
Emotional Orientation	0.604	0.612	0.613	0.611	0.613	0.611
Social Orientation	0.609	0.613	0.611	0.611	0.614	0.615
Cognitive Orientation	0.596	0.606	0.602	0.606	0.607	0.609
Interpersonal Orientation	0.596	0.638	0.637	0.631	0.647	0.636
Task Orientation	0.613	<u>0.612</u>	0.614	<u>0.611</u>	<u>0.612</u>	<u>0.612</u>
	0.603	**0.616**	**0.615**	**0.614**	**0.619**	**0.617**

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Cross-validity coefficients. Table 4.4 presents the cross-validity coefficients for the linear regression and neural network equations by dataset and predictor. The results are averaged across the twenty partitions within each dataset. The same presentation format as in Table 4.3 is used. Once again the linear equations generally outperfomed the neural networks. Table 4.4 indicates that the linear equations had higher cross-validity coefficients than the average of the four hidden unit levels for 24 of the 31 predictors, and outperformed all four hidden unit levels for 22 of the predictors. Furthermore, the nine instances where at least one hidden unit level outperformed linear regression were spread across personality constructs and datasets, and occurred about as often for variables that were not theoretically relevant for the dataset as it did for those that were. In the cases where neural networks did obtain a higher average cross-validity coefficient than the associated linear equations the difference was typically small and rarely exceeded .05 of a coefficient. The only exception occurred for the Intellectance measure in Dataset 5 for which the neural networks obtained an average cross-validity coefficient ($r = .10$) that was .15 of a coefficient higher than that obtained by the linear regression equations ($r = -.05$). This was also the only contrast that yielded a statistically significant difference, corrected $t(19) = -4.63$, $p < .01$ (see Appendix D). However note that Intellectance was not a theoretically relevant predictor for this dataset and therefore it had not been expected that neural networks would outperform linear regression for this predictor.

Table 4.4

Cross-validity coefficients for the linear regression (LR) and artificial neural network (ANN) equations, by dataset and predictor.

Dataset and predictor	LR	ANN	H1	H2	H3	H4
1. University students						
Neuroticism	.02	.06	.06	.06	.06	.06
Extraversion	.16	.15	.15	.15	.15	.15
Openness	.13	.08	.12	.08	.07	.05
Agreeableness	-.08	-.09	-.08	-.09	-.10	-.10
Conscientiousness	.22	.20	.19	.20	.20	.20
	.09	**.08**	**.09**	**.08**	**.07**	**.07**
2. Police recruits						
Neuroticism	.13	.09	.10	.09	.09	.09
Extraversion	.16	.12	.11	.12	.12	.13
Openness	.05	.04	.03	.04	.04	.04
Agreeableness	.09	.04	.05	.05	.04	.03
Conscientiousness	.26	.25	.25	.25	.24	.24
	.14	**.11**	**.11**	**.11**	**.11**	**.11**
3. Flight attendants						
Neuroticism	.12	.08	.08	.08	.07	.07
Extraversion	.15	.13	.13	.13	.13	.13
Openness	.12	.09	.10	.10	.09	.09
Agreeableness	.17	.13	.14	.13	.13	.13
Conscientiousness	.10	.08	.09	.09	.09	.08
	.13	**.10**	**.11**	**.10**	**.10**	**.10**
4. Managers						
Neuroticism	.19	.13	.17	.14	.11	.11
Extraversion	.10	.14	.13	.14	.14	.14
Openness	-.07	-.02	-.06	-.03	.01	.00
Agreeableness	.01	-.09	-.06	-.10	-.10	-.11
Conscientiousness	.26	.25	.25	.25	.25	.24
	.10	**.08**	**.09**	**.08**	**.08**	**.08**
5. Bus drivers						
Adjustment	.08	.06	.06	.06	.06	.05
Ambition	.08	.07	.08	.07	.07	.07
Sociability	-.06	-.09	-.09	-.09	-.10	-.09
Intellectance	-.05	.10	.09	.10	.09	.10
Likeability	.09	.07	.07	.07	.07	.07
Prudence	.14	.14	.15	.15	.14	.14
	.05	**.06**	**.06**	**.06**	**.06**	**.06**
6. Professionals						
Emotional Orientation	-.04	-.01	-.03	-.02	.01	.00
Social Orientation	-.03	-.07	-.05	-.04	-.08	-.09
Cognitive Orientation	.10	.01	.05	-.01	-.01	.01
Interpersonal Orientation	.03	.02	.04	.02	.03	.01
Task Orientation	-.13	-.10	-.12	-.12	-.08	-.08
	-.01	**-.03**	**-.02**	**-.03**	**-.03**	**-.03**

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Training set correlations. A possible reason for the poor predictive performance of the neural networks may be the failure to learn the relationships that are present in the training set. Despite the large representational capability of neural networks, it is possible for the training process to get trapped in a local minimum in the error function and consequently produce a final set of weights that does not fit the training data well (Reed & Marks, 1999). To examine this possibility the correlation between predicted and observed criterion scores in the training set was used as an index of the fit of the network to the training data. The training set correlation coefficients, averaged across the twenty data partitions, are presented in Table 4.5. It can be seen that the neural networks provided a better fit to the training data than the linear equations for all 31 predictors and across all four hidden unit levels. Therefore, the typically poor test set performance of the neural networks cannot be attributed to the failure to learn the relationships present in the training set. Rather, it is more likely that the neural networks are performing poorly because they are overfitting the datasets. This is supported by the fact that there was a tendency for the training set correlations to increase as the number of hidden units increased, whereas the test set measures of predictive performance tended to decrease with increasing hidden units.

Table 4.5

Training set correlations for the linear regression (LR) and artificial neural network (ANN) equations, by dataset and predictor.

Dataset and predictor	LR	ANN	H1	H2	H3	H4
1. University students						
Neuroticism	.06	.14	.13	.14	.14	.14
Extraversion	.12	.15	.14	.15	.15	.15
Openness	.11	.13	.12	.12	.13	.13
Agreeableness	.06	.08	.07	.08	.08	.09
Conscientiousness	.19	.21	.21	.22	.22	.22
	.11	**.14**	**.13**	**.14**	**.14**	**.15**
2. Police recruits						
Neuroticism	.17	.19	.18	.19	.19	.19
Extraversion	.16	.23	.23	.23	.24	.24
Openness	.13	.16	.15	.16	.16	.16
Agreeableness	.12	.14	.14	.14	.14	.15
Conscientiousness	.28	.29	.28	.28	.29	.29
	.17	**.20**	**.20**	**.20**	**.20**	**.20**
3. Flight attendants						
Neuroticism	.12	.13	.13	.13	.13	.13
Extraversion	.15	.16	.16	.16	.16	.17
Openness	.16	.18	.17	.17	.18	.18
Agreeableness	.17	.19	.19	.19	.20	.20
Conscientiousness	.12	.13	.13	.13	.13	.13
	.15	**.16**	**.16**	**.16**	**.16**	**.16**
4. Managers						
Neuroticism	.21	.22	.21	.21	.22	.22
Extraversion	.11	.20	.19	.21	.20	.19
Openness	.04	.13	.08	.12	.16	.16
Agreeableness	.08	.12	.10	.13	.13	.13
Conscientiousness	.28	.29	.28	.29	.29	.29
	.14	**.19**	**.17**	**.19**	**.20**	**.20**
5. Bus drivers						
Adjustment	.09	.10	.10	.10	.10	.10
Ambition	.12	.12	.12	.12	.12	.12
Sociability	.03	.06	.05	.05	.06	.06
Intellectance	.03	.15	.14	.15	.15	.15
Likeability	.10	.12	.12	.12	.12	.12
Prudence	.14	.17	.17	.17	.17	.17
	.08	**.12**	**.12**	**.12**	**.12**	**.12**
6. Professionals						
Emotional Orientation	.10	.16	.12	.16	.19	.18
Social Orientation	.08	.10	.08	.09	.10	.11
Cognitive Orientation	.12	.19	.15	.17	.22	.22
Interpersonal Orientation	.09	.19	.14	.17	.22	.21
Task Orientation	.06	.12	.07	.11	.16	.16
	.09	**.15**	**.11**	**.14**	**.18**	**.18**

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Differences in Predictive Performance Between Predictors

The findings reported in Tables 4.3 and 4.4 can also be used to evaluate differences in predictive performance between the predictors within each dataset, and to therefore draw some conclusions about effectiveness of the different personality variables for predicting work performance. Conscientiousness generally emerged as the best predictor of work performance. Referring to first two columns of numbers in Tables 4.3 and 4.4, it can be seen that Conscientiousness produced the lowest MAE values and highest cross-validity coefficients in Datasets 1, 2 and 4 regardless of the type of prediction equation, with cross-validity coefficients of approximately .20 or higher under both the linear regression and neural network methods. Similarly, in Dataset 5 the Prudence scale (which was constructed with the Conscientiousness factor of the five-factor model in mind, Hogan & Hogan, 1995) obtained lower MAE and higher cross-validity coefficients than any other predictor under both methods of generating prediction equations.

The predictive performance of measures of the other factors varied by dataset and the results were generally consistent with the expectations that were outlined in Table 4.1. Neuroticism was the best predictor of work performance after Conscientiousness in Dataset 4 (the managers). Furthermore, under the linear method low though non-trivial cross-validity coefficients were observed for this factor in Datasets 2 and 3 (the police recruit and flight attendant trainee samples). Extraversion predicted work performance for the three samples that were collected on training programs (the university students, police recruits and flight attendant trainees), and also for the sample of managers. Openness obtained low non-trivial cross-validity coefficients in two of the samples collected on training programs (the university students and flight attendant trainees) though only under the linear regression method.

Agreeableness obtained the highest cross-validity coefficients in Dataset 3 (the flight attendants), although in most of the datasets it was a poor predictor of the work performance criterion.

Finally, it should be noted that in Dataset 6 the predictive performance of all the predictors was low. One possible reason for this is that the work performance criterion that was available for this dataset consisted of a single item and is therefore probably less reliable than the criteria used in the other datasets. Furthermore, prediction in this dataset occurred over a longer time span than in the other datasets and this too may have contributed to the lower levels of predictive performance.

Nonlinear Relationships Between Personality Variables and Work Performance

That the neural networks failed to outperform the linear regression equations for all but one predictor (namely Intellectance in Dataset 5) provides some indirect evidence against the idea that personality variables are nonlinearly related to work performance. Nevertheless, neural networks can perform poorly compared to linear regression even in the presence of systematic sources of nonlinearity, for example as might occur if the neural networks detect highly nonlinear (and sample-specific) relationships in the training set when the functions underlying the data are only mildly nonlinear. Consequently, as a further test of some specific forms of nonlinearity, polynomial regression was applied to the total data in each dataset to test for the presence of quadratic and cubic relationships. For each of the 31 personality measures two new variables were defined by raising the relevant personality variable to the power of two (in order to test the quadratic effect) and to the power of three (to test the cubic effect). Following Pedhazur (1997), the quadratic effect was tested by regressing the performance criterion on to the quadratic variable while holding the linear effect constant; and the cubic effect was tested by regressing performance on to the cubic variable while holding the linear and quadratic effects constant. Table 4.6 presents the change in R^2 associated with the linear, quadratic, and cubic components, and the sum of the change in R^2 associated with the quadratic and cubic components.[11]

Table 4.6 indicates that the evidence for nonlinear personality-performance relationships was not strong. For most of the predictors, the quadratic and cubic components together accounted for less than 1% additional variance above what was already accounted for by the linear component. The only statistically significant nonlinear effect was a cubic relationship between Intellectance and work performance in Dataset 5, $t(482) = 2.75$, $p < .01$. A graphical representation of the third degree polynomial equation (see Figure 4.1) suggests that at the lower end of Intellectance there is a relatively steep increase in work performance as Intellectance increases. Performance peaks then decreases slightly through the mid-range before increasing again at the higher end of the personality variable. This is similar to the type of nonlinearity suggested by theories of traitedness (see Sinclair et al., 1999) where personality is more strongly related to performance at the extremes than in the middle. Nevertheless, Intellectance had not been expected to be a relevant variable in this sample, and it is not clear why a low level of this variable (which assesses creativity, brightness, and interest in intellectual matters) has such a detrimental effect on the performance of bus drivers, some of whom had been on the job for over 20 years. Given the relatively large number of significance tests that were conducted, and that nonlinearity had not been hypothesised for this predictor, it is possible that the effect was obtained purely by chance.

[11] Complete results for the polynomial regression analyses are provided in Appendix E.

Table 4.6

Change in R^2 associated with the linear, quadratic, cubic, and quadratic + cubic terms.

Dataset and predictor	Linear	Quadratic	Cubic	Quadratic + Cubic
Psychology students				
Neuroticism	.003	.011	.001	.012
Extraversion	.017	.007	.000	.007
Openness	.015	.000	.000	.000
Agreeableness	.001	.001	.000	.001
Conscientiousness	.040**	.004	.004	.008
Police recruits				
Neuroticism	.026**	.001	.001	.002
Extraversion	.026**	.006	.012	.018
Openness	.011	.000	.008	.008
Agreeableness	.012	.001	.002	.003
Conscientiousness	.075**	.000	.001	.001
Flight attendant trainees				
Neuroticism	.014*	.000	.001	.001
Extraversion	.023**	.002	.000	.002
Openness	.022**	.002	.001	.003
Agreeableness	.030**	.000	.006	.006
Conscientiousness	.013*	.002	.000	.002
Managers				
Neuroticism	.039**	.000	.003	.003
Extraversion	.011	.017	.004	.021
Openness	.000	.005	.004	.009
Agreeableness	.006	.005	.000	.005
Conscientiousness	.076**	.000	.001	.001
Bus drivers				
Adjustment	.007	.000	.000	.000
Ambition	.011*	.001	.000	.001
Sociability	.000	.000	.000	.000
Intellectance	.000	.005	.015**	.020**
Likeability	.009*	.002	.002	.004
Prudence	.021**	.005	.001	.006
Professionals				
Emotional Orientation	.007	.014	.000	.014
Social Orientation	.005	.000	.000	.000
Cognitive Orientation	.016	.000	.022	.022
Interpersonal Orientation	.012	.005	.000	.005
Task Orientation	.000	.004	.000	.004

* $p < .05$, ** $p < .01$

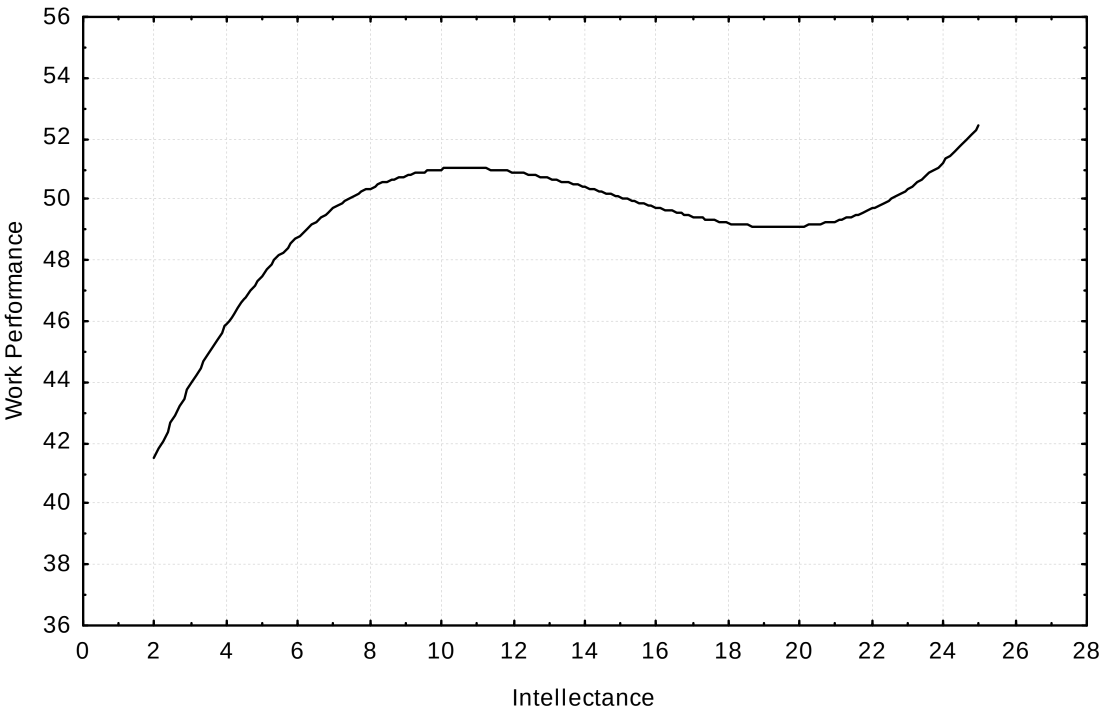


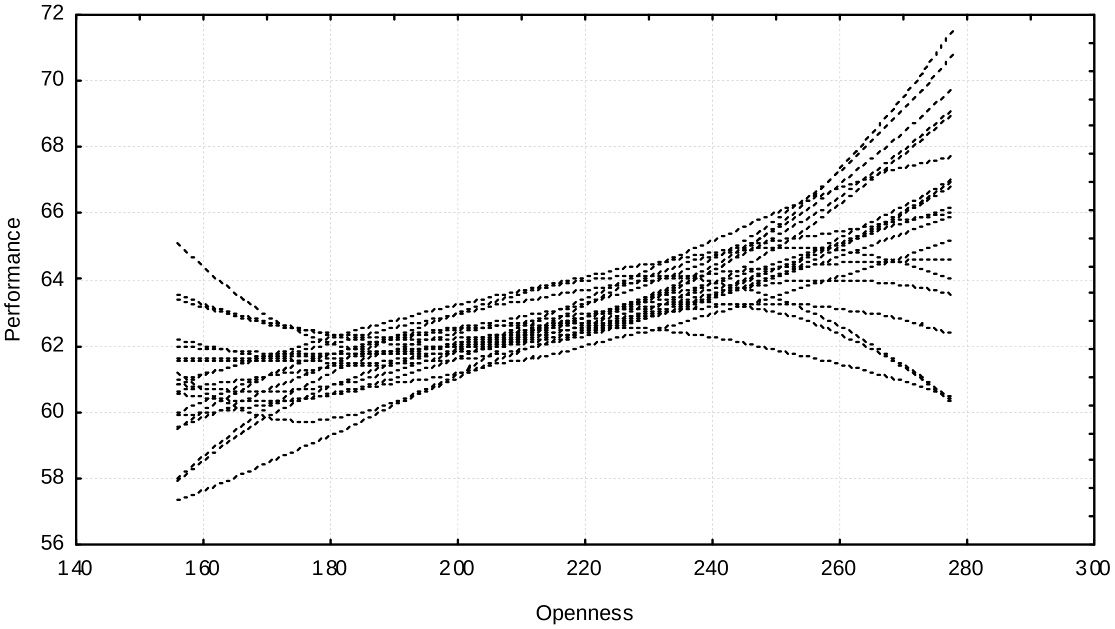
Figure 4.1

A graphical representation of the third-degree polynomial equation relating Intellectance to work performance in the bus driver sample.

In light of the relatively weak nonlinear structure of the data it is not surprising that the linear regression equations typically outperformed the neural networks. In the absence of a systematic nonlinear relationship between the predictor and the criterion any additional representational capacity of neural networks over linear regression will simply capture the noise in the training set and therefore result in lower predictive performance. To illustrate, Figure 4.2 depicts the relationships between the Openness scale and the performance criterion in Dataset 1 that were detected by the H4 networks (panel a) and the linear regression equations (panel b) from the training cases in the twenty partitions. The sum of the quadratic and cubic components for this predictor over the entire data was approximately zero. Figure 4.2 shows that the H4 networks are very sensitive to the particular training sets such that the functions detected across training sets show high levels of variability (see panel a). The linear equations also show some sensitivity to the training sets, though to a far lesser extent than the neural networks. The greater variability of the neural network functions arises directly as a result of their greater representational capability and has the result of decreasing the expected predictive accuracy for new cases (see Bishop, 1995; Geman et al., 1992).

However, the current results also provide some optimism about the benefits of neural networks in the presence of nonlinearity. Figure 4.3 depicts the relationships between Intellectance and the work performance criterion in Dataset 5 that were detected by the H4 networks and the linear regression equations. As described above, the data was characterised by a steep increase in work performance as one moves from low levels of Intellectance to the mid-range that then decreases slightly through the mid-range before increasing again at the upper extreme. Figure 4.3 shows that the neural networks were able to detect this relationship from the training sets (see panel a), whereas the linear regression equations – which are restricted to maintaining a constant gradient over the entire range of the predictor scores – could not (see panel b). Thus, in this case, the tendency of neural networks to represent noise in the data was more than offset by the benefits of being able to capture the systematic nonlinearity in the data, as reflected by the higher predictive performance of the networks. As further support for the benefits of neural networks in the presence of nonlinearity, note that the networks outperformed the linear regression equations with respect to cross-validity coefficients on four of the six occasions where the change in R^2 of the quadratic and cubic components of a predictor was more than .01 (67% success rate), compared to only three out of 25 occasions when it was less than .01 (12% success rate; refer to Tables 4.4 and 4.6).

a) H4 Neural Networks



b) Linear Regression Equations

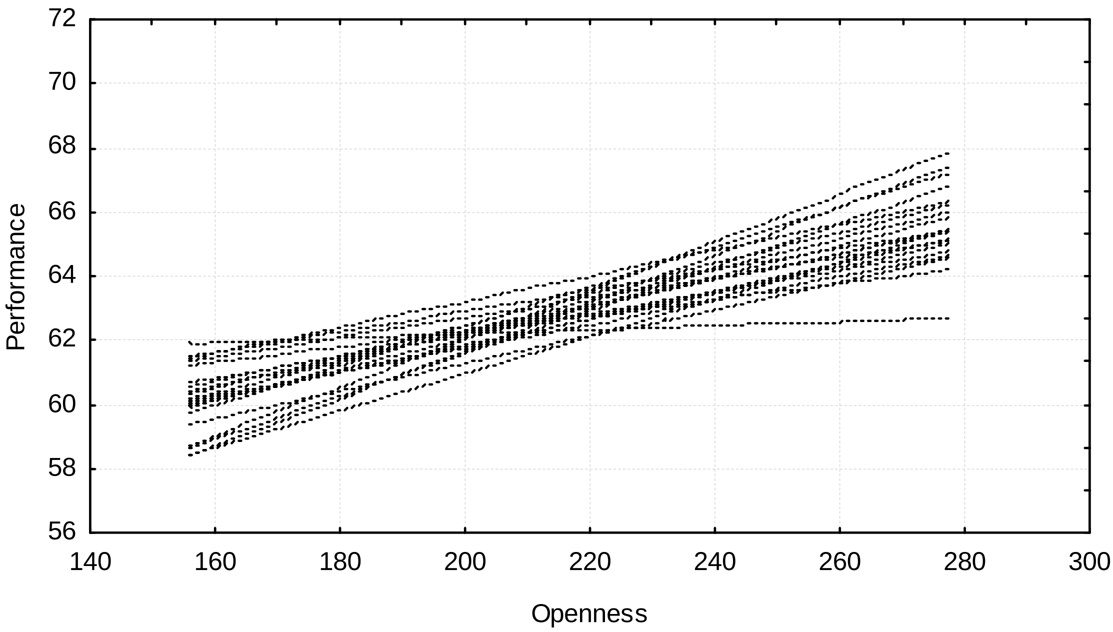
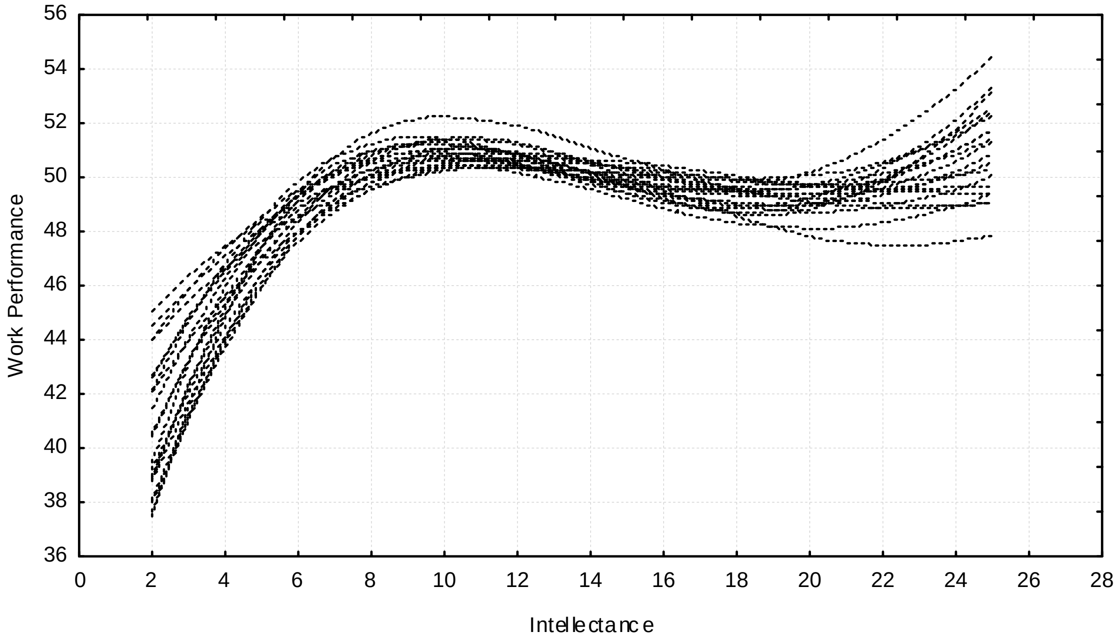


Figure 4.2

Relationships between Openness and performance that were detected by the twenty H4 neural networks (panel a) and twenty linear regression equations (panel b) in the university student sample (Dataset 1).

a) H4 Neural Networks



b) Linear Regression Equations

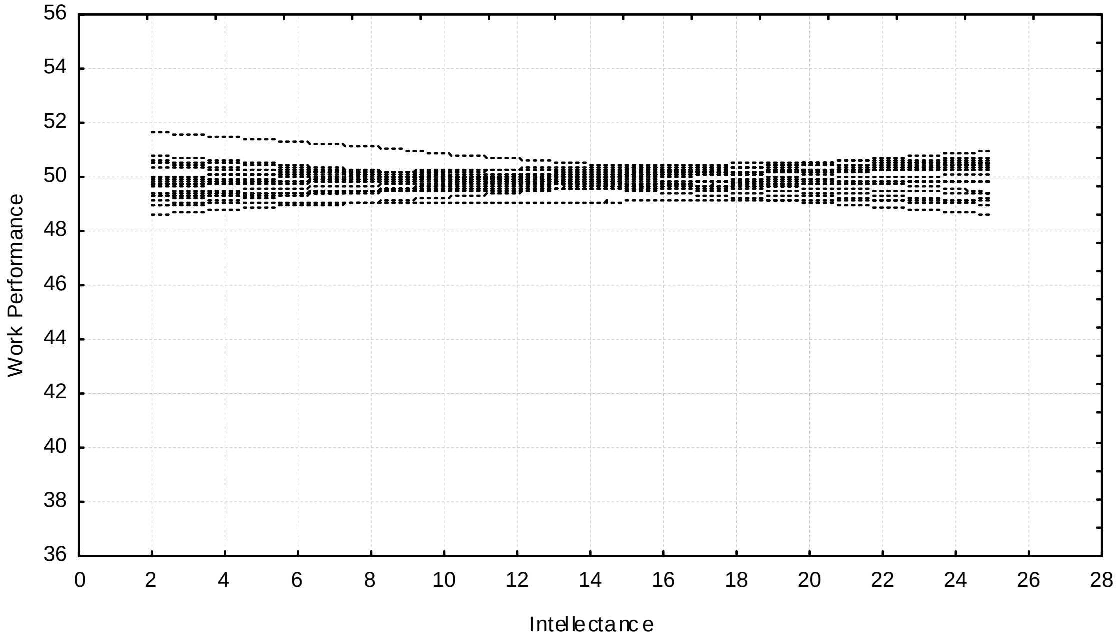


Figure 4.3

Relationships between Intellectance and work performance that were detected by the twenty H4 neural networks (panel a) and twenty linear regression equations (panel b) in the bus driver sample (Dataset 5).

To summarise, this study found little evidence that neural networks produce better predictions of work performance than linear regression equations when personality variables are used to predict work performance. The poor predictive performance of neural networks was attributed to the lack of systematic nonlinearity in the data; in this context the additional representational capability of neural networks over linear regression only serves to fit the noise in the data. Conversely, there was some evidence that neural networks could improve predictive performance, at least with respect to cross-validity coefficients, on the small number of occasions where there was some evidence for reliable nonlinearity in the data.

In Study 2, I experimented with a different procedure for developing neural networks, one that provides feedback during the training process about the extent to which the network is fitting the noise in the data, and thus enables training to be stopped once the network starts to overfit the training set. It was anticipated that if this could be successfully implemented then the neural networks would match the predictive performance of linear regression for the many instances where nonlinear relationships were absent, and outperform linear regression for the small number of instances that were characterised by nonlinearity. Consequently, the neural networks would be slightly more accurate than the linear equations on average. The results of this study are reported next.

Study 2

In Study 2 neural networks were developed using the early stopping procedure described in chapter 2, and the predictive performance of the networks was compared to that of the linear equations and neural networks developed in Study 1. It has been shown that the complexity of the relationships detected by neural networks increases as a function of the number of training iterations (Caruana et al., 2000). Hence, during the initial stages of training, the expected predictive performance of the network for unseen cases improves as the complexity of the network approaches the complexity of the underlying function; however as training progresses the network starts to fit the noise in the training data and consequently the network's expected performance for unseen cases deteriorates. The early stopping procedure monitors prediction error on a separate validation set during the training process and stops training once this error starts to rise. Thus, if the relationship underlying the data is linear then error on the validation set can be expected to rise once the network begins to fit the nonlinear idiosyncrasies of the training set and training is stopped. In this way the network will provide a very close approximation to a linear solution. On the other hand if there are systematic nonlinear components underlying the data then this too should be reflected in the validation set and error on the validation set should continue to fall until the network has fitted these components. One may therefore expect neural networks trained with early stopping to provide a close match to the predictive performance of linear regression when the underlying relationships are linear, and to exceed the predictive performance of linear regression when there are generalisable nonlinearities in the data.

Method

There are a number of issues that are relevant to the implementation of the early stopping procedure. First, early stopping requires some of the training data to be allocated to a validation set that is used to monitor prediction error during the training process, but which cannot be used to set the weights of the network. As a result the procedure has been criticised for its inability to use all the training data for training the network (e.g., Utans, 1997). This limitation can be partially mitigated by training multiple networks using different random divisions of the training data into training and validation sets, and taking the average prediction of the multiple networks as the predicted score for any new cases presented to the committee of networks (e.g., Varsta, Heikkonen, Millan, & Mourino, 2000).

A second issue relates to the proportion of cases that should be set aside for the validation set. Sarle (1995) used artificial datasets to experiment with proportions ranging from 10% to 50% of the total data available for training but did not obtain clear-cut results. Nevertheless, allocating 25% of the training data to a validation set provided good results across his datasets.

Third, a large number of hidden units should be used with the early stopping (Sarle, 1995). It has been shown that (given enough hidden units to model the underlying function) the test set performance of a neural network developed with early stopping varies little as a function of increasing hidden units, even when the number of free parameters far exceeds the number of training cases (e.g., Tetko et al., 1995). Moreover, the use of a large number of hidden units decreases the likelihood of the training algorithm converging to a poor local minimum (Reed & Marks, 1999).

Following from the above, all neural networks in this study were developed using the largest number of hidden units that were included in Study 1 (the H4 level). In order to facilitate comparisons with the equations developed in Study 1 the same twenty training/test partitions that were used in that study were also used here. However, each network was developed using only 75% of the training data to adjust the weights, and the remaining 25% of the training data was allocated to the validation set which provided feedback on when to stop training. Error for the validation cases was computed after each iteration, and training was stopped once this error failed to decrease (or increased) for 50 consecutive iterations. The weights that produced the lowest error on the validation set were used to define the resulting neural network. For each of the twenty partitions defined in Study 1 the above process was repeated 15 times using different random divisions of the training data into training and validation sets. The 15 resulting networks were then combined to form one committee for each partition and the predictive performance of the committee was assessed on the relevant test set (which was not involved in training in any way). Specifically, for each test case a predicted criterion score was generated by taking the average of the scores predicted by the 15 networks. The predictions were compared to the observed criterion scores for the test cases in order to derive an MAE value and cross-validity coefficient for each committee. This entire process was repeated separately for each of the 31 predictors.[12]

[12] All other specifications for developing the networks (such as the training algorithm, activation functions, and the use of weight regularisation) were the same as those used in Study 1. It could be argued that the use of weight regularisation is not necessary for this experiment as the early stopping procedure provides protection against overfitting. Nevertheless, there is some previous evidence that the performance of the early stopping procedure is enhanced when combined with weight regularisation (e.g., Finnoff et al., 1993). Furthermore, post-hoc analyses conducted on the present datasets suggested that network performance was impaired when weight regularisation was turned off. These analyses are not reported here.

Results and Discussion

Table 4.7 presents the MAE values and cross-validity coefficients for the early stopping neural network committees developed in this study (labelled ANN2). The results are averaged across the twenty partitions. For comparison the corresponding values for the Study 1 linear regression equations and neural networks are also presented (labelled LR and ANN1). Underlined values indicate that the neural networks outperformed the corresponding linear regression equations. Values that are marked by an asterisk indicate that the committees of early stopping neural networks outperformed the networks developed in Study 1.

It is clear from the number of asterisks in Table 4.7 that the use of the early stopping procedure and committee formation improved the predictive performance of neural networks for almost all predictors. Of the 31 predictors there were only 4 instances where MAE did not decrease and 6 instances where cross-validity coefficients did not improve. However, there were few predictors for which the early stopping committees were able to outperform the linear equations. Similar to the findings of Study 1, the network committees were more likely to outpredict linear regression when the predictor-criterion relationships were characterised by a certain degree of nonlinearity (e.g., Intellectance in Dataset 5). However, the network committees typically performed worse than the linear equations for the majority of predictors in which the quadratic and cubic components of the data were small. Consequently, when the results were averaged across the predictors in each dataset the linear equations obtained lower MAE values and higher cross-validity coefficients than the committees in five of the six datasets (see bold values in Table 4.7).[13]

[13] As in Study 1, the MAE and cross-validity coefficients of the network committees developed in this study were compared to the linear regression equations using the corrected t-test outlined in the Data Analysis section of Study 1. All differences were statistically nonsignificant (see Appendix D).

Table 4.7

MAE values and cross-validity coefficients for the linear regression equations and artificial neural networks generated in Study 1 (LR and ANN1), and artificial neural networks generated in Study 2 (ANN 2).

Dataset and predictor	MAE			Cross-validity coefficients		
	LR	ANN1	ANN2	LR	ANN 1	ANN 2
1. University students						
Neuroticism	7.395	7.394	7.379*	.02	.06	.05
Extraversion	7.289	7.292	7.275*	.16	.15	.17*
Openness	7.329	7.362	7.338*	.13	.08	.12*
Agreeableness	7.401	7.444	7.400*	-.08	-.09	-.08*
Conscientiousness	7.226	7.257	7.231*	.22	.20	.21*
	7.328	**7.350**	**7.324***	**.09**	**.08**	**.09***
2. Police recruits						
Neuroticism	32.39	32.61	32.42*	.13	.09	.12*
Extraversion	32.53	32.57	32.60	.16	.12	.14*
Openness	32.97	33.35	32.96*	.05	.04	.05*
Agreeableness	32.53	32.89	32.67*	.09	.04	.07*
Conscientiousness	31.55	31.68	31.60*	.26	.25	.25*
	32.40	**32.62**	**32.45***	**.14**	**.11**	**.13***
3. Flight attendants						
Neuroticism	1.063	1.069	1.066*	.12	.08	.10*
Extraversion	1.063	1.068	1.063*	.15	.13	.14*
Openness	1.070	1.075	1.072*	.12	.09	.11*
Agreeableness	1.061	1.075	1.067*	.17	.13	.16*
Conscientiousness	1.063	1.064	1.067	.10	.08	.09*
	1.064	**1.070**	**1.067***	**.13**	**.10**	**.12***
4. Managers						
Neuroticism	0.488	0.494	0.490*	.19	.13	.18*
Extraversion	0.505	0.506	0.504*	.10	.14	.14*
Openness	0.507	0.514	0.508*	-.07	-.02	-.06
Agreeableness	0.509	0.514	0.510*	.01	-.09	-.05*
Conscientiousness	0.480	0.483	0.482*	.26	.25	.26*
	0.498	**0.502**	**0.499***	**.10**	**.08**	**.09***
5. Bus drivers						
Adjustment	6.155	6.164	6.163*	.08	.06	.07*
Ambition	6.140	6.152	6.148*	.08	.07	.08*
Sociability	6.165	6.195	6.187*	-.06	-.09	-.09*
Intellectance	6.160	6.105	6.133	-.05	.10	.04
Likeability	6.146	6.162	6.151*	.09	.07	.04
Prudence	6.081	6.094	6.093*	.14	.14	.15*
	6.141	**6.145**	**6.146**	**.05**	**.06**	**.05**
6. Professionals						
Emotional Orientation	0.604	0.612	0.605*	-.04	-.01	-.01*
Social Orientation	0.609	0.613	0.608*	-.03	-.07	-.07
Cognitive Orientation	0.596	0.606	0.598*	.10	.01	.06*
Interpersonal Orientation	0.596	0.638	0.603*	.03	.02	.07*
Task Orientation	0.613	0.612	0.615	-.13	-.10	-.15
	0.603	**0.616**	**0.606***	**-.01**	**-.03**	**-.02***

Note: Underlined values indicate that the neural networks outperformed the associated linear equations. An asterisk (*) indicates that the ANN2 networks outperformed the corresponding ANN1 networks.

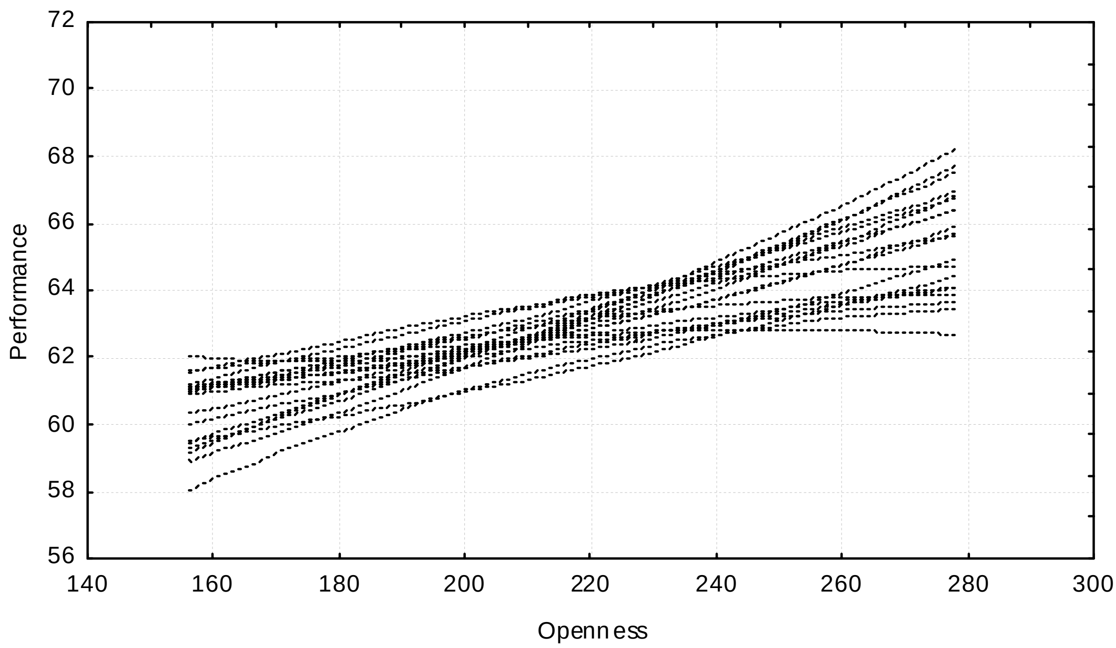


Figure 4.4

Relationships between Openness and performance that were detected by the twenty early stopping neural network committees in the university student sample (Dataset 1).

Figure 4.4 plots the relationships between the Openness scale and the performance criterion in Dataset 1 that were detected by the twenty neural network committees. Comparing this figure to Figure 4.2 (which plots the corresponding relationships detected by the H4 neural networks of Study 1) it can be seen that the use of early stopping and committee formation has reduced the complexity of the relationships that are detected, and that the networks are less likely to fit the noise of the training sets. However, the committees are still slightly more sensitive to the training sets than the linear equations (as reflected in the slightly greater variability and non-linearity of the functions detected) and this may explain the marginally poorer performance of the committees relative to the linear equations for this predictor.

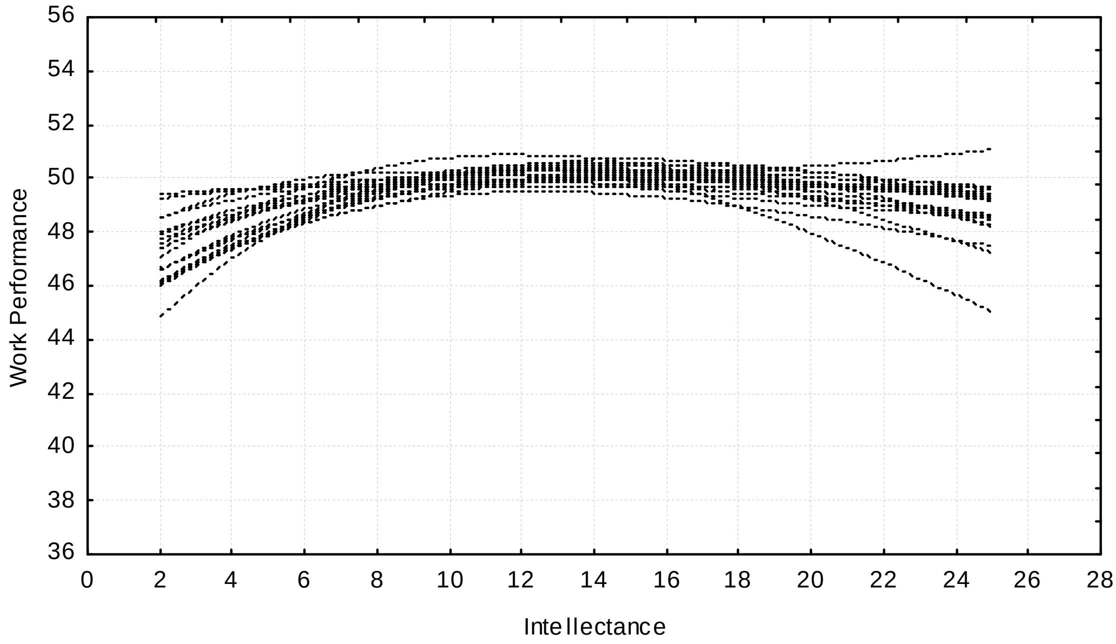


Figure 4.5

Relationships between Intellectance and work performance that were detected by the twenty early stopping neural network committees in the bus driver sample (Dataset 5).

Figure 4.5 plots the relationships between the Intellectance scale and the work performance criterion in Dataset 5 that were detected by the twenty early stopping neural network committees. While the committees detect some of the nonlinearity in the training sets, the relationships show far less curvature than the corresponding H4 networks from Study 1 (refer to Figure 4.3). In this case the greater simplicity of the functions detected by the network committees impairs their predictive performance as they fail to fully capture the steep increase in work performance at the lower extreme of the predictor or the slight increase at the higher end of the predictor. This is reflected in the higher MAE values and lower cross-validity coefficients of the committees relative to the Study 1 networks for this predictor.

Conclusions

The main findings resulting from the research reported in this chapter can be summarised as follows. First, it was found the for the majority of personality variables the neural networks failed to improve predictive performance compared to the predictive performance achieved by linear regression. This finding generalised across two measures of predictive performance, two different procedure for developing neural networks, and four hidden unit levels.

Second, for the majority of the datasets the predictive performance of the measure of the Conscientiousness factor was higher than that of the other factors that comprise the five-factor model. This occurred regardless of whether neural networks or linear regression was used to generate prediction equations.

Third, there was little support for nonlinear relationships between personality variables and work performance; the only personality variable for which a statistically significant nonlinear relationship was obtained had not been hypothesised to be relevant for work performance in the sample in which it was assessed.

Thus, the findings reported in this chapter provide little justification for the adoption of neural networks as a tool for deriving prediction equations for individual personality variables. However, in practice it is often useful to take into account the information provided by multiple predictors when attempting to predict a criterion variable, and it is in these types of situations that neural networks are typically implemented. Consequently, the next chapter examines the extent to which differences in predictive performance between artificial neural networks and linear regression generalise to instances in which multiple personality variables are included in the prediction equations.

CHAPTER 5: Analyses Using Combinations of Personality Variables

Introduction

In the previous chapter neural networks were compared to linear regression in the context of single-predictor equations. To the extent that different personality variables provide nonredundant information relating to the criterion, improvements in prediction can often be gained by considering equations that combine the information from multiple personality variables (e.g., Goldberg, in press). Therefore, to obtain a more complete picture of the usefulness of the neural network methodology for the types of datasets considered here, it is also important to compare neural networks to linear regression within the context of multiple-predictor equations.

A key determinant of the usefulness of the neural network method within the context of multiple predictors concerns the nature of the information provided by the predictors. On the one hand, if the underlying predictor-criterion relationships are linear and additive then this can be represented by a linear regression equation and any additional representational capacity provided by neural networks can only serve to fit the noise in the datasets. On the other hand, if the relationship between a particular predictor and the criterion depends on the level of other predictors then this cannot be represented by a linear additive equation; hence improvements in predictive performance may be realised by combining the predictor information using neural networks. In chapter 1 conceptual perspectives that were consistent with such configural relationships between personality variables and work performance were presented. Based on those arguments it was hypothesised that neural networks would produce more accurate predictions than linear regression for combinations of personality variables. The present chapter reports the findings from studies that were conducted to test this hypothesis.

An important consideration in conducting such studies involves deciding which combinations of personality variables to test. A number of researchers, both in the personnel psychology domain (e.g., Guion, 1998) and more generally in the behavioural sciences (e.g., Babyak, 2004), endorse an approach in which the predictors to be included are specified a priori based on their hypothesised relevance to the criterion being predicted. The advantage of such an approach is that it avoids the capitalisation on chance that occurs when many different combinations of predictors are considered (Cohen, 1990), or when predictors are pulled in and out of the equation based on their relation with the criterion (Babyak, 2004). On the other hand, it has been argued that it is wasteful to limit the analysis of data to that which has been preplanned (e.g., Tukey, 1969); and even among proponents of a hypothesis driven approach it is acknowledged that there is some value in going beyond the a priori analyses in order to further explore the data, although less confidence should be ascribed to such post-hoc analyses (e.g., Babyak, 2004). Accordingly, a two-phase approach was adopted for testing combinations of personality variables. In the first study reported in this chapter (labelled Study 3) prediction equations were developed and compared for the combination of personality variables that had been identified as theoretically relevant in each dataset. Configural relationships between personality variables and work performance were also examined as part of this study. The second study reported in this chapter (labelled Study 4) adopted an exploratory approach in which a predictor-selection procedure was used to empirically identify other combinations of the personality variables that were potentially useful for predicting work performance, and the neural network and linear regression methods were also compared for these combinations.

Study 3

The primary aim of this study was to compare linear regression and neural networks for the combination of personality variables that had been specified to be theoretically relevant for work performance in each dataset. These variables were identified in the previous chapter based on arguments presented in the literature, and are reproduced below in Table 5.1. Based on the conceptual arguments for configural relationships between theoretically relevant personality variables and work performance that were presented in chapter 1, it was hypothesised that neural networks would outperform linear regression for these datasets. A second aim of Study 3 was to empirically examine the extent and nature of configural relationships between personality variables and work performance.

Table 5.1

The personality variables in each dataset that had been specified to be theoretically relevant for work performance.

Dataset	Theoretically relevant personality variables
1. University students	Neuroticism, Openness, Conscientiousness
2. Police recruits	Neuroticism, Extraversion, Openness, Conscientiousness
3. Flight attendants	Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness
4. Managers	Neuroticism, Extraversion, Conscientiousness
5. Bus drivers	Adjustment, Likeability, Prudence
6. Professionals	Emotional Orientation, Cognitive Orientation, Task Orientation

Method

All six datasets described in chapter 3 were used in this study. The prediction equations for the present study were developed using the same procedures and the same twenty training/test set partitions as in Studies 1 and 2. The only difference being that in the present study prediction equations were developed using the combinations of predictors specified in Table 5.1, whereas in the study reported in the previous chapter prediction equations were developed separately for each predictor. Thus, for each of the twenty training set partitions within each dataset I used the specified predictors to obtain one linear regression equation, four neural networks that were developed without early stopping (with the number of hidden units corresponding to the H1, H2, H3, and H4 levels used in Study 1), and a committee of 15 early stopping neural networks. All prediction equations were tested on the relevant test cases in order to obtain MAE values and cross-validity coefficients that formed the basis of the subsequent analyses.[1]

Results and Discussion

Artificial Neural Networks Versus Linear Regression

Table 5.2 presents the MAE values and cross-validity coefficients for the linear regression equations, the neural networks developed without early stopping (labelled ANN1), and the early stopping neural network committees (labelled ANN2). The results are averaged across the twenty partitions within each dataset. Furthermore, the results for the ANN1 networks are averaged across the four hidden unit levels. Underlined values are used to indicate instances where neural networks outperformed linear regression.

[1] Tables C32 to C37 of Appendix C provide the predictive performance results for the individual prediction equations developed in Study 3.

Table 5.2

MAE values and cross-validity coefficients for the linear regression (LR) equations and artificial neural networks (ANN1 and ANN2), by dataset.

Dataset	MAE			Cross-validity coefficient		
	LR	ANN1	ANN2	LR	ANN 1	ANN 2
1. University students	7.121	7.274	<u>7.120</u>	.26	.21	.25
2. Police recruits	32.24	32.68	<u>32.03</u>	.22	<u>.23</u>	<u>.23</u>
3. Flight attendants	1.079	1.139	1.095	.09	.01	.08
4. Managers	0.484	0.502	0.485	.25	.22	<u>.26</u>
5. Bus drivers	6.125	6.185	6.138	.10	.08	.10
6. Professionals	0.603	0.623	<u>0.596</u>	-.02	<u>.08</u>	<u>.00</u>

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Table 5.2 indicates that the neural networks developed without early stopping (ANN1) performed poorly compared to the linear regression equations. The ANN1 equations produced higher MAE values than linear regression for all six datasets, and lower cross-validity coefficients for four of the six datasets. However, it should be noted that for Dataset 6 the ANN1 equations obtained a cross-validity coefficient that was on average .10 of a unit higher than that of the corresponding linear regression equations. This is a non-trivial improvement.

The predictive performance of the early stopping neural network committees (ANN2) was similar to that of the linear regression equations, and any differences between the two methods were small in magnitude. For example, the MAE difference between the linear regression and ANN2 equations for Dataset 6 represents a reduction of only approximately 1% (as a percentage of the linear regression MAE), yet this was the biggest reduction in favour of the ANN2 equations across all the datasets. Similarly, the difference in the average cross-validity coefficients between the linear regression and ANN2 equations did not exceed .02 for any of the datasets.

The corrected t-test procedure outlined in the Data Analysis section of Study 1 was used to perform significance tests for the differences in MAE values and cross-validity coefficients between the linear regression equations and neural networks. For each dataset, the ANN1 and ANN2 networks were separately compared to the linear regression equations. These were analyses that had been planned prior to examining the data, and were designed to be consistent with the contrasts tested in Studies 1 and 2. The only statistically significant difference occurred for Dataset 3, in which the linear regressions equations produced significantly lower MAE than the ANN1 equations, corrected $t(19) = -2.32$, $p < .05$.[2]

Table 5.3

MAE values and cross-validity coefficients for the neural networks developed without early stopping, by dataset and hidden unit level (linear regression results are also presented for comparison).

Predictive performance measure and dataset	LR	H1	H2	H3	H4
MAE					
1. Tertiary students	7.121	7.185	7.272	7.299	7.341
2. Police recruits	32.24	32.98	32.89	32.61	32.26
3. Flight attendants	1.079	1.123	1.136	1.153	1.142
4. Managers	0.484	0.490	0.502	0.508	0.506
5. Bus drivers	6.125	6.164	6.187	6.197	6.192
6. Professionals	0.603	0.631	0.628	0.603	0.629
Cross-validity coefficients					
1. Tertiary students	.26	.24	.20	.20	.20
2. Police recruits	.22	.20	.23	.24	.26
3. Flight attendants	.09	.05	.01	-.01	.00
4. Managers	.25	.25	.21	.21	.21
5. Bus drivers	.10	.09	.09	.08	.08
6. Professionals	-.02	-.03	.08	.14	.13

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

[2] See Tables D5 and D6 in Appendix D for complete results of these statistical tests.

Table 5.3 provides a breakdown of the ANN1 results by hidden unit level. The linear regression results are also presented for comparison. The MAE associated with almost all hidden unit levels in all datasets was higher than that of the corresponding linear regression equations. The only exception occurred for the H3 networks in Dataset 6, which produced MAE approximately equal to that of linear regression. Thus, regardless of the number of hidden units, neural networks developed without early stopping were unable to outperform the linear regression equations with respect to the absolute measure of predictive performance. However, there were two datasets in which some of the networks outperformed linear regression with respect to cross-validity coefficients. In Dataset 2 the H2, H3, and H4 networks obtained cross-validity coefficients that were .01, .02, and .04 of a coefficient larger than that of the linear regression equations. More noticeably, in Dataset 6 the H2, H3, and H4 networks obtained cross-validity coefficients that were .10, .16, and .15 of a coefficient larger than that of the associated linear regression equations.

Overall then, contrary to what was expected, the present study found little evidence that neural networks generally outperform linear regression for the types of datasets considered here. With the exception of Dataset 6, differences in predictive performance between linear regression and neural networks were either approximately equivalent or else in favour of the linear equations.

Configural Relationships Between Personality Variables and Work Performance

To examine the evidence for specific forms of configurality in personality-performance relationships moderated multiple regression was applied to the total data in each dataset. Only personality variables that had been hypothesised to be theoretical relevant for performance in the dataset in question were included in the analyses. For each dataset all possible product terms (two-way, three-way, and where applicable four- and five-way) between the theoretically relevant personality variables were calculated and each product term was represented as a separate variable. Following Pedhazur (1997), the analyses were conducted hierarchically: Each of the two-way product variables (e.g., x_1x_2) was tested by regressing the work performance criterion onto the variable while holding the two individual personality variables (x_1 and x_2) constant; each three-way variable (e.g., $x_1x_2x_3$) was tested by regressing performance on to it while holding the three individual personality variables (x_1, x_2, and x_3) and three two-way product terms (x_1x_2, x_1x_3, and x_2x_3) constant; and so on. Thus, across the six datasets a total of 53 product terms were tested for significance (four in each of the four datasets that contained three theoretically relevant personality variable, 11 in the dataset with four theoretically relevant personality variables, and 26 in the dataset with five relevant variables).[3]

Table 5.4 presents the change in R^2 associated with each step of the analysis for the highest product term in each dataset. The change in R^2 associated with entering all the product terms simultaneously is presented in the final column.[4]

---

[3] Note that the multiplicative functions tested here do not encompass all possible forms of configural relationships that could exist between predictors and a criterion, or that could be detected by a neural network. Nevertheless, the analyses are likely to provide a good representation of many plausible configural relationships between personality variables and work performance.
[4] The results for all the other product terms are provided in Appendix F.

Table 5.4

Change in R^2 associated with entry of the theoretically relevant personality variables (Step 1), the two-way product terms (Step 2), the three-way product term(s) (Step3), the four-way product term(s) (Step 4), the five-way product term (Step 5), and the sum of all product terms (Steps 2-5).

Dataset	Step1	Step 2	Step 3	Step 4	Step 5	Steps 2-5
1. University students	.074**	.002	.004	NA	NA	.006
2. Police recruits	.085**	.032	.051**	.001	NA	.084**
3. Flight attendants	.040*	.019	.027	.026	.001	.073
4. Managers	.087**	.024	.008	NA	NA	.032
5. Bus drivers	.022*	.007	.001	NA	NA	.008
6. Professionals	.022	.070*	.026	NA	NA	.096*

* $p < .05$, ** $p < .01$, NA = Not Applicable.

Table 5.4 indicates that in five of the six datasets the theoretically relevant personality variables, when combined linearly and additively, accounted for a statistically significant proportion of the criterion variance (step 1). In contrast, the evidence for multiplicative relationships was limited. The increment in the proportion of the variance accounted for by the product terms was statistically significant in Datasets 2 and 6, but not in any other datasets. The specific nature of the configural relationships is indicated by the results of the tests that were applied to the separate product terms, as outlined above. Of the 53 separate product terms that were tested, only five were statistically significant at $\alpha = .05$. In Dataset 6 there were two statistically significant effects. First, the Emotional Orientation x Cognitive Orientation product term accounted for a significant proportion of the variance above that accounted for by the additive effects of these two scales, $t(116) = -2.30$, $p < .05$; and second, there was a significant Emotional Orientation x Task Orientation effect $t(116) = -2.40$, $p < .05$. Figures 5.1 and 5.2 plot the moderated multiple regression equations associated with these two effects.

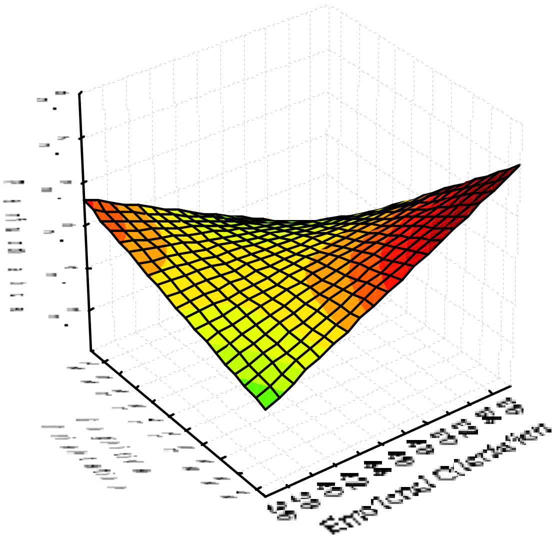


Figure 5.1

The moderated multiple regression equation relating Emotional Orientation and Cognitive Orientation to work performance in Dataset 6.

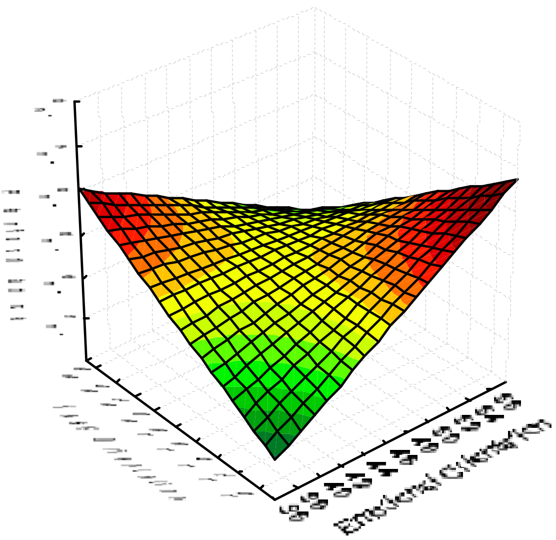


Figure 5.2

The moderated multiple regression equation relating Emotional Orientation and Task Orientation to work performance in Dataset 6.

Figure 5.1 shows that the direction of the relationship between Cognitive Orientation and work performance changes as the level of Emotional Orientation increases. Specifically, the relationship between Cognitive Orientation and work performance is positive at lower levels of Emotional Orientation, but negative at higher levels of Emotional Orientation. Figure 5.2 indicates a similar moderating effect of Emotional Orientation on the relationship between Task Orientation and work performance. Alternatively, the two effects could be interpreted as the moderating effects of Cognitive Orientation and Task Orientation on the relationship between Emotional Orientation and work performance. Regardless of the interpretation adopted, the type of moderation effect detected here – where the *direction* (rather than strength) of the relationship between a personality variable and work performance changes as a function of another personality variable – does not correspond to the theoretical rationale for moderator effects presented in chapter 1. While the positive effects of Task orientation and Cognitive Orientation on performance at low levels of Emotional Orientation are not unexpected, it is not clear why either variable would be *negatively* related to performance at high levels of Emotional Orientation.

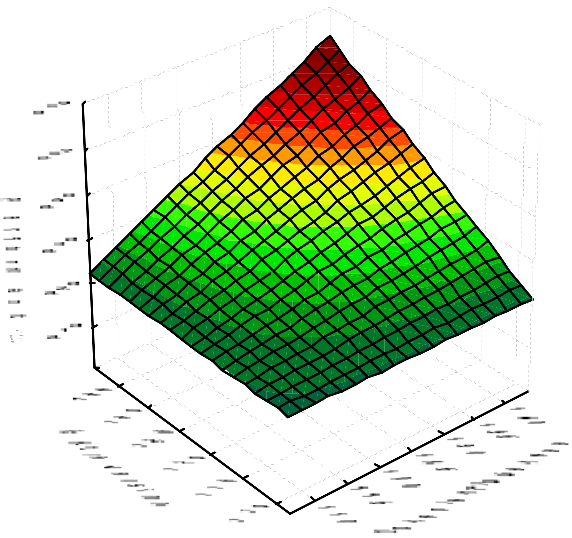
Additionally, in Dataset 2, three of the four three-way product terms were statistically significant: namely, the Neuroticism x Extraversion x Conscientiousness effect, $t(278) = 2.53$, and the Neuroticism x Openness x Conscientiousness effect, $t(278) = 2.12$, were both significant at $\alpha = .05$, and the Extraversion x Openness x Conscientiousness effect, $t(278) = -2.69$, was significant at $\alpha = .01$. These effects were typically difficult to interpret, and given the large number of product terms that were tested for this dataset some of the effects may simply be due to chance. Nevertheless, the latter effect showed some similarity to the findings of Witt (2002), and could be meaningfully interpreted in terms of these findings. The two graphs in Figure 5.3

provide a graphical representation of the moderated multiple regression equation for which the Extraversion x Openness x Conscientiousness term was significant. The multiplicative effect of Extraversion x Conscientiousness is plotted at low (panel a) and high (panel b) levels of Openness, where low and high were defined as one standard deviation above or below the mean of the Openness scale.

Panel a of Figure 5.3 shows that, at low levels of Openness, Conscientiousness is more strongly related to performance when Extraversion is high than when it is low. Witt (2002) obtained a similar Extraversion x Conscientiousness effect on work performance in three separate samples. He argued that Extraversion reflects a tendency to interact with others, and that without this tendency an employee will fail to fully capitalise on the benefits of Conscientiousness in jobs involving interpersonal interaction. Likewise, for interactive training programs such as the police academy training considered here, it could be the case that the effects of Conscientiousness are enhanced when the individual engages in the type of behaviours associated with Extraversion such as asking more questions and actively interacting with other trainees.

In contrast, panel b of Figure 5.3 shows that, at high levels of Openness, the strength of the relationship between Conscientiousness and performance varies little at different levels of Extraversion. In other words, the moderating effect of Extraversion on Conscientiousness no longer occurs when Openness is high. This could reflect the idea outlined in chapter 1 that high levels of certain attributes compensate for low levels of other attributes. Specifically, within the context of a training program, high levels of Openness may provide the motivation that is normally provided by Extraversion to actively participate in the program and ask questions. Therefore, the relatively weak relationship between Conscientiousness and performance only eventuates for individuals who are low on both Openness and Extraversion.

a) Low Openness (1 SD below mean of Openness scale)



a) High Openness (1 SD above mean of Openness scale)

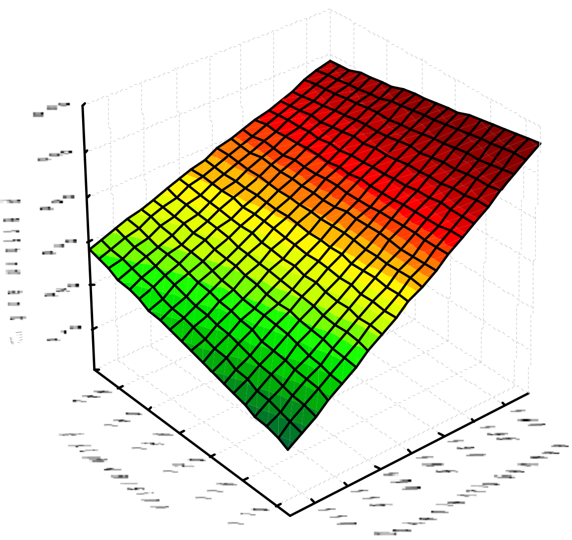


Figure 5.3

The moderated multiple regression equation relating Extraversion and Conscientiousness to work performance for low (panel a) and high (panel b) levels of Openness in Dataset 6.

Additionally, it can be noted that the two datasets in which there was some evidence for configural relationships between personality and work performance are also the same datasets in which the neural networks performed the best relative to linear regression. Specifically, Datasets 2 and 6 were the only ones in which any of the hidden unit levels obtained cross-validity coefficients that were more than .02 of a coefficient higher than that of the corresponding linear equations. In the latter dataset the difference was as high as .16 for the larger hidden unit levels. Therefore, despite the overall negative conclusion, the present study provides some grounds for optimism about the benefits of neural networks when configural relationships are present. Furthermore, it motivates a consideration of other combinations of personality variables, some of which may be configurally related to work performance. This was examined in Study 4.

Study 4

The combination of personality variables that includes those variables that are considered theoretically relevant to performance and excludes those that are not serves as a useful starting point for comparing neural networks and linear regression. Nevertheless, there are a number of reasons why it is also desirable to compare the two methods for other combinations of personality variables. First, as alluded to in the previous study, the failure of neural networks to outperform linear regression may have resulted from the choice of predictor combinations that were, in general, not configurally related to the criterion. By exploring other combinations of personality variables one can obtain a better indication of the extent of predictive performance that can be achieved by neural networks for the present datasets.

Furthermore, predictive performance can sometimes be improved by omitting relevant variables, even when these variables are empirically related to the criterion and uncorrelated with other predictors in the equation (see Blum & Langley, 1997 for a discussion of different conceptions of relevance in the context of predictor selection). This occurs because there is less tendency to fit the noise in the training set when there are fewer parameters to be estimated (e.g., Reed & Marks, 1999). Hence, the omission of a predictor may improve predictive performance for unseen cases if the resulting decrease in overfitting more than offsets the loss of the information provided by the predictor.

Based on the above reasoning, it is possible that the performance of the neural networks in Study 3 was attenuated either as a result of the inclusion of too many variables, or else due to the omission of variables that contain important configural information relating to the criterion. In the present study the predictive performance of neural networks was explored for other combinations of personality variables. As it is not feasible (nor necessarily desirable) to test all possible combinations of the available predictor variables, a predictor-selection procedure was employed to identify subsets of predictors that were potentially useful for predicting the criterion, and that varied in the number of included predictors. The neural network and linear regression methods were compared for these combinations. Given the exploratory nature of the study, no hypotheses were developed about which combination of predictors would produce the most accurate predictions, or whether neural networks would outperform linear regression for any particular combination.[5]

[5] It should be kept in mind that the use of an automated predictor-selection procedure does not avoid the capitalisation on chance that occurs when many different combinations of predictors are considered (Cohen, 1990), and therefore the results from such an analysis need to be interpreted tentatively.

Method

Certain decisions need to be made as part of the implementation of a predictor-selection procedure. First, one must decide on the type of prediction equation that will guide the selection process. The simplest alternative is to use a linear equation, however this approach will fail to identify combinations of predictors that are useful because of the configural and/or nonlinear nature of their relationship with the criterion. As it is in the presence of these types of relationships that neural networks are of most use, and as the focus of this study is on the predictive performance that can be achieved by neural networks for the present datasets, the identification of potentially useful subsets of predictor variables was guided by neural network equations.

Second, a search algorithm needs to be chosen. Two popular choices are *forward selection*, in which one commences with an empty equation and predictors are added to the equation one at a time, and *backward elimination*, in which one commences with an equation that contains all the available predictors and then sequentially removes predictors (Pedhazur, 1997). In both procedures the decision about which variable to add/remove at each step is determined by the contribution of each variable to prediction in the context of the variables currently in the equation. It has been argued that forward selection, though computationally more efficient than backward elimination, results in weaker subsets because the importance of each predictor is not evaluated in the context of predictors that are not included yet (Guyon & Elisseeff, 2003). Consequently this algorithm can fail to identify key subsets that consist of predictors that are interdependently related to the criterion. For this reason, backward elimination was used in the present study.

Another issue concerns the strategy used to determine the relative importance of predictors at each step of the analysis, and consequently to decide which variable to remove at that step. Chapter 2 introduced sensitivity analysis as one method of assessing the relative importance of predictors. At each step of the analysis the information provided by each predictor is in turn made unavailable by clamping its value to a typical value (such as the mean of the predictor), and test set error is computed. The variable whose ‘unavailability’ is associated with the smallest deterioration (or largest improvement) in test set predictive accuracy is eliminated at that step.

The predictor-selection procedure adopted in the present study was implemented with the above considerations in mind. First, for each dataset prediction equations were developed and tested using all of the personality measures in the dataset. The equations were developed and tested using the same twenty training/test partitions, the same procedures, and the same type of equations (linear, H1, H2, H3, H4, and early stopping committees) as in Study 3. Thus, the only difference between the equations of Study 3 and the initial equations developed here was that the former combined the information from the personality variables that were identified as theoretically relevant, whereas the latter combined the information from all of the personality variables.

Sensitivity analysis was then conducted on the H4 networks. These networks were used as the basis of the sensitivity analysis as they have the greatest capacity for representing any nonlinear and configural relationships that may be present. For each of the twenty H4 networks in each dataset the sensitivity of each predictor was determined separately by clamping its value to the mean of the predictor and computing the network’s test set error in the absence of the information provided by the predictor. Across the twenty H4 networks in each dataset, the predictor whose absence most often resulted in the smallest decrement (or greatest improvement) in predictive accuracy was eliminated. New sets of equations (linear, H1, H2, H3, H4, and early stopping committees) were then developed and tested with the combination of the remaining predictors, and backward elimination via sensitivity analysis was once again implemented to determine which predictor to next eliminate. The above process was repeated until there was only one predictor remaining within each dataset.[6]

Results and Discussion

Table 5.5 presents the combinations of personality variables that resulted from the backward elimination procedure for each of the datasets. Consistent with the previous conclusions about the importance of Conscientiousness relative to the other personality factors, note that in the majority of datasets measures of this factor emerged as the final variable remaining in the equation. Tables 5.6 and 5.7 present the MAE values and cross-validity coefficients of the prediction equations for the different combinations. The label ANN2 is used to denote the early stopping committees. The results for the different hidden unit levels of the networks developed without early stopping are presented separately (labelled as H1, H2, H3, and H4), hence the absence of the ANN1 label (which was used in previous studies to denote the average of the four hidden unit levels). The results can be summarised in terms of three main findings, namely:

1. Differences between neural networks and linear regression equations.
2. The effect of variations in the number of predictors.
3. The interaction between equation type and the number of predictors.

[6] Tables C38 to C59 of Appendix C provide the predictive performance results for each prediction equation developed in Study 4. Some of the combinations that resulted from the backward elimination procedure were those that had been developed and tested in the previous studies. The statistics for these combinations can be found in the Appendix C tables relating to those studies.

Table 5.5

The combinations of personality variables that resulted from the backward elimination procedure, by dataset and number of predictors.

Dataset and number of predictors	Personality variables remaining after backward elimination
1. University students	
One predictor	Conscientiousness
Two predictors	Conscientiousness, Extraversion
Three predictors	Conscientiousness, Extraversion, Openness
Four predictors	Conscientiousness, Extraversion, Openness, Neuroticism
Five predictors	Conscientiousness, Extraversion, Openness, Neuroticism, Agreeableness
2. Police recruits	
One predictor	Conscientiousness
Two predictors	Conscientiousness, Extraversion
Three predictors	Conscientiousness, Extraversion, Neuroticism
Four predictors	Conscientiousness, Extraversion, Neuroticism, Openness
Five predictors	Conscientiousness, Extraversion, Neuroticism, Openness, Agreeableness
3. Flight attendants	
One predictor	Agreeableness
Two predictors	Agreeableness, Conscientiousness
Three predictors	Agreeableness, Conscientiousness, Neuroticism
Four predictors	Agreeableness, Conscientiousness, Neuroticism, Openness
Five predictors	Agreeableness, Conscientiousness, Neuroticism, Openness, Extraversion
4. Managers	
One predictor	Conscientiousness
Two predictors	Conscientiousness, Openness
Three predictors	Conscientiousness, Openness, Agreeableness
Four predictors	Conscientiousness, Openness, Agreeableness, Extraversion
Five predictors	Conscientiousness, Openness, Agreeableness, Extraversion, Neuroticism
5. Bus drivers	
One predictor	Prudence
Two predictors	Prudence, Ambition
Three predictors	Prudence, Ambition, Adjustment
Four predictors	Prudence, Ambition, Adjustment, Likeability
Five predictors	Prudence, Ambition, Adjustment, Likeability, Intellectance
Six predictors	Prudence, Ambition, Adjustment, Likeability, Intellectance, Sociability
6. Professionals	
One predictor	Cognitive O.
Two predictors	Cognitive O., Emotional O.
Three predictors	Cognitive O., Emotional O., Task O.
Four predictors	Cognitive O., Emotional O., Task O., Interpersonal O.
Five predictors	Cognitive O., Emotional O., Task O., Interpersonal O., Social O.

Table 5.6

MAE values for the linear regression (LR) equations and artificial neural networks (H1, H2, H3, H4, and ANN2), by dataset and number of predictors.

Dataset and number of predictors	LR	H1	H2	H3	H4	ANN2
1. University students						
One predictor	7.226	7.273	7.260	7.253	7.243	7.231
Two predictors	7.109	7.175	**7.193**	7.207	7.184	7.121
Three predictors	**6.928**	**7.096**	7.193	**7.164**	**7.216**	**7.011**
Four predictors	6.958	7.199	7.404	7.501	7.480	7.030
Five predictors	7.018	7.279	7.619	7.824	8.081	7.105
2. Police recruits						
One predictor	**31.55**	**31.66**	**31.65**	**31.71**	**31.71**	**31.60**
Two predictors	31.83	32.33	32.29	32.35	32.39	32.05
Three predictors	31.98	32.06	32.02	31.94	31.95	31.93
Four predictors	32.11	32.94	33.26	33.10	33.04	32.03
Five predictors	32.43	34.01	34.19	34.76	33.89	32.45
3. Flight attendants						
One predictor	**1.061**	**1.072**	**1.075**	**1.076**	**1.077**	**1.067**
Two predictors	1.064	1.079	1.088	1.082	1.083	1.067
Three predictors	1.068	1.089	1.086	1.090	1.093	1.071
Four predictors	1.074	1.098	1.117	1.113	1.125	1.085
Five predictors	1.079	1.123	1.136	1.153	1.142	1.095
4. Managers						
One predictor	**0.480**	**0.482**	**0.483**	**0.483**	**0.484**	**0.482**
Two predictors	0.482	0.487	0.492	0.496	0.494	0.482
Three predictors	0.488	0.493	0.504	0.506	0.503	0.485
Four predictors	0.491	0.509	0.543	0.532	0.539	0.487
Five predictors	0.490	0.509	0.546	0.566	0.556	0.486
5. Bus drivers						
One predictor	**6.081**	**6.094**	**6.090**	**6.096**	**6.094**	**6.093**
Two predictors	6.108	6.142	6.146	6.139	6.138	6.107
Three predictors	6.124	6.169	6.167	6.185	6.194	6.134
Four predictors	6.143	6.210	6.241	6.269	6.260	6.146
Five predictors	6.164	6.287	6.280	6.321	6.353	6.174
Six predictors	6.182	6.290	6.326	6.388	6.347	6.190
6. Professionals						
One predictor	**0.596**	**0.602**	0.606	0.607	0.609	0.598
Two predictors	0.599	0.616	**0.582**	**0.580**	**0.583**	**0.594**
Three predictors	0.603	0.631	0.628	0.603	0.629	0.596
Four predictors	0.599	0.640	0.647	0.714	0.828	0.601
Five predictors	0.608	0.651	0.660	0.840	1.193	0.608

Note: Underlined values indicate that the neural networks outperformed the associated linear equations. Boldface indicates the best performing combination of predictors within each type of equation and dataset.

Table 5.7

Cross-validity coefficients for the linear equations and artificial neural networks (H1, H2, H3, H4, and ANN2), by dataset and number of predictors.

Dataset and number of predictors	LR	H1	H2	H3	H4	ANN2
<u>1. University students</u>						
One predictor	.22	.19	.20	.20	.20	.21
Two predictors	.28	.25	.25	.24	.25	<u>.28</u>
Three predictors	**.34**	**.28**	**.25**	**.27**	**.25**	**.32**
Four predictors	.32	.27	.21	.19	.19	.31
Five predictors	.29	.27	.23	.21	.18	.29
<u>2. Police recruits</u>						
One predictor	**.26**	**.25**	**.25**	**.24**	.24	**.25**
Two predictors	.25	.22	.22	.22	.21	.24
Three predictors	.24	.24	<u>.24</u>	<u>.24</u>	.24	.23
Four predictors	.22	.20	<u>.23</u>	<u>.24</u>	<u>**.26**</u>	<u>.23</u>
Five predictors	.20	.16	.18	.17	.20	.21
<u>3. Flight attendants</u>						
One predictor	**.17**	**.14**	**.13**	**.13**	**.13**	**.16**
Two predictors	.15	.11	.10	.11	.10	.14
Three predictors	.13	.09	.10	.09	.09	.12
Four predictors	.11	.07	.06	.06	.05	.08
Five predictors	.09	.05	.01	-.01	.00	.08
<u>4. Managers</u>						
One predictor	**.26**	**.25**	**.25**	**.25**	**.24**	.26
Two predictors	.25	.22	.21	.19	.19	.24
Three predictors	.25	.23	.20	.21	.21	<u>.26</u>
Four predictors	.25	.21	.17	.18	.18	<u>.26</u>
Five predictors	.25	.18	.15	.18	.17	<u>**.27**</u>
<u>5. Bus drivers</u>						
One predictor	**.14**	<u>**.15**</u>	<u>**.15**</u>	**.14**	<u>**.14**</u>	<u>**.15**</u>
Two predictors	.12	<u>.13</u>	.12	.12	.12	<u>.13</u>
Three predictors	.10	.10	.09	.09	.09	<u>.11</u>
Four predictors	.09	.09	.08	.09	.09	<u>.11</u>
Five predictors	.08	.07	.07	.07	.07	<u>.09</u>
Six predictors	.07	.05	.05	.05	.06	<u>.07</u>
<u>6. Professionals</u>						
One predictor	**.10**	**.05**	-.01	-.01	.01	.06
Two predictors	.01	<u>.01</u>	<u>**.14**</u>	<u>**.20**</u>	<u>**.19**</u>	<u>**.07**</u>
Three predictors	-.02	-.03	<u>.08</u>	<u>.14</u>	<u>.13</u>	<u>.00</u>
Four predictors	-.03	-.04	-<u>.01</u>	-.03	-<u>.01</u>	-.03
Five predictors	-.07	-.11	-.08	-<u>.06</u>	-<u>.03</u>	-<u>.07</u>

Note: Underlined values indicate that the neural networks outperformed the associated linear equations. Boldface indicates the best performing combination of predictors within each type of equation and dataset.

Artificial Neural Networks Versus Linear Regression

The neural network equations developed without early stopping (H1 to H4) were rarely more accurate than the linear regression equations. Table 5.6 indicates that across all datasets there were only two combinations of predictors for which any of the neural networks developed without early stopping produced lower MAE than the corresponding linear equations. Specifically, the H3 and H4 networks produced lower MAE than linear regression for the three-predictor combination in Dataset 2, and the H2, H3, and H4 networks produced lower MAE than linear regression for the two-predictor combination in Dataset 6. As documented in the previous study, both these predictor combinations were characterised by a statistically significant multiplicative effect with the criterion, and therefore these results reflect the capability of neural networks to incorporate information about the configural relationship between the predictors and the criterion. A similar pattern of findings was obtained when cross-validity coefficients were used as the measure of predictive performance (see Table 5.7). Once again the linear regression equations typically outperformed the networks that were developed without early stopping, and the instances where this did not occur were mostly clustered in Datasets 2 and 6. Of particular note, for the two-predictor combination in Dataset 6, the H3 and H4 networks produced cross-validity coefficients that were on average .19 and .18 higher than the coefficients produced by the corresponding linear equations.

In contrast, differences between the early stopping committees (ANN2) and linear regression equations were less apparent. For the 31 combinations of predictors listed in Tables 5.6 and 5.7, the early stopping committees produced lower MAE than the corresponding linear regression equations 9 times, and higher cross-validity coefficients 14 times. Moreover the magnitude of the differences were smaller than those between the networks developed without early stopping and linear regression. For example, for all 31 predictor combinations, the difference in MAE between the early stopping committees and linear regression equations was approximately 1% or less of the linear regression MAE. Thus, any reduction in MAE that was achieved by the use of early stopping neural network committees could be described as trivially small. Similarly, for 28 of the 31 predictor-combinations the difference between the average cross-validity coefficient of the linear equations and early stopping networks was no more than .02 of a coefficient. The largest difference occurred for the combination of two predictors in Dataset 6. In this case the average cross-validity coefficient of the early stopping committees was .07, which was .06 of a unit higher than the average coefficient obtained by the corresponding linear equations, but not as high as the cross-validity coefficients obtained for this predictor-combination by the neural networks developed without early stopping.

Overall then, the pattern of findings with respect to differences between neural networks and linear regression was similar to the results obtained in Study 3: For the majority of predictor-combinations the neural networks performed worse than or similarly to the linear regression equations, whereas for the small number of combinations that were characterised by configural relationships the networks with a large number of hidden units were able to outperform linear regression, at least in terms of the obtained cross-validity coefficients.

The Effect of Variations in the Number of Predictors

To best illustrate the effect of the number of predictors on predictive performance, the values presented in Tables 5.6 and 5.7 have been plotted graphically in Figures 5.4 and 5.5.

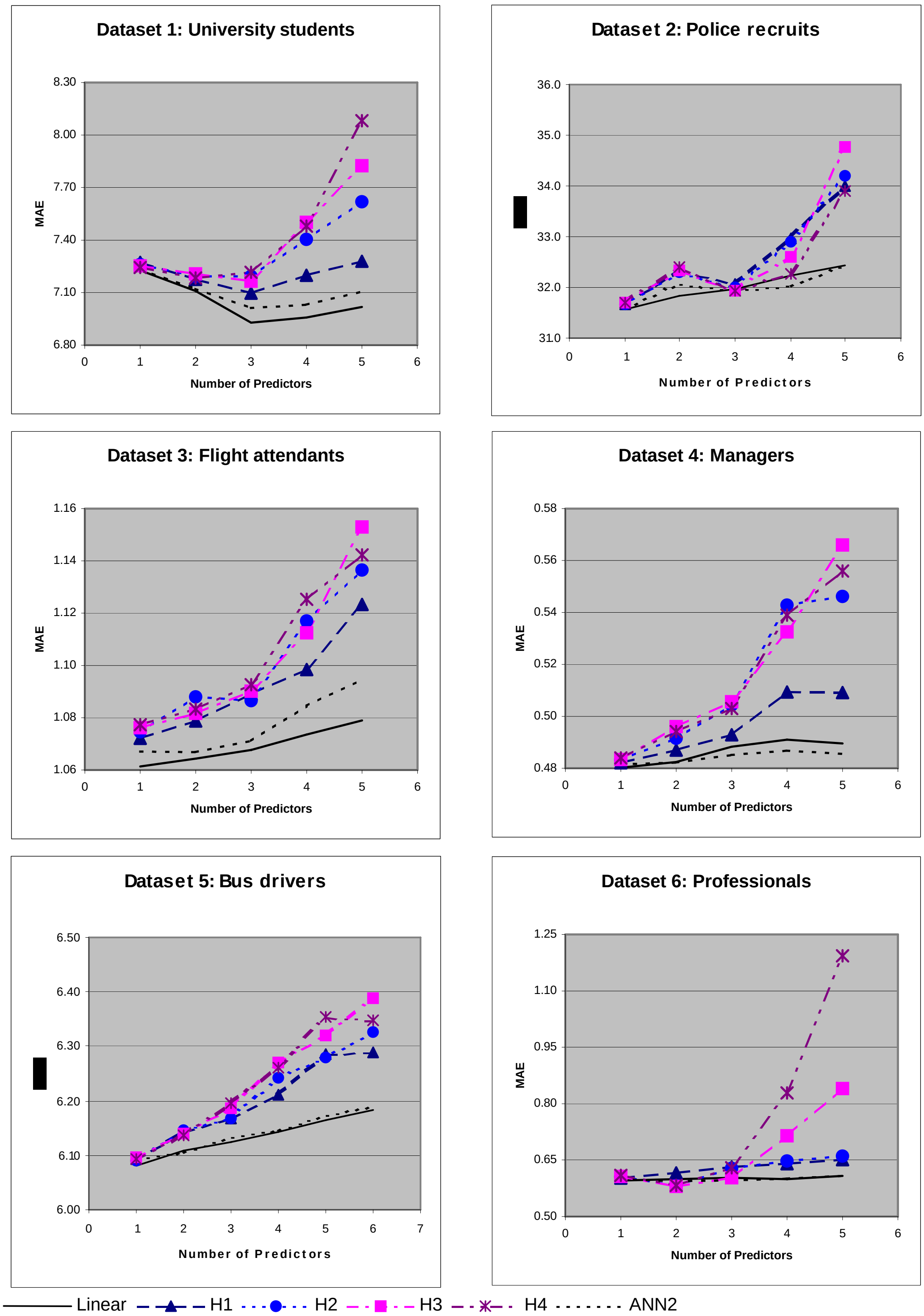


Figure 5.4

The relationship between the number of predictors included in the equation and MAE, by equation type and dataset.

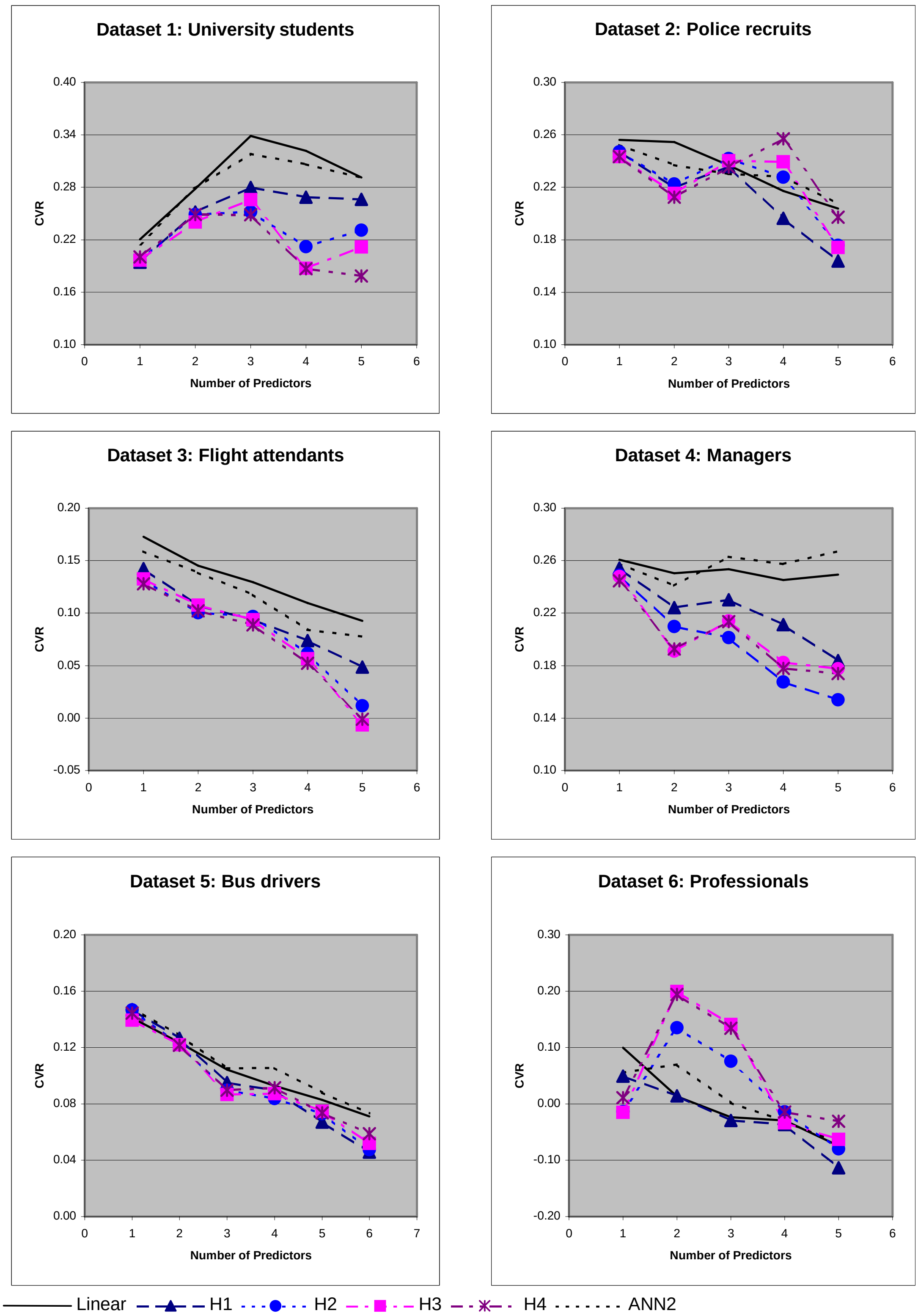


Figure 5.5

The relationship between the number of predictors included in the equation and cross-validity coefficient (CVR), by equation type and dataset.

Figure 5.4 plots the relationship between the number of predictors included in the equation and MAE for each type of equation within each dataset. It can be seen that across the datasets there tended to be a positive relationship between the number of predictors in the equation and MAE. In Datasets 2, 3, 4, and 5 the lowest MAE was obtained by single-predictor equations. In Dataset 6 the lowest MAE was obtained by the two-predictor equations, although this only applied to the H2, H3, H4, and ANN2 equations; the linear regression equations and H1 networks did best with a single predictor. This presumably occurred because the bulk of the information about the criterion carried by the second predictor was due to its configural relationship with the first predictor. As the linear equations cannot represent configural relationships (and the H1 networks only have a limited capability to do so), it is not surprising that the second predictor was of little use in the context of these equations. Finally, in Dataset 1, three predictors were optimal.

A similar relationship occurred between the number of predictors and predictive performance measured in terms of cross-validity coefficients. Figure 5.5 indicates that in Datasets 2, 3, 4, 5, and 6 cross-validity coefficients tended to be decrease as the number of predictors in the equation increased. In Dataset 1 cross-validity coefficients peaked for the three-predictor combination.

The inverse relationship between the number of predictors and predictive accuracy can be interpreted in terms of the increased capitalisation on chance that occurs as the number of predictors is increased. Specifically, increasing the number of predictors also increases the number of parameters that need to be estimated and – given a fixed training set – decreases the ratio of training cases to free parameters. A consequence of trying to estimate a greater number of parameters with a fixed number of training cases is that one is more likely to capture relationships that are idiosyncratic to the training sample but do not generalise to the population from which the sample is drawn (Babyak, 2004). Of course additional predictors can also improve predictive performance to the extent that they provide nonredundant information about the criterion. However, the above findings suggest that this was generally not the case for the datasets considered here – most if not all of the benefits of personality variables for predicting work performance were realised by including one or two personality variables, and the inclusion of additional variables beyond that typically resulted in poorer predictions.

The Interaction Between Equation Type and Number of Predictors

Figure 5.4 also indicates that the increase in MAE as a function of an increase in the number of predictors occurred at a faster rate for the neural networks developed without early stopping (especially the H3 and H4 networks) than for the early stopping committees or the linear regression equations. This trend occurred in all six datasets. Similarly, cross-validity coefficients decreased at a faster rate for the larger neural networks that were developed without early stopping than for the early stopping committees or the linear regression equations, although this is not as apparent for all six datasets (see Figure 5.5). Thus, in general, the predictive performance of the neural networks developed without early stopping was hampered more than that of the linear regression equations as the number of predictors included in the equations increased.

It could be argued that the above finding is an artefact of the procedure used for selecting the combination of personality variables to test at each level of the number of predictors. Specifically, the backward elimination process was guided by the H4 networks, and was therefore biased towards eliminating predictors that produced the greatest decrease in error for these types of equations, but not necessarily for the linear regression equations. Yet note that the predictor combination tested at the higher extreme of the number of predictors included all the available predictors, and therefore was not influenced by the predictor-selection procedure. Furthermore, for all datasets the predictor tested at the lower extreme turned out to be the best single predictor regardless of whether the H4 networks or linear regression was used to generate the equations (this can be verified by referring back to the results of Study 1, specifically see Tables 4.3 and 4.4). Therefore, the predictor tested at this point could also be said to be independent of the procedure used to select combinations of predictors. If the interaction between equation type and number of predictors was purely an artefact of the predictor-selection process, one would expect that differences between the H4 networks and linear equations to be approximately equivalent at the higher and lower extremes (given that the predictor combinations tested at both these points were independent of the predictor-selection procedure). However, it can be seen from Figure 5.4 that this is clearly not the case in any of the datasets.

A more likely explanation for the interactive effect of equation type and number of predictors on predictive performance relates to the growth in the number of parameters that occurs as predictors are increased. Specifically, the increase in the number of parameters (as predictors are added to the equation) occurs at a faster rate for neural networks than for equations developed using linear regression. In the case of linear regression, the addition of each predictor increases the number of parameters to be estimated by one, whereas for neural networks the increase in the number of parameters is equal to the number of hidden units.[7] To the extent that additional predictors carry configural information about the criterion, the extra parameters of neural networks can be useful for representing such information. However, the extra

[7] This occurs because each input unit (i.e., predictor) is connected to every hidden unit in the neural network, and each connection constitutes a parameter that needs to be estimated.

parameters also increase the likelihood of capturing the noise in the training sets (Reed & Marks, 1999). Therefore, in the absence of configural relationships (or if such relationships are weak), the predictive performance of neural networks can be expected to deteriorate at a faster rate than that of linear regression as the number of predictors is increased.

However, note that the use of early stopping committees provides a certain level of protection against the rapid deterioration in predictive performance that occurs for the neural networks developed without early stopping. For example, it can be seen in Figure 5.4 that the increase in MAE (as a function of increasing predictors) occurs at a slower rate for the early stopping committees than for the H4 networks developed without early stopping. This occurs despite the fact that the networks that comprise the committees contained the same number of hidden units as the H4 networks developed without early stopping, and therefore had the same potential capacity for fitting the training sets.

The robustness of the early stopping procedure against the adverse effects of having a large number of parameters (relative to the number of available training cases) has been previously documented (e.g., Caruana et al., 2000; Sarle, 2001c; Tetko et al., 1995), and can been explained in terms of the notion of effective complexity. The effective complexity of a network refers to the extent to which the network uses its representational capacity to fit the training set, and is related to both the actual number of parameters and the size of the parameters (e.g., Wang et al., 1994). Whereas the number of parameters is fixed prior to training and reflects the network's potential for fitting the training set, the size of the parameters – and hence the effective complexity of the network – increases as a function of the number of training iterations (Weigend, 1994). Specifically, it has been shown that at the initial stages of training, when the weights of the network (the parameters to be estimated) are small in size, neural networks learn relatively simple relationships; and that it is only after training has progressed that the networks develop the large weights required to more closely fit the training sets. By monitoring prediction error on a separate validation set, the early stopping procedures allows the training process to be stopped before the network has an opportunity to use its large number of parameters to fit the noise in the training set.

To illustrate the above points, the correlation between the predicted and observed criterion scores in the training set (which indicates the extent to which an equation fits the training data) was calculated for the H4 networks developed without early stopping, the early stopping committees (ANN2), and the linear regression equations. The training set correlations, averaged across the twenty partitions, are presented in Figure 5.6. First, it can be seen that the extent to which the equations fit the training data increases as the number of predictors is increased. Second, the increase occurs at a faster rate for both types of neural network equations than for linear regression equations, which reflects the fact that the representational capacity of neural networks increases more quickly than that of linear equations as predictors are added. Finally, the increase occurs at a slower rate for the early stopping committees (ANN2) than the H4 networks (developed without early stopping) despite the fact that both types of equations have the same representational capability. The early stopping procedure limits the extent to which the networks make use of their representational capability.

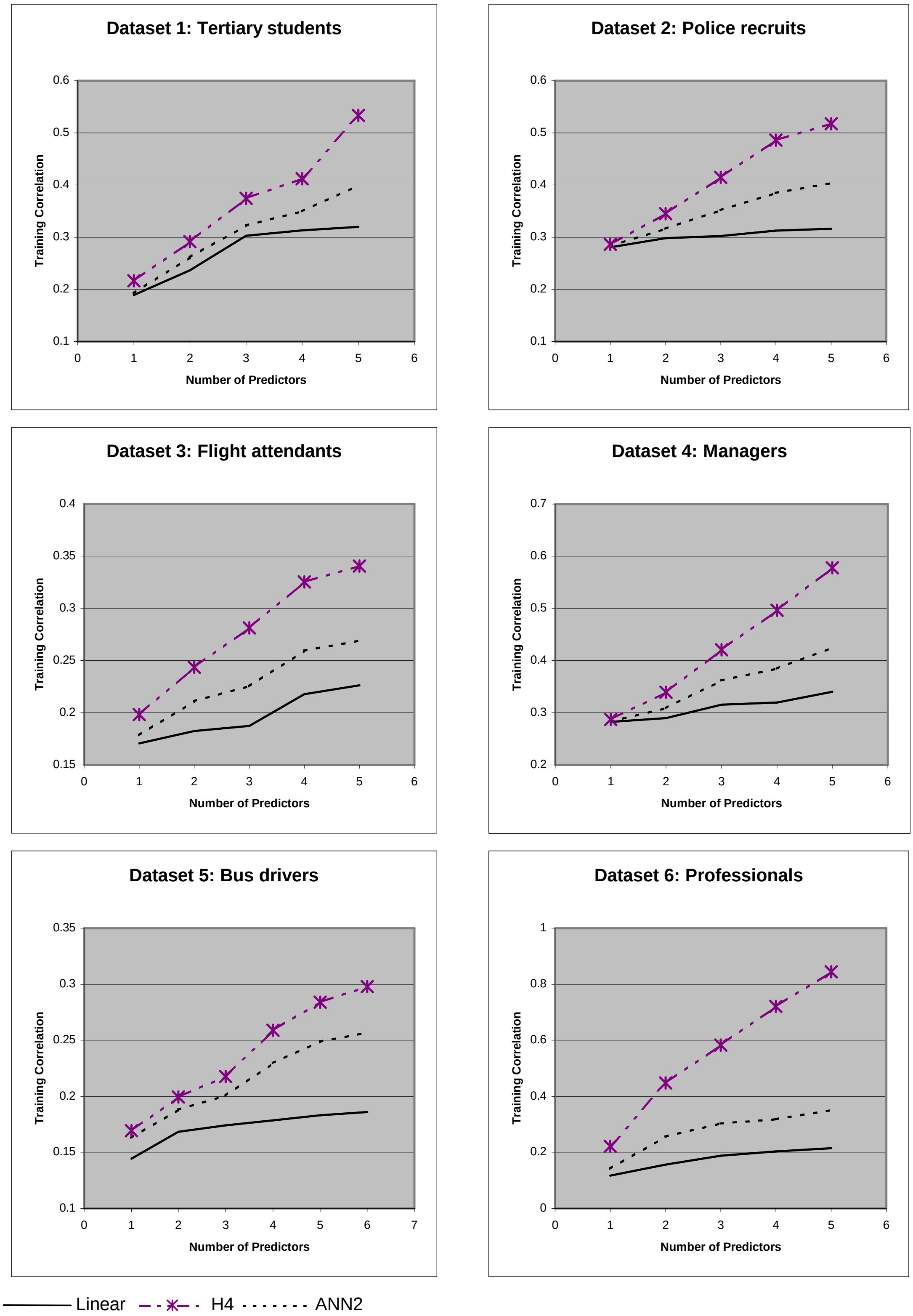


Figure 5.6

The relationship between the number of predictors included in the equation and training set correlation, by equation type and dataset.

Conclusions

A number of conclusions can be drawn based on the results presented in this chapter. First, there was little evidence that neural networks outperform linear regression when combinations of personality variables are used as the predictors. For the majority of predictor combinations that were considered in this chapter the linear equations performed as well if not better than the various neural networks. This occurred for combinations that were identified a priori based on theoretical grounds, as well as combinations that resulted from the backward elimination procedure. Furthermore, it should be noted that these results occurred despite the fact that the backward elimination procedure was directed towards identifying predictor combinations that produced the best predictive performance for neural networks, but not necessarily for the linear equations.

Second, there were a small number of predictor combinations for which neural networks were able to outperform linear regression with respect to cross-validity coefficients, and these combinations were characterised by statistically significant configural relationships between the predictors and the criterion. However, configural relationships occurred infrequently, and when they did occur they could not be easily interpreted in terms of the theoretical rationale for configurality that had been proposed.

Finally, the results highlight the dangers of capitalisation on chance when multiple personality variables are used to predict work performance. For most of the datasets considered here the inclusion of a single predictor, typically Conscientiousness, was optimal for predictive purposes. The detrimental effect of increasing the number of predictors occurred for both neural networks and linear regression, however the effect was stronger for the networks developed without early stopping than for the linear equations. The use of the early stopping procedure of developing neural networks provided a certain level of protection against this effect.

Taken together, the findings presented in this chapter and the previous chapter provided little support for the hypothesis that the effectiveness of measures of the five personality factors for predicting work performance would be higher when using artificial neural networks than when using linear regression equations. However, as pointed out in chapter 1, personality variables can be conceptualised at different levels of breath, and the broad dimensions of the five-factor model represent one level within this hierarchy. It has been argued that there are benefits of considering personality variables at a lower level within the hierarchy for the purposes of prediction (e.g., Mount et al., 2003; Tett, Steele, & Beauregard, 2003). Consequently, additional analyses were conducted to determine whether the conclusions drawn in relation to measures of the five broad dimensions generalised to narrower measures of personality traits, and these analyses are reported in the next chapter.

CHAPTER 6: Analyses Using Narrow Personality Variables

Introduction

The personality scales that were used as predictors in chapters 4 and 5 assessed each participant's standing on the five broad personality dimensions. However, there are other ways that the items included in personality inventories can be aggregated to form personality scales. Particularly, as outlined in chapter 1, in the first instance personality items can be combined to obtain measures of narrower personality traits that assess specific facets of the broader constructs represented by the five factors. In recent years some researchers have argued that representing personality variables in this way forms a better basis for the purposes of prediction than aggregating them into broad measures (e.g., Paunonen, Haddock, Forsterling, & Keinonen, 2003; Paunonen, Rothstein, & Douglas, 1999). The purpose of the research reported in the present chapter was two-fold. The first aim was to contribute to the literature on the relative merits of broad versus narrow personality variables by empirically comparing the predictive performance of personality variables at the broad and narrow levels. The second aim was to compare artificial neural networks and linear regression within the context of narrow personality measures in order to determine whether the conclusions drawn about the comparative predictive performance of these two methods for broad personality variables could also be extended to narrower measures.

One difficulty that arises when studying narrow personality constructs is that there has been little agreement among researchers about the number or nature of factors at the lower trait level (Mount et al., 2003). For example, Barrick et al. (2001) note that the number of lower-level constructs in four popular inventories based on the five-factor model range from 12 to 45. In the study reported in this chapter, personality scales that were designed to assess the 30 lower-level constructs of the NEO PI-R were employed. This is a good choice as the instrument itself is a popular measure that is often used in both research and practice, and there is a good deal of evidence to support the reliability and validity of its lower-level facets (see Costa & McCrae, 1992). Furthermore, given the overlap between the constructs assessed by the scales from various personality inventories, the results obtained here are also likely to have some generality to scales from other inventories.

A second issue relates to the type of criterion that is employed. Broad and narrow personality variables have been compared with respect to the prediction of many socially significant criteria, such as dating frequency, drug use, alcohol consumption, occupational profiles, workplace delinquency, and traffic violations (Ashton, 1998; Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Goldberg, in press; Mershon & Gorsuch, 1988; Paunonen, 1998; Paunonen et al., 2003). Consistent with the focus of this thesis, here we are concerned with broad work performance criteria, such as overall measures of academic performance, training performance and job performance. Broad performance measures are frequently used in organisational psychology research, and often form the basis for important decisions about employees. Moreover, it has been contended that narrow personality variables are not suited to the prediction of such broad criteria (e.g., Ones & Viswesvaran, 1996), a claim that requires empirical verification.

In the study reported in this chapter (Study 5), I first review the arguments and evidence relating to the merits of broad versus narrow personality variables for the purposes of predicting broad work performance criteria. Following this a rationale is provided for comparing linear regression and neural networks in the context of narrow personality variables. The findings from Study 5 are then presented and discussed.

Study 5

Facet scores were available for three of the datasets presented in chapter 3 (Datasets 1, 2, and 4), and therefore these samples were used as the basis of the analyses in the present study. One aim of the study was to compare the predictive performance of broad versus narrow personality measures with respect to the prediction of broad work performance criteria. The second aim was to compare the predictive performance of artificial neural networks and linear regression for predicting work performance when narrow personality variables are used as predictors. The background literature and expectations in relation to each of these aims is outlined in more detail below.

Broad Versus Narrow Representation of Personality

In discussing the relative merits of representing personality variables either broadly or narrowly it is useful to first note that the total variance in a set of predictor variables can be attributed to a) variance that is common to two or more variables, b) non-error variance that is specific to a variable, and c) variance due to measurement error. Whereas measurement error limits the accuracy of predictions that can be achieved, predictive performance is facilitated to the extent that either the common or specific non-error sources of variance overlap with the criterion.

As broad personality scales are typically derived by summing scores on a number of narrower facet scales, they contain more items than their narrower counterparts and are therefore more reliable measures of the common variance underlying the facets (e.g., Ones & Viswesvaran, 1996). To the extent that it is the common variance that overlaps with the criterion, the greater reliability of broad measures facilitates their ability to predict the criterion. Furthermore, based on proposals regarding the appropriate bandwidth of predictors (see Cronbach, 1960), proponents of the broad approach have argued that optimal prediction is achieved when there is a match between the breadth of the predictor and that of the criterion, and therefore that any single personality facet is unlikely to correlate highly with a multifaceted criterion such as job performance (e.g., Hogan & Roberts, 1996; Ones & Viswesvaran, 1996). For example, Ones & Viswesvaran (1996) suggest that the factorial purity of a narrower personality measure will lower correlations with broad criteria because the narrow personality measure is likely to share its variance with only one of the dimensions of the factorially complex criterion, whereas a broad predictor is likely to tap into multiple dimensions of the criterion.

On the other hand, one loses information about the specific non-error variance associated with lower-level facets when they are aggregated to the level of the five broad dimensions (Ashton et al., 1995). Correlations between facet scales within the same broad personality dimension are often only low to moderate in magnitude (see Costa & McCrae, 1992 for examples from the NEO PI-R), suggesting that much of the variance captured by the facet scales is specific to those scales. To the extent that the specific variance components of the facets overlap differentially with the criterion, predictive performance may be diminished through aggregation (Paunonen et al., 1999; Van Iddekinge, Taylor, & Eidson; 2005). For example, take a situation where some of the facets within a broad dimension are positively related to the work performance criterion (i.e., are desirable traits for that occupation) and others are negatively related to the criterion (i.e., are undesirable traits). An individual can obtain the same score on the broad personality scale either as a result of scoring highly on the desirable facets and moderately on the undesirable ones, or else moderately on the desirable facets and highly on the undesirable ones. An individual with the former personality profile is likely to perform better than an individual with the latter profile, yet one cannot distinguish between the two based purely on their score on the broad measure. Similarly, when only a subset of the facets are related to the criterion, aggregating the relevant and irrelevant facets into a broad measure can make it difficult to distinguish between those who possess the relevant lower-level traits and those who do not.

The discussion above suggests that the usefulness of the narrow level of representation compared to the broad approach depends on the extent to which the facets within a broad dimension tap into specific aspects of the construct that in turn are differentially related to the criterion. As this is likely to differ across dimensions and across different performance contexts, it is unlikely that one approach will be more useful than the other for all five factors in all performance contexts. Consequently, the conceptual arguments and empirical evidence regarding the relative usefulness of the broad and narrow approaches are summarised separately for each factor below, with particular emphasis on the performance contexts investigated in the present study.

Neuroticism. The lower-level facet scales of the Neuroticism factor tend to be highly intercorrelated, at least with respect to the NEO PI-R instrument.[1] This suggests that the common source of variance captured by these scales is large relative to the specific variance captured by each facet, and therefore that there will be few criteria that are differentially related to the facets of this factor. Indeed it has recently been argued that the Neuroticism factor itself is too narrowly defined for the purposes of predicting job performance, and that the construct should be broadened to incorporate traits such as locus of control, self-esteem, and generalised self-efficacy (Judge, Van Vianen, & De Pater, 2004). Possibly for these reasons Neuroticism and its lower-level facets have received little empirical attention with respect to the broad versus narrow issue within

[1] The average correlation between the facet scales of the Neuroticism factor (r = .48), calculated using the intercorrelation matrix presented in Costa and McCrae (1992), was higher than that of any other factor. The average facet intercorrelations for Extraversion, Openness, Agreeableness, and Conscientiousness were .35, .28, .33, and .45.

performance contexts. In one of the few studies that included measures of this factor, Stewart, Barrick, and Parks (2003) compared the narrow and broad measures of the Personal Characteristics Inventory using a sample of 110 participants from a wide dispersion of job types, and concluded that there was generally no evidence that the narrow facets were superior predictors for predicting job performance ratings. On the other hand, Chamorro-Premuzic and Furnham (2003) used the aggregated examination grades of British university students as the criterion, and obtained some support for the usefulness of the Neuroticism facets from the NEO PI-R. The validity coefficients for these facets ranged from r = .13 for the Self-Consciousness scale to r = -.29 for the Anxiety scale, suggesting that at least some of the facets are differentially related to academic performance, and therefore that there may be a benefit associated with preserving the specific variance captured by the facet scales. Moreover, the latter correlation was larger in magnitude than that obtained by the broad Neuroticism measure (r = -.16). The strong effect of the Anxiety scale relative to the broad Neuroticism measure in this context is perhaps due to the nature of the criterion, which consisted entirely of performance in examinations. The negative effects of anxiety under exam conditions are well established, whereas there is little evidence to suggest that subclinical levels of other negative emotions such as depression or anger are related to test performance in this way (see Zeidner & Matthews, 2000).

Extraversion. In contrast to the Neuroticism factor, the facets of Extraversion tend to only be moderately correlated (average r = .35, see footnote 1 on page 5). This raises the possibility that these facets are more strongly defined by their specific variances, which in turn may overlap differentially with a given criterion. Within performance contexts such differential relationships between the facets of the Extraversion factor would often be expected on rational grounds. For example, traits such as being energetic and assertive are likely to facilitate performance in many contexts, whereas craving excitement is not; and the importance of being sociable is likely to vary from one context to the next. In support of this assertion, when Costa, McCrae, and Kay (1995) asked ten experts to rate the desirability of the various characteristics assessed by the NEO PI-R facets for entry-level police they found that within the Extraversion domain assertiveness was rated on average as between somewhat desirable to very desirable, whereas excitement-seeking was rated as somewhat to very undesirable. The other facets within this domain were rated as being between irrelevant to somewhat desirable.

There is also empirical support for the claim that the facets of Extraversion are differentially related to work performance. Hough (1992) distinguished between the traits of Extraversion related to *potency* (such as impact, energy, and influence) and the traits associated with *affiliation* (such as sociability and friendliness). In a review of 237 studies, she found that measures of potency were positively related to overall job performance across occupations, whereas measures of affiliation were unrelated to job performance. The findings from subsequent studies have largely been consistent with Hough's results. Rothstein, Paunonen, Rush, and King (1994) found that the Dominance and Exhibition scales of the Personality Research Form were significantly and positively associated with the GPA of graduate business school students, whereas the Affiliation scale and a broad Extraversion measure were not. Vinchur et al. (1998) included measures of the affiliation and potency components of Extraversion in their meta-analysis of the predictors of sales performance, and found that the validity of potency measures for predicting ratings of sales performance (corrected $r = .28$) was higher than that of the global Extraversion measure (corrected $r = .18$), whereas the validity coefficient for affiliation was lower (corrected $r = .12$). Stewart, Barrick, and

Parks (2003) failed to obtain statistically significant relations between overall job performance and any of the Extraversion facets of the Personal Characteristics Inventory, however the Surgency and Sociability scales were related to job performance in the opposite direction ($r = .09$ and $r = -.16$), and both facets were more strongly related to job performance than the broad Extraversion composite ($r = -.03$). Chamorro-Premuzic and Furnham (2003) found that the Activity and Gregariousness facet scales of the NEO PI-R were negatively related to university examination marks, whereas the other facets and the broad Extraversion measure were unrelated to academic performance. Additionally, based on a review of 43 studies that had used the Hogan Personality Inventory to predict job performance, Hogan and Holland (2003) concluded that job performance is predicted by the Ambition scale (which assesses characteristics associated with the potency component of Extraversion), but not with the Sociability scale. Therefore, taken together, the results with respect to the Extraversion factor suggest that the component of this factor to do with potency is typically more strongly related to performance criteria than the component to do with affiliation, and that often the former produces a higher validity coefficient than the corresponding broad Extraversion measure.

Openness. Of the five factors, the facets of the Openness factor are the least intercorrelated (average $r = .28$, see footnote 1 on page 5), and therefore may possess enough specific non-error variance to outpredict their sum. Openness is particularly relevant in situations where performance is contingent on learning new skills and acquiring knowledge, such as on training programs or in academic settings. Within such contexts one might expect intellectual curiosity to be more important for predicting performance than other aspects of the factor, such as receptivity to one’s inner feelings or sensitivity to art and beauty. Indeed, intellectual curiosity has been rated as the most desirable traits of the Openness factor for entry-level police, whereas other traits such as being open to fantasy have been rated as somewhat undesirable (Costa et al., 1995). Similarly, based on rational grounds, expert judges have rated measures of the intellectual curiosity facet of Openness, such as the Understanding scale of the Personality Research Form and the Ideas scale of the NEO PI-R, as being among the most useful lower-level traits for the purposes of predicting academic performance (e.g., Paunonen & Ashton, 2001a). Paunonen and Ashton (2001b) empirically compared the Understanding scale of the Personality Research Form to a broad Openness composite, and found that the validity coefficient of the narrow measure ($r = .23$) was significantly greater than that of the broad measure ($r = -.04$) when predicting university course grades. In contrast, Chamorro-Premuzic and Furnham (2003) obtained low and nonsignificant validity coefficients for the broad Openness measure and all of its facets when predicting university examination marks.

Outside of academic contexts, one may still expect differential relationships between the various facets of Openness and job performance, although facets of the Openness factor other than intellectual curiosity probably also play an important role. For example, taxonomies of managerial competencies emphasise the importance of various aspects of open-mindedness such as tolerance, adaptability, and creative thinking (Tett, Guterman, Bleier, & Murphy, 2000), not just intellectual curiosity. In a recent study, Griffin and Hesketh (2004) hypothesised that openness to external experiences (as measured by a composite of the Adventurousness, Intellect, and Liberalism facet scales of the IPIP-NEO) would be positively related to job performance whereas openness to internal experiences (measured by a composite of the Imagination, Artistic interests, and Emotionality facet scales) would be negatively related to job performance. They tested their hypotheses using a combined sample of 186 employees from two organisations. Neither of the two subdimensions nor the Openness composite was significantly correlated with job performance ratings, however the Intellect facet scale was significantly related to both task performance and adaptive performance ratings. Interestingly then, this study provided further support for the importance of the intellectual curiosity facet of Openness for predicting performance, even outside of academic contexts.

Agreeableness. Of the five factors, Agreeableness is of least interest in the present study. This factor had not been hypothesised to be theoretically relevant in any of the datasets used in this study (see Table 4.1 in chapter 4), and the few empirical studies that have compared the broad and narrow scales within this factor have found little evidence that any of the narrow facets is more valid than the broad measure for predicting either academic performance or job performance (e.g., Chamorro-Premuzic & Furnham, 2003; Stewart et al., 2003). Consequently, it was not expected that representing these facets narrowly would provide any advantage over a broad representation for the datasets used in this study.

Conscientiousness. Most of the research that has compared broad and narrow personality measures for predicting performance criteria has concentrated on the Conscientiousness factor. This is perhaps surprising given that the traits associated with this factor – such as competence, orderliness, dutifulness, achievement-striving, self-discipline, and deliberation – would in most performance contexts be expected to be related to performance in the same direction, namely positively. Furthermore, the facets are highly intercorrelated (average r = .45, see footnote 1 on page 5), suggesting that they are predominantly tapping into a common source of variance. Hough (1992) presented evidence that measures of the achievement facet of Conscientiousness were more strongly related to overall job performance than measures of the dependability facet. Her study, however, did not include a broad measure of Conscientiousness and therefore it was not possible to determine whether the narrow facets outperformed the broad measure. A number of subsequent meta-analyses have included a broad Conscientiousness measure and have found that this measure typically results in validity coefficients that are comparable to the best performing narrow measure. For example, Mount and Barrick (1995) compared measures of the Conscientiousness factor to measures of the narrower traits of achievement and dependability across a number of occupations, and found that the validity coefficients of the narrow traits were similar to each other and to that of the broad Conscientiousness when the criterion was a broad measure of work performance. Similarly, a meta-analysis of the predictors of salesperson performance conducted by Vinchur et al. (1998) obtained comparable validity coefficients for measures of the Conscientiousness factor and the achievement trait when predicting job performance ratings (corrected validity coefficients = .21 and .25), although the narrow trait outperformed the broad dimension when specifically predicting sales volume (corrected coefficients = .31 and .41). A more recent meta-analysis by Dudley, Orvis, and Lebiecki (2003) derived validity coefficients for four narrow traits of Conscientiousness (achievement, dependability, order, and cautiousness) in four occupational categories (sales, customer service, managers, and skilled/semi-skilled workers). Although the narrow trait with the highest validity coefficient varied across the four occupational categories, in all four categories the validity coefficients for the four narrow traits were smaller than or at best similar to that of a global Conscientiousness measure when predicting overall job performance.[2]

---

[2] It should be noted that the findings reported here do not imply that the specific variance associated with the narrow traits of Conscientiousness are unrelated to overall job performance. On the contrary, it has been shown that the narrow measures account for statistically significant amounts of variance in work performance above that accounted for by the broad measure (e.g., Dudley et al., 2003; Stewart, 1999). However, the results do suggest that if one is to select a single predictor for predicting overall

Similar results have been obtained when the Conscientiousness factor has been studied within academic settings. Paunonen and Ashton (2001a) found that the validity coefficient of the Achievement scale of the Personality Research Form (r = .26) was higher than that of a Conscientiousness composite derived from the same inventory (r = .21) when predicting course grades, although this difference was not statistically significant. Gray and Watson (2002) found that the validity of the Achievement-Striving (r = .39) and Self-Discipline (r = .36) scales of the NEO PI-R for predicting university GPA were significantly greater than those of the other Conscientiousness facets, although the former was only slightly more accurate than the Conscientiousness composite (r = .36). Lievens, Coetsier, Fruyt, & De Maesseneer (2002) concluded that the Achievement-striving and Self-Discipline scales predict medical student performance better than some of the other Conscientiousness facet scales from the NEO PI-R, yet they found little evidence that either facet scale outpredicts the broader domain scale. Chamorro-Premuzic and Furnham (2003) found that within the Conscientiousness domain the Dutifulness and Achievement-Striving scales were the facets that were the most strongly related to performance (r = .38 and .35). The validity coefficients for these two scales were approximately the same as those for the Conscientiousness composite (r = .36).

To summarise, the empirical studies that have compared broad and narrow personality measures as predictors of performance criteria have largely focused on the Conscientiousness factor, and to a lesser extent on the Extraversion factor. Although differences between the narrow facets of Conscientiousness have been found, the validity of broad Conscientiousness is typically comparable to that of the best

---

performance, there is little if any additional gain associated with selecting a narrow facet of Conscientiousness over a more global measure of the construct.

performing narrow facet. In contrast, often at least one of the lower-level traits of Extraversion – typically a trait associated with the potency component – produces a higher validity coefficient than the corresponding broad measure. Less research has been conducted using the other factors, however of the work that has been done there is some indication that the facet of Openness that is related to intellectual curiosity may be more valid for predicting performance than broad measures of Openness, and that the anxiety facet of Neuroticism may be a more valid predictor of performance in academic settings (where performance is often assessed by exams) than the corresponding broad measure.

A methodological difficulty that arises when comparing broad and narrow variables, and that is a limitation of much of the research cited above, stems from the disparity between the number of predictors at the broad and narrow levels. The larger number of variables for the latter increases the likelihood that a narrow variable will outperform the corresponding broad counterpart, though purely due to chance. One could address this problem by reducing the number of narrow traits that are examined, for example by hypothesising about and selecting the narrow facet within each domain that would be expected to be the best predictor of the criterion based on previous empirical findings or on rational grounds (see Paunonen & Ashton, 2001b). This hypothesis-driven approach, however, entails the loss of the specific information captured by all but the selected facet, and can be problematic if there is little reason for preferring one facet over the others, or else if multiple facets within a domain are expected to be related to the criterion though in the opposite direction. A second approach is to combine the information from the narrow facets using multiple regression, and to then cross-validate these equations on unseen data (Paunonen et al., 1999). This approach has the advantage of retaining and incorporating the specific information captured by the multiple narrow facets. However, it has the drawback of increasing the number of parameters that need to be estimated. As discussed in chapter 5, increasing the number of parameters to be estimated also increases the likelihood of capitalising on chance, and therefore that predictive performance will be lower upon cross-validation. Consequently, under this methodological approach the broad versus narrow debate takes on a different flavour: The benefit of retaining the specific information associated with lower-level facet scales needs to outweigh the risk of increased capitalisation on chance if the narrow approach is to yield more accurate predictions for unseen cases. With limited sample sizes this is an issue that needs to be assessed empirically (see Goldberg, 1993).

Variants of both approaches were used in the present study. Consistent with the latter approach, prediction equations that included all the facets within a broad domain were developed and subsequently tested on unseen data, and these results were compared to the findings for broad personality variables that were presented in the previous chapters of this thesis. Additionally, consistent with the hypothesis-driven approach, there was an expectation that specific facets within some of the domains would outperform the broad domain measure based on the conceptual arguments and empirical studies that were summarised above, and consequently analyses were performed to test these specific hypotheses:

1. It was hypothesised that, within the Neuroticism domain, the anxiety facet of this factor would be a more valid predictor of performance than the broad measure in Datasets 1 and 2, as the criteria in these datasets were primarily composed of performance under exam conditions.

2. It was hypothesised that, within the Extraversion domain, facets associated with the potency component would be more valid predictors of performance than the broad measure in all three datasets.
3. It was hypothesised that, within the Openness domain, the facet of this factor related to intellectual curiosity would be a more valid predictor of performance than the broad measure in all three datasets.

Based on the literature that was reviewed, it was not expected that any of the facets of the Agreeableness or Conscientiousness factor would yield significantly greater validity coefficients than the corresponding broad measure, and therefore no specific hypotheses were proposed for these factors.

Neural networks and Narrow Personality Variables

Given the limited evidence for nonlinear and configural relationships between broad personality measures and work performance, and the consequent poor predictive performance of neural networks in the context of broad personality measures, is there any reason to expect neural networks to outperform linear regression when narrow personality variables are used as the predictors? One possibility is that nonlinear and/or configural relationships exist at a narrower level of personality representation but that these relationships are submerged when the narrow traits are aggregated into broad composites. For example, Murphy's (1996) arguments that Conscientiousness and Extraversion may be quadratically related to performance are expressed in terms of specific facets within each domain, rather than the domain in its entirety. Namely, for certain occupations job performance may be hampered if an individual is excessively rule-bound (a behavioural response that is presumably associated with high levels of the cautiousness facet of Conscientiousness), or else if they spend too much time interacting with others at the expense of their own work (a behavioural response that is presumably associated with high levels of the sociability facet of Extraversion). The same type of reasoning is not easily applied to other facets within these domains, such as the achievement facet of Conscientiousness or the activity facet of Extraversion, which are more likely to be monotonically related to performance. Similarly, it is specifically the anxiety trait of Neuroticism that has been proposed to be related to academic performance and job performance in an inverted-U fashion (e.g., Braden, 1995; Stewart et al., 2003). This is because this trait is associated with arousal, which is most useful in moderate amounts for the purposes of facilitating performance (Matthews & Deary, 1998). Other traits of Neuroticism, such as depression or hostility, are unlikely to be related to performance in this way. If the broad approach was used for the examples cited above, the variability added by the other facets that are aggregated to obtain the broad measure would obscure the nonlinear relationship between the specific facet and work performance, and consequently there would be little benefit in using neural networks rather than linear regression. On the other hand, if the personality variables were represented at the facet level, then a neural network with an adequately large number of hidden units could detect the specific nonlinear relationships and potentially improve predictive performance.

It is also not difficult to think of circumstances where configural relationships are submerged by the aggregation of narrow traits into broad measures. This can occur in at least two ways. First, if the configural relationship occurs between two facets within the same personality dimension then such a relationship may only be detected and exploited for predictive purposes by representing each facet separately. For example, it may be that success in managerial jobs requires at least moderate levels of *both* sociability and assertiveness, and conversely that a low level of either facet results in the failure to perform well (regardless of the level of the other facet). As these two facets of the Extraversion domain are not strongly correlated (see Costa & McCrae, 1992), it is possible to obtain the same broad Extraversion score either by scoring moderately on both facets or by scoring high on one facet and low on the other. However, it is only the former configuration that is associated with successful performance in this case, and therefore representing the Extraversion factor broadly would not allow one to exploit this configural relationship in order to distinguish between the high and low performing manager.

Second, configurality may exist between specific facets that are from different broad dimensions. For example, in the previous chapter it was suggested that in the context of training programs Openness and Extraversion represent alternative motivations to actively participate in the program and ask questions, and that high levels of one factor may compensate for low levels of the other such that optimal performance is achievable as long as at least one factor is high. However, within the Openness domain, it is specifically intellectual curiosity that is likely to motivate the individual in this way; whereas other attributes associated with this factor such as being open to one's feelings or having a deep appreciation for art and beauty are unlikely to provide this motivation. Similarly, within the Extraversion domain, some facets (eg., activity and assertiveness) are likely to be more relevant as motivating factors in training contexts than other facets (e.g., warmth). To the extent that it is the specific facets that form the basis of the configural relationships, such relationships are more easily discernible if personality variable are represented at the narrower facet level rather than at the broad level, and consequently the narrow level of representation may well provide a context in which neural networks outpredict linear regression.

Of course there is also a cost associated with the neural network method in that it is more likely to capitalise on chance than linear regression, and this cost is likely to be greater in the context of narrow predictors than broad ones given the larger number of narrow predictors and the typically lower reliability of these measures. Consequently the benefit of detecting any systematic nonlinearity or configurality in the data needs to be strong enough to outweigh this cost in order for neural networks to be useful in this context. This too, therefore, is an issue that needs to be assessed empirically.

Method

Scores on the 30 lower-level facet scales of the IPIP-NEO or NEO PI-R were available for each of the participants in Datasets 1, 2 and 4, and these scores were used as the basis of the analyses performed in this study.[3] Prediction equations were developed and tested for each of the combination of personality facets listed below, using the same procedures and the same twenty training/test set partitions that were used in all previous studies, with the exception that in the present study only one level of hidden units (the H4 level) was used for the neural networks developed without early stopping. Thus, for each combination listed below, the twenty training/test sets were used to develop and test linear regression equations, neural networks without early stopping (with H4 hidden units), and early stopping committees (also with H4 hidden units).

Within each dataset prediction equations were developed using the following six combinations of the facet scales:

1. All facets of the Neuroticism factor.
2. All facets of the Extraversion factor.

[3] The labels for the facet scales of these instruments can be found in Table 3.2 of chapter 3.

3. All facets of the Openness factor.
4. All facets of the Agreeableness factor.
5. All facets of the Conscientiousness factor.
6. All facets of the combination of theoretically relevant factors.

The first five combinations consisted of six predictor variables each, and allowed me to address the broad versus narrow issue separately for each of the five factors. However, it is not possible to capture configural relationships that may exist between facets from different factors with these combinations, which is relevant for the analyses that compared linear regression and neural networks. For this reason a sixth combination was developed for each dataset that consisted of all the facets from the combination of theoretically relevant factors tested in Study 3. For Datasets 1 and 4 this latter combination consisted of 18 predictors, whereas for Dataset 2 it consisted of 24 predictors.[4]

To address the first aim of this study, the predictive performance of the linear regression equations developed here (which used the facet scales as predictors) were compared to the predictive performance of the corresponding linear regression equations that had been developed in chapters 4 and 5 (which used the broad domain measures as predictors). Furthermore, the validity coefficients for individual facets were compared to the corresponding broad domain-level validity coefficients in order to test the specific hypotheses that had been proposed. To address the second aim of the study, the predictive performance of the neural networks and linear regression equations developed in the present study were compared to each other.

[4] Tables C60 to C77 of Appendix C provide the predictive performance results for each prediction equation developed in Study 5.

Results and Discussion

Broad Versus Narrow Representation of Personality Variables

Table 6.1 presents the MAE values and cross-validity coefficients for the linear regression equations that were developed using the narrow facet variables of the NEO inventories. The results for the corresponding broad linear regression equations, which were initially documented in chapters 4 and 5, are also presented here for the purposes of comparison. The findings are presented for each of the five factors within each dataset, and for the combination of theoretically relevant factors within each dataset.

Table 6.1

MAE values and cross-validity coefficients for the linear regression equations that were developed using either the broad (domain) variable(s) or the narrow (facet) variables within each domain.

Dataset and	MAE		Cross-validity coefficient	
predictor domain	Broad	Narrow	Broad	Narrow
1. University students				
Neuroticism	7.395	7.166	.02	.19
Extraversion	7.289	7.321	.16	.16
Openness	7.329	7.232	.13	.20
Agreeableness	7.401	7.595	-.08	-.03
Conscientiousness	7.226	7.325	.22	.16
Theoretical Combination	7.121	7.078	.26	.29
2. Police recruits				
Neuroticism	32.39	33.06	.13	.08
Extraversion	32.53	32.09	.16	.23
Openness	32.97	32.69	.05	.14
Agreeableness	32.53	32.84	.09	.07
Conscientiousness	31.55	32.62	.26	.18
Theoretical Combination	32.24	33.65	.22	.19
4. Managers				
Neuroticism	0.488	0.498	.19	.15
Extraversion	0.505	0.510	.10	.14
Openness	0.507	0.496	-.07	.19
Agreeableness	0.509	0.520	.01	.02
Conscientiousness	0.480	0.484	.26	.26
Theoretical Combination	0.484	0.503	.25	.20

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Table 6.1 indicates that of the 18 comparisons between the broad and narrow levels of representation, the narrow approach produced lower MAE than the broad approach 6 times, and higher cross-validity coefficients 9 times. In Dataset 1 the largest gains in cross-validity coefficients occurred for the Neuroticism and Openness factors, and in Datasets 2 and 4 the largest gains occurred for the Openness and Extraversion factors. Importantly, therefore, the results for the Openness factor, and to a lesser extent the Extraversion factor, showed some consistency across the datasets. Consequently, although the results cannot be taken as evidence for the general usefulness of representing personality variables at the facet level, they do indicate that for these two domains a narrow level of representation may be more appropriate, and that within academic settings the Neuroticism factor too may benefit from the narrow approach. In contrast, there was no benefit associated with the narrow approach within the Conscientiousness domain – in all three datasets MAE was higher and average cross-validity coefficients were lower for the narrow level of representation. Finally, with respect to the Agreeableness factor, the narrow approach resulted in small improvements in cross-validity coefficients (but not MAE) in two of the datasets, however despite these improvements the magnitude of the coefficients remained trivially small.

To address each of the specific hypotheses that were previously outlined, the validity coefficient for the relevant narrow measure was compared to validity coefficient for the corresponding broad measure (both derived using the entire data in each dataset) and tested for significance using Meng, Rosenthal and Rubin’s (1992) procedure for comparing correlated correlation coefficients. Given the directionality of the hypotheses, one-tailed tests were employed.

Hypothesis 1. To test the hypothesis that the narrow anxiety trait is a more valid predictor than broad Neuroticism when performance is predominantly assessed under exam conditions, the validity coefficient for the Anxiety facet scale and Neuroticism domain scale were compared in Datasets 1 and 2. In Dataset 1 the validity coefficient for the Anxiety scale ($r = .07$) was slightly higher in magnitude than the validity coefficient for Neuroticism ($r = .05$), although this difference was not statistically significant ($z = -0.58$, $p > .05$). In Dataset 2 the validity coefficient for Anxiety ($r = -.10$) was smaller in magnitude than that of Neuroticism ($r = -.16$). Hypothesis 1 was therefore not supported.

Hypothesis 2. The second hypothesis was that the potency component of Extraversion would better predict performance than the corresponding broad Extraversion measure. There is no one facet within the NEO PI-R that assesses potency, although the Assertiveness and Activity scales closely resemble descriptions of the potency component (see Hough, 1992), and therefore this hypothesis was tested by aggregating scores on these two facets within each dataset to obtain a measure of potency.[5] In Dataset 1 the magnitude of the validity coefficient of the potency measure ($r = -.02$) was smaller than that of Extraversion ($r = -.13$). However, in Datasets 2 and 4 the narrow measure was more strongly related to performance ($r = .25$ and $.23$) than was the broad measure ($r = .16$ and $.11$), and both these differences were statistically significant ($z = -2.11$ and -1.99, $p < .05$). Therefore, there was partial support for hypothesis 2.

Hypothesis 3. The third hypothesis was that the facet of Openness specifically to do with intellectual curiosity would better predict performance than the corresponding

[5] Additional support for the appropriateness of summing these two scales comes from principal components analyses of the Extraversion facet scales within the three datasets considered here. In each dataset there was justification for the extraction of two components, one of which was most strongly defined by the Assertiveness and Activity facets after direct oblimin rotation.

broad measure. Within the NEO PI-R, intellectual curiosity is assessed by the Ideas scale, and within the IPIP NEO it is assessed by the scale labelled Intellect. In Dataset 1, the validity coefficient for the Intellect scale ($r = .20$) was higher in magnitude than that of the broad Openness measure ($r = .12$), and this difference was statistically significant ($z = -1.80$, $p < .05$). Similarly, in Datasets 2 and 4, the Ideas scale was more strongly related to performance ($r = .24$ and $.19$) than was broad Openness ($r = .10$ and $.02$), and both differences were statistically significant ($z = -2.79$ and -2.96, $p < .01$). Hypothesis 3 was therefore supported.

Exploratory analyses. To further explore the data, Table 6.2 presents the validity coefficients for each personality scale within the three datsets. The validity coefficients for the broad scales are presented in bold. The validity coefficients provide some insight into why the narrow approach outpredicted the broad approach for the factors in which it did. For example, for the Neuroticism factor in Dataset 1, it can be seen that the superior predictive performance of the narrow approach was not driven by the Anxiety scale, as had been hypothesised, but rather by a positive correlation between the Self-Consciousness scale and the criterion (see Table 6.2). The validity coefficients for the other facets within this factor were trivially small, and so it is not surprising that when the six facets were combined into a broad measure the coefficient for this measure was also trivially small. One can speculate about why the Self-Consciousness scale was more strongly related to performance than the other facets of Neuroticism. One possibility is that the results were obtained purely due to chance. However, Chamorro-Premuzic and Furnham (2003) also obtained a significant positive correlation between this facet and academic performance that did not occur for any of the other Neuroticism facets, and that was strongest for first year university students (the same population examined in Dataset 1). Therefore, it is plausible that this facet is capturing specific

Table 6.2

Validity coefficients for the personality scales within each dataset.

Personality scale	Dataset 1 University students	Dataset 2 Police recruits	Dataset 4 Managers
Neuroticism	**.05**	**-.16**	**-.20**
N1: Anxiety	.07	-.10	-.11
N2: Anger	.02	-.13	.00
N3: Depression	.01	-.12	-.17
N4: Self-Consciousness	.19	-.07	-.18
N5: Immoderation/Impulsiveness	-.03	-.17	-.17
N6:Vulnerability	.00	-.17	-.26
Extraversion	**-.13**	**.16**	**.11**
E1: Friendliness/Warmth	-.15	.02	.03
E2: Gregariousness	-.17	.05	.00
E3: Assertiveness	-.04	.19	.25
E4: Activity Level	.01	.24	.14
E5: Excitement-Seeking	-.13	.07	.03
E6: Cheerfulness/Positive Emotions	-.06	.07	.03
Openness	**.12**	**.10**	**.02**
O1: Imagination/Fantasy	.16	-.02	-.09
O2: Artistic Interests/Aesthetics	-.01	.01	-.03
O3: Emotionality/Feelings	.05	.05	.04
O4: Adventurousness/Actions	-.03	.05	-.03
O5: Intellect/Ideas	.20	.24	.19
O6: Liberalism/Values	.07	.05	-.01
Agreeableness	**.03**	**.11**	**-.08**
A1: Trust	-.04	.18	-.04
A2: Morality/Straightforwardness	.11	.07	-.06
A3: Altruism	-.03	.07	.03
A4: Cooperation/Compliance	.02	.06	-.15
A5: Modesty	.02	.02	-.08
A6: Sympathy/Tender-Mindedness	.05	.01	-.01
Conscientiousness	**.20**	**.27**	**.28**
C1: Self-Efficacy/Competence	.14	.24	.18
C2: Orderliness/Order	.15	.20	.06
C3: Dutifulness	.10	.22	.21
C4: Achievement-Striving	.14	.22	.27
C5: Self-Discipline	.12	.25	.22
C6: Cautiousness/Deliberation	.21	.15	.26

Note: The validity coefficients for the broad measures are presented in boldface.

information that is relevant for academic performance in the first year of university and that is not captured by the other facets of this factor. Unlike the other traits of the Neuroticism factor, self-consciousness is specifically characterised by negative emotions towards social situations, having been described as akin to social anxiety or shyness (Costa & McCrae, 1992). The Self-Consciousness scale may therefore provide information about participation in social activities that is relevant to the academic success of first-year university students. Researchers have noted that the developmental transitions experienced by first-year university students can result in a distracting social environment, and that the extent to which students engage in this environment is negatively related to their first-year academic performance (Bauer & Liang, 2003). Thus, the positive correlation between this scale and academic performance may be due to the dissuading effects of being highly self-conscious on participation in social activities. As support for this claim, note that facet scales of Extraversion that are also likely to provide information about participation in social activities (though in the opposite direction to the Self-Consciousness scale), such as the Gregariousness, Friendliness, and Excitement-Seeking scales, were negatively related to academic performance in Dataset 1.

In contrast to the Neuroticism factor, within the Extraversion and Openness domains the superior predictive performance of the narrow approach was accounted for by the facets that had been hypothesised to be the strongest predictors of work performance. Thus, it can be seen from Table 6.2 that in Datasets 2 and 4 the scales that assess traits of Extraversion to do with potency, namely the Assertiveness and Activity scales, were more strongly related to the criterion than the other facets of Extraversion. In Dataset 2 it was the Activity scale that was the strongest predictor of performance, which is consistent with Barrick et al.'s (2001) theorising that Extraversion is related to performance in training programs because of the greater activity of highly extraverted trainees during training. In Dataset 4 the Assertiveness scale was the strongest predictor of performance, which is also not surprising given that high scorers on this facet often become group leaders (Costa & McCrae, 1992), and that leadership in turn has been rated as one of the most important attributes of good managers (see Raymark, Schmit, & Guion, 1997). Furthermore, given the positive effects of the Activity and Aseertiveness scales in Dataset 2 and 4, and that conceptually an active and assertive student might be expected to perform better than a passive student, one can speculate about why neither of these scales was positively related to performance in Dataset 1. A possible reason for this finding is that part of the variance captured by these scales is that which is common to the various facets of Extraversion; as alluded to above, being highly extraverted can be an undesirable quality for academic success in the first year of university, given the socially distracting environment. In support of this explanation, when the effect of broad Extraversion was statistically controlled in Dataset 1, the potency measure (the aggregation of the Assertiveness and Activity scales) was positively related to academic performance, $t(224) = 1.95$, $p = .05$, $\beta = 2.39$.

With respect to the Openness factor, and as had been hypothesised, the Ideas/Intellect facet was the strongest predictor of performance in all three datasets (see Table 6.2). The results for Datasets 1 and 2, therefore, support the intuition that intellectual curiosity is more relevant than other traits of Openness for performance in academic and training contexts, whereas the result for Dataset 4 suggests that it may also be more relevant outside of such contexts, within organisational settings. The superior validity of measures of intellectual curiosity for predicting performance may be explained by the relationship of this trait to cognitive ability. The Openness factor has traditionally been viewed as distinct from but related to cognitive ability, especially crystallised intelligence (Goff & Ackerman, 1992). Moreover, it is specifically the intellectual curiosity facet of Openness that is most strongly related to scores on cognitive ability tests (e.g., Ackerman & Goff, 1994; Ashton, Lee, Vernon, & Lang, 2000), and that has been suggested to contribute to the development of intellectual potential (Costa & McCrae, 1992). In light of the fact that cognitive ability tests are among the most valid methods for predicting performance across many different contexts (Schmidt & Hunter, 1998), it is not surprising that measures of intellectual curiosity too are useful for predicting performance across contexts. This point is discussed in greater detail in the general discussion of chapter 7.

Finally, within the Agreeableness and Conscientiousness domains, there was little evidence for differential validity coefficients among the facet scales. In all three datasets the validity coefficients for the broad Agreeableness measure and its narrower facet scales were mostly trivially small. Therefore these results are consistent with the earlier claim that this personality factor is not relevant for performance in the datasets considered here. On the other hand Conscientiousness is relevant for performance in all three datasets, as evidenced by the positive non-trivial validity coefficients for the narrow and broad scales of this factor (see Table 6.2). However, differences between the facets were typically small, and in all three datasets the broad measure obtained a correlation coefficient that was higher than or at least similar to that of the best narrow facet. This finding is consistent with much of the previous research on this factor (e.g., Mount & Barrick, 1995; Dudley et al., 2003), and suggests that it is the common variance underlying the facets of Conscientiousness that predominantly overlaps with work performance criteria.

Artificial Neural Networks Versus Linear Regression

Table 6.3 presents the MAE values and cross-validity coefficients for the neural networks that were developed using the six facet scales within each factor. The results for the corresponding linear regression equations, which were initially presented in Table 6.1, are also reproduced here. Consistent with the presentation in previous studies, the neural networks that were developed without early stopping are labelled ANN1, and the early stopping committees are labelled ANN2. However note that in previous chapters the former represented an average over different hidden unit levels whereas in the present study only the H4 level was included.

Table 6.3

MAE values and cross-validity coefficients for the linear regression equations and neural networks that were developed using the six facet variables within each factor.

Dataset and predictors	MAE			Cross-validity coefficient		
	LR	ANN1	ANN2	LR	ANN1	ANN2
1. University students						
Neuroticism facets	7.166	8.022	7.199	.19	.14	.19
Extraversion facets	7.321	8.223	7.342	.16	.05	.14
Openness facets	7.232	7.689	7.216	.20	.17	.19
Agreeableness facets	7.595	8.964	7.704	-.03	-.07	-.06
Conscientiousness facets	7.325	8.601	7.393	.16	.00	.14
Theoretical combination	7.078	16.254	7.174	.29	.09	.27
2. Police recruits						
Neuroticism facets	33.06	34.84	33.06	.08	.11	.08
Extraversion facets	32.09	35.58	32.38	.23	.10	.18
Openness facets	32.69	35.94	33.11	.14	.07	.12
Agreeableness facets	32.84	36.10	33.29	.07	.03	.03
Conscientiousness facets	32.62	35.38	32.71	.18	.11	.16
Theoretical combination	33.65	60.15	33.01	.19	.12	.20
4. Managers						
Neuroticism facets	0.498	0.762	0.501	.15	-.05	.12
Extraversion facets	0.510	0.674	0.509	.14	.04	.12
Openness facets	0.496	0.600	0.496	.19	.09	.17
Agreeableness facets	0.520	0.641	0.517	.02	.06	.08
Conscientiousness facets	0.484	0.557	0.487	.26	.06	.25
Theoretical combination	0.503	0.996	0.493	.20	.02	.20

Note: Underlined values indicate that the neural networks outperformed the associated linear equations.

Table 6.3 indicates that the networks developed without early stopping (ANN1) performed poorly compared to the linear regression equations. They produced higher MAE values than the linear regression equations for all 18 comparisons, and also produced lower cross-validity coefficients for 16 of the 18 comparisons. Differences between the two types of equations were particularly large for the comparisons involving the combinations of theoretically relevant factors, especially with respect to MAE. In all three datasets the difference in MAE was statistically significant, Dataset 1: corrected $t(19) = -7.10$, $p < .01$; Dataset 2: corrected $t(19) = -4.24$, $p < .01$, Dataset 4: corrected $t(19) = -4.66$, $p < .01$.[6] In these cases the number of predictors included in the equations was substantially greater than when each factor was considered separately, and therefore these results once again reflect the rapid deterioration in predictive performance that occurs for neural networks trained without early stopping when the number of predictors is increased. The result for the combination of theoretically relevant factors in Dataset 2 is an interesting case in point. In Study 3, where each of the four theoretically relevant factors was represented in terms of one global domain measure, it was found that the average cross-validity coefficient of the H4 networks was .04 of a unit higher than that of the linear regression equations (see Table 5.3 in Chapter 5). Whereas in this study, each of the four factors was represented in terms of six facets (that together comprised the same items that were used to derive the domain measure), and the average cross-validity coefficient for the networks was .07 lower than that of the linear regression equations (see Table 6.3). It would seem then that, for neural networks developed without early stopping, the increased capitalisation on chance that results

[6] These were the only MAE differences that were statistically significant at $\alpha = .01$, however a number of other statistically significant MAE differences (all in favour of the linear regression equations) were obtained at the less conservative type I error rate of $\alpha = .05$. None of the cross-validity coefficient differences were significant at the more conservative rate (see Tables D9 and D10 in Appendix D for complete results).

from the greater number of parameters to be estimated when representing personality variables narrowly more than offsets any benefit that arises from retaining the facet-specific information, at least in the present datasets.

The early stopping network committees (ANN2) were also typically less accurate than the linear regression equations. Of the 18 comparisons, they performed worse than linear regression 11 times with respect to MAE, and 15 times with respect to cross-validity coefficients. However, the differences in MAE values and cross-validity coefficients were generally smaller in magnitude than the differences between the ANN1 networks and linear regression equations. Indeed all comparisons between the ANN2 networks and linear regression were statistically nonsignificant with respect to both MAE and cross-validity coefficients.[7] Furthermore, the ANN2 networks were less likely than the ANN1 networks to be prone to the deterioration in predictive performance that occurs as a function of an increase in the number of parameters. Indeed, as can be seen in Table 6.3, if anything the ANN2 networks performed slightly better than the linear regression equations (and a lot better than the ANN1 networks) for the theoretically relevant combinations in Datasets 2 and 3, which involved 24 and 18 predictors respectively.

Conclusions

The study reported in this chapter obtained evidence that for some factors the predictive benefit of retaining the specific information associated with facets of the five broad factors outweighs the predictive cost of increased capitalisation on chance. For the Extraversion and Openness factors, and for Neuroticism within an academic context, predictive performance was higher when personality variables were represented at the

[7] The results of these statistical tests are provided in Table D11 and D12 of Appendix D.

narrower facet level than at the broad domain level. As had been hypothesised, within the Extraversion domain it was the scales associated with potency – namely the Assertiveness and Activity scales – that were typically more strongly related to the work performance criterion than the broad measure. Similarly, the hypothesis that the intellectual curiosity facet of Openness would outpredict the corresponding broad measure was also supported. However, for the Neuroticism factor within an academic setting, it was not the anxiety trait that accounted for the success of the narrow approach as had been hypothesised but rather the self-consciousness trait. It was suggested that the positive relationship between this trait and academic performance in the first year of university may be due to the dissuading effects of self-consciousness on participation in social activities.

The present study also provided evidence in relation to the usefulness of neural networks compared to linear regression in the context of narrow personality variables as predictors of work performance. As in the previous studies reported in this thesis, the neural networks generated predictions that were less accurate or at best similar in accuracy to those generated by linear regression, and these results held across different combinations of personality facets and different datasets. Furthermore, as in previous studies, the early stopping neural network committees tended to perform better than the networks developed without early stopping, and were also less prone to the adverse effects of capitalisation on chance that occurs when the number of parameters to be estimated is increased. Consequently, it would seem that the conclusions drawn about the comparative accuracy of linear regression and neural networks for broad personality variables also generalise to instances where personality variables are represented narrowly.

CHAPTER 7: General Discussion

Introduction

The research reported in this thesis compared artificial neural networks and linear regression as two methods of evaluating the effectiveness of personality variables for predicting work performance. Additionally, the nature and extent of the relationship between different personality variables and work performance was examined with special emphasis on the role of nonlinearity, configurality, and predictor breadth. It was expected that the findings would have implications for two broad areas of research, namely, the literature on organisational research methods and the literature on personality and work performance. This final chapter is divided into two parts in order to reflect this dichotomy. In the first section, the results relevant to the comparison between artificial neural networks and linear regression are reviewed and compared to previous findings. Some of the methodological strengths of the thesis that contribute to a better understanding of the findings are also discussed. The second section discusses the findings in terms of three issues relevant to the literature on personality and work performance. First, the evidence for complex relationships between personality and work performance is evaluated, and the theoretical and practical implications of the conclusions are considered. Second, differences between personality variables in the extent to which they predict performance and in their importance within theories of job performance are discussed, especially with respect to differences between broad and narrow variables. Third, the optimal number of personality variables to include in prediction equations is addressed. The chapter concludes with some recommendations for future applications of artificial neural networks to organisational psychology research.

Artificial Neural Networks Versus Linear Regression

Principal Findings

The results of the present research did not support the hypothesis that neural networks would produce more accurate predictions than the linear regression method. The networks for the most part performed poorly or at best similarly compared to linear regression equations. This finding generalised across hidden unit levels, two procedures of developing neural networks, and two measures of predictive performance. It occurred when individual personality variables were used as predictors and also when combinations of personality variables were used; in the latter case it occurred regardless of whether the combinations were selected based on theoretical grounds or by an empirical predictor-selection procedure. It occurred for broad personality variables as well as for variables that were represented at the narrower facet level. Furthermore, the results generalised across multiple datasets. Consequently, the present findings were inconsistent with the results of several previous studies that have found neural networks perform better than traditional linear methods when predicting the behaviour of employees in organisations (e.g., Scarborough, 1996; Somers, 1999, 2001).

There are at least two plausible explanations for the discrepancy in findings. First, the total number of cases in the six datasets used in this thesis ranged from 120 to 486. Although this represents a relatively large range that encompasses the sample sizes typically available in organisational psychology research, the size of the datasets used here are smaller than those in some of the previous studies in which neural networks have outperformed linear methods (e.g., Scarborough, 1996, Somers, 1999). Larger samples are likely to be of more benefit for complex methods such as neural networks than simpler methods such as linear regression (see Reed & Marks, 1999); in fact, theoretically, when sample sizes are unlimited a neural network will necessarily perform no worse than linear regression (see Geman et al., 1992). As a result, the findings of the present research are only generalisable to datasets that contain fewer than 500 participants. However, it should be noted that neural networks have previously been successfully implemented within organisational settings with far fewer than this many cases (Somers, 2001), and conversely that neural networks have failed to outperform linear methods in datasets with more than 18,000 participants (Paik, 2000). Therefore, factors other than sample size need to be considered to explain the poor performance of neural networks in the datasets used here.

A second explanation for the discrepancy in findings relates to differences in the predictors that were employed. In Chapter 2 I noted that previous successful implementations of neural networks employed predictors that were nonlinearly or configurally related to the criterion. For example, Somers' (2001) application of neural networks was based on the rationale that work attitudes measures are nonlinearly related to job performance. On the other hand, Griffin (1998) found that neural networks performed no better than linear regression for predicting naval aviator flight training performance and this was attributed to the linear structure of the underlying relations between the predictors and the criterion. As outlined in the introductory chapters of this thesis, a major motivation for the application of the neural network method in the present research was the expectation of nonlinear and configural relationships between personality variables and work performance criteria. Given the lack of evidence for complex relationships in the present datasets (see the discussion in the next section), it is not surprising that the neural networks generally failed to outperform the linear regression equations. In the absence of such relationships the hidden layer of a neural network can only serve to capture the noise in the data, and therefore has an adverse effect on predictive performance.

In support of this explanation, note that the neural networks outpredicted the linear regression equations in the small number of instances where statistically significant nonlinear or configural effects were present, such as for the measure of Intellectance in Dataset 5, the combination of Emotional Orientation, Cognitive Orientation, and Task Orientation in Dataset 6, and the combination of Neuroticism, Extraversion, Openness, and Conscientiousness in Dataset 2. Importantly, the neural networks were able to outperform linear regression in these cases without the researcher having to make assumptions about the nature of the underlying nonlinearity or configurality, thus highlighting the value of this methodology in situations where complex relationships occur but existing theories are not precise enough to specify the exact functional relationships. Therefore, although the usefulness of neural networks for improving the predictive performance of personality variables in the present datasets was disappointing, in a broader sense the present findings provide some grounds for optimism about the use of neural network methodology in organisational psychology research: The results demonstrate that neural networks can cope with the level of noise that is present in organisational data such that when the selected predictors are nonlinearly or configurally related to the criterion the networks are able to detect such relationships and exploit them for predictive purposes.

Further Methodological Contributions

There were three methodological strengths of the present research that together provided a better understanding of the generality of the findings, and facilitated the provision of a number of guidelines for the implementation of neural networks in future studies. These relate to the use of two alternate procedures for developing networks, the use of multiple measures of predictive performance, and the use of a resampling procedure.

As part of the development of neural networks, I experimented with two procedures for controlling the complexity of networks, namely weight regularisation with varying numbers of hidden units and early stopping committees. In their comparison of various training procedures on artificial datasets, Finnoff et al. (1993) found that both weight regularisation and early stopping resulted in higher predictive performance than when no precautions were taken to avoid overfitting other than varying the number of hidden units. In the present research, it was demonstrated that the superiority (or equivalency) of linear regression held across both of these training procedures. Moreover, important differences were observed between the two training procedures that are relevant for future applications of neural networks. For the vast majority of predictors and predictor combinations, the use of early stopping and committee formation improved the predictive performance of neural networks and resulted in predictions that were similar or only slightly inferior in their accuracy to the predictions made by linear regression equations. Furthermore, networks trained in this way were less prone to the adverse effects of capitalisation on chance that occurs as the number of predictors is increased. Thus, the use of early stopping committees is recommended when training neural networks, especially if one is considering the inclusion of large numbers of predictors. On the other hand, although the early stopping committees outpredicted linear regression in the presence of nonlinear or configural relationships, the extent of this gain was smaller than that associated with simply using weight regularisation with a large number of hidden units. By reducing the effective complexity of networks, early stopping fails to fully exploit the predictive power of neural networks in the presence of complexity. Therefore, in situations where there is a strong expectation of nonlinear or configural relationships, and only a small number of predictors are involved, it may be more beneficial to simply use weight regularisation with a large number of hidden units.

A second methodological strength of this thesis relates to the use of both absolute and relational measures of predictive performance. Previous studies that have applied neural networks to organisational psychology have for the most part only relied on the latter (e.g., Collins & Clark, 1993; Griffin, 1998; Somers, 1999, 2001). However, there are good reasons why one should include both types of measures. For one, it is entirely possible that neural networks will perform better than linear methods using one measure of predictive performance but worse using another (see Paik, 2000). Moreover, depending on the type of decisions for which the predictions are used, one type of predictive performance may be more relevant than the other. If the actual criterion level is to be predicted, for example as might occur when a recruiter is interested in estimating the actual amount of sales volume that a prospective salesperson is likely to generate, then absolute measures of accuracy are more relevant. On the other hand if an employer is interested in the potential performance of a prospective employee relative to other prospects then a relational measure may be appropriate.

In the present research linear regression outperformed neural networks on both the absolute measure (MAE) and the relational measure (cross-validity coefficient). However, for the neural networks developed without early stopping, predictive performance tended to be especially poor for the absolute measure. Indeed, an inspection of Tables 4.7, 5.2, and 6.3 (in chapters 4, 5, and 6) reveals that such networks rarely obtained a lower MAE than linear regression, even when cross-validity coefficients were higher. A possible reason for this finding is that absolute measures of predictive performance, unlike relational measures, are sensitive to differences in the scale of actual and predicted criterion scores (see Kirlik & Strauss, 2003); consequently overfitting is likely to be reflected to a greater extent in such measures. Consistent with this interpretation, note that the early stopping committees, which were less susceptible to overfitting than networks developed without early stopping, did not display the same discrepancy between the relative and absolute measures – they obtained lower MAE than linear regression about as often as they obtained higher cross-validity coefficients (see Tables 4.7, 5.2, and 6.3). The implication of these results is that neural networks developed without early stopping are less likely to be of use in situations where absolute accuracy is required.

A third methodological strength of the current research relates to the use of a resampling procedure. Previous applications of neural networks to organisational psychology have predominantly relied on the hold-out method of assessing predictive performance, in which the data is randomly partitioned into a training set and a test set once (e.g., Collins & Clark, 1993; Griffin, 1998; Scarborough, 1996; Somers, 1999, 2001). A shortfall of this method is that it does not take into account the variability in the performance of neural networks relative to linear regression that occurs across different partitions of the data (see Dietterich, 1998; Neal, 1998). When this variability is large, one cannot be confident that the hold-out method will yield reliable estimates of the true differences between neural networks and linear regression.

The present research employed a resampling procedure in which the partitioning of the data was repeated 20 times, and predictive performance was estimated by averaging over the twenty partitions. Consequently, estimates of the differences between neural networks and linear regression were more stable than if only one partition had been used. To provide a concrete example of the dangers of the hold-out method, consider the comparison between linear regression and the neural networks developed without early stopping for the Interpersonal Orientation scale in Dataset 6. The cross-validity coefficients for the two methods were .03 and .02 when averaged across the twenty partitions (see Table 4.4), suggesting little difference between them on average. However, the differences in cross-validity coefficients ranged from .31 to -.32. Thus, if the hold-out method had been used, one could have arrived at one of two contradictory conclusions based on the random choice of the partition, namely that linear regression provides a large advantage over neural networks (given the first partition), or else that neural networks are far superior to linear regression (given the second partition).

As previously noted, however, a limitation of the resampling procedure employed here is that the overlap between training sets from different partitions creates dependencies that violate assumptions required for formal statistical tests (see Neal, 1998). In the present research a corrected t-test proposed by Nadeau and Bengio (2003) was used to partially address this problem, although this test too is based on assumptions that are not necessarily met. Martin and Hirschberg (1996) have argued that such approximate statistical tests can only ever be heuristic and should be presented as such. Consequently, the absence of a robust test for assessing the statistical significance of the findings is a threat to the statistical conclusion validity of the present results (see Cook & Campbell, 1979).[1] Nevertheless, that the direction of the effects was for the most part in favour of linear regression, and the consistency with which this occurred across predictors and datasets, suggests confidence in the conclusion that neural networks typically do not improve the effectiveness of personality variables for predicting work performance within the range of sample sizes considered here.

---

[1] A more robust procedure that would have mitigated this threat is the use of bootstrapping (see Dybowski & Roberts, 2001). This is a computationally intensive procedure that was beyond the resources available for the present research.

Personality and Work Performance

Evidence for Complex Relationships

Previous studies addressing complex relationships between personality variables and performance criteria have predominantly focused on the Conscientiousness factor, and have provided mixed findings. Some studies have found evidence for a nonlinear relationship between Conscientiousness and work performance (e.g., Cucina & Vasilopoulos, 2005; La Huis et al., 2005), whereas others have not (e.g., Robie & Ryan, 1999). Similarly, although evidence from a number of studies suggests that the joint effects of Conscientiousness and other personality dimensions on work performance may be configural in nature (see Witt, 2003 for a review), the findings do not necessarily generalise across all work settings and occupations (e.g., Witt et al., 2002). The present research examined complex relationships for all five dimensions of the five-factor model, not just Conscientiousness. Furthermore, such relationships were examined across a relatively large number of occupations. Little evidence was obtained in support of complexity in the relationship between personality variables and work performance criteria. Of the large number of tests for nonlinearity that were conducted in Study 1, only one statistically significant nonlinear effect was obtained and this occurred for a predictor that had not been hypothesised to be nonlinearly related to the criterion. In Study 3, configural relationships occurred infrequently and could not be easily interpreted in terms of the theoretical rationale for configurality. Furthermore, the neural network analyses provided little indication of other nonlinear or configural relationships that generalised to unseen cases.

There are several possible explanations for why some studies have obtained statistically significant findings whereas others have not. First, it may be that the present research and other studies that obtained nonsignificant results lacked the statistical power to detect nonlinear and configural relationships. A number of authors have noted the difficulty of detecting such effects in applied settings (e.g., Aguinis, Beaty, Boik, & Pierce, 2005; Aiken & West, 1991; Lubinski & Humphreys, 1990; McClelland & Judd, 1993). Small sample sizes, measurement error, and restriction of range in the predictors are some of the factors that have been identified as detrimental to statistical power (Aguinis, 1995). Aiken and West (1991) showed that for large, moderate, and small effect sizes, N = 26, 55, and 392 cases are required to detect a multiplicative effect between two predictors with 80% probability (setting $\alpha = .05$). Based on these figures, all of the datasets used in this thesis had adequate power for detecting moderate and large effects, and one dataset had adequate power for detecting small effects.

The above calculations are based on the assumption that constructs are assessed with no measurement error. In practice, predictors and criteria are typically measured with scales that have less than perfect reliability, and this is likely to adversely affect statistical power (Aguinis, 1995). Measurement error is further exacerbated when predictor variables are multiplied to produce product and power terms (McLelland & Judd, 1993), and consequently complex relationships are especially difficult to detect. Stone-Romero and Anderson (1994) conducted a Monte Carlo study in which they evaluated the power of detecting multiplicative effects when predictors contain measurement error. With a sample size of 120 and predictor reliabilities of .80, the power of detecting a multiplicative effect between two predictors was estimated to be over 90% for moderate and large effect sizes, and 28% for a small effect. Given that the majority of the five-factor scales employed in this thesis had reliability coefficients over .80 (see chapter 3), and that all datasets contained at least 120 cases, the above finding suggests that the present analyses had adequate power for at least detecting moderate and large effects in spite of the less than perfect predictor reliability.

Nevertheless, these latter findings are also possibly an optimistic estimate of the statistical power of the tests conducted in the present research as they do not reflect the adverse effects of restriction of range in the predictors. McLelland and Judd (1993) showed that range restriction has a greater negative impact on the detection of nonlinear and multiplicative effects than on the detection of linear effects. Moreover, if the relationship between the predictor and criterion is nonlinear over the entire range of scores but linear over a restricted range, then it will not be possible to detect the nonlinearity when only the restricted range is represented in the sample. In chapter 3, it was noted that the standard deviations of the personality variables in the present samples were sometimes smaller than those reported in the test manuals, although it is difficult to determine whether this is due to range restriction or to the greater homogeneity of the population of applicants within an occupation compared to the general population. There were no datasets in which personality variables were used to select individuals into the occupations, and therefore range restriction due to direct truncation is unlikely to be a problem here. In some of the datasets individuals were selected using variables that potentially correlate with personality variables, such as performance in interviews or on cognitive ability tests, thus raising the possibility of indirect range restriction. Nevertheless, the nonsignificant findings in the present research occurred across datasets, including datasets that are unlikely to suffer from low statistical power due to the large samples involved. Consequently, although it is possible that some of the negative results may be due to the combined effects of measurement error and restriction of range on statistical power, Type II errors at best only partially explain the present findings.

Alternatively, it may be that the lack of complexity in personality-performance relationships obtained here represents the true state of affairs, and that the findings from studies that have obtained significant nonlinear or multiplicative effects represent Type I errors. However, this too is unlikely to be an entirely satisfactory explanation given that some of the previous findings have been replicated in multiple datasets (e.g., La Huis et al., 2005; Witt, 2002). Possibly a more plausible explanation for discrepancy across studies is that complex relationships are not pervasive, but rather occur in certain situations and under certain circumstances. Characteristics of the occupation, the organisation, or the context in which testing occurs may influence the presence of such relationships. For example, La Huis et al. (2005) noted that their nonlinear findings may be due to several characteristics associated with the clerical positions they examined, such as the nature of the tasks to be performed and low autonomy. Witt (2003) suggested that personality-performance relationships, multiplicative or otherwise, are more likely to occur when supervisors are ineffective, as this creates a weak situation in which personality has greatest impact. Haaland and Christiansen (1998) proposed that departures from linearity tend to occur in situations where there is a motivation to fake personality test scores. This is because distorted responses are more likely to be found in the upper ranges of the score distribution, and therefore the relationship between test scores and the criterion is likely to be weaker in this range than in other regions of the distribution. The present research did not include a clerical sample, nor did it assess the effectiveness of supervisors. Furthermore, participants in all datasets completed personality inventories after they had been employed and therefore had less motivation to fake. An area for future research will be to measure or where possible manipulate situational and circumstantial factors such as these in order to gain a better understanding of the conditions under which complex relationships are likely to occur, if at all. Nevertheless, based on the present findings, one can only concur with the view that if there is any complexity in the relationship between personality and work performance then this does not occur systematically across performance contexts (e.g., Robie & Ryan, 1999).

The above conclusion has important theoretical and practical implications. From a theoretical perspective, it lends support to the assumptions of linearity and additivity that are implicit in most general theories of personality and work performance (e.g., Barrick, et al., 2003; Motowidlo et al., 1997). Thus, one can be confident that the importance of personality variables within such theories is unlikely to have been attenuated as a result of the use of linear and additive methods. With respect to linearity, the results suggest that when a personality variable is related to work performance the strength and direction of the relationship typically remains constant across the entire range of the personality variable. A given increase in a particular attribute will have the same effect on work performance regardless of whether the person is low or high on that characteristic. This runs counter to the "deficiency-sufficiency" view, whereby an increase in an attribute results in improved performance up to a given point after which the relationship ceases to exist; it also contradicts the idea that optimal performance occurs at moderate levels of an attribute (e.g., Murphy, 1996). Furthermore, the results do not provide support for the view, derived from theories of traitedness, that personality is less strongly related to performance in the mid-range than at the extremes (e.g., Sinclair et al., 1999).

With respect to configurality, the results suggest that the strength and direction of the relationship between a given personality variable and work performance does not in general depend on the individual's standing on other personality variables. Thus, a given increase in a particular attribute will have the same effect on work performance regardless of whether the person is low or high on other personality characteristics. This implies that deficiencies on a particular attribute that is relevant for performance cannot be fully compensated for by high levels of another relevant attribute. Furthermore, it implies that the joint effect of two attributes is limited to their additive effects, and consequently that high levels of one attribute do not enhance the effects of high scores on other relevant attributes.

From a practical perspective, the results provide guidance in relation to the selection strategy to adopt when using personality scores as the basis of selection decisions. Specifically, the present results provide some justification for the use of a top-down strategy where candidates are rank-ordered and selected sequentially, starting with the highest scorers. This is because when the relationship between scores on the selection test and the performance criterion is monotonically increasing the expected performance of individuals selected via a top-down strategy will be higher than the expected performance of those selected by any other strategy. Of course the actual choice of a selection strategy will also depend on other considerations including administrative and legal arguments (see Campion et al., 2001). Nevertheless, the present results do not provide the statistical justification for forsaking the top-down approach, as might have been justified if inverted-U relationships had been found.

Differences Between Personality Variables

Consistent with the results of previous meta-analyses (e.g., Hurtz & Donovan, 2000; Barrick et al., 2001), the findings reported in this thesis highlight the importance of measures of Conscientiousness relative to other personality variables for the purposes of predicting work performance. The predictive performance of this dimension was higher than that of the other dimensions of the five-factor model in four of the six datasets considered here. Furthermore, whereas previous research has predominantly established the superiority of Conscientiousness using linear methodology, the present research showed that this is also true when using a method that is capable of detecting nonlinear and configural relationships. Thus, in Study 1, where personality variables were considered individually, Conscientiousness was typically the most valid predictor of work performance under both linear regression and neural networks; and in Study 4, where predictor combinations were selected using backward elimination, Conscientiousness typically emerged as the final variable after all others had been eliminated. Finally, it was found that the superiority of broad Conscientiousness also extended to instances in which narrower personality variables were considered as predictors. In two of the three datasets employed in Study 5, the validity of the broad Conscientiousness measure was higher than that of all 30 of the facet scales of the NEO PI-R, and in the third dataset only one of the 30 facets obtained a slightly higher validity than this broad measure.

Theoretically, the importance of broad Conscientiousness is not surprising given that the essence of this factor has been suggested to be self-control (Costa & McCrae, 1992), an attribute that is likely to be desirable across performance contexts. From a practical perspective, the results suggest that when selecting among personality variables to predict work performance, Conscientiousness will usually be the predictor of choice. However, one should also be careful not to overstate the results. In all six datasets the observed validity and cross-validity coefficients associated with the measure of Conscientiousness were less than .30, a value that is itself substantially less than that associated with the most valid selection methods, such as cognitive ability tests (Schmidt & Hunter, 1998). Nevertheless, even in the presence of cognitive ability tests, the use of Conscientiousness as a selection method may be justified given that it has been shown to have practically useful levels of incremental validity (Schmidt & Hunter, 1998).

The predictive performance of the other four dimensions within the five-factor model was typically substantially lower than of Conscientiousness, and was to a large extent dependent on the dataset in question. Neuroticism was the best predictor of performance in the managerial sample after Conscientiousness, and also predicted performance in the police and flight attendant samples. These results are consistent with the previous meta-analytic finding that Emotional Stability is associated with superior performance in some occupations but not others (Barrick et al., 2001). Additionally, the Self-Consciousness scale (a facet of Neuroticism) was one of the strongest predictors of performance in the sample of university students, and this was attributed to the ability of this predictor to provide information relating to the social activity of first-year university students. Future research can test this explanation by, for example, assessing whether the relationship between this scale and first-year university performance is mediated by the extent of participation in social activities.

Extraversion predicted work performance in a number of the samples, although work performance tended to be more strongly related to specific facets of this factor than to the composite measure. In particular, Study 5 found that a measure of the potency component of Extraversion, derived by aggregating scores on the Activity and Assertiveness facet scales, was among the strongest predictors of work performance in two of the three datasets examined in that study, and had a validity coefficient that was significantly higher than that of broad Extraversion in both those studies. This finding has some precedence in previous work (e.g., Hough, 1992), and supports the claim that the Extraversion dimension is too broad for the purposes of predicting work performance (e.g., Hogan & Holland, 2003). Furthermore, the finding can be explained in terms of job performance theory. Specifically, Extraversion has been suggested to affect job performance through status striving, a motivational construct which reflects actions directed toward obtaining power and dominance (e.g., Barrick et al., 2003; Barrick, Stewart, & Piotrowski, 2002; see also Hogan & Holland, 2003). Conceptually, status striving is more similar to the potency component than the affiliation component of Extraversion, and therefore the present findings are consistent with this theory.

Openness has traditionally been found to predict performance on training programs, but not performance on the job (e.g., Barrick & Mount, 1991; Salgado, 1997). Consequently this personality dimension has sometimes been omitted from theories of job performance (e.g., Barrick et al., 2003). The present results suggest that the usefulness of Openness as a predictor of performance may be greater than has been previously acknowledged, although specificity within the Openness domain needs to be taken into account in order to exploit the predictive power of this factor. In particular, the results of Study 5 consistently indicated that work performance was specifically related to the facet of Openness that assessed intellectual curiosity, and that combining this facet with the other facets of Openness resulted in statistically significant reductions in validity. It should be noted that the present finding is not without precedence, as Griffin and Hesketh (2004) also obtained a significant relationship between a measure of intellectual curiosity and job performance.

An issue that remains unclear though is the reason for the relationship between intellectual curiosity and job performance. This has implications for both the role of intellectual curiosity in theories of job performance and its usefulness for selecting personnel. One possible explanation is that this personality trait is providing information about cognitive ability, which as noted above has been shown to be one of the most valid predictors of job performance (e.g., Schmidt & Hunter, 1998). This explanation is supported by the relatively large correlation between measures of intellectual curiosity and crystallised intelligence (e.g., Ackerman & Goff, 1994; Ashton et al., 2000). If the relationship between intellectual curiosity and job performance is in fact due the confounding effect of cognitive ability, then this would place in doubt the role of this personality construct in job performance theory and limit its usefulness for selecting personnel in contexts where cognitive ability test scores are available. Alternatively, it may be that the significant effect of intellectual curiosity occurs independently of its relationship with scores on cognitive ability tests. Although these two constructs are related, they are also distinct in that intellectual curiosity assesses interest in intellectual pursuits (Costa & McCrae, 1992), and may therefore tap into motivational constructs that are important for job performance, such as the motivation to solve problems and acquire job-related knowledge. Furthermore, intellectual curiosity corresponds closely to the construct of typical intellectual engagement (Ackerman & Goff, 1994), which has been argued to be a better predictor of long-term performance than the maximal intellectual engagement assessed by cognitive ability tests (Goff & Ackerman, 1992). To the extent that it is these unique aspects of intellectual curiosity that relate to job performance, this personality variable will play a legitimate role in explaining job performance, and will potentially be of use for selecting personnel even when cognitive ability test scores are available. Future research can distinguish between these two explanations by testing whether the relationship between intellectual curiosity and work performance remains statistically significant when the effect of cognitive ability is held constant.

Finally, the predictive performance of Agreeableness tended to be trivially small in most of the datasets employed in this study. This is possibly because customer service was only an explicit measure of the work performance criterion in two of the six samples, the flight attendants and the bus drivers. In the former sample, however, Agreeableness was the best predictor of work performance.

Optimal Number of Personality Variables

The results of the present research also provide insight into the optimal number of personality variables to include in prediction equations when predicting work performance. This is an area that has received little research attention, yet is an important issue as it has recently been claimed that the failure to consider personality variables jointly has resulted in an underestimate of the validity of personality for predicting work performance (e.g., Barrick & Mount, 2005). The present findings do not support this view. With respect to both MAE and cross-validity coefficients, optimal prediction occurred with a single personality variable for the majority of the datasets. This occured regardless of whether neural networks or linear regression was used to derive the prediction equations.

In considering the generality of this finding it should be kept in mind that the optimal number of predictors is influenced by sample size (Goldberg, 1972), and therefore larger samples than used here may support the inclusion of a greater number of personality variables as predictors. Furthermore, the present analysis was conducted using the broad (higher-level) scales within each inventory. Given that there are a greater number of lower-level scales to select from, some of which may be more strongly related to the criterion than their higher-level counterparts, it may be that the optimal number of predictors is larger when personality is operationalised in terms of the lower-level facet scales. Therefore, one needs to be careful not to overgeneralise the present findings, which suggest that a single personality variable tends to be optimal for predicting work performance when personality is represented in terms of the five broad personality dimensions and fewer than 500 cases are available. The generality of these findings to instances where personality variables are represented at the facet level is an area for future research. Moreover, future studies may wish to consider prediction equations that combine both broad and narrow personality variables. This could yield cross-validity coefficients that are higher than those based on only one level of representation, and therefore contribute to the utility of personality inventories in applied settings.

Conclusions and Recommendations

The topic of the research reported in this thesis was in line with recent calls for a move towards more complex methodology for understanding and predicting work-related phenomena (e.g., Hanges et al., 2002; Mount et al., 2003; Somers, 1999). The findings provide some grounds for optimism about the application of artificial neural networks in organisational psychology research, but mainly highlight limitations of this method. The strength of the neural network method lies in its ability to detect complex nonlinear and configural relationships among variables without the need for prior specification of functional relationships. Nevertheless, this capability does not necessarily translate into predictive or explanatory gains compared to traditional linear techniques, as the usefulness of the neural network methodology is contingent on the availability of adequate sample sizes, data that is not overly noisy, and the presence of complex relationships between variables. On a positive note, then, the current research demonstrated that for sample sizes that are typical of those available in organisational psychology research, and when the variables are related in moderately complex ways, neural networks can cope with the level of noise present in the data such that they are able to detect and exploit the complex relationships for predictive purposes.

Importantly, however, the results also illustrate that the increased representational capability of complex methods comes with a cost, namely the propensity to capture features of the data that are idiosyncratic to the training sample.

This is most noticeable when the relationships are either linear and additive, or only weakly nonlinear and configural, as in these instances the risk of overfitting the training data is likely to outweigh any benefit of being able to capture complex relationships. In the case of personality and work performance these instances are the rule rather than the exception, and therefore it would seem that within this domain simpler methods are better than more complex ones.

For researchers interested in applying neural networks to organisational psychology research, a number of recommendations can be provided. First, overfitting is less likely to occur when sample sizes are large (Babyak, 2004). The actual number of cases required to successfully implement a given application is domain dependent, and cannot easily be determined a priori. Nevertheless, the benefit of an increase in sample size is likely to be greater for neural networks than for linear methods, and therefore researchers employing neural networks should make every effort to maximise sample size. Furthermore, in choosing between two potential applications, the number of cases that are likely to be available should be a relevant consideration.

Second, overfitting is less likely to occur when there are fewer parameters to be estimated. The number of parameters can be reduced through procedures designed to control the complexity of neural networks, for example by reducing the number of hidden units. Other procedures such as weight regularisation, early stopping, and committee formation are also of help. Additionally, however, one can preserve degrees of freedom by including only a limited number of predictors in the analysis, for example based on theoretical considerations (Tabachnick & Fidell, 2001). Furthermore, the number of parameters to be estimated as a function of increasing predictors grows at a faster rate for neural networks than for linear regression, and consequently one needs to be especially careful about including a large number of speculative variables when using neural networks.

Third, noise in the data increases the risk of overfitting, and limits the level of generalisation that can be achieved (Sarle, 2001c). One source of noise that is likely to be relevant to organisational psychology researchers is that associated with measurement error in predictors and criteria. It is therefore particularly important to ensure that the measures that are included in neural network analyses have adequate levels of reliability. This can be achieved, for example, by using longer tests, administering inventories under standardised conditions, and training raters.

Finally, neural networks are more likely to provide benefits over traditional methods when there is some expectation of nonlinearity or configurality between the predictors and the criterion. With this in mind, two suggestions are provided in relation to future applications of artificial neural networks in organisational psychology. First, the applications to date have primarily used individual difference variables as the inputs, yet it is known that to best predict behaviour from trait measures the systematic effects of situations and their interaction with the person must also be taken into account (e.g., Kenrick & Funder, 1988). Recent theoretical efforts have outlined some of the main situational influences on trait-behaviour relations within organisations (e.g., Barrick et al., 2003), and it has been shown empirically that the joint effect of these situational demands and individual difference variables on performance is configural in nature (e.g., Barrick & Mount, 1993; Stewart, 1996). Therefore, neural network methodology may well be of more benefit to research that incorporates both situational and personal variables into the design than research that only includes individual difference variables as the inputs. Second, neural networks could make a meaningful contribution to research aimed at modeling judgments about work-related criteria, such as those made by managers and HR personnel. It has been shown that configurality is often a characteristic of judgments about a criterion, but not necessarily of the criterion itself (Ganzach, 1997). For example, substantial configurality has been found in the way in which managers combine cues to arrive at judgments about job applicant favourability (Hitt & Barr, 1989). Moreover, the inherent difficulty of explicitly specifying the nonlinear processes underlying judgments can result in low levels of nonlinear variance accounted for when in fact judgments are highly nonlinear and configural in nature (Ganzach, 2001). The ability of neural networks to capture complex relationships without the need for prior model specification could be of value within this domain.

To conclude, then, artificial neural networks have the potential to become an important part of the research methods used by organisational psychologists. However, it should be kept in mind that in some situations simpler is better, and therefore that neural networks should be seen as a complement to the traditional methodologies used within organisational psychology rather than a replacement to them.

REFERENCES

Ackerman, P. L. & Goff, M. (1994). Typical intellectual engagement and personality: Reply to Rocklin (1994). *Journal of Educational Psychology*, 86, 150-153.

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. (1995). *Journal of Management*, 21, 1141-1158.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, 90, 94-107.

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Allport, G. W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1, Whole No. 211).

American Management Association. (2001). *AMA survey on workplace testing: Basic skills, job skills, psychological measurement.* New York: American Management Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior,* 19, 289-303.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, 29, 432-442.

Ashton, M. C., Lee, K., Vernon, P. A., & Lang, K. L. (2000). Fluid intelligence, crystallized intelligence, and the Openness/Intellect factor. *Journal of Research in Personality*, 34, 198-207.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine,* 66, 411-421.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* 51, 1173-1182.

Barrick, M. R., Mitchell, T. R., & Stewart, G. L. (2003). Situational and motivational influences on trait-behavior relationships. In M. R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp. 60-82). San Francisco: Jossey-Bass.

Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.

Barrick, M. R. & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111-118.

Barrick, M. R. & Mount, M. K. (2003). Impact of meta-analysis methods on understanding personality-performance relations. In K. M. Murphy (Ed.), *Validity generalization: A critical review* (pp. 197-219). Mahwah, NJ: Lawrence Erlbaum.

Barrick, M. R. & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance,* 18, 359-372.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.

Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87, 43-51.

Bauer, K. W. & Liang, Q. (2003). The effect of personality and precollege characteristics on first-year activities and academic performance. *Journal of College Student Development*, 44, 277-290.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Black, J. (2000). Personality testing and police selection: Utility of the 'Big Five'. *New Zealand Journal of Psychology*, 29, 2-9.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Charles C Thomas.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187-215.

Blum, A. L. & langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97, 245-271.

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science*, 9, 8-17.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp 271-326). Palo Alto, CA: Consulting Psychologists Press.

Bounds, D. G., Lloyd, P. J., & Mathew, B. G. (1990). A comparison of neural network and other pattern recognition approaches to the diagnosis of low back disorders. *Neural Networks*, 3, 583-590.

Braden, J. P. (1995). Intelligence and personality in school and educational psychology. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 621-650). New York: Plenum Press.

Burke, L. A. & Witt, L. A. (2002). Moderators of the Openness to Experience-performance relationship. *Journal of Managerial Psychology*, 17, 712-721.

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 180-207). New York: Guilford.

Campion, M. A., Outtz, J. L., Zedeck, S., Schmidt, F. L., Kehoe, J. R., Murphy, K. R., & Guion, R. M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology*, 54, 149-185.

Caruana, R., Lawrence, S., & Giles, C. L. (2000, November). *Overfitting in neural networks: Backpropagation, conjugate gradient, and early stopping.* Paper presented at the Neural Information Processing Systems meeting, Denver, CO.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, 12, 197-220.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the 16 Personality Factor Questionnaire*. Champaign, IL: IPAT.

Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the five factor model. *Journal of Applied Psychology*, 81, 694-704.

Chamorro-Premuzic, T. & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, 17, 237-250.

Church, A. T., Katigbak, M. S., & Reyes, J. A. S. (1996). Toward a taxonomy of trait adjectives in Filipino: Comparing personality lexicons across cultures. *European Journal of Personality*, 10, 3-24.

Cohen, J. (1977). *Statistical power analysis for the behavioural sciences* (Rev. ed.). New York: Academic Press.

Cohen, J. (1978). Partialed products *are* interactions; partialed powers *are* curve components. *Psychological Bulletin*, 85, 858-866.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement* (3rd ed.). Mountain View, CA: Mayfield.

Collins, J. M. & Clark, M. R. (1993). An application of the theory of neural computation to the prediction of workplace behavior: An illustration and assessment of network analysis. *Personnel Psychology*, 46, 503-524.

Cook, T. D & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Person *r*s and Fischer's *z* transformations. *The Journal of General Psychology*, 125, 245-261.

Costa, P. T., Jr., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI factor scales with measures of the five factor model of personality. *Journal of Personality Assessment*, 50, 640-650.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, 55, 258-265.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., McCrae, R. R., & Kay, G. G. (1995). Persons, places, and personality: Career assessment using the Revised NEO Personality Inventory. *Journal of Career Assessment*, 3, 123-139.

Costa, P. T., Jr., Zonderman, A. B., McCrae, R. R., & Williams, R. B., Jr. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology*, 48, 925-933.

Coward, W. M. & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology*, 75, 297-300.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Brothers.

Cucina, J. M. & Vasilopoulos, N. L. (2005). Non-linear personality-performance relationships and the spurious moderating effects of traitedness. *Journal of Personality*, 73, 227-259.

Dalton, M., Ernst, C., Leslie, J., & Deal, J. (2002). Effective global management: Established constructs and novel contexts. *European Journal of Work and Organizational Psychology*, 11, 443-468.

Day, D. V. & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, 42, 25-36.

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10, 1895-1923.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.

Digman, J. M. & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research*, 16, 149-170.

Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. Cincinnati, Ohio: College Division South Western.

Driskell, J. E., Hogan, J., Salas, E., & Hoskin, B. (1994). Cognitive and personality predictors of training performance. *Military Psychology*, 6, 31-46.

Dudley, N. M., Orvis, K. A., & Lebiecki, J. E. (2003, April). A meta-analytic investigation of Conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. In J. M. Cortina (Chair), N. M. Dudley (Chair), K. A. Orvis (Chair), & M. R. Barrick (Discussant), *Spotting the trees: Beyond the Big Five in predicting performance*. Symposium conducted at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Dybowski, R. & Roberts, S. J. (2001). Confidence intervals and prediction intervals for feed-forward neural networks. In R. Dybowski & V. Gant (Eds.), *Clinical applications of artificial neural networks* (pp. 298-326). Cambridge, UK: Cambridge University Press.

Edwards, P. J. & Murray, A. F. (2000). A study of early stopping and model selection applied to the papermaking industry. *International Journal of Neural Systems,* 10, 9-18.

Ennett, C. M. & Frize, M. (2003). Weight-elimination neural networks applied to coronary surgery mortality prediction. *IEEE Transactions on Information Technology in Biomedicine*, 7, 86-92.

Eysenck, H. J. (1947). *Dimensions of personality*. London: Routledge & Kegan Paul.

Eysenck, H. J. & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum Press.

Finnoff, W., Hergert, F., & Zimmermann, G. (1993). Improving model selection by nonconvergent methods. *Neural Networks,* 6, 771-783.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, 329-344.

Ganzach, Y. (1997). Configurality in judgment: Is it a bias? *Psychonomic Bulletin and Review*, 4, 382-386.

Ganzach, Y. (2001). Nonlinear models of clinical judgment: Communal nonlinearity and nonlinear accuracy. *Psychological Science*, 12, 403-407.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1-58.

Gençay, R. & Qi, M. (2001). Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12, 726-734.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.

Ghiselli, E. E. & Barthol, R. P. (1953). The validity of personality inventories in the selection of employees. *Journal of Applied Psychology,* 37, 18-20.

Goff, M. & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, 84, 537-552.

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*, 7, No. 72-2.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26-42.

Goldberg, L. R. (1992). An alternative "Description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology,* 59, 1216-1229.

Goldberg, L. R. (1993). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 169-188). Washington, DC: American Psychological Association.

Goldberg, L. R. (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, The Netherlands: Tilburg University Press.

Goldberg, L. R. (in press). The comparative validity of adult personality inventories: Applications of a consumer-testing framework. In S. R. Briggs, J. M. Cheek, &

E. M. Donahue (Eds.), *Handbook of adult personality inventories*. New York: Plenum. Retrieved from http://ipip.ori.org/newInventoriesText.htm

Goldberg, L. R. & Somer, O. (2000). The hierarchical structure of common Turkish person-descriptive adjectives. *European Journal of Personality*, 14, 497-531.

Gorman, R. P. & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1, 75-89.

Gough, H. G. (1987). *California Psychological Inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Gray, E. K. & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, 70, 177-206.

Griffin, B. & Hesketh, B. (2004). Why Openness to Experience is not a good predictor of job performance. *International Journal of Selection and Assessment*, 12, 243-251.

Griffin, G. R. (1998). Predicting naval aviator flight training performance using multiple regression and an artificial neural network. *The International Journal of Aviation Psychology*, 8, 121-135.

Guion, R. M. (1991). Personnel assessment, selection, and placement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp 327-397). Palo Alto, CA: Consulting Psychologists Press.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum.

Guion, R. M. & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.

Haaland, D. & Christiansen, N. D. (1998, June). *Departures from linearity in the relationship between applicant personality test scores and performance as evidence of response distortion*. Paper presented at the 22nd Annual International Personnel Management Association Assessment Council Conference, Chicago, IL.

Hagiwara, K. (2002). Regularization learning, early stopping and biased estimator. *Neurocomputing*, 48, 937-955.

Hanges, P. J., Lord, R. G., Godfrey, E. G., & Raver, J. L. (2002). Modeling nonlinear relationships: Neural networks and catastrophe analysis. In S. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 431-455). Malden, MA: Blackwell.

Hartstone, M. & Kirby, N. (1998). Australian personnel managers and organisational psychology: An update. *Australian Psychologist*, 33, 148-154.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction.* New York: Springer.

Higgins, L. T. & Sun, C. H. (2002). The development of psychological testing in China. *International Journal of Psychology*, 37, 246-254.

Hinton, G. E. (1992, September). How neural networks learn from experience. *Scientific American*, 104-109.

Hitt, M. A. & Barr, S. H. (1989). Managerial selection decision models: Examination of configural cue processing. *Journal of Applied Psychology,* 74, 53-61.

Hochwarter, W. A., Witt, L. A., & Kacmar, K. M. (2000). Perceptions of organizational politics as a moderator of the relationship between conscientiousness and job performance. *Journal of Applied Psychology*, 85, 472-478.

Hodgkinson, G. P. & Payne, R. L. (1998). Graduate selection in three European countries. *Journal of Occupational and Organizational Psychology*, 359-365.

Hofstee, W. K. B., De Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146-163.

Hofstee, W. K. B., Kiers, H. A. L., De Raad, B., Goldberg, L. R., & Ostendorf, F. (1997). A comparison of the Big-Five structures of personality traits in Dutch, English, and German. *European Journal of Personality*, 11, 15-31.

Hogan, R. (1996). A socio-analytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 163-179). New York: Guilford.

Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88, 100-112.

Hogan, J. & Roberts, B. W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, 17, 627-637.

Hogan, R. & Hogan, J. (1995). *Hogan Personality Inventory manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.

Hough, L. M. (1992). The “Big Five” personality variables – construct confusion: description versus prediction. *Human Performance*, 5, 139-155.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hurtz, G. M. & Donovan, J. J. (2000). Personality and performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869-879.

International Personality Item Pool (2001). A scientific collaboratory for the development of advanced measures of personality traits and other individual differences. Retrieved February, 2003 from http://ipip.ori.org/ipip/.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.

Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.

Jacobs, R. R., Conte, J. M., Day, D. V., Silva, J. M., & Harris, R. (1996). Selecting bus drivers: Multiple predictors, multiple perspectives on validity, and multiple estimates of utility. *Human Performance*, 9, 199-217.

John, O. P. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: Guilford Press.

Johnson, J. A. (1994). Clarification of factor five with the help of the AB5C model. *European Journal of Personality*, 8, 311-334.

Johnson, J. A. (2000, March). *Web-based personality assessment*. Poster session presented at the 71st Annual Meeting of the Eastern Psychological Association, Baltimore, MD. Retrieved February 20, 2003 from http://www.personal.psu.edu/faculty/j/5/j5j/papers/web.html

Johnson, J. M., Null, C., J. H., Butcher, & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology*, 47, 105-114.

Judge, T. A., Van Vianen, A. E. M., & De Pater, I. E. (2004). Emotional Stability, core self-evaluations, and job outcomes: A review of the evidence and an agenda for future research. *Human Performance*, 17, 325-346.

Kenrick, D. T. & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23-34.

Kirlik, A. & Strauss, R. (2003). *A systems perspective on situation awareness I: Conceptual framework, modelling, and quantitative measurement* (Technical Report AHFD-03-12/NTSC-02-2). Savoy, IL: University of Illinois, Institute of Aviation, Aviation Human Factors Division.

Kline, P. & Barrett, P. (1983). The factors in personality questionnaires among normal subjects. *Advances in Behaviour Research and Therapy*, 5, 141-202.

Krug, S. E. & Johns, E. F. (1986). A large scale cross-validation of second-order personality structure defined by the 16PF. *Psychological Reports*, 59, 683-693.

Kuo, R. J., Wu, P., & Wang, C.P. (2002). An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination. *Neural Networks*, 15, 909-925.

LaHuis, D. M., Martin, N. R. & Avis, J. M. (2005). Investigating nonlinear Conscientiousness-job performance relations for clerical employees. *Human Performance*, 18, 199-212.

LeBaron, B. & Weigend, A. S. (1998). A bootstrap evaluation of the effect of data splitting on financial time series. *IEEE Transactions on Neural Networks*, 9, 213-220.

Lévy-Leboyer, C. (1994). Selection and assessment in Europe. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 4, pp 173-190). Palo Alto, CA: Consulting Psychologists Press.

Lievens, F., Coetsier, P., De Fruyt, F., & De Maeseneer, J. (2002). Medical students' personality characteristics and academic performance: A five-factor model perspective. *Medical Education*, 36, 1050-1056.

Lubinski, D. & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, 107, 385-393.

Marshall, D. B. & English, D. J. (2000). Neural network modelling of risk assessment in child protective services. *Psychological Methods*, 5, 102-124.

Martin, J. K. & Hirschberg, D. S. (1996). *Small sample statistics for classification error rates II: Confidence intervals and significance tests* (Technical Report No. 96-22). University of California, Irvine. Department of Information and Computer Science.

Masters, T. (1993). *Practical neural network recipes in C++*. Boston: Academic Press.

Matthews, G. & Deary, I. J. (1998). *Personality traits.* Cambridge, UK: Cambridge University Press.

McClelland, G. H. & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376-390.

McCrae, R. R. & Costa, P. T., Jr. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, 6, 587-597.

McCrae, R. R. & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality*, 57, 17-40.

McCrae, R. R., Costa, P. T., Jr., & Busch, C. M. (1986). Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. *Journal of Personality*, 54, 430-446.

McCrae, R. R., Costa, P. T., Jr., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the five-factor model. *Journal of Personality*, 61, 1-26.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.

Mershon, B. & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, 55, 675-680.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71-83.

Mount, M. K. & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resource management. *Research in Personnel and Human Resources Management*, 13, 153-200.

Mount, M. K., Barrick, M. R., & Ryan, A. M. (2003). Research themes for the future. In M. R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp. 326-344). San Francisco: Jossey-Bass.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interaction. *Human Performance*, 11, 145-165.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management*, 25, 707-721.

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than g. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3-30). San Francisco: Jossey-Bass.

Murphy, K. R. & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18, 343-357.

Myers, I. B. & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto: Consulting Psychologists Press.

Nadeau, C. & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, 52, 239-281.

Neal, R. M. (1998). Assessing relevance determination methods using DELVE. In C. M. Bishop (Ed.), *Neural Networks and Machine Learning* (pp. 97-129). Springer-Verlag.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

Ones, D. S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17, 609-626.

Paik, H. (2000). Comments on neural networks. *Sociological Methods and Research*, 28, 425-453.

Paunonen, S. V. (1988). Trait relevance and the differential predictability of behavior. *Journal of Personality*, 56, 599-619.

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, 74, 538-556.

Paunonen, S. V. & Ashton, M. C. (2001a). Big Five factors and the prediction of behavior. *Journal of Personality and Social Psychology*, 81, 524-539.

Paunonen, S. V. & Ashton, M. C. (2001b). Big Five predictors of academic achievement. *Journal of Research in Personality*, 35, 78-90.

Paunonen, S. V., Haddock, G., Fosterling, F., & Keinonen, M. (2003). Broad versus narrow personality measures and the prediction of behaviour across cultures. *European Journal of Personality*, 17, 413-433.

Paunonen, S. V., Jackson, D. N., Trzebinski, J., & Fosterling, F. (1992). Personality structure across cultures: A multimethod evaluation. *Journal of Personality and Social Psychology*, 62, 447-456.

Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior*, 20, 389-405.

Pedhazur, E. J. (1997). *Multiple regression in behavioural research: Prediction and explanation* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Prechelt, L. (1998). Automatic early stopping using cross-validation: Quantifying the criteria. *Neural Networks*, 11, 761-767.

Price, R. K., Spitznagel, E. L., Downey, T. J., Meyer, D. J., Risk, N. K. & El-Ghazzawy, O. G. (2000). Applying artificial neural network models to clinical decision making. *Psychological Assessment*, 12, 40-51.

Pryor, R. & Taylor, N. (2000). *Congruence Personality Scale – Form 2*. Sydney, Australia: Congruence.

Rajavelu, A., Musavi, M. T., & Shirvaikar, M. V. (1989). A neural network approach to character recognition. *Neural Networks*, 2, 387-393.

Raymark, P. H., Schmit, M. J., & Guion, R. M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology*, 50, 723-736.

Reed, R. D. & Marks, R. J., II (1999). *Neural smithing: Supervised learning in feedforward artificial neural networks*. Cambridge, MA: The MIT Press.

Robie, C. & Ryan, A. M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment*, 7, 157-169.

Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, 86, 516-530.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359-391.

Sackett, P. R., Gruys, M. L., & Ellingson, J. E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology*, 83, 545-556.

Salgado, J. F. (1997). The five-factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30-43.

Sarle, W. S. (1995). Stopped training and other remedies for overfitting. *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pp. 352-360. Retrieved from ftp://ftp.sas.com/pub/neural/inter95.ps.Z

Sarle, W. S. (2000). *How to measure importance of inputs?* Retrieved January 5, 2006 from ftp://ftp.sas.com/pub/neural/importance.html

Sarle, W. S. (Ed.). (2001a). *Neural Network FAQ, part 1 of 7: Introduction* [Periodic posting to the Usenet newsgroup comp.ai.neural-nets]. Retrieved September 2001 from the World Wide Web: ftp://ftp.sas.com/pub/neural/FAQ.html

Sarle, W. S. (Ed.). (2001b). *Neural Network FAQ, part 2 of 7: Learning* [Periodic posting to the Usenet newsgroup comp.ai.neural-nets]. Retrieved September 2001 from the World Wide Web: ftp://ftp.sas.com/pub/neural/FAQ2.html

Sarle, W. S. (Ed.). (2001c). *Neural Network FAQ, part 3 of 7: Generalization* [Periodic posting to the Usenet newsgroup comp.ai.neural-nets]. Retrieved September 2001 from the World Wide Web: ftp://ftp.sas.com/pub/neural/FAQ3.html

Saucier, G. (1994). Trapnell versus the lexical factor: More ado about nothing? *European Journal of Personality*, 8, 291-298.

Scarborough, D. J. (1996). An evaluation of backpropagation neural network modelling as an alternative methodology for criterion validation of employee selection testing. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 56(8-B), 4624.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology,* 37, 407-422.

Shackleton, V. & Newell, S. (1991). Management selection: A comparative survey of methods used in top British and French companies. *Journal of Occupational Psychology*, 64, 23-36.

Shmelyov, A. G. & Pokhil'ko, V. I. (1993). A taxonomy-oriented study of Russian personality-trait names. *European Journal of Personality*, 7, 1-17.

Siem, F. M. (1998). Metatraits and self-schemata: Same or different? *Journal of Personality*, 66, 783-803.

Silverman, R. H. & Noetzel, A. S. (1990). Image processing and pattern recognition in ultrasonograms by backpropogation. *Neural Networks*, 3, 593-603.

Sinclair, R. R., Banas, C., & Radwinsky, R. (1999, April). *Non-linearity in personality-performance relations: Theory, assessment methods, and empirical evidence.* Paper presented at the 14th annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Sinclair, R. R. & Lyne, R. (1997, April). Non-linearity in personality-performance relations: models, methods, and initial evidence. In J. Hogan (Chair), *Non-cognitive measures of job performance*. Symposium conducted at the annual conference of the Southwestern Psychological Association, Fort Worth, TX.

Smith, G. M. (1967). Usefulness of peer ratings of personality in educational research. *Educational and Psychological Measurement*, 27, 967-984.

Smith, M. (1993). *Neural networks for statistical modelling*. New York: Van Nostrand Reinhold.

Smith, M. & Abrahamsen, M. (1992). Patterns of selection in six countries. *The Psychologist: Bulletin of the British Psychological Society*, 5, 205-207.

Somers, M. J. (1999). Application of two neural network paradigms to the study of voluntary employee turnover. *Journal of Applied Psychology*,84, 177-185.

Somers, M. J. (2001). Thinking differently: Assessing nonlinearities in the relationship between work attitudes and job performance using a Bayesian neural network. *Journal of Occupational and Organizational Psychology*, 74, 47-61.

Statsoft, Inc. (1998). *Statistica Neural Networks*. Tulsa, OK: Statsoft.

Statsoft, Inc. (1999). *Statistica Neural Networks addendum for version 4*. Tulsa, OK: Statsoft.

Stewart, G. L. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology*, 81, 619-627.

Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for Conscientiousness and its subtraits. *Journal of Applied Psychology*, 84, 959-968.

Stewart, G. L., Barrick, M. R., & Parks, L. (2003, April). A theoretical and empirical analysis of broad and narrow traits. In H. Moon (Chair), *Lumpers and splitters: The utility of personality beyond the FFM*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Stone-Romero, E. F. & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology*, 79, 354-359.

Stumpf, H. (1993). The factor structure of the Personality Research Form: A cross-national evaluation. *Journal of Personality*, 61, 27-48.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organizations and recruitment firms. *New Zealand Journal of Psychology*, 31, 8-18.

Tetko, I. V., Livingstone, D. J., & Luik, A. I. (1995). Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences*, 35, 826-833.

Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a “hyperdimensional” taxonomy of managerial competence. *Human Performance*, 13, 205-251.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology,* 44, 703-742.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior*, 24, 335-356.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, 24, 83-91.

Tupes, E. C. & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, 60, 225-251.

Utans, J. (1997). Stopped training via algebraic on-line estimation of the expected test-set error. *Proceedings of the 1997 IEEE International Conference on Neural Networks*, 1088-1092.

Van Iddekinge, C. H, Taylor, M. A., & Eidson, C. E., Jr. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. *Human Performance*, 18, 151-177.

Varsta, M., Heikkonen, J., Millán, J. R., Mouriño, J. (2000). Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP committee. *Proceedings of the 15th International Conference on Pattern Recognition*, 2, 911-915.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., III, & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83, 586-597.

Wang, C., Venkatesh, S. S., & Judd, J. S. (1994). Optimal stopping and effective machine complexity in learning. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in Neural Information Processing Systems 6* (pp. 303-310). San Mateo, CA: Morgan Kaufmann.

Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.

Weigend, A. S. (1994). On overfitting and the effective number of hidden units. In M. C. Mozer, P. Smolensky, D. S. Touretzky, J. L. Elman, A. S. Weigend (Eds.), *Proceedings of the 1993 Connectionist Models Summer School*, (pp.335-342). Hillsdale, NJ: Lawrence Erlbaum.

Weigend, A. S., Huberman, B. A., & Rumelhart, D. E. (1990). Predicting the future: A connectionist approach. *International Journal of Neural Systems*, 1, 193-209.

Weiss, S. M. & Kulikowski, C. A. (1991). *Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems.* San Mateo, CA: Morgan Kaufmann.

Williams, R. S. (1994, January). Occupational testing: Contemporary British practice. *The Psychologist*, 11-13.

Witt, L. A. (2002). The interactive effects of Extraversion and Conscientiousness on performance. *Journal of Management*, 28, 835-851.

Witt, L. A. (2003, April). Conscientiousness may not be enough. In J. M. Cortina (Chair), N. M. Dudley (Chair), K. A. Orvis (Chair), & M. R. Barrick (Discussant), *Spotting the trees: Beyond the Big Five in predicting performance*. Symposium conducted at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of Conscientiousness and Agreeableness on job performance. *Journal of Applied Psychology*, 87, 164-169.

Wright, P. M., Kacmar, K. M., McMahan, G. C., & Deleeuw, K. (1995). P=f(M X A): Cognitive ability as a moderator of the relationship between personality and job performance. *Journal of Management*, 21, 1129-1139.

Zeidner, M. & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence*. Cambridge, UK: Cambridge University Press.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116, 16-32.

Appendix A: Correlations Matrices

Table A1.

Dataset 1: Correlation matrix.

	N	E	O	A	C	Perf.
Neuroticism	1.00					
Extraversion	-.45**	1.00				
Openness	-.15*	.43**	1.00			
Agreeableness	-.16*	.08	.14*	1.00		
Conscientiousness	-.35**	.11	.04	.20**	1.00	
Performance	.05	-.13	.12	.03	.20**	1.00

* p < .05, ** p < .01

Table A2.

Dataset 2: Correlation matrix.

	N	E	O	A	C	Perf.
Neuroticism	1.00					
Extraversion	-.25**	1.00				
Openness	-.16**	.38**	1.00			
Agreeableness	-.31**	.13*	.06	1.00		
Conscientiousness	-.52**	.28**	.10	.33**	1.00	
Performance	-.16**	.16**	.10	.11	.27**	1.00

* p < .05, ** p < .01

Table A3.

Dataset 3. Correlation matrix.

	N	E	O	A	C	Perf.
Neuroticism	1.00					
Extraversion	-.39**	1.00				
Openness	-.13*	.47**	1.00			
Agreeableness	-.53**	.51**	.46**	1.00		
Conscientiousness	-.49**	.36**	.19**	.48**	1.00	
Performance	-.12*	.15**	.15**	.17**	.11*	1.00

* p < .05, ** p < .01

Table A4.

Dataset 4: Correlation matrix.

	N	E	O	A	C	Perf.
Neuroticism	1.00					
Extraversion	-.24**	1.00				
Openness	-.05	.39**	1.00			
Agreeableness	-.18*	.08	.22**	1.00		
Conscientiousness	-.41**	.18*	-.12	.09	1.00	
Performance	-.20**	.11	.02	-.08	.28**	1.00

* p < .05,   ** p < .01

Table A5.

Dataset 5: Correlation matrix.

	Adj.	Amb.	Soc.	Int.	Lik.	Pru.	Perf.
Adjustment	1.00						
Ambition	.46**	1.00					
Sociability	-.08	.32**	1.00				
Intellectance	.09	.28**	.43**	1.00			
Likeability	.45**	.38**	.19**	.19**	1.00		
Prudence	.54**	.29**	-.26**	.02	.42**	1.00	
Performance	.08	.11*	-.01	.00	.10*	.14**	1.00

* p < .05,   ** p < .01

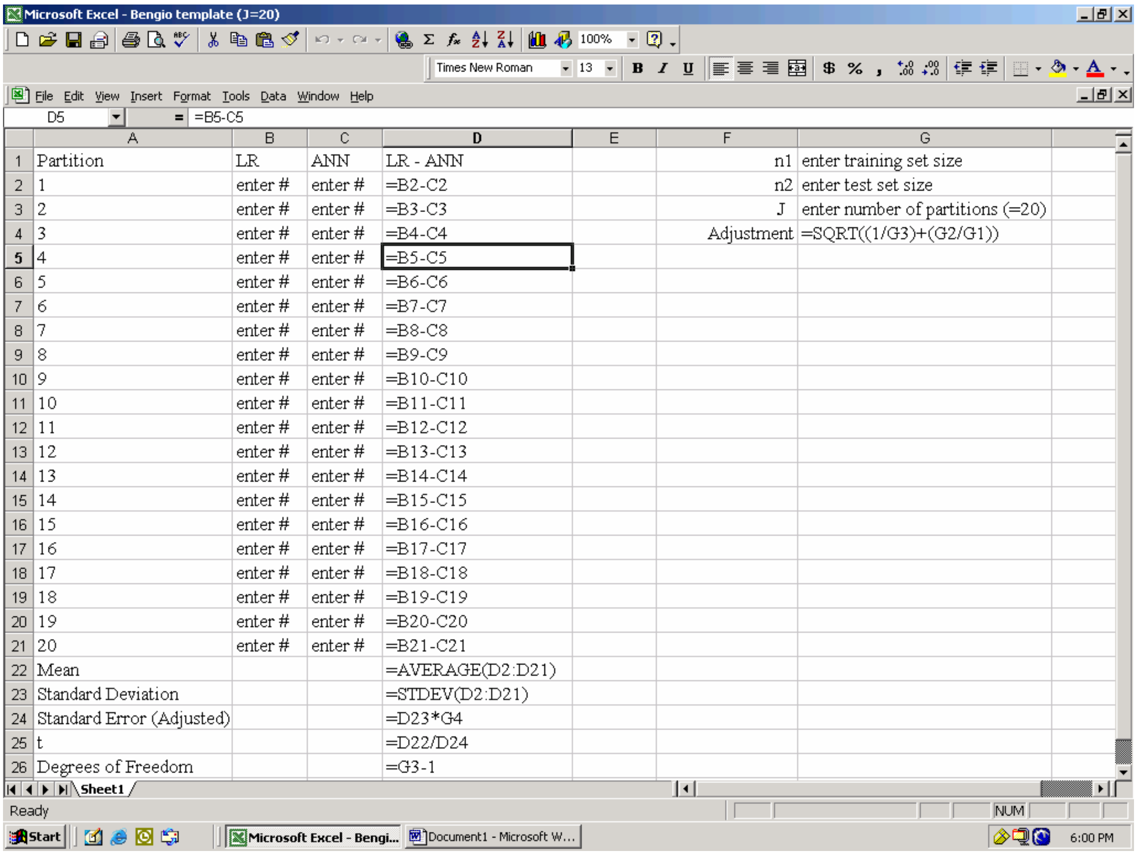
Table A6.

Dataset 6: Correlation matrix.

	EO	SO	CO	IO	TO	Perf.
Emotional Orientation	1.00					
Social Orientation	-.47**	1.00				
Cognitive Orientation	-.20*	.40**	1.00			
Interpersonal Orientation	-.19*	.11	.26**	1.00		
Task Orientation	-.27**	.26**	.25**	.20*	1.00	
Performance	-.09	.07	.13	.11	.00	1.00

* p < .05,   ** p < .01

Appendix B: Corrected t-Test Template


Microsoft Excel - Bengio template (J=20)

D5 = =B5-C5

	A	B	C	D	E	F	G
1	Partition	LR	ANN	LR - ANN		n1	enter training set size
2	1	enter #	enter #	=B2-C2		n2	enter test set size
3	2	enter #	enter #	=B3-C3		J	enter number of partitions (=20)
4	3	enter #	enter #	=B4-C4		Adjustment	=SQRT((1/G3)+(G2/G1))
5	4	enter #	enter #	=B5-C5			
6	5	enter #	enter #	=B6-C6			
7	6	enter #	enter #	=B7-C7			
8	7	enter #	enter #	=B8-C8			
9	8	enter #	enter #	=B9-C9			
10	9	enter #	enter #	=B10-C10			
11	10	enter #	enter #	=B11-C11			
12	11	enter #	enter #	=B12-C12			
13	12	enter #	enter #	=B13-C13			
14	13	enter #	enter #	=B14-C14			
15	14	enter #	enter #	=B15-C15			
16	15	enter #	enter #	=B16-C16			
17	16	enter #	enter #	=B17-C17			
18	17	enter #	enter #	=B18-C18			
19	18	enter #	enter #	=B19-C19			
20	19	enter #	enter #	=B20-C20			
21	20	enter #	enter #	=B21-C21			
22	Mean			=AVERAGE(D2:D21)			
23	Standard Deviation			=STDEV(D2:D21)			
24	Standard Error (Adjusted)			=D23*G4			
25	t			=D22/D24			
26	Degrees of Freedom			=G3-1			

Sheet1

Ready

Appendix C: Predictive Performance Statistics

Table C1

Dataset 1: Predictive performance statistics for the prediction equations developed using Neuroticism (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.457	7.346	7.340	7.337	7.337	7.340	7.400	0.13	0.14	0.14	0.14	0.14	0.14	0.15
2	6.755	6.709	6.747	6.721	6.715	6.723	6.649	-0.06	0.04	-0.04	0.02	0.03	0.01	0.01
3	6.768	6.630	6.613	6.619	6.623	6.621	6.694	0.07	0.17	0.17	0.17	0.17	0.17	0.14
4	7.055	7.075	7.089	7.083	7.081	7.082	7.016	-0.24	-0.02	-0.02	-0.02	-0.02	-0.02	-0.19
5	7.096	7.048	7.070	7.147	7.146	7.103	7.025	0.10	0.12	0.11	0.09	0.10	0.11	0.14
6	7.563	7.551	7.555	7.555	7.569	7.557	7.583	-0.04	0.01	0.01	0.01	-0.01	0.01	-0.01
7	7.287	7.252	7.169	7.166	7.160	7.187	7.313	0.02	0.04	0.11	0.11	0.12	0.10	0.05
8	7.135	7.319	7.332	7.322	7.340	7.328	7.215	0.08	0.02	0.02	0.02	0.03	0.02	0.03
9	7.518	7.544	7.548	7.843	7.872	7.702	7.528	0.00	0.04	0.04	-0.02	-0.03	0.01	0.03
10	7.859	7.915	8.006	7.985	8.005	7.978	7.904	0.01	-0.03	-0.05	-0.05	-0.05	-0.04	-0.02
11	7.948	7.906	7.906	7.904	7.904	7.905	7.914	0.05	0.11	0.11	0.10	0.11	0.11	0.10
12	7.638	7.531	7.533	7.517	7.543	7.531	7.576	0.00	0.08	0.08	0.08	0.06	0.08	0.06
13	7.040	6.939	6.938	6.952	6.966	6.949	7.007	0.06	0.20	0.20	0.20	0.19	0.19	0.12
14	6.874	6.858	6.824	6.837	6.826	6.836	6.927	0.10	0.11	0.17	0.17	0.17	0.15	0.17
15	7.223	7.152	7.136	7.132	7.136	7.139	7.106	0.03	0.08	0.06	0.06	0.06	0.06	0.06
16	7.701	7.664	7.677	7.676	7.668	7.671	7.669	0.13	0.12	0.12	0.11	0.12	0.12	0.17
17	7.664	7.580	7.571	7.570	7.622	7.586	7.595	-0.03	0.06	0.06	0.06	0.06	0.06	0.04
18	7.458	7.495	7.509	7.557	7.485	7.511	7.409	0.10	0.00	-0.01	-0.04	-0.01	-0.01	0.06
19	8.152	8.176	8.175	8.177	8.177	8.176	8.250	-0.10	-0.05	-0.05	-0.05	-0.05	-0.05	-0.07
20	7.699	7.944	7.945	7.953	7.946	7.947	7.792	0.00	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
Mean	**7.395**	**7.382**	**7.384**	**7.403**	**7.406**	**7.394**	**7.379**	**0.02**	**0.06**	**0.06**	**0.06**	**0.06**	**0.06**	**0.05**
SD	*0.396*	*0.428*	*0.437*	*0.445*	*0.448*	*0.438*	*0.425*	*0.09*	*0.07*	*0.08*	*0.08*	*0.08*	*0.08*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C2

Dataset 1: Predictive performance statistics for the prediction equations developed using Extraversion (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.292	7.276	7.226	7.218	7.222	7.236	7.230	0.16	0.18	0.21	0.22	0.21	0.20	0.19
2	6.447	6.419	6.347	6.334	6.327	6.357	6.386	0.20	0.21	0.17	0.19	0.19	0.19	0.22
3	6.687	6.578	6.559	6.569	6.593	6.575	6.637	0.19	0.21	0.21	0.21	0.20	0.21	0.20
4	7.136	7.321	7.494	7.381	7.447	7.411	7.248	-0.02	-0.06	-0.07	-0.07	-0.07	-0.07	-0.05
5	6.989	6.994	6.990	6.996	6.991	6.993	6.977	0.11	0.11	0.10	0.10	0.10	0.10	0.12
6	7.456	7.362	7.366	7.365	7.359	7.363	7.438	0.11	0.15	0.15	0.15	0.15	0.15	0.14
7	7.108	7.106	7.095	7.082	7.114	7.099	6.984	0.16	0.17	0.20	0.21	0.19	0.19	0.18
8	7.235	7.429	7.484	7.412	7.448	7.443	7.383	0.08	0.06	0.07	0.06	0.06	0.06	0.07
9	7.445	7.440	7.433	7.435	7.435	7.436	7.428	0.06	0.05	0.06	0.06	0.06	0.05	0.07
10	7.746	7.798	7.802	7.797	7.803	7.800	7.709	0.21	0.00	0.00	0.01	0.00	0.00	0.10
11	7.901	7.867	7.859	7.878	7.880	7.871	7.879	0.28	0.29	0.29	0.28	0.24	0.28	0.29
12	7.447	7.443	7.465	7.490	7.511	7.477	7.459	0.24	0.25	0.22	0.17	0.15	0.20	0.22
13	6.951	6.980	6.981	6.992	6.974	6.982	6.970	0.21	0.17	0.17	0.16	0.20	0.17	0.25
14	6.787	6.750	6.760	6.743	6.747	6.750	6.844	0.28	0.33	0.32	0.33	0.33	0.33	0.29
15	7.111	7.041	7.056	7.051	7.052	7.050	7.081	0.16	0.19	0.17	0.18	0.18	0.18	0.18
16	7.608	7.540	7.530	7.516	7.532	7.529	7.483	0.28	0.29	0.28	0.27	0.27	0.28	0.29
17	7.511	7.487	7.409	7.395	7.389	7.420	7.492	0.20	0.22	0.24	0.24	0.23	0.23	0.21
18	7.285	7.247	7.253	7.249	7.252	7.250	7.265	0.20	0.18	0.17	0.17	0.17	0.17	0.20
19	8.000	8.005	7.990	8.000	7.989	7.996	7.989	0.06	0.07	0.09	0.09	0.09	0.08	0.08
20	7.632	7.734	7.851	7.721	7.933	7.810	7.616	0.06	0.05	0.02	0.05	0.03	0.04	0.06
Mean	**7.289**	**7.291**	**7.297**	**7.281**	**7.300**	**7.292**	**7.275**	**0.16**	**0.15**	**0.15**	**0.15**	**0.15**	**0.15**	**0.17**
SD	*0.401*	*0.417*	*0.434*	*0.427*	*0.441*	*0.429*	*0.401*	*0.08*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C3

Dataset 1: Predictive performance statistics for the prediction equations developed using Openness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.394	7.443	7.452	7.462	7.475	7.458	7.437	0.10	0.05	0.04	0.03	0.02	0.03	0.09
2	6.781	6.783	6.810	6.845	6.885	6.831	6.751	0.05	0.04	0.04	0.01	-0.01	0.02	0.04
3	6.880	6.890	6.888	6.901	6.913	6.898	6.834	0.00	-0.01	0.00	-0.01	-0.04	-0.01	0.00
4	6.833	6.834	6.833	6.830	6.833	6.832	6.826	0.15	0.15	0.15	0.14	0.15	0.15	0.15
5	7.494	7.485	7.457	7.473	7.570	7.496	7.466	0.01	0.01	0.01	0.01	-0.02	0.00	0.01
6	7.503	7.519	7.575	7.614	7.579	7.572	7.528	0.17	0.17	0.11	0.07	0.05	0.10	0.17
7	7.360	7.358	7.285	7.339	7.314	7.324	7.318	-0.01	-0.01	-0.03	-0.01	-0.03	-0.02	-0.01
8	7.120	7.142	7.153	7.160	7.162	7.154	7.186	0.12	0.08	0.07	0.06	0.06	0.07	0.09
9	7.388	7.389	7.397	7.397	7.398	7.395	7.440	0.15	0.15	0.12	0.11	0.11	0.12	0.14
10	7.818	7.831	7.879	7.886	7.895	7.873	7.732	0.29	0.27	-0.10	-0.11	-0.12	-0.02	0.24
11	7.866	7.866	7.876	7.894	7.895	7.883	7.842	0.13	0.13	0.13	0.12	0.12	0.12	0.13
12	7.452	7.450	7.462	7.490	7.545	7.487	7.484	0.24	0.24	0.21	0.16	0.06	0.17	0.23
13	6.942	6.943	6.941	6.942	6.942	6.942	6.972	0.20	0.20	0.19	0.19	0.19	0.19	0.20
14	6.776	6.804	6.815	6.799	6.800	6.805	6.726	0.16	0.15	0.14	0.15	0.15	0.15	0.15
15	7.075	7.096	7.099	7.092	7.095	7.095	7.139	0.21	0.19	0.18	0.19	0.19	0.19	0.19
16	7.568	7.571	7.566	7.561	7.574	7.568	7.648	0.09	0.09	0.10	0.10	0.10	0.10	0.10
17	7.496	7.533	7.611	7.648	7.642	7.608	7.523	0.23	0.17	0.02	-0.04	-0.06	0.03	0.18
18	7.513	7.576	7.601	7.607	7.582	7.591	7.529	0.11	0.07	0.05	0.04	0.06	0.06	0.11
19	7.774	7.804	7.876	7.892	7.877	7.862	7.814	0.14	0.11	0.01	0.00	0.01	0.03	0.09
20	7.554	7.556	7.578	7.572	7.586	7.573	7.565	0.11	0.10	0.08	0.09	0.08	0.09	0.10
Mean	**7.329**	**7.344**	**7.358**	**7.370**	**7.378**	**7.362**	**7.338**	**0.13**	**0.12**	**0.08**	**0.07**	**0.05**	**0.08**	**0.12**
SD	*0.347*	*0.349*	*0.361*	*0.365*	*0.363*	*0.359*	*0.354*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C4

Dataset 1: Predictive performance statistics for the prediction equations developed using Agreeableness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.466	7.458	7.458	7.453	7.459	7.457	7.477	-0.09	0.00	0.01	0.01	0.00	0.01	0.03
2	6.671	6.672	6.680	6.679	6.678	6.677	6.641	0.02	0.04	0.05	0.05	0.05	0.05	0.03
3	6.752	6.750	6.800	6.777	6.928	6.814	6.755	0.01	0.01	-0.03	-0.02	-0.05	-0.02	0.01
4	6.982	6.981	6.979	6.979	6.982	6.980	6.971	-0.12	-0.12	-0.12	-0.12	-0.12	-0.12	-0.13
5	7.115	7.127	7.155	7.150	7.152	7.146	7.135	0.01	0.00	-0.01	-0.01	-0.01	-0.01	0.01
6	7.625	7.615	7.609	7.612	7.607	7.610	7.564	-0.14	-0.13	-0.12	-0.12	-0.11	-0.12	-0.07
7	7.185	7.193	7.211	7.262	7.279	7.236	7.172	-0.02	-0.03	-0.05	-0.10	-0.12	-0.07	-0.09
8	7.153	7.154	7.153	7.163	7.163	7.158	7.130	-0.03	-0.03	-0.03	-0.06	-0.08	-0.05	-0.04
9	7.522	7.520	7.523	7.522	7.522	7.522	7.541	-0.03	-0.03	-0.04	-0.02	-0.03	-0.03	-0.03
10	8.074	8.042	8.034	8.041	8.043	8.040	7.971	-0.08	-0.12	-0.13	-0.12	-0.13	-0.12	-0.08
11	8.078	8.229	8.220	8.226	8.235	8.227	8.104	-0.22	-0.19	-0.18	-0.18	-0.18	-0.18	-0.24
12	7.574	7.574	7.569	7.572	7.587	7.576	7.560	-0.21	-0.21	-0.20	-0.20	-0.21	-0.20	-0.22
13	7.068	7.065	7.053	7.051	7.116	7.071	7.096	-0.03	-0.03	-0.02	-0.01	-0.06	-0.03	-0.05
14	6.871	6.873	6.934	7.070	7.190	7.017	6.868	0.08	0.08	-0.08	-0.10	-0.11	-0.05	0.05
15	7.315	7.314	7.543	7.480	7.667	7.501	7.333	-0.24	-0.24	-0.15	-0.17	-0.13	-0.17	-0.24
16	7.780	7.822	7.885	7.952	7.859	7.879	7.818	-0.03	-0.04	-0.07	-0.11	-0.09	-0.08	-0.04
17	7.656	7.691	7.758	7.737	7.738	7.731	7.675	-0.16	-0.20	-0.23	-0.25	-0.23	-0.23	-0.10
18	7.599	7.599	7.633	7.623	7.636	7.623	7.627	-0.21	-0.21	-0.23	-0.23	-0.24	-0.23	-0.22
19	7.948	7.935	8.126	8.004	8.036	8.025	7.951	-0.09	-0.10	-0.11	-0.13	-0.11	-0.11	-0.10
20	7.592	7.583	7.587	7.583	7.588	7.585	7.608	-0.12	-0.08	-0.05	-0.08	-0.11	-0.08	-0.12
Mean	**7.401**	**7.410**	**7.446**	**7.447**	**7.473**	**7.444**	**7.400**	**-0.08**	**-0.08**	**-0.09**	**-0.10**	**-0.10**	**-0.09**	**-0.08**
SD	*0.418*	*0.430*	*0.440*	*0.427*	*0.408*	*0.424*	*0.417*	*0.09*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C5

Dataset 1: Predictive performance statistics for the prediction equations developed using Conscientiousness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.319	7.541	7.517	7.497	7.479	7.508	7.321	0.22	0.14	0.14	0.16	0.16	0.15	0.21
2	6.481	6.502	6.476	6.500	6.498	6.494	6.497	0.25	0.20	0.22	0.20	0.21	0.21	0.24
3	6.613	6.816	6.773	6.787	6.800	6.794	6.718	0.21	0.13	0.14	0.14	0.14	0.14	0.18
4	7.567	7.625	7.635	7.656	7.635	7.638	7.609	0.02	0.05	0.01	0.01	0.04	0.03	0.03
5	7.139	7.146	7.139	7.116	7.053	7.114	7.102	0.22	0.23	0.19	0.20	0.23	0.21	0.23
6	7.381	7.627	7.636	7.643	7.606	7.628	7.445	0.25	0.14	0.14	0.13	0.14	0.14	0.21
7	6.843	7.062	7.091	7.085	7.052	7.072	6.980	0.30	0.17	0.16	0.16	0.16	0.16	0.26
8	6.857	6.855	6.806	6.808	6.782	6.813	6.918	0.26	0.26	0.27	0.27	0.27	0.27	0.26
9	7.225	7.225	7.213	7.210	7.209	7.214	7.201	0.29	0.29	0.30	0.31	0.31	0.30	0.29
10	7.779	7.779	7.809	7.772	7.787	7.786	7.883	0.16	0.17	0.17	0.19	0.19	0.18	0.17
11	7.753	7.806	7.814	7.791	7.790	7.800	7.654	0.32	0.19	0.19	0.22	0.20	0.20	0.29
12	7.151	7.142	7.132	7.131	7.109	7.128	7.179	0.28	0.28	0.28	0.28	0.29	0.28	0.28
13	7.103	7.082	7.090	7.084	7.082	7.084	7.096	0.12	0.13	0.12	0.12	0.13	0.12	0.12
14	6.739	6.704	6.698	6.680	6.646	6.682	6.601	0.20	0.22	0.22	0.24	0.25	0.23	0.20
15	7.059	7.052	7.034	7.037	7.036	7.040	7.004	0.22	0.23	0.25	0.24	0.25	0.24	0.23
16	7.481	7.474	7.421	7.404	7.410	7.427	7.458	0.20	0.20	0.20	0.20	0.20	0.20	0.20
17	7.428	7.424	7.390	7.388	7.400	7.400	7.431	0.22	0.23	0.25	0.26	0.25	0.25	0.23
18	7.440	7.470	7.433	7.368	7.413	7.421	7.412	0.16	0.16	0.17	0.18	0.17	0.17	0.16
19	7.608	7.600	7.572	7.581	7.558	7.577	7.571	0.26	0.27	0.28	0.19	0.21	0.24	0.27
20	7.562	7.527	7.529	7.527	7.526	7.528	7.541	0.23	0.21	0.21	0.22	0.22	0.21	0.23
Mean	**7.226**	**7.273**	**7.260**	**7.253**	**7.243**	**7.257**	**7.231**	**0.22**	**0.19**	**0.20**	**0.20**	**0.20**	**0.20**	**0.21**
SD	*0.373*	*0.370*	*0.378*	*0.371*	*0.375*	*0.373*	*0.369*	*0.07*	*0.06*	*0.07*	*0.07*	*0.06*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C6

Dataset 2: Predictive performance statistics for the prediction equations developed using Neuroticism (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	26.991	26.995	26.984	27.163	27.029	27.043	26.978	0.15	0.15	0.15	0.12	0.15	0.14	0.16
2	34.226	34.311	34.551	34.368	34.467	34.424	34.327	0.24	0.23	0.17	0.21	0.18	0.20	0.24
3	32.758	32.804	32.687	32.744	32.740	32.744	32.352	0.07	0.07	0.07	0.07	0.06	0.07	0.07
4	30.883	31.521	31.808	31.445	32.030	31.701	30.787	0.17	0.14	0.09	0.14	0.07	0.11	0.18
5	33.146	32.960	32.919	32.918	32.993	32.948	33.047	0.05	0.05	0.05	0.05	0.05	0.05	0.05
6	30.927	30.758	30.759	30.813	30.857	30.797	30.893	0.07	0.07	0.07	0.07	0.08	0.07	0.07
7	34.123	34.447	34.234	34.122	34.019	34.205	33.850	0.04	0.04	0.04	0.05	0.05	0.04	0.05
8	33.150	33.128	33.116	33.112	33.118	33.119	33.108	0.24	0.20	0.20	0.20	0.20	0.20	0.22
9	34.687	34.485	34.312	34.272	34.453	34.380	34.520	-0.02	-0.02	-0.03	-0.03	-0.04	-0.03	-0.02
10	33.753	34.219	34.740	34.238	34.343	34.385	33.880	0.14	0.08	0.03	0.07	0.06	0.06	0.12
11	28.534	28.542	28.490	28.503	28.469	28.501	28.484	0.10	0.10	0.11	0.10	0.10	0.10	0.11
12	35.198	35.123	35.159	35.107	35.072	35.115	35.299	0.06	0.06	0.06	0.06	0.06	0.06	0.06
13	32.994	33.366	33.298	33.295	33.251	33.302	33.273	0.11	0.06	0.07	0.07	0.07	0.06	0.09
14	33.167	33.115	33.370	33.222	33.281	33.247	33.764	0.11	0.12	0.10	0.09	0.10	0.10	0.11
15	34.064	34.675	34.587	35.280	34.668	34.802	34.259	0.20	0.09	0.10	0.02	0.09	0.07	0.16
16	29.345	29.322	29.339	29.357	29.362	29.345	29.220	0.08	0.08	0.08	0.08	0.07	0.08	0.08
17	34.587	34.588	34.597	34.563	34.611	34.590	34.590	0.24	0.24	0.24	0.24	0.23	0.24	0.24
18	31.861	33.288	33.200	33.703	33.440	33.408	32.421	0.32	0.02	0.03	0.00	0.01	0.01	0.19
19	30.313	30.318	30.309	30.315	30.320	30.315	30.251	0.17	0.16	0.16	0.16	0.16	0.16	0.17
20	33.172	33.495	33.882	33.720	34.041	33.784	33.192	0.08	0.06	0.03	0.04	0.03	0.04	0.07
Mean	**32.394**	**32.573**	**32.617**	**32.613**	**32.628**	**32.608**	**32.425**	**0.13**	**0.10**	**0.09**	**0.09**	**0.09**	**0.09**	**0.12**
SD	*2.226*	*2.272*	*2.297*	*2.278*	*2.267*	*2.275*	*2.267*	*0.09*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C7

Dataset 2: Predictive performance statistics for the prediction equations developed using Extraversion (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.491	27.800	27.773	27.832	27.786	27.798	27.818	0.13	0.10	0.10	0.10	0.11	0.10	0.11
2	33.974	33.767	33.614	33.156	33.242	33.445	33.621	0.26	0.30	0.33	0.34	0.33	0.32	0.30
3	31.371	31.771	31.775	31.806	31.800	31.788	32.229	0.33	0.06	0.06	0.05	0.06	0.06	0.16
4	32.259	32.036	32.227	31.928	31.754	31.986	32.132	0.05	0.08	0.06	0.08	0.09	0.08	0.05
5	34.357	34.438	34.681	34.717	34.588	34.606	34.357	0.05	0.06	0.07	0.07	0.07	0.07	0.04
6	31.654	32.372	32.344	32.407	32.526	32.412	31.708	-0.01	-0.01	-0.01	-0.01	-0.05	-0.02	-0.02
7	34.683	36.128	36.094	35.938	36.079	36.060	35.071	0.08	-0.01	-0.01	0.00	-0.01	-0.01	0.02
8	33.474	33.469	33.480	33.662	33.562	33.543	33.719	0.17	0.16	0.16	0.18	0.17	0.17	0.17
9	34.328	33.764	33.396	32.841	32.760	33.190	33.989	-0.03	0.00	0.01	0.06	0.06	0.03	-0.02
10	34.766	34.486	34.438	34.239	33.938	34.275	34.749	0.04	0.09	0.10	0.12	0.15	0.11	0.05
11	27.755	27.534	27.575	27.765	27.731	27.651	27.542	0.38	0.17	0.17	0.14	0.17	0.16	0.28
12	34.382	34.049	33.908	33.892	33.932	33.945	34.780	0.16	0.12	0.13	0.13	0.13	0.13	0.13
13	34.006	33.859	33.831	33.625	33.642	33.739	33.736	0.04	0.08	0.11	0.11	0.10	0.10	0.05
14	32.880	32.878	32.835	32.836	32.690	32.810	32.529	0.27	0.10	0.09	0.08	0.12	0.10	0.26
15	34.617	34.010	34.168	33.754	33.838	33.942	34.777	0.22	0.29	0.26	0.29	0.28	0.28	0.23
16	28.489	29.799	29.836	29.919	30.072	29.907	28.749	0.16	-0.05	-0.05	-0.06	-0.05	-0.05	0.06
17	34.609	34.355	34.517	34.414	34.302	34.397	34.759	0.35	0.33	0.30	0.32	0.32	0.32	0.33
18	32.079	32.066	32.438	31.795	31.957	32.064	32.167	0.31	0.19	0.18	0.23	0.22	0.21	0.28
19	30.435	30.419	30.397	30.618	30.720	30.539	30.314	0.15	0.18	0.19	0.18	0.17	0.18	0.16
20	33.050	33.356	33.343	33.362	33.226	33.322	33.299	0.20	0.08	0.08	0.08	0.09	0.08	0.13
Mean	**32.533**	**32.618**	**32.633**	**32.525**	**32.507**	**32.571**	**32.602**	**0.16**	**0.11**	**0.12**	**0.12**	**0.13**	**0.12**	**0.14**
SD	*2.353*	*2.235*	*2.227*	*2.136*	*2.119*	*2.175*	*2.338*	*0.12*	*0.10*	*0.10*	*0.11*	*0.11*	*0.10*	*0.11*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C8

Dataset 2: Predictive performance statistics for the prediction equations developed using Openness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.159	27.788	27.827	27.876	27.863	27.839	27.412	0.13	0.06	0.05	0.05	0.05	0.05	0.11
2	34.712	34.727	34.732	34.777	34.751	34.747	34.699	0.14	0.15	0.15	0.17	0.17	0.16	0.15
3	32.144	32.191	32.192	32.154	32.177	32.178	31.952	0.18	0.08	0.10	0.10	0.09	0.09	0.16
4	31.411	31.557	31.764	31.726	31.592	31.660	31.354	-0.03	-0.05	-0.01	0.00	-0.02	-0.02	-0.03
5	33.019	33.179	33.088	33.480	33.217	33.241	33.193	0.13	0.07	0.10	0.09	0.03	0.07	0.10
6	32.500	34.228	33.798	33.743	33.834	33.901	32.710	-0.04	-0.10	-0.08	-0.08	-0.09	-0.09	-0.06
7	34.846	34.824	34.865	35.384	35.359	35.108	34.855	-0.02	-0.01	-0.01	0.01	0.01	0.00	-0.01
8	33.464	33.452	33.425	33.698	33.483	33.515	33.350	0.14	0.14	0.17	0.12	0.18	0.15	0.13
9	34.709	34.719	34.693	34.566	34.752	34.682	34.681	-0.03	-0.03	-0.03	-0.02	-0.02	-0.03	-0.03
10	35.104	35.096	35.021	35.035	35.020	35.043	34.647	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.01
11	30.088	31.083	31.023	31.076	30.770	30.988	29.958	-0.08	-0.09	-0.09	-0.10	-0.09	-0.09	-0.08
12	35.651	35.916	35.836	35.978	36.118	35.962	35.859	0.02	0.02	0.04	0.03	0.02	0.03	0.02
13	33.144	34.561	35.753	34.641	34.468	34.856	33.190	0.11	0.03	-0.05	-0.03	-0.02	-0.02	0.10
14	33.870	33.776	33.705	33.714	33.678	33.718	33.683	0.06	0.05	0.06	0.06	0.07	0.06	0.05
15	34.779	34.770	34.968	34.760	34.761	34.815	35.207	0.08	0.08	0.09	0.09	0.09	0.09	0.08
16	29.344	29.575	30.771	30.848	30.965	30.540	29.280	0.02	0.02	-0.01	-0.01	-0.01	0.00	0.02
17	34.949	34.982	35.026	34.963	34.932	34.976	34.866	0.20	0.19	0.17	0.18	0.19	0.18	0.20
18	33.836	33.930	33.718	33.912	34.040	33.900	33.595	0.02	0.02	0.03	0.02	0.05	0.03	0.02
19	31.180	31.190	31.220	31.590	31.393	31.348	30.935	0.05	0.05	0.06	0.07	0.07	0.06	0.05
20	33.424	33.982	33.843	33.853	33.893	33.893	33.820	0.01	-0.04	-0.01	0.01	-0.02	-0.02	0.00
Mean	**32.967**	**33.276**	**33.363**	**33.389**	**33.353**	**33.345**	**32.962**	**0.05**	**0.03**	**0.04**	**0.04**	**0.04**	**0.04**	**0.05**
SD	*2.204*	*2.089*	*2.025*	*1.962*	*2.001*	*2.009*	*2.207*	*0.08*	*0.08*	*0.08*	*0.07*	*0.08*	*0.08*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C9

Dataset 2: Predictive performance statistics for the prediction equations developed using Agreeableness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.690	27.706	27.693	27.732	27.621	27.688	27.774	0.10	0.09	0.09	0.07	0.08	0.09	0.10
2	34.460	34.463	34.493	34.370	34.466	34.448	34.529	0.14	0.14	0.12	0.15	0.12	0.13	0.14
3	31.807	32.129	32.175	32.113	32.171	32.147	31.875	0.06	0.05	0.05	0.05	0.05	0.05	0.06
4	31.195	32.079	32.253	32.186	32.250	32.192	31.732	0.20	-0.07	-0.08	-0.07	-0.11	-0.08	-0.03
5	32.874	32.768	32.943	33.367	33.750	33.207	32.862	0.08	0.07	0.09	0.01	-0.02	0.04	0.09
6	30.679	31.351	31.671	31.354	31.349	31.431	30.972	0.04	-0.02	-0.02	0.00	-0.01	-0.01	0.03
7	33.551	34.010	34.074	34.043	34.058	34.046	33.851	0.09	0.07	0.07	0.07	0.07	0.07	0.09
8	33.373	33.420	33.412	33.376	33.346	33.388	33.279	0.13	0.13	0.13	0.11	0.09	0.11	0.13
9	32.731	33.459	33.365	33.267	33.353	33.361	33.021	0.16	0.02	0.03	0.04	0.03	0.03	0.09
10	34.131	34.256	34.397	34.295	34.286	34.308	34.150	0.09	0.05	0.02	0.04	0.04	0.04	0.09
11	28.356	28.436	28.579	28.541	28.582	28.535	28.487	0.05	0.04	0.02	0.02	0.01	0.02	0.05
12	35.325	35.271	35.278	35.281	35.170	35.250	35.358	0.07	0.07	0.06	0.05	0.04	0.05	0.07
13	32.927	33.770	33.764	33.806	33.786	33.781	32.880	0.15	0.04	0.04	0.03	0.03	0.04	0.08
14	33.479	33.516	33.532	33.504	33.501	33.513	33.338	0.02	0.03	0.03	0.03	0.03	0.03	0.02
15	36.541	37.946	37.794	37.625	38.016	37.845	37.376	-0.03	-0.09	-0.11	-0.11	-0.11	-0.11	-0.07
16	29.146	29.202	29.302	29.428	29.590	29.381	29.116	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03
17	36.537	37.058	37.149	36.928	36.873	37.002	36.725	-0.03	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02
18	32.712	32.802	32.879	32.856	32.866	32.851	32.791	0.10	0.10	0.09	0.08	0.07	0.09	0.10
19	30.289	30.281	30.311	30.405	30.218	30.304	30.430	0.17	0.17	0.17	0.11	0.15	0.15	0.17
20	32.882	32.893	32.972	33.097	33.257	33.055	32.754	0.21	0.21	0.21	0.19	0.17	0.20	0.21
Mean	**32.534**	**32.841**	**32.902**	**32.879**	**32.925**	**32.887**	**32.665**	**0.09**	**0.05**	**0.05**	**0.04**	**0.03**	**0.04**	**0.07**
SD	*2.433*	*2.582*	*2.548*	*2.502*	*2.547*	*2.542*	*2.486*	*0.07*	*0.08*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C10

Dataset 2: Predictive performance statistics for the prediction equations developed using Conscientiousness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.029	27.041	27.000	27.038	27.042	27.030	27.074	0.23	0.24	0.24	0.24	0.24	0.24	0.24
2	32.913	32.912	32.888	32.982	33.002	32.946	32.942	0.35	0.35	0.34	0.33	0.33	0.34	0.35
3	30.499	30.511	30.518	30.523	30.555	30.527	30.458	0.29	0.29	0.28	0.28	0.28	0.28	0.29
4	29.976	29.990	29.970	30.493	30.635	30.272	29.914	0.29	0.29	0.29	0.25	0.25	0.27	0.29
5	32.283	33.332	33.268	33.264	33.361	33.306	32.603	0.24	0.16	0.16	0.16	0.16	0.16	0.21
6	30.495	30.541	30.553	30.542	30.495	30.533	30.427	0.16	0.16	0.16	0.16	0.16	0.16	0.16
7	32.864	33.231	33.346	33.619	33.498	33.423	33.261	0.19	0.17	0.17	0.15	0.17	0.16	0.18
8	32.040	32.111	32.051	32.054	32.058	32.068	31.909	0.29	0.28	0.28	0.28	0.28	0.28	0.28
9	33.470	33.731	33.668	33.646	33.573	33.655	33.513	0.18	0.18	0.18	0.18	0.18	0.18	0.18
10	34.831	34.881	34.851	34.840	34.838	34.852	34.710	0.18	0.18	0.18	0.18	0.18	0.18	0.18
11	27.512	27.470	27.416	27.448	27.420	27.439	27.669	0.32	0.32	0.32	0.32	0.32	0.32	0.32
12	34.750	34.775	34.886	34.869	34.974	34.876	34.774	0.17	0.17	0.16	0.17	0.16	0.16	0.17
13	31.847	32.099	32.141	32.119	32.040	32.100	31.877	0.30	0.27	0.27	0.27	0.27	0.27	0.29
14	31.465	31.373	31.408	31.469	31.426	31.419	31.479	0.29	0.29	0.29	0.29	0.29	0.29	0.29
15	33.490	33.493	33.542	33.449	33.437	33.480	34.022	0.29	0.28	0.28	0.28	0.27	0.28	0.28
16	28.349	28.409	28.361	28.450	28.441	28.415	28.298	0.21	0.21	0.21	0.21	0.21	0.21	0.21
17	34.572	34.543	34.533	34.535	34.530	34.535	34.560	0.25	0.25	0.25	0.25	0.25	0.25	0.25
18	31.906	31.919	31.988	31.978	31.938	31.956	32.042	0.30	0.29	0.29	0.30	0.29	0.29	0.30
19	29.757	29.859	29.804	29.845	29.926	29.859	29.621	0.24	0.23	0.23	0.23	0.23	0.23	0.23
20	30.974	30.980	30.852	31.005	30.988	30.956	30.926	0.37	0.36	0.37	0.36	0.36	0.36	0.36
Mean	**31.551**	**31.660**	**31.652**	**31.709**	**31.709**	**31.682**	**31.604**	**0.26**	**0.25**	**0.25**	**0.24**	**0.24**	**0.25**	**0.25**
SD	*2.264*	*2.316*	*2.340*	*2.311*	*2.310*	*2.318*	*2.302*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C11

Dataset 3: Predictive performance statistics for the prediction equations developed using Neuroticism (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.060	1.060	1.060	1.060	1.060	1.060	1.056	0.06	0.06	0.06	0.06	0.06	0.06	0.06
2	1.091	1.111	1.112	1.114	1.112	1.112	1.105	0.17	0.05	0.05	0.04	0.05	0.05	0.10
3	1.102	1.102	1.103	1.104	1.103	1.103	1.099	0.19	0.19	0.18	0.18	0.17	0.18	0.19
4	1.031	1.033	1.034	1.037	1.035	1.035	1.031	0.21	0.19	0.18	0.15	0.17	0.17	0.20
5	1.054	1.056	1.056	1.057	1.056	1.056	1.057	0.10	0.09	0.08	0.08	0.08	0.09	0.10
6	0.956	0.956	0.957	0.957	0.957	0.957	0.958	0.07	0.07	0.07	0.06	0.07	0.07	0.07
7	1.111	1.119	1.119	1.119	1.119	1.119	1.114	0.23	0.13	0.13	0.12	0.12	0.12	0.20
8	1.050	1.059	1.062	1.064	1.075	1.065	1.054	0.18	0.11	0.09	0.08	0.05	0.08	0.14
9	1.063	1.066	1.066	1.065	1.066	1.066	1.068	0.06	0.03	0.03	0.03	0.03	0.03	0.06
10	1.065	1.065	1.065	1.067	1.068	1.066	1.066	0.07	0.07	0.07	0.06	0.06	0.06	0.07
11	1.032	1.031	1.032	1.033	1.033	1.032	1.036	0.14	0.14	0.14	0.13	0.13	0.14	0.13
12	1.044	1.046	1.049	1.049	1.049	1.048	1.050	0.22	0.17	0.13	0.12	0.14	0.14	0.18
13	1.122	1.123	1.123	1.123	1.123	1.123	1.119	0.04	0.04	0.05	0.05	0.04	0.05	0.05
14	1.100	1.118	1.116	1.117	1.122	1.118	1.110	0.19	0.08	0.09	0.07	0.05	0.07	0.15
15	1.162	1.162	1.162	1.162	1.162	1.162	1.157	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00
16	1.038	1.038	1.039	1.040	1.040	1.039	1.039	0.06	0.05	0.05	0.04	0.05	0.05	0.05
17	0.964	0.967	0.968	0.968	0.969	0.968	0.967	0.17	0.13	0.13	0.13	0.12	0.13	0.15
18	1.065	1.089	1.088	1.088	1.087	1.088	1.076	0.11	0.02	0.02	0.02	0.02	0.02	0.04
19	1.070	1.071	1.071	1.080	1.072	1.073	1.073	0.11	0.11	0.10	0.06	0.09	0.09	0.11
20	1.080	1.081	1.083	1.086	1.078	1.082	1.076	-0.03	-0.04	-0.04	-0.05	-0.05	-0.04	-0.04
Mean	**1.063**	**1.068**	**1.068**	**1.069**	**1.069**	**1.069**	**1.066**	**0.12**	**0.08**	**0.08**	**0.07**	**0.07**	**0.08**	**0.10**
SD	*0.048*	*0.050*	*0.050*	*0.050*	*0.050*	*0.050*	*0.048*	*0.08*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C12

Dataset 3: Predictive performance statistics for the prediction equations developed using Extraversion (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.039	1.040	1.040	1.039	1.039	1.039	1.042	0.21	0.22	0.22	0.22	0.22	0.22	0.21
2	1.089	1.089	1.087	1.085	1.084	1.086	1.090	0.16	0.16	0.17	0.17	0.17	0.17	0.16
3	1.117	1.114	1.149	1.152	1.143	1.139	1.114	0.06	0.05	-0.06	-0.06	-0.04	-0.03	0.06
4	1.022	1.020	1.020	1.019	1.019	1.019	1.032	0.19	0.20	0.20	0.20	0.20	0.20	0.19
5	1.053	1.066	1.068	1.068	1.063	1.066	1.057	0.14	0.09	0.08	0.08	0.08	0.08	0.14
6	0.947	0.945	0.945	0.944	0.942	0.944	0.948	0.25	0.26	0.26	0.26	0.26	0.26	0.24
7	1.100	1.099	1.099	1.100	1.099	1.099	1.101	0.30	0.30	0.30	0.30	0.30	0.30	0.30
8	1.086	1.085	1.085	1.088	1.095	1.088	1.083	0.06	0.06	0.06	0.05	0.05	0.06	0.05
9	1.078	1.083	1.083	1.085	1.085	1.084	1.090	0.12	0.07	0.07	0.06	0.06	0.07	0.10
10	1.048	1.045	1.042	1.042	1.043	1.043	1.042	0.20	0.19	0.19	0.19	0.19	0.19	0.19
11	1.015	1.016	1.017	1.021	1.025	1.019	1.019	0.25	0.24	0.24	0.20	0.16	0.21	0.25
12	1.065	1.067	1.067	1.067	1.068	1.067	1.066	0.11	0.11	0.11	0.11	0.11	0.11	0.11
13	1.099	1.098	1.096	1.096	1.094	1.096	1.102	0.16	0.16	0.16	0.16	0.16	0.16	0.16
14	1.120	1.119	1.116	1.116	1.117	1.117	1.098	0.13	0.13	0.14	0.14	0.14	0.14	0.13
15	1.129	1.127	1.127	1.128	1.128	1.128	1.126	0.10	0.10	0.09	0.09	0.10	0.10	0.10
16	1.046	1.064	1.066	1.052	1.057	1.060	1.057	0.08	0.02	0.02	0.05	0.04	0.03	0.07
17	0.969	1.022	1.009	0.997	0.999	1.006	0.983	0.17	0.06	0.07	0.09	0.09	0.08	0.12
18	1.062	1.068	1.068	1.066	1.067	1.067	1.061	0.18	0.14	0.13	0.14	0.14	0.14	0.16
19	1.069	1.078	1.072	1.073	1.073	1.074	1.062	0.08	0.03	0.06	0.06	0.06	0.05	0.07
20	1.103	1.102	1.097	1.111	1.130	1.110	1.097	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Mean	**1.063**	**1.067**	**1.068**	**1.067**	**1.068**	**1.068**	**1.063**	**0.15**	**0.13**	**0.13**	**0.13**	**0.13**	**0.13**	**0.14**
SD	*0.048*	*0.043*	*0.046*	*0.048*	*0.049*	*0.046*	*0.044*	*0.07*	*0.09*	*0.09*	*0.09*	*0.08*	*0.09*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C13

Dataset 3: Predictive performance statistics for the prediction equations developed using Openness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.049	1.052	1.046	1.046	1.052	1.049	1.050	0.19	0.18	0.20	0.20	0.18	0.19	0.19
2	1.102	1.119	1.119	1.122	1.119	1.120	1.103	0.07	0.04	0.03	0.03	0.03	0.03	0.05
3	1.105	1.111	1.112	1.111	1.109	1.111	1.110	0.16	0.10	0.10	0.11	0.11	0.10	0.14
4	1.030	1.047	1.047	1.045	1.045	1.046	1.038	0.15	0.09	0.09	0.10	0.10	0.10	0.13
5	1.070	1.073	1.075	1.074	1.075	1.074	1.070	0.11	0.09	0.08	0.08	0.08	0.08	0.10
6	0.964	0.964	0.965	0.964	0.965	0.965	0.969	0.16	0.17	0.16	0.17	0.16	0.16	0.17
7	1.113	1.110	1.111	1.111	1.112	1.111	1.112	0.20	0.21	0.21	0.21	0.21	0.21	0.20
8	1.048	1.057	1.057	1.057	1.057	1.057	1.054	0.23	0.09	0.09	0.09	0.09	0.09	0.15
9	1.092	1.092	1.093	1.099	1.105	1.097	1.086	-0.01	0.00	0.00	-0.02	-0.03	-0.01	-0.01
10	1.051	1.050	1.050	1.044	1.046	1.048	1.048	0.33	0.33	0.33	0.32	0.33	0.33	0.33
11	1.081	1.087	1.089	1.090	1.090	1.089	1.083	-0.05	-0.06	-0.06	-0.07	-0.06	-0.06	-0.06
12	1.064	1.066	1.066	1.066	1.065	1.066	1.065	0.14	0.11	0.11	0.11	0.12	0.11	0.14
13	1.103	1.100	1.102	1.097	1.096	1.099	1.122	0.18	0.19	0.18	0.19	0.18	0.18	0.18
14	1.127	1.129	1.128	1.125	1.125	1.127	1.119	0.05	0.04	0.05	0.05	0.05	0.05	0.05
15	1.123	1.121	1.121	1.122	1.122	1.121	1.123	0.10	0.10	0.11	0.10	0.11	0.10	0.11
16	1.075	1.078	1.079	1.081	1.080	1.080	1.069	0.02	0.03	0.03	0.03	0.03	0.03	0.02
17	0.995	0.998	0.998	1.000	0.994	0.997	0.996	0.06	0.05	0.05	0.04	0.06	0.05	0.05
18	1.087	1.087	1.087	1.094	1.087	1.089	1.087	0.09	0.10	0.10	0.08	0.10	0.09	0.09
19	1.072	1.081	1.083	1.086	1.086	1.084	1.082	0.08	0.05	0.05	0.04	0.04	0.04	0.06
20	1.043	1.078	1.075	1.075	1.081	1.077	1.056	0.14	0.00	0.01	0.01	0.00	0.00	0.06
Mean	**1.070**	**1.075**	**1.075**	**1.075**	**1.076**	**1.075**	**1.072**	**0.12**	**0.10**	**0.10**	**0.09**	**0.09**	**0.09**	**0.11**
SD	*0.041*	*0.041*	*0.041*	*0.041*	*0.041*	*0.041*	*0.041*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C14

Dataset 3: Predictive performance statistics for the prediction equations developed using Agreeableness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.037	1.038	1.038	1.038	1.038	1.038	1.048	0.22	0.21	0.21	0.21	0.21	0.21	0.22
2	1.090	1.130	1.124	1.123	1.130	1.127	1.102	0.26	0.15	0.14	0.17	0.14	0.15	0.21
3	1.088	1.088	1.094	1.090	1.105	1.094	1.088	0.19	0.19	0.18	0.19	0.15	0.18	0.19
4	1.020	1.020	1.022	1.021	1.021	1.021	1.029	0.31	0.30	0.27	0.32	0.30	0.30	0.31
5	1.070	1.069	1.069	1.069	1.070	1.069	1.069	0.06	0.06	0.06	0.06	0.06	0.06	0.06
6	0.967	0.967	0.993	0.992	0.980	0.983	0.968	0.13	0.13	0.07	0.08	0.11	0.10	0.13
7	1.109	1.114	1.112	1.120	1.116	1.115	1.110	0.26	0.24	0.23	0.23	0.23	0.23	0.25
8	1.062	1.071	1.074	1.075	1.075	1.074	1.068	0.17	0.14	0.13	0.12	0.12	0.13	0.15
9	1.045	1.047	1.047	1.052	1.053	1.050	1.052	0.15	0.15	0.15	0.14	0.14	0.14	0.15
10	1.057	1.064	1.065	1.063	1.066	1.064	1.062	0.19	0.16	0.15	0.16	0.15	0.16	0.17
11	1.035	1.036	1.036	1.060	1.054	1.047	1.036	0.17	0.17	0.17	0.09	0.11	0.13	0.17
12	1.056	1.057	1.056	1.055	1.056	1.056	1.057	0.11	0.11	0.11	0.12	0.11	0.11	0.11
13	1.130	1.135	1.135	1.135	1.136	1.135	1.137	0.15	0.10	0.08	0.08	0.07	0.09	0.12
14	1.122	1.157	1.144	1.153	1.157	1.153	1.134	0.13	0.06	0.09	0.08	0.05	0.07	0.11
15	1.114	1.119	1.128	1.132	1.136	1.129	1.118	0.25	0.24	0.20	0.18	0.16	0.20	0.25
16	1.025	1.042	1.041	1.042	1.041	1.042	1.026	0.17	0.03	0.04	0.03	0.04	0.03	0.11
17	0.964	1.007	1.028	1.022	1.017	1.019	0.969	0.20	0.11	0.09	0.10	0.11	0.10	0.16
18	1.071	1.096	1.106	1.102	1.109	1.103	1.084	0.17	0.11	0.10	0.11	0.09	0.10	0.14
19	1.088	1.088	1.083	1.085	1.080	1.084	1.100	0.11	0.11	0.11	0.11	0.12	0.11	0.11
20	1.078	1.099	1.095	1.093	1.104	1.098	1.083	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Mean	**1.061**	**1.072**	**1.075**	**1.076**	**1.077**	**1.075**	**1.067**	**0.17**	**0.14**	**0.13**	**0.13**	**0.13**	**0.13**	**0.16**
SD	*0.045*	*0.047*	*0.042*	*0.043*	*0.046*	*0.045*	*0.047*	*0.06*	*0.07*	*0.06*	*0.07*	*0.07*	*0.07*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C15

Dataset 3: Predictive performance statistics for the prediction equations developed using Conscientiousness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.044	1.045	1.057	1.055	1.055	1.053	1.047	0.07	0.07	0.05	0.05	0.05	0.05	0.07
2	1.094	1.096	1.096	1.097	1.097	1.097	1.100	0.14	0.13	0.13	0.12	0.12	0.13	0.14
3	1.100	1.100	1.101	1.097	1.099	1.099	1.104	0.22	0.14	0.14	0.14	0.14	0.14	0.22
4	1.028	1.027	1.026	1.026	1.027	1.027	1.037	0.19	0.19	0.20	0.20	0.19	0.20	0.18
5	1.053	1.058	1.057	1.056	1.057	1.057	1.058	0.17	0.10	0.09	0.10	0.10	0.10	0.14
6	0.950	0.949	0.950	0.950	0.950	0.950	0.953	0.09	0.09	0.09	0.09	0.09	0.09	0.08
7	1.110	1.110	1.108	1.108	1.108	1.108	1.113	0.14	0.14	0.13	0.13	0.13	0.13	0.14
8	1.052	1.049	1.051	1.048	1.056	1.051	1.054	0.19	0.20	0.20	0.21	0.07	0.17	0.19
9	1.066	1.074	1.074	1.074	1.074	1.074	1.071	0.01	-0.04	-0.04	-0.04	-0.04	-0.04	-0.01
10	1.065	1.063	1.060	1.059	1.059	1.060	1.064	0.08	0.09	0.09	0.09	0.09	0.09	0.09
11	1.029	1.028	1.028	1.029	1.028	1.028	1.029	0.03	0.02	0.02	0.02	0.02	0.02	0.03
12	1.050	1.049	1.048	1.047	1.048	1.048	1.054	0.12	0.13	0.13	0.13	0.13	0.13	0.12
13	1.110	1.108	1.106	1.112	1.113	1.110	1.105	0.15	0.15	0.14	0.07	0.04	0.10	0.15
14	1.127	1.125	1.124	1.124	1.125	1.124	1.123	-0.03	-0.02	-0.02	-0.03	-0.03	-0.03	-0.03
15	1.141	1.140	1.138	1.139	1.140	1.139	1.141	0.03	0.04	0.04	0.04	0.04	0.04	0.04
16	1.032	1.031	1.031	1.031	1.031	1.031	1.028	0.10	0.11	0.11	0.11	0.11	0.11	0.10
17	0.960	0.957	0.956	0.956	0.956	0.956	0.970	0.04	0.04	0.04	0.04	0.04	0.04	0.04
18	1.060	1.060	1.060	1.060	1.060	1.060	1.066	0.12	0.12	0.12	0.12	0.12	0.12	0.11
19	1.069	1.066	1.065	1.064	1.064	1.065	1.091	0.13	0.14	0.14	0.14	0.14	0.14	0.13
20	1.123	1.140	1.168	1.140	1.160	1.152	1.133	-0.06	-0.06	-0.05	-0.05	-0.05	-0.05	-0.06
Mean	**1.063**	**1.064**	**1.065**	**1.064**	**1.065**	**1.064**	**1.067**	**0.10**	**0.09**	**0.09**	**0.09**	**0.08**	**0.08**	**0.09**
SD	*0.050*	*0.052*	*0.054*	*0.052*	*0.054*	*0.053*	*0.050*	*0.08*	*0.07*	*0.07*	*0.07*	*0.06*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C16

Dataset 4: Predictive performance statistics for the prediction equations developed using Neuroticism (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.487	0.488	0.488	0.488	0.487	0.487	0.482	0.07	0.07	0.07	0.07	0.07	0.07	0.08
2	0.431	0.435	0.435	0.436	0.436	0.436	0.440	0.30	0.26	0.26	0.25	0.25	0.26	0.28
3	0.513	0.513	0.513	0.513	0.513	0.513	0.514	0.13	0.13	0.13	0.13	0.13	0.13	0.13
4	0.494	0.493	0.493	0.498	0.507	0.498	0.493	0.29	0.29	0.29	0.26	0.18	0.26	0.29
5	0.482	0.485	0.485	0.489	0.487	0.486	0.482	0.17	0.12	0.13	0.08	0.11	0.11	0.16
6	0.487	0.486	0.486	0.486	0.486	0.486	0.485	0.18	0.18	0.18	0.18	0.18	0.18	0.18
7	0.546	0.549	0.548	0.548	0.548	0.548	0.546	-0.05	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
8	0.538	0.541	0.540	0.543	0.548	0.543	0.542	0.26	0.25	0.25	0.15	0.12	0.19	0.26
9	0.599	0.598	0.597	0.598	0.621	0.604	0.598	0.19	0.19	0.18	0.16	0.09	0.16	0.18
10	0.454	0.455	0.455	0.456	0.455	0.455	0.455	0.23	0.23	0.23	0.23	0.23	0.23	0.23
11	0.569	0.569	0.568	0.567	0.568	0.568	0.572	0.13	0.13	0.13	0.11	0.12	0.12	0.13
12	0.454	0.457	0.500	0.491	0.509	0.489	0.460	0.25	0.21	-0.10	-0.10	-0.16	-0.04	0.22
13	0.464	0.465	0.467	0.466	0.466	0.466	0.466	0.26	0.22	0.20	0.20	0.21	0.21	0.24
14	0.489	0.494	0.510	0.535	0.513	0.513	0.485	0.29	0.26	0.16	-0.04	0.13	0.13	0.29
15	0.514	0.514	0.514	0.514	0.513	0.514	0.518	0.22	0.22	0.22	0.22	0.22	0.22	0.22
16	0.450	0.450	0.450	0.450	0.450	0.450	0.452	0.15	0.14	0.14	0.14	0.14	0.14	0.14
17	0.443	0.443	0.443	0.448	0.446	0.445	0.450	0.16	0.16	0.16	0.12	0.13	0.15	0.16
18	0.429	0.429	0.429	0.428	0.427	0.428	0.436	0.13	0.14	0.14	0.14	0.14	0.14	0.14
19	0.426	0.434	0.447	0.476	0.477	0.459	0.433	0.23	0.13	-0.06	-0.20	-0.20	-0.08	0.14
20	0.497	0.499	0.500	0.501	0.500	0.500	0.497	0.15	0.15	0.14	0.14	0.15	0.14	0.15
Mean	**0.488**	**0.490**	**0.493**	**0.497**	**0.498**	**0.494**	**0.490**	**0.19**	**0.17**	**0.14**	**0.11**	**0.11**	**0.13**	**0.18**
SD	*0.048*	*0.047*	*0.045*	*0.045*	*0.048*	*0.046*	*0.046*	*0.09*	*0.08*	*0.11*	*0.12*	*0.12*	*0.10*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C17

Dataset 4: Predictive performance statistics for the prediction equations developed using Extraversion (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.485	0.485	0.490	0.494	0.494	0.491	0.478	0.16	0.17	0.19	0.18	0.18	0.18	0.21
2	0.466	0.478	0.473	0.477	0.479	0.477	0.463	0.04	0.07	0.06	0.06	0.05	0.06	0.06
3	0.520	0.518	0.525	0.533	0.531	0.527	0.520	0.11	0.07	0.05	0.02	0.03	0.04	0.07
4	0.523	0.523	0.507	0.506	0.510	0.511	0.519	0.11	0.11	0.16	0.17	0.18	0.15	0.20
5	0.513	0.505	0.506	0.504	0.505	0.505	0.511	0.08	0.20	0.19	0.20	0.20	0.20	0.13
6	0.515	0.512	0.513	0.513	0.517	0.514	0.509	-0.02	0.07	0.08	0.05	0.05	0.06	0.04
7	0.543	0.540	0.557	0.539	0.541	0.544	0.537	-0.08	0.04	-0.03	0.04	0.02	0.02	-0.03
8	0.555	0.569	0.570	0.566	0.567	0.568	0.560	0.17	0.08	0.09	0.10	0.09	0.09	0.17
9	0.610	0.610	0.598	0.597	0.601	0.601	0.605	0.13	0.13	0.18	0.21	0.18	0.18	0.17
10	0.462	0.464	0.468	0.466	0.470	0.467	0.461	0.14	0.12	0.11	0.11	0.10	0.11	0.13
11	0.575	0.573	0.573	0.570	0.574	0.573	0.578	0.19	0.16	0.17	0.19	0.19	0.18	0.22
12	0.475	0.479	0.479	0.479	0.479	0.479	0.476	0.14	0.17	0.16	0.17	0.17	0.17	0.17
13	0.488	0.475	0.477	0.476	0.474	0.476	0.482	0.04	0.18	0.17	0.18	0.19	0.18	0.13
14	0.521	0.523	0.524	0.523	0.526	0.524	0.526	0.10	0.10	0.09	0.08	0.05	0.08	0.13
15	0.530	0.530	0.522	0.523	0.525	0.525	0.527	0.14	0.15	0.24	0.23	0.21	0.21	0.18
16	0.462	0.462	0.462	0.462	0.462	0.462	0.463	0.15	0.16	0.24	0.23	0.23	0.22	0.22
17	0.453	0.449	0.449	0.448	0.448	0.449	0.449	0.22	0.24	0.24	0.25	0.25	0.25	0.24
18	0.448	0.467	0.467	0.470	0.471	0.469	0.453	0.04	0.04	0.05	0.09	0.11	0.07	0.10
19	0.439	0.453	0.451	0.459	0.460	0.456	0.442	0.10	0.17	0.18	0.15	0.15	0.17	0.14
20	0.516	0.507	0.510	0.513	0.513	0.511	0.512	0.10	0.21	0.20	0.18	0.17	0.19	0.14
Mean	**0.505**	**0.506**	**0.506**	**0.506**	**0.507**	**0.506**	**0.504**	**0.10**	**0.13**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**
SD	*0.045*	*0.043*	*0.043*	*0.041*	*0.041*	*0.042*	*0.045*	*0.07*	*0.06*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C18

Dataset 4: Predictive performance statistics for the prediction equations developed using Openness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.504	0.504	0.515	0.505	0.513	0.509	0.496	-0.26	-0.24	-0.09	-0.11	-0.16	-0.15	-0.24
2	0.464	0.464	0.464	0.463	0.462	0.463	0.462	-0.19	-0.18	-0.18	-0.17	-0.07	-0.15	-0.20
3	0.522	0.524	0.521	0.523	0.521	0.522	0.520	-0.09	-0.08	-0.04	-0.06	-0.03	-0.05	-0.06
4	0.527	0.527	0.521	0.521	0.521	0.522	0.527	0.04	0.04	0.04	0.07	0.07	0.06	0.07
5	0.517	0.517	0.517	0.547	0.523	0.526	0.519	-0.07	-0.07	-0.07	-0.03	-0.03	-0.05	-0.08
6	0.497	0.502	0.500	0.521	0.517	0.510	0.499	0.00	-0.19	-0.16	-0.02	-0.05	-0.10	-0.15
7	0.514	0.514	0.512	0.518	0.517	0.515	0.514	-0.04	-0.04	0.06	0.13	0.10	0.06	-0.04
8	0.562	0.562	0.562	0.557	0.558	0.560	0.565	-0.06	-0.05	-0.03	0.10	0.07	0.02	-0.07
9	0.616	0.617	0.627	0.628	0.631	0.626	0.617	-0.08	-0.08	0.09	0.09	0.08	0.05	-0.08
10	0.475	0.482	0.482	0.480	0.484	0.482	0.470	-0.08	0.00	-0.03	0.02	-0.06	-0.02	-0.08
11	0.582	0.600	0.601	0.600	0.600	0.600	0.584	-0.10	-0.06	-0.05	-0.03	-0.04	-0.04	-0.08
12	0.479	0.475	0.479	0.477	0.481	0.478	0.477	-0.06	0.00	-0.05	0.07	0.04	0.01	-0.05
13	0.479	0.480	0.475	0.475	0.475	0.476	0.479	-0.07	-0.06	0.00	0.00	-0.01	-0.01	0.10
14	0.525	0.525	0.525	0.523	0.523	0.524	0.526	0.06	0.06	0.06	0.17	0.17	0.11	0.02
15	0.542	0.545	0.542	0.541	0.542	0.543	0.548	-0.17	-0.15	-0.07	-0.06	-0.07	-0.09	-0.14
16	0.484	0.481	0.482	0.483	0.483	0.482	0.481	-0.20	-0.08	-0.10	-0.09	-0.10	-0.09	-0.16
17	0.459	0.459	0.489	0.475	0.490	0.478	0.458	0.04	0.03	-0.07	-0.01	-0.03	-0.02	0.02
18	0.449	0.489	0.500	0.487	0.494	0.492	0.473	-0.08	-0.03	-0.08	-0.05	-0.08	-0.06	-0.09
19	0.433	0.433	0.431	0.479	0.478	0.455	0.433	0.00	0.01	0.07	-0.01	0.00	0.02	0.00
20	0.516	0.516	0.514	0.519	0.515	0.516	0.516	0.04	0.04	0.15	0.09	0.10	0.09	0.07
Mean	**0.507**	**0.511**	**0.513**	**0.516**	**0.516**	**0.514**	**0.508**	**-0.07**	**-0.06**	**-0.03**	**0.01**	**0.00**	**-0.02**	**-0.06**
SD	*0.045*	*0.045*	*0.045*	*0.043*	*0.042*	*0.043*	*0.045*	*0.09*	*0.08*	*0.09*	*0.09*	*0.08*	*0.08*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C19

Dataset 4: Predictive performance statistics for the prediction equations developed using Agreeableness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.487	0.489	0.488	0.489	0.489	0.489	0.491	0.08	-0.07	-0.05	-0.07	-0.06	-0.06	0.06
2	0.457	0.453	0.462	0.464	0.461	0.460	0.457	0.17	0.15	-0.03	-0.07	-0.02	0.01	0.16
3	0.528	0.528	0.543	0.566	0.557	0.548	0.532	-0.33	-0.33	-0.17	-0.17	-0.21	-0.22	-0.25
4	0.525	0.530	0.536	0.537	0.538	0.535	0.531	0.16	-0.09	-0.20	-0.20	-0.20	-0.17	-0.16
5	0.517	0.517	0.527	0.532	0.533	0.527	0.521	0.14	0.12	-0.25	-0.25	-0.24	-0.16	-0.16
6	0.516	0.523	0.566	0.564	0.563	0.554	0.529	-0.10	-0.16	-0.32	-0.33	-0.33	-0.29	-0.24
7	0.516	0.516	0.522	0.521	0.521	0.520	0.516	-0.27	-0.28	-0.23	-0.18	-0.20	-0.22	-0.25
8	0.559	0.560	0.560	0.561	0.561	0.561	0.557	0.15	-0.05	-0.04	-0.06	-0.12	-0.07	0.13
9	0.616	0.616	0.615	0.615	0.615	0.615	0.616	0.10	0.10	0.13	0.14	0.13	0.12	0.11
10	0.489	0.488	0.487	0.487	0.487	0.487	0.489	-0.05	-0.05	-0.05	-0.06	-0.06	-0.05	-0.05
11	0.582	0.582	0.584	0.583	0.584	0.583	0.580	-0.28	-0.28	-0.28	-0.28	-0.27	-0.28	-0.28
12	0.476	0.477	0.504	0.501	0.503	0.496	0.485	0.22	-0.22	-0.31	-0.33	-0.32	-0.30	-0.34
13	0.480	0.480	0.481	0.480	0.480	0.480	0.481	0.01	0.02	0.02	0.02	0.02	0.02	0.02
14	0.530	0.529	0.531	0.532	0.531	0.531	0.528	0.08	-0.10	-0.09	-0.09	-0.09	-0.09	0.01
15	0.534	0.533	0.532	0.533	0.532	0.532	0.531	0.10	0.11	0.13	0.08	0.12	0.11	0.10
16	0.471	0.471	0.471	0.471	0.471	0.471	0.471	0.11	0.10	0.06	0.04	0.04	0.06	0.09
17	0.456	0.465	0.461	0.466	0.466	0.465	0.456	0.12	0.01	0.00	0.02	0.02	0.01	0.10
18	0.449	0.446	0.446	0.446	0.446	0.446	0.443	0.11	-0.06	-0.06	-0.07	-0.07	-0.07	0.09
19	0.463	0.460	0.461	0.461	0.460	0.461	0.456	-0.18	-0.16	-0.17	-0.16	-0.16	-0.16	-0.17
20	0.523	0.521	0.521	0.521	0.521	0.521	0.521	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	-0.06
Mean	**0.509**	**0.509**	**0.515**	**0.516**	**0.516**	**0.514**	**0.510**	**0.01**	**-0.06**	**-0.10**	**-0.10**	**-0.11**	**-0.09**	**-0.05**
SD	*0.044*	*0.044*	*0.046*	*0.046*	*0.046*	*0.045*	*0.044*	*0.17*	*0.14*	*0.14*	*0.13*	*0.14*	*0.13*	*0.16*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C20

Dataset 4: Predictive performance statistics for the prediction equations developed using Conscientiousness (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.474	0.475	0.475	0.476	0.476	0.476	0.472	0.14	0.14	0.14	0.14	0.14	0.14	0.14
2	0.424	0.424	0.423	0.423	0.422	0.423	0.429	0.24	0.24	0.24	0.24	0.24	0.24	0.24
3	0.547	0.550	0.549	0.550	0.550	0.550	0.540	0.08	0.08	0.08	0.08	0.08	0.08	0.09
4	0.494	0.494	0.494	0.494	0.494	0.494	0.495	0.31	0.30	0.30	0.30	0.30	0.30	0.31
5	0.479	0.485	0.493	0.495	0.493	0.491	0.484	0.29	0.28	0.25	0.24	0.25	0.25	0.28
6	0.460	0.461	0.466	0.463	0.466	0.464	0.458	0.25	0.25	0.22	0.23	0.22	0.23	0.25
7	0.502	0.503	0.502	0.503	0.502	0.502	0.504	0.22	0.22	0.22	0.22	0.22	0.22	0.22
8	0.534	0.534	0.534	0.534	0.534	0.534	0.534	0.30	0.30	0.30	0.30	0.29	0.30	0.30
9	0.580	0.582	0.582	0.582	0.583	0.583	0.581	0.25	0.25	0.25	0.25	0.25	0.25	0.25
10	0.450	0.451	0.454	0.453	0.454	0.453	0.451	0.19	0.19	0.18	0.18	0.17	0.18	0.19
11	0.537	0.535	0.537	0.538	0.539	0.537	0.541	0.32	0.31	0.29	0.29	0.29	0.29	0.31
12	0.441	0.442	0.443	0.444	0.443	0.443	0.443	0.34	0.33	0.33	0.31	0.32	0.32	0.34
13	0.463	0.466	0.467	0.467	0.467	0.467	0.465	0.23	0.22	0.21	0.21	0.21	0.21	0.23
14	0.501	0.502	0.502	0.502	0.503	0.502	0.511	0.25	0.24	0.24	0.24	0.23	0.24	0.24
15	0.500	0.500	0.500	0.500	0.499	0.500	0.498	0.35	0.35	0.35	0.35	0.35	0.35	0.35
16	0.435	0.434	0.437	0.438	0.438	0.437	0.435	0.31	0.31	0.31	0.30	0.30	0.31	0.31
17	0.452	0.453	0.453	0.454	0.455	0.454	0.452	0.19	0.18	0.18	0.18	0.16	0.18	0.18
18	0.442	0.456	0.456	0.454	0.457	0.456	0.442	0.23	0.19	0.19	0.20	0.20	0.20	0.22
19	0.406	0.409	0.411	0.413	0.414	0.412	0.407	0.38	0.37	0.35	0.35	0.34	0.35	0.37
20	0.484	0.487	0.486	0.487	0.488	0.487	0.488	0.34	0.34	0.34	0.34	0.33	0.34	0.34
Mean	**0.480**	**0.482**	**0.483**	**0.483**	**0.484**	**0.483**	**0.482**	**0.26**	**0.25**	**0.25**	**0.25**	**0.24**	**0.25**	**0.26**
SD	*0.045*	*0.044*	*0.044*	*0.044*	*0.044*	*0.044*	*0.045*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C21

Dataset 5: Predictive performance statistics for the prediction equations developed using Adjustment (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.849	6.841	6.834	6.849	6.844	6.842	6.892	0.01	0.02	0.02	0.01	0.02	0.02	0.01
2	6.268	6.249	6.254	6.252	6.253	6.252	6.268	0.06	0.07	0.07	0.07	0.07	0.07	0.06
3	6.200	6.228	6.232	6.231	6.228	6.230	6.195	0.08	0.04	0.06	0.04	0.05	0.05	0.08
4	6.231	6.245	6.247	6.246	6.246	6.246	6.335	0.09	0.08	0.08	0.08	0.08	0.08	0.09
5	5.892	5.923	5.932	5.930	5.943	5.932	5.889	0.13	0.08	0.07	0.06	0.04	0.06	0.11
6	6.417	6.444	6.438	6.444	6.500	6.456	6.455	0.08	0.04	0.04	0.03	-0.01	0.02	0.07
7	5.753	5.755	5.754	5.754	5.759	5.756	5.752	0.07	0.06	0.06	0.06	0.06	0.06	0.07
8	6.375	6.374	6.367	6.376	6.394	6.378	6.371	-0.01	-0.04	-0.04	-0.04	-0.04	-0.04	-0.03
9	6.324	6.325	6.325	6.325	6.324	6.325	6.344	0.08	0.08	0.08	0.08	0.08	0.08	0.06
10	5.783	5.782	5.783	5.782	5.784	5.783	5.757	0.15	0.15	0.15	0.15	0.15	0.15	0.15
11	5.771	5.766	5.764	5.787	5.772	5.772	5.813	0.21	0.20	0.17	0.06	0.14	0.15	0.22
12	5.855	5.872	5.867	5.867	5.869	5.869	5.881	0.01	-0.02	-0.02	-0.01	-0.02	-0.02	0.00
13	5.613	5.643	5.645	5.646	5.645	5.645	5.635	0.14	0.08	0.08	0.08	0.08	0.08	0.11
14	5.808	5.802	5.798	5.795	5.805	5.800	5.757	0.13	0.12	0.12	0.13	0.11	0.12	0.14
15	6.192	6.187	6.184	6.192	6.195	6.190	6.188	0.11	0.12	0.12	0.12	0.11	0.12	0.12
16	6.095	6.094	6.092	6.093	6.092	6.093	6.120	0.02	0.02	0.03	0.03	0.02	0.02	0.03
17	6.505	6.497	6.494	6.499	6.507	6.499	6.488	0.16	0.16	0.16	0.16	0.13	0.15	0.17
18	6.493	6.491	6.486	6.487	6.492	6.489	6.432	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
19	6.139	6.136	6.138	6.140	6.138	6.138	6.147	0.09	0.09	0.09	0.09	0.09	0.09	0.08
20	6.545	6.584	6.595	6.588	6.591	6.590	6.549	0.01	0.00	0.00	0.00	0.00	0.00	0.01
Mean	**6.155**	**6.162**	**6.162**	**6.164**	**6.169**	**6.164**	**6.163**	**0.08**	**0.06**	**0.06**	**0.06**	**0.05**	**0.06**	**0.07**
SD	*0.329*	*0.328*	*0.327*	*0.328*	*0.331*	*0.328*	*0.334*	*0.07*	*0.07*	*0.07*	*0.06*	*0.06*	*0.06*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C22

Dataset 5: Predictive performance statistics for the prediction equations developed using Ambition (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.795	6.805	6.801	6.803	6.802	6.803	6.794	0.10	0.10	0.10	0.10	0.10	0.10	0.10
2	6.196	6.227	6.219	6.227	6.220	6.223	6.209	0.10	0.06	0.07	0.05	0.05	0.06	0.08
3	6.174	6.172	6.173	6.174	6.173	6.173	6.197	0.11	0.11	0.11	0.11	0.11	0.11	0.11
4	6.301	6.295	6.288	6.284	6.291	6.290	6.271	0.04	0.04	0.04	0.04	0.04	0.04	0.04
5	5.868	5.865	5.868	5.862	5.870	5.866	5.871	0.07	0.08	0.07	0.08	0.07	0.08	0.07
6	6.362	6.358	6.359	6.359	6.359	6.359	6.353	0.13	0.13	0.13	0.12	0.10	0.12	0.13
7	5.833	5.841	5.828	5.840	5.827	5.834	5.789	0.03	0.04	0.04	0.04	0.03	0.03	0.03
8	6.391	6.393	6.395	6.432	6.430	6.412	6.370	-0.01	-0.01	-0.01	-0.03	-0.02	-0.02	0.00
9	6.288	6.280	6.284	6.287	6.280	6.283	6.286	0.24	0.25	0.24	0.24	0.25	0.24	0.25
10	5.757	5.745	5.747	5.744	5.764	5.750	5.752	0.07	0.07	0.07	0.07	0.07	0.07	0.07
11	5.750	5.812	5.813	5.814	5.809	5.812	5.783	0.15	0.08	0.08	0.08	0.08	0.08	0.13
12	5.990	6.038	6.031	6.023	6.032	6.031	5.948	-0.03	-0.04	-0.04	-0.05	-0.04	-0.04	-0.03
13	5.674	5.677	5.674	5.674	5.672	5.674	5.692	0.04	0.04	0.04	0.04	0.04	0.04	0.04
14	5.812	5.823	5.823	5.823	5.821	5.822	5.964	0.08	0.08	0.08	0.08	0.08	0.08	0.09
15	6.161	6.162	6.159	6.162	6.160	6.161	6.177	0.15	0.15	0.15	0.15	0.15	0.15	0.14
16	6.058	6.094	6.098	6.101	6.115	6.102	6.078	0.06	0.04	0.04	0.04	0.04	0.04	0.05
17	6.502	6.503	6.503	6.513	6.514	6.508	6.554	0.11	0.11	0.11	0.10	0.09	0.10	0.11
18	6.271	6.270	6.272	6.268	6.274	6.271	6.227	0.07	0.07	0.06	0.06	0.06	0.06	0.07
19	6.145	6.165	6.162	6.164	6.172	6.166	6.169	0.05	0.03	0.04	0.04	0.03	0.03	0.04
20	6.470	6.494	6.500	6.496	6.496	6.496	6.482	0.10	0.07	0.06	0.06	0.06	0.06	0.07
Mean	**6.140**	**6.151**	**6.150**	**6.152**	**6.154**	**6.152**	**6.148**	**0.08**	**0.08**	**0.07**	**0.07**	**0.07**	**0.07**	**0.08**
SD	0.297	*0.293*	*0.294*	*0.296*	*0.295*	*0.295*	*0.292*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C23

Dataset 5: Predictive performance statistics for the prediction equations developed using Sociability (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.870	6.868	6.869	6.868	6.868	6.868	6.870	-0.03	-0.03	-0.04	-0.03	-0.03	-0.03	-0.04
2	6.246	6.228	6.227	6.226	6.227	6.227	6.237	-0.01	-0.03	-0.03	-0.04	-0.03	-0.03	-0.03
3	6.202	6.261	6.262	6.259	6.260	6.260	6.236	-0.01	-0.14	-0.14	-0.14	-0.14	-0.14	-0.12
4	6.231	6.282	6.282	6.278	6.294	6.284	6.226	-0.03	-0.12	-0.12	-0.13	-0.10	-0.11	-0.10
5	5.967	5.965	5.967	5.965	5.964	5.965	5.986	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
6	6.492	6.504	6.539	6.506	6.529	6.520	6.517	-0.14	-0.16	-0.15	-0.16	-0.14	-0.16	-0.15
7	5.737	6.013	6.036	6.028	6.010	6.022	5.861	-0.08	-0.28	-0.28	-0.28	-0.28	-0.28	-0.28
8	6.352	6.352	6.343	6.347	6.366	6.352	6.414	-0.13	-0.13	-0.12	-0.12	-0.10	-0.12	-0.14
9	6.433	6.446	6.447	6.459	6.442	6.448	6.413	-0.12	-0.13	-0.13	-0.14	-0.12	-0.13	-0.13
10	5.828	5.832	5.829	5.832	5.830	5.831	5.860	-0.03	-0.07	-0.07	-0.06	-0.05	-0.06	-0.04
11	5.797	5.794	5.796	5.794	5.794	5.794	5.803	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
12	5.807	5.803	5.806	5.808	5.805	5.805	5.811	-0.09	-0.08	-0.09	-0.07	-0.06	-0.08	-0.09
13	5.658	5.667	5.666	5.681	5.668	5.670	5.680	-0.03	-0.05	-0.05	-0.06	-0.05	-0.05	-0.03
14	5.828	5.845	5.821	5.862	5.815	5.836	5.891	-0.09	-0.10	-0.09	-0.12	-0.08	-0.10	-0.10
15	6.244	6.294	6.301	6.294	6.338	6.307	6.263	-0.04	-0.10	-0.11	-0.14	-0.15	-0.13	-0.06
16	6.058	6.056	6.056	6.056	6.060	6.057	6.103	-0.04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.04
17	6.533	6.587	6.592	6.597	6.596	6.593	6.554	-0.04	-0.13	-0.13	-0.13	-0.14	-0.13	-0.12
18	6.292	6.280	6.287	6.293	6.301	6.290	6.267	-0.08	-0.05	-0.07	-0.03	-0.07	-0.06	-0.09
19	6.132	6.130	6.132	6.133	6.130	6.132	6.123	-0.03	0.00	-0.03	-0.05	-0.04	-0.03	-0.01
20	6.588	6.628	6.631	6.660	6.634	6.638	6.618	-0.06	-0.07	-0.07	-0.09	-0.09	-0.08	-0.07
Mean	**6.165**	**6.192**	**6.195**	**6.197**	**6.197**	**6.195**	**6.187**	**-0.06**	**-0.09**	**-0.09**	**-0.10**	**-0.09**	**-0.09**	**-0.09**
SD	*0.328*	*0.322*	*0.325*	*0.322*	*0.327*	*0.324*	*0.317*	*0.04*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C24

Dataset 5: Predictive performance statistics for the prediction equations developed using Intellectance (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.857	6.827	6.666	6.790	6.803	6.771	6.799	-0.01	0.15	0.16	0.12	0.12	0.13	0.07
2	6.283	6.202	6.168	6.116	6.140	6.156	6.254	-0.09	0.10	0.11	0.14	0.14	0.12	0.01
3	6.196	6.116	6.124	6.063	6.125	6.107	6.192	-0.01	0.16	0.19	0.20	0.17	0.18	0.09
4	6.229	6.087	6.103	6.097	6.113	6.100	6.173	-0.02	0.16	0.14	0.14	0.13	0.14	0.09
5	5.969	6.109	6.112	6.115	6.113	6.112	6.108	-0.12	-0.07	-0.06	-0.08	-0.04	-0.06	-0.13
6	6.442	6.401	6.427	6.406	6.409	6.411	6.422	-0.08	0.05	0.04	0.02	0.05	0.04	0.02
7	5.744	5.712	5.691	5.714	5.689	5.702	5.732	-0.07	0.07	0.11	0.06	0.10	0.09	0.03
8	6.348	6.338	6.359	6.309	6.311	6.329	6.371	-0.10	0.01	0.00	0.03	0.03	0.02	-0.06
9	6.356	6.300	6.276	6.325	6.297	6.300	6.302	-0.04	0.08	0.08	0.08	0.07	0.07	0.04
10	5.824	5.787	5.773	5.778	5.773	5.778	5.804	-0.01	0.08	0.08	0.08	0.08	0.08	0.05
11	5.790	5.729	5.744	5.745	5.717	5.734	5.773	-0.02	0.11	0.10	0.10	0.14	0.11	0.04
12	5.798	5.764	5.777	5.787	5.748	5.769	5.805	-0.04	0.05	0.04	0.04	0.09	0.05	-0.06
13	5.638	5.591	5.605	5.573	5.582	5.588	5.606	-0.02	0.12	0.11	0.14	0.14	0.13	0.06
14	5.911	5.982	6.000	5.984	5.991	5.989	5.809	-0.16	-0.01	0.00	0.01	-0.01	0.00	-0.07
15	6.236	6.073	6.065	6.131	6.073	6.086	6.214	-0.07	0.18	0.19	0.16	0.18	0.18	0.19
16	6.061	5.990	5.964	6.001	5.960	5.979	6.020	-0.01	0.13	0.15	0.15	0.16	0.15	0.08
17	6.538	6.449	6.469	6.445	6.442	6.451	6.485	-0.06	0.10	0.06	0.07	0.11	0.09	0.03
18	6.291	6.215	6.216	6.218	6.142	6.198	6.194	-0.02	0.14	0.17	0.15	0.19	0.16	0.13
19	6.139	6.057	6.015	6.008	6.006	6.021	6.084	-0.04	0.11	0.14	0.14	0.14	0.13	0.14
20	6.558	6.525	6.524	6.525	6.507	6.520	6.516	-0.04	0.11	0.11	0.12	0.11	0.11	0.07
Mean	**6.160**	**6.113**	**6.104**	**6.106**	**6.097**	**6.105**	**6.133**	**-0.05**	**0.09**	**0.10**	**0.09**	**0.10**	**0.10**	**0.04**
SD	*0.318*	*0.310*	*0.294*	*0.305*	*0.309*	*0.304*	*0.309*	*0.04*	*0.06*	*0.07*	*0.07*	*0.06*	*0.06*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C21

Dataset 5: Predictive performance statistics for the prediction equations developed using Likeability (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.865	6.862	6.858	6.853	6.857	6.858	6.846	0.04	0.05	0.06	0.06	0.06	0.06	0.05
2	6.368	6.370	6.361	6.420	6.387	6.385	6.359	0.02	0.00	0.00	0.01	0.01	0.00	0.02
3	6.157	6.137	6.136	6.136	6.137	6.136	6.192	0.09	0.11	0.11	0.11	0.11	0.11	0.11
4	6.178	6.200	6.201	6.173	6.202	6.194	6.186	0.12	0.10	0.10	0.12	0.11	0.11	0.11
5	5.851	5.877	5.984	5.976	5.903	5.935	5.873	0.14	0.12	0.09	0.08	0.12	0.10	0.13
6	6.352	6.347	6.334	6.339	6.340	6.340	6.348	0.17	0.13	0.15	0.14	0.14	0.14	0.16
7	5.739	5.741	5.744	5.742	5.747	5.744	5.787	0.12	0.12	0.11	0.11	0.11	0.11	-0.21
8	6.272	6.314	6.312	6.301	6.312	6.310	6.292	0.08	0.05	0.05	0.06	0.05	0.06	0.07
9	6.341	6.326	6.329	6.326	6.331	6.328	6.333	0.04	0.02	0.02	0.02	0.02	0.02	0.03
10	5.721	5.715	5.708	5.716	5.747	5.721	5.733	0.14	0.15	0.17	0.16	0.12	0.15	0.16
11	5.766	5.756	5.752	5.753	5.747	5.752	5.752	0.21	0.21	0.22	0.22	0.22	0.22	-0.17
12	5.860	5.840	5.835	5.839	5.836	5.837	5.830	0.03	0.05	0.05	0.05	0.05	0.05	0.09
13	5.645	5.646	5.647	5.646	5.649	5.647	5.612	0.10	0.10	0.10	0.10	0.10	0.10	0.10
14	5.872	5.908	5.887	5.921	5.904	5.905	5.878	0.07	0.02	0.02	0.01	0.04	0.03	0.03
15	6.165	6.268	6.207	6.272	6.228	6.244	6.213	0.19	-0.03	0.00	-0.03	0.02	-0.01	0.00
16	6.141	6.143	6.139	6.139	6.143	6.141	6.148	0.03	0.03	0.04	0.03	0.03	0.03	0.03
17	6.501	6.496	6.496	6.493	6.495	6.495	6.508	0.08	0.10	0.10	0.10	0.10	0.10	0.09
18	6.523	6.633	6.609	6.611	6.627	6.620	6.532	-0.04	-0.07	-0.07	-0.07	-0.08	-0.07	-0.07
19	6.124	6.122	6.121	6.124	6.126	6.123	6.126	0.09	0.10	0.10	0.10	0.10	0.10	0.08
20	6.472	6.524	6.522	6.549	6.509	6.526	6.471	0.17	-0.02	-0.01	-0.02	0.00	-0.01	0.04
Mean	**6.146**	**6.161**	**6.159**	**6.166**	**6.161**	**6.162**	**6.151**	**0.09**	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**	**0.04**
SD	*0.326*	*0.337*	*0.331*	*0.333*	*0.332*	*0.333*	*0.325*	*0.06*	*0.07*	*0.07*	*0.07*	*0.06*	*0.07*	*0.10*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C26

Dataset 5: Predictive performance statistics for the prediction equations developed using Prudence (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.683	6.757	6.768	6.778	6.783	6.771	6.719	0.17	0.09	0.08	0.09	0.08	0.08	0.13
2	6.185	6.148	6.151	6.151	6.143	6.148	6.174	0.09	0.13	0.13	0.13	0.13	0.13	0.11
3	6.146	6.146	6.146	6.147	6.146	6.146	6.146	0.13	0.15	0.14	0.15	0.15	0.15	0.14
4	6.120	6.197	6.122	6.177	6.151	6.162	6.090	0.16	0.18	0.19	0.18	0.18	0.18	0.18
5	5.944	5.900	5.904	5.906	5.903	5.903	5.914	0.10	0.15	0.15	0.15	0.15	0.15	0.13
6	6.264	6.279	6.279	6.278	6.280	6.279	6.281	0.15	0.16	0.16	0.16	0.16	0.16	0.16
7	5.803	5.857	5.857	5.873	5.889	5.869	5.820	0.10	0.05	0.06	0.05	0.04	0.05	0.07
8	6.159	6.137	6.144	6.136	6.138	6.139	6.327	0.18	0.21	0.21	0.21	0.21	0.21	0.19
9	6.354	6.344	6.346	6.347	6.349	6.346	6.398	0.10	0.12	0.12	0.13	0.10	0.12	0.11
10	5.676	5.698	5.699	5.696	5.697	5.698	5.672	0.23	0.18	0.18	0.18	0.18	0.18	0.21
11	5.738	5.707	5.709	5.711	5.708	5.709	5.739	0.14	0.18	0.18	0.18	0.18	0.18	0.16
12	5.695	5.708	5.699	5.696	5.697	5.700	5.678	0.19	0.19	0.19	0.20	0.20	0.19	0.21
13	5.591	5.660	5.654	5.654	5.659	5.657	5.602	0.09	0.11	0.11	0.11	0.11	0.11	0.10
14	5.743	5.732	5.726	5.735	5.736	5.732	5.698	0.18	0.19	0.19	0.19	0.19	0.19	0.19
15	6.187	6.199	6.198	6.198	6.198	6.199	6.190	0.07	0.10	0.10	0.11	0.10	0.10	0.09
16	5.911	5.912	5.916	5.918	5.914	5.915	5.902	0.21	0.20	0.20	0.20	0.20	0.20	0.21
17	6.424	6.448	6.449	6.481	6.444	6.455	6.451	0.22	0.23	0.23	0.09	0.22	0.19	0.24
18	6.413	6.477	6.482	6.474	6.480	6.478	6.468	0.04	0.04	0.04	0.04	0.04	0.04	0.04
19	6.032	6.042	6.041	6.039	6.041	6.041	6.031	0.17	0.16	0.16	0.16	0.16	0.16	0.17
20	6.549	6.523	6.512	6.530	6.526	6.523	6.559	0.08	0.11	0.12	0.10	0.10	0.11	0.09
Mean	**6.081**	**6.094**	**6.090**	**6.096**	**6.094**	**6.094**	**6.093**	**0.14**	**0.15**	**0.15**	**0.14**	**0.14**	**0.14**	**0.15**
SD	*0.313*	*0.318*	*0.319*	*0.323*	*0.320*	*0.320*	*0.332*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C27

Dataset 6: Predictive performance statistics for the prediction equations developed using Emotional Orientation (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.610	0.609	0.615	0.612	0.616	0.613	0.611	-0.13	-0.13	-0.07	-0.06	-0.07	-0.08	-0.11
2	0.638	0.637	0.637	0.636	0.639	0.637	0.636	-0.21	-0.20	-0.20	-0.18	-0.21	-0.20	-0.21
3	0.586	0.586	0.611	0.612	0.612	0.605	0.595	0.13	0.13	0.03	0.02	0.01	0.05	0.13
4	0.611	0.613	0.613	0.613	0.613	0.613	0.611	-0.26	-0.25	-0.25	-0.24	-0.24	-0.24	-0.26
5	0.583	0.582	0.582	0.581	0.580	0.581	0.597	0.02	0.06	0.04	0.12	0.14	0.09	0.00
6	0.628	0.740	0.713	0.719	0.713	0.721	0.613	-0.34	-0.25	-0.21	-0.23	-0.23	-0.23	-0.20
7	0.594	0.593	0.596	0.598	0.597	0.596	0.601	0.18	0.18	0.08	0.06	0.07	0.10	0.06
8	0.586	0.576	0.578	0.584	0.583	0.580	0.569	-0.36	-0.36	-0.36	-0.05	-0.07	-0.21	0.28
9	0.581	0.644	0.650	0.653	0.649	0.649	0.597	0.10	0.02	-0.03	-0.07	-0.04	-0.03	0.06
10	0.596	0.597	0.589	0.633	0.596	0.604	0.594	0.15	0.15	0.15	0.06	0.07	0.11	0.17
11	0.602	0.602	0.602	0.598	0.599	0.600	0.603	-0.03	-0.02	-0.03	0.05	0.04	0.01	-0.03
12	0.594	0.608	0.587	0.587	0.586	0.592	0.591	-0.03	0.07	0.06	0.06	0.06	0.06	0.02
13	0.609	0.609	0.609	0.605	0.602	0.606	0.619	0.10	0.10	0.10	0.16	0.17	0.13	0.04
14	0.657	0.655	0.648	0.647	0.646	0.649	0.652	-0.08	-0.07	-0.03	-0.02	-0.02	-0.04	-0.08
15	0.563	0.557	0.556	0.559	0.559	0.558	0.570	-0.03	0.02	0.06	0.06	0.06	0.05	-0.01
16	0.607	0.614	0.613	0.619	0.616	0.616	0.611	0.14	0.14	-0.05	-0.05	-0.05	-0.01	0.01
17	0.624	0.624	0.616	0.616	0.616	0.618	0.618	0.22	0.22	0.18	0.16	0.16	0.18	0.25
18	0.628	0.629	0.628	0.628	0.647	0.633	0.634	-0.25	-0.21	-0.25	0.01	-0.10	-0.14	-0.33
19	0.566	0.566	0.560	0.560	0.560	0.561	0.576	0.13	0.13	0.14	0.14	0.14	0.14	0.14
20	0.609	0.610	0.610	0.599	0.597	0.604	0.604	-0.27	-0.26	0.27	0.23	0.08	0.08	-0.10
Mean	**0.604**	**0.613**	**0.611**	**0.613**	**0.611**	**0.612**	**0.605**	**-0.04**	**-0.03**	**-0.02**	**0.01**	**0.00**	**-0.01**	**-0.01**
SD	*0.024*	*0.039*	*0.035*	*0.036*	*0.035*	*0.036*	*0.021*	*0.19*	*0.18*	*0.16*	*0.13*	*0.12*	*0.13*	*0.16*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C28

Dataset 6: Predictive performance statistics for the prediction equations developed using Social Orientation (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.609	0.609	0.608	0.613	0.613	0.611	0.610	-0.13	-0.13	-0.10	-0.15	-0.15	-0.13	-0.13
2	0.626	0.644	0.649	0.652	0.653	0.649	0.629	-0.04	-0.12	-0.17	-0.15	-0.19	-0.16	-0.07
3	0.599	0.599	0.599	0.607	0.607	0.603	0.596	0.21	0.21	0.19	-0.17	-0.17	0.01	0.22
4	0.601	0.601	0.601	0.602	0.603	0.602	0.609	-0.06	-0.05	-0.06	-0.05	-0.05	-0.05	-0.06
5	0.603	0.603	0.603	0.603	0.603	0.603	0.594	-0.11	-0.10	-0.11	-0.10	-0.10	-0.10	-0.11
6	0.599	0.601	0.600	0.603	0.607	0.603	0.602	0.17	-0.17	0.16	-0.02	-0.10	-0.03	-0.12
7	0.601	0.602	0.602	0.618	0.617	0.610	0.606	0.06	0.05	0.05	-0.14	-0.12	-0.04	-0.02
8	0.578	0.576	0.578	0.580	0.580	0.579	0.583	-0.30	-0.30	-0.29	-0.26	-0.19	-0.26	-0.31
9	0.585	0.587	0.586	0.590	0.591	0.589	0.577	0.13	0.13	0.11	-0.02	-0.11	0.03	0.04
10	0.618	0.618	0.618	0.618	0.618	0.618	0.606	-0.03	-0.04	-0.05	-0.04	-0.03	-0.04	-0.06
11	0.620	0.623	0.634	0.633	0.632	0.631	0.628	-0.20	-0.21	-0.23	-0.22	-0.22	-0.22	-0.20
12	0.619	0.626	0.623	0.626	0.640	0.629	0.621	-0.14	-0.11	-0.12	-0.11	-0.09	-0.11	-0.15
13	0.624	0.626	0.628	0.630	0.630	0.629	0.623	0.17	0.17	0.05	0.02	0.01	0.06	0.17
14	0.661	0.661	0.661	0.661	0.661	0.661	0.646	-0.15	-0.16	-0.12	-0.13	-0.03	-0.11	-0.17
15	0.563	0.567	0.564	0.573	0.574	0.569	0.566	-0.05	-0.07	-0.06	-0.09	-0.09	-0.08	-0.06
16	0.621	0.621	0.618	0.619	0.620	0.620	0.638	-0.26	-0.29	-0.27	-0.19	-0.30	-0.26	-0.29
17	0.625	0.626	0.625	0.624	0.628	0.626	0.623	0.20	0.20	0.20	0.21	0.00	0.15	0.01
18	0.622	0.622	0.622	0.622	0.622	0.622	0.622	0.15	0.15	0.16	0.15	0.15	0.15	0.12
19	0.597	0.596	0.596	0.597	0.597	0.597	0.579	-0.30	-0.29	-0.29	-0.30	-0.13	-0.25	-0.29
20	0.607	0.608	0.607	0.607	0.609	0.608	0.607	0.16	0.16	0.16	0.15	0.11	0.14	0.15
Mean	**0.609**	**0.611**	**0.611**	**0.614**	**0.615**	**0.613**	**0.608**	**-0.03**	**-0.05**	**-0.04**	**-0.08**	**-0.09**	**-0.07**	**-0.07**
SD	*0.021*	*0.022*	*0.023*	*0.022*	*0.022*	*0.022*	*0.021*	*0.17*	*0.17*	*0.17*	*0.14*	*0.11*	*0.13*	*0.15*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C29

Dataset 6: Predictive performance statistics for the prediction equations developed using Cognitive Orientation (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.591	0.592	0.591	0.589	0.589	0.590	0.587	0.16	0.15	0.16	0.17	0.13	0.15	0.16
2	0.607	0.607	0.607	0.607	0.606	0.607	0.608	0.15	0.15	0.15	0.15	0.25	0.17	0.15
3	0.587	0.586	0.597	0.599	0.597	0.595	0.590	0.25	0.12	0.02	-0.06	-0.02	0.02	0.11
4	0.589	0.589	0.590	0.579	0.582	0.585	0.589	0.07	0.08	0.07	0.18	0.17	0.12	0.07
5	0.578	0.620	0.624	0.607	0.605	0.614	0.575	0.04	-0.02	-0.03	0.00	-0.01	-0.01	0.00
6	0.598	0.598	0.598	0.583	0.583	0.591	0.600	0.05	0.05	0.05	0.13	0.12	0.09	0.05
7	0.592	0.591	0.592	0.587	0.590	0.590	0.596	0.27	0.26	0.25	0.07	0.16	0.19	0.28
8	0.566	0.589	0.614	0.614	0.616	0.608	0.586	0.20	-0.11	-0.26	-0.28	-0.29	-0.23	-0.13
9	0.565	0.575	0.567	0.593	0.600	0.584	0.575	0.07	0.04	0.07	0.03	0.04	0.05	0.08
10	0.606	0.607	0.608	0.623	0.610	0.612	0.603	-0.01	-0.01	-0.01	0.03	0.02	0.01	-0.01
11	0.585	0.585	0.586	0.624	0.626	0.605	0.585	0.15	0.15	0.14	-0.19	-0.18	-0.02	0.14
12	0.609	0.610	0.610	0.600	0.599	0.605	0.606	-0.08	-0.08	-0.08	-0.07	-0.07	-0.08	-0.08
13	0.618	0.623	0.620	0.615	0.618	0.619	0.621	0.10	0.12	0.10	0.10	0.10	0.11	0.10
14	0.657	0.656	0.670	0.669	0.669	0.666	0.665	0.27	0.26	-0.25	-0.24	-0.21	-0.11	0.20
15	0.548	0.567	0.572	0.556	0.556	0.563	0.558	0.14	-0.04	-0.04	0.07	0.05	0.01	0.11
16	0.600	0.600	0.603	0.607	0.673	0.621	0.595	0.18	0.18	0.20	0.19	0.21	0.19	0.15
17	0.617	0.616	0.643	0.644	0.645	0.637	0.622	0.26	0.26	-0.26	-0.24	-0.27	-0.13	0.04
18	0.623	0.622	0.622	0.624	0.614	0.621	0.625	0.05	0.06	0.06	0.10	0.12	0.09	0.06
19	0.560	0.588	0.591	0.591	0.593	0.591	0.563	0.06	-0.20	-0.20	-0.10	-0.10	-0.15	0.01
20	0.616	0.618	0.613	0.625	0.609	0.616	0.620	-0.42	-0.41	-0.41	-0.35	-0.02	-0.30	-0.37
Mean	**0.596**	**0.602**	**0.606**	**0.607**	**0.609**	**0.606**	**0.598**	**0.10**	**0.05**	**-0.01**	**-0.01**	**0.01**	**0.01**	**0.06**
SD	*0.025*	*0.020*	*0.024*	*0.025*	*0.028*	*0.022*	*0.024*	*0.16*	*0.17*	*0.18*	*0.17*	*0.16*	*0.14*	*0.14*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C30

Dataset 6: Predictive performance statistics for the prediction equations developed using Interpersonal Orientation (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.582	0.582	0.586	0.639	0.645	0.613	0.574	0.13	0.14	0.16	0.01	-0.02	0.07	0.15
2	0.614	0.614	0.613	0.612	0.614	0.613	0.614	0.15	0.15	0.15	0.13	0.12	0.14	0.14
3	0.590	0.590	0.590	0.590	0.587	0.589	0.593	0.21	0.21	0.21	0.21	0.22	0.21	0.23
4	0.594	0.593	0.593	0.587	0.587	0.590	0.584	0.18	0.19	0.19	0.18	0.16	0.18	0.20
5	0.588	0.587	0.591	0.595	0.593	0.591	0.596	-0.05	-0.05	-0.02	-0.01	0.01	-0.02	-0.03
6	0.590	0.587	0.751	0.747	0.760	0.711	0.605	0.26	0.27	-0.06	-0.06	-0.06	0.02	0.02
7	0.606	0.801	0.707	0.717	0.784	0.752	0.633	-0.36	-0.12	-0.07	-0.07	-0.09	-0.09	-0.13
8	0.579	0.580	0.579	0.574	0.587	0.580	0.569	-0.46	-0.46	-0.46	0.13	-0.14	-0.23	0.35
9	0.582	0.618	0.612	0.559	0.609	0.600	0.597	-0.03	-0.04	-0.07	-0.05	-0.09	-0.06	-0.05
10	0.588	0.587	0.588	0.590	0.592	0.589	0.593	-0.02	0.00	-0.01	0.02	0.02	0.01	0.01
11	0.580	0.581	0.581	0.627	0.640	0.607	0.589	0.09	0.09	0.09	0.02	0.00	0.05	0.11
12	0.597	0.641	0.642	0.641	0.646	0.643	0.602	-0.12	-0.15	-0.15	-0.13	-0.12	-0.14	-0.11
13	0.618	0.617	0.619	0.586	0.636	0.615	0.623	0.23	0.22	0.22	0.33	0.04	0.20	0.21
14	0.660	0.928	0.848	0.854	0.779	0.852	0.673	0.17	-0.15	-0.14	-0.14	-0.14	-0.14	-0.09
15	0.539	0.541	0.540	0.545	0.545	0.543	0.556	0.21	0.17	0.18	0.09	0.09	0.13	0.24
16	0.596	0.596	0.597	0.743	0.667	0.651	0.595	0.19	0.19	0.19	-0.08	-0.04	0.06	0.20
17	0.610	0.609	0.611	0.801	0.620	0.660	0.607	0.08	0.07	0.09	0.07	0.11	0.09	0.06
18	0.615	0.615	0.621	0.631	0.641	0.627	0.617	0.17	0.17	0.05	0.00	-0.02	0.05	0.10
19	0.570	0.568	0.564	0.564	0.565	0.565	0.575	-0.02	-0.01	0.01	0.01	0.01	0.01	0.00
20	0.621	0.898	0.775	0.746	0.621	0.760	0.667	-0.40	-0.14	-0.11	-0.11	0.04	-0.08	-0.14
Mean	**0.596**	**0.637**	**0.631**	**0.647**	**0.636**	**0.638**	**0.603**	**0.03**	**0.04**	**0.02**	**0.03**	**0.01**	**0.02**	**0.07**
SD	*0.024*	*0.107*	*0.079*	*0.089*	*0.067*	*0.077*	*0.030*	*0.22*	*0.18*	*0.17*	*0.12*	*0.10*	*0.12*	*0.14*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C31

Dataset 6: Predictive performance statistics for the prediction equations developed using Task Orientation (Studies 1 and 2).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.610	0.611	0.613	0.657	0.651	0.633	0.604	-0.10	-0.10	-0.03	0.02	0.04	-0.02	0.00
2	0.625	0.625	0.624	0.614	0.619	0.621	0.624	-0.01	-0.01	0.04	0.01	-0.09	-0.01	0.02
3	0.604	0.605	0.610	0.615	0.613	0.611	0.608	-0.13	-0.14	-0.15	-0.17	-0.17	-0.16	-0.18
4	0.616	0.642	0.620	0.619	0.621	0.625	0.619	-0.19	-0.10	-0.06	-0.04	-0.04	-0.06	-0.17
5	0.590	0.590	0.594	0.593	0.594	0.593	0.589	-0.11	-0.12	-0.19	-0.17	-0.19	-0.17	-0.21
6	0.612	0.615	0.628	0.628	0.629	0.625	0.620	-0.14	-0.15	-0.25	-0.13	-0.13	-0.16	-0.22
7	0.599	0.599	0.565	0.577	0.582	0.581	0.604	-0.06	-0.08	-0.07	-0.02	-0.04	-0.05	-0.20
8	0.583	0.584	0.583	0.567	0.596	0.583	0.582	-0.17	-0.18	-0.17	0.03	-0.14	-0.11	-0.20
9	0.600	0.601	0.599	0.599	0.601	0.600	0.608	-0.18	-0.18	-0.18	-0.17	-0.19	-0.18	-0.18
10	0.603	0.604	0.604	0.615	0.617	0.610	0.600	-0.12	-0.12	-0.13	-0.28	-0.29	-0.20	-0.26
11	0.603	0.602	0.602	0.599	0.598	0.601	0.605	-0.02	-0.02	0.03	0.19	0.19	0.10	-0.10
12	0.608	0.609	0.609	0.610	0.612	0.610	0.612	-0.19	-0.19	-0.19	-0.21	-0.23	-0.21	-0.21
13	0.628	0.628	0.628	0.619	0.611	0.621	0.628	-0.06	-0.06	-0.05	0.03	0.07	0.00	-0.06
14	0.664	0.664	0.662	0.667	0.655	0.662	0.683	0.00	0.00	0.00	0.02	0.10	0.03	0.00
15	0.560	0.560	0.560	0.560	0.558	0.559	0.557	-0.12	-0.12	-0.12	-0.13	-0.08	-0.11	-0.15
16	0.619	0.619	0.619	0.617	0.619	0.619	0.628	-0.06	-0.07	-0.07	0.00	0.02	-0.03	-0.08
17	0.628	0.627	0.613	0.594	0.588	0.606	0.629	-0.13	-0.13	-0.09	0.00	-0.04	-0.06	-0.13
18	0.653	0.652	0.652	0.646	0.646	0.649	0.653	-0.27	-0.24	-0.24	-0.20	-0.19	-0.22	-0.26
19	0.581	0.581	0.581	0.579	0.579	0.580	0.579	-0.04	-0.05	-0.04	0.07	0.09	0.02	-0.06
20	0.672	0.667	0.655	0.667	0.657	0.661	0.661	-0.43	-0.38	-0.35	-0.37	-0.36	-0.37	-0.43
Mean	**0.613**	**0.614**	**0.611**	**0.612**	**0.612**	**0.612**	**0.615**	**-0.13**	**-0.12**	**-0.12**	**-0.08**	**-0.08**	**-0.10**	**-0.15**
SD	*0.027*	*0.027*	*0.027*	*0.030*	*0.027*	*0.027*	*0.029*	*0.10*	*0.09*	*0.10*	*0.14*	*0.14*	*0.11*	*0.11*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C32

Dataset 1: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Neuroticism, Openness, and Conscientiousness (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.117	7.332	7.438	7.593	7.696	7.515	7.330	0.29	0.21	0.17	0.16	0.15	0.18	0.25
2	6.538	6.517	6.691	6.490	6.719	6.604	6.520	0.18	0.20	0.13	0.20	0.13	0.16	0.18
3	6.657	6.545	6.854	6.972	6.933	6.826	6.613	0.22	0.25	0.21	0.17	0.18	0.21	0.23
4	7.258	7.334	7.192	7.473	7.477	7.369	7.267	0.14	0.10	0.14	0.12	0.13	0.12	0.13
5	7.340	7.586	7.421	7.912	8.315	7.808	7.306	0.20	0.16	0.17	0.09	0.06	0.12	0.21
6	7.181	7.341	7.440	7.549	7.673	7.501	7.493	0.26	0.24	0.22	0.19	0.15	0.20	0.21
7	6.959	6.868	7.160	6.876	7.256	7.040	6.812	0.20	0.23	0.13	0.22	0.16	0.19	0.19
8	6.865	7.089	7.078	7.094	7.125	7.096	6.923	0.33	0.25	0.24	0.23	0.21	0.23	0.28
9	7.121	7.416	7.357	7.411	7.349	7.383	7.085	0.31	0.23	0.25	0.24	0.27	0.25	0.31
10	7.504	7.562	7.792	7.560	7.609	7.631	7.371	0.24	0.19	0.13	0.20	0.18	0.17	0.22
11	7.506	7.399	7.591	7.511	7.475	7.494	7.500	0.38	0.35	0.34	0.29	0.34	0.33	0.36
12	7.117	7.088	7.053	7.141	7.079	7.090	7.054	0.32	0.33	0.34	0.32	0.31	0.32	0.35
13	6.906	6.982	6.929	7.453	6.867	7.058	6.929	0.21	0.22	0.23	0.13	0.22	0.20	0.21
14	6.576	6.643	6.748	6.794	6.708	6.723	6.624	0.31	0.31	0.25	0.24	0.27	0.27	0.27
15	6.834	6.920	6.913	6.917	6.805	6.889	6.869	0.32	0.25	0.26	0.26	0.30	0.27	0.32
16	7.155	7.068	7.236	6.957	7.233	7.124	7.014	0.28	0.27	0.21	0.29	0.23	0.25	0.29
17	7.342	7.325	7.367	7.330	7.305	7.332	7.287	0.27	0.29	0.19	0.22	0.22	0.23	0.29
18	7.271	7.350	7.543	7.533	7.670	7.524	7.170	0.24	0.18	0.09	0.08	0.05	0.10	0.25
19	7.618	7.612	7.692	7.638	7.713	7.664	7.608	0.24	0.25	0.19	0.23	0.19	0.22	0.23
20	7.547	7.722	7.942	7.783	7.813	7.815	7.619	0.26	0.21	0.10	0.18	0.16	0.16	0.23
Mean	**7.121**	**7.185**	**7.272**	**7.299**	**7.341**	**7.274**	**7.120**	**0.26**	**0.24**	**0.20**	**0.20**	**0.20**	**0.21**	**0.25**
SD	*0.319*	*0.354*	*0.348*	*0.372*	*0.421*	*0.355*	*0.331*	*0.06*	*0.06*	*0.07*	*0.06*	*0.08*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C33

Dataset 2: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Neuroticism, Extraversion, Openness, and Conscientiousness (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.195	27.556	29.407	28.265	28.516	28.436	27.122	0.25	0.23	0.23	0.27	0.28	0.25	0.25
2	32.683	31.995	33.342	32.930	33.244	32.878	32.497	0.37	0.39	0.32	0.32	0.32	0.34	0.37
3	30.935	33.403	34.845	33.148	33.857	33.813	31.481	0.23	0.14	0.11	0.16	0.15	0.14	0.19
4	30.605	35.029	33.976	34.970	31.004	33.745	29.904	0.22	-0.01	0.21	0.16	0.25	0.15	0.26
5	32.948	33.869	33.986	32.933	32.819	33.402	32.619	0.20	0.18	0.20	0.25	0.25	0.22	0.23
6	32.320	32.801	31.498	31.080	31.020	31.600	31.748	0.10	0.14	0.24	0.25	0.28	0.23	0.16
7	33.671	34.800	36.181	34.616	33.842	34.860	33.520	0.18	0.19	0.11	0.19	0.22	0.18	0.17
8	32.619	33.325	34.112	33.628	34.095	33.790	32.893	0.30	0.27	0.24	0.25	0.24	0.25	0.28
9	35.477	35.324	35.818	35.409	32.939	34.873	35.169	0.08	0.10	0.05	0.11	0.23	0.12	0.08
10	35.178	36.024	34.383	35.456	34.491	35.088	34.658	0.15	0.08	0.23	0.11	0.18	0.15	0.16
11	28.498	29.363	30.851	29.291	29.312	29.704	29.052	0.16	0.23	0.16	0.29	0.27	0.24	0.22
12	35.725	35.893	35.818	35.557	35.208	35.619	35.554	0.15	0.15	0.21	0.19	0.20	0.19	0.15
13	32.954	32.318	32.085	32.457	32.133	32.248	32.774	0.22	0.19	0.25	0.28	0.30	0.25	0.22
14	31.924	33.195	32.215	32.737	31.700	32.462	31.197	0.25	0.20	0.23	0.22	0.29	0.23	0.28
15	33.656	34.321	31.741	32.321	32.349	32.683	32.463	0.32	0.31	0.36	0.33	0.30	0.33	0.33
16	29.178	29.468	30.285	30.047	29.694	29.873	29.055	0.19	0.17	0.17	0.17	0.20	0.18	0.18
17	34.583	36.495	34.226	34.009	34.701	34.858	33.894	0.25	0.18	0.32	0.35	0.31	0.29	0.33
18	32.865	32.618	31.240	31.457	32.299	31.903	33.054	0.23	0.24	0.31	0.30	0.25	0.28	0.23
19	30.298	30.312	30.747	30.561	31.044	30.666	30.302	0.24	0.27	0.29	0.29	0.29	0.28	0.25
20	31.563	31.485	30.990	31.263	30.896	31.158	31.722	0.26	0.29	0.32	0.30	0.33	0.31	0.23
Mean	**32.244**	**32.980**	**32.887**	**32.607**	**32.258**	**32.683**	**32.034**	**0.22**	**0.20**	**0.23**	**0.24**	**0.26**	**0.23**	**0.23**
SD	*2.295*	*2.426*	*2.029*	*2.102*	*1.862*	*2.001*	*2.158*	*0.07*	*0.09*	*0.08*	*0.07*	*0.05*	*0.06*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C34

Dataset 3: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Neuroticism, Openness, Agreeableness, and Conscientiousness (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.040	1.057	1.055	1.120	1.091	1.081	1.046	0.18	0.15	0.12	-0.02	0.09	0.09	0.20
2	1.085	1.127	1.219	1.200	1.237	1.196	1.132	0.14	0.08	-0.01	0.01	-0.03	0.02	0.06
3	1.127	1.281	1.273	1.290	1.225	1.267	1.160	0.05	-0.04	-0.05	-0.11	-0.05	-0.06	-0.01
4	1.016	1.078	1.084	1.071	1.067	1.075	1.016	0.20	0.02	0.01	0.04	0.05	0.03	0.22
5	1.080	1.104	1.108	1.105	1.100	1.104	1.078	0.08	0.03	0.04	0.04	0.05	0.04	0.08
6	0.955	0.965	0.964	1.011	0.994	0.984	0.952	0.14	0.13	0.17	0.08	0.14	0.13	0.18
7	1.115	1.141	1.143	1.159	1.140	1.146	1.101	0.20	0.10	0.15	0.10	0.15	0.13	0.24
8	1.104	1.165	1.149	1.179	1.150	1.161	1.098	0.03	0.01	-0.01	0.01	0.00	0.00	0.06
9	1.080	1.099	1.111	1.096	1.111	1.104	1.071	0.03	-0.03	-0.05	-0.01	-0.05	-0.03	0.04
10	1.052	1.048	1.119	1.099	1.089	1.089	1.060	0.16	0.17	0.00	0.02	-0.04	0.04	0.13
11	1.080	1.109	1.127	1.126	1.121	1.121	1.084	-0.06	0.00	-0.02	-0.01	-0.04	-0.02	-0.02
12	1.078	1.092	1.101	1.134	1.113	1.110	1.096	0.10	0.04	0.07	0.02	0.05	0.05	0.06
13	1.116	1.123	1.123	1.202	1.163	1.153	1.122	0.14	0.17	0.08	-0.02	-0.02	0.05	0.11
14	1.144	1.180	1.225	1.206	1.243	1.214	1.176	0.03	0.01	-0.09	-0.08	-0.13	-0.07	0.01
15	1.154	1.174	1.205	1.176	1.192	1.187	1.163	-0.01	0.02	-0.05	-0.01	0.00	-0.01	0.00
16	1.063	1.125	1.159	1.171	1.148	1.151	1.069	0.06	0.02	-0.10	-0.09	-0.09	-0.07	0.01
17	0.986	1.136	1.058	1.118	1.135	1.112	1.079	0.09	-0.04	-0.04	-0.07	-0.06	-0.05	0.01
18	1.078	1.112	1.157	1.148	1.166	1.146	1.089	0.15	0.07	0.01	0.01	-0.02	0.02	0.11
19	1.074	1.106	1.132	1.165	1.162	1.141	1.090	0.12	0.05	-0.01	-0.04	-0.01	0.00	0.08
20	1.152	1.242	1.219	1.283	1.197	1.236	1.209	0.00	0.01	-0.01	0.00	-0.03	-0.01	-0.01
Mean	**1.079**	**1.123**	**1.136**	**1.153**	**1.142**	**1.139**	**1.095**	**0.09**	**0.05**	**0.01**	**-0.01**	**0.00**	**0.01**	**0.08**
SD	*0.052*	*0.068*	*0.071*	*0.066*	*0.061*	*0.063*	*0.057*	*0.07*	*0.07*	*0.07*	*0.05*	*0.07*	*0.06*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C35

Dataset 4: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Neuroticism, Extraversion, and Conscientiousness (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.471	0.478	0.478	0.487	0.481	0.481	0.471	0.14	0.23	0.24	0.22	0.23	0.23	0.17
2	0.436	0.462	0.491	0.511	0.500	0.491	0.451	0.23	0.16	0.14	0.11	0.08	0.12	0.18
3	0.543	0.576	0.574	0.582	0.579	0.578	0.560	0.12	0.07	0.08	0.07	0.05	0.07	0.09
4	0.493	0.498	0.501	0.543	0.528	0.517	0.489	0.31	0.28	0.25	0.18	0.20	0.23	0.33
5	0.475	0.464	0.466	0.481	0.492	0.476	0.474	0.28	0.34	0.33	0.27	0.25	0.29	0.29
6	0.471	0.473	0.497	0.487	0.500	0.489	0.468	0.25	0.25	0.17	0.23	0.25	0.22	0.27
7	0.541	0.542	0.521	0.520	0.518	0.525	0.537	0.05	0.08	0.16	0.18	0.19	0.15	0.06
8	0.530	0.551	0.526	0.521	0.530	0.532	0.526	0.33	0.23	0.36	0.36	0.35	0.33	0.35
9	0.580	0.576	0.560	0.556	0.572	0.566	0.572	0.28	0.22	0.34	0.35	0.29	0.30	0.32
10	0.450	0.477	0.490	0.490	0.492	0.487	0.460	0.22	0.12	0.04	0.04	0.01	0.05	0.18
11	0.545	0.572	0.577	0.659	0.569	0.594	0.545	0.25	0.20	0.16	-0.03	0.23	0.14	0.27
12	0.439	0.445	0.453	0.447	0.501	0.461	0.434	0.37	0.35	0.27	0.28	0.06	0.24	0.37
13	0.464	0.471	0.508	0.521	0.499	0.500	0.461	0.25	0.21	0.19	0.15	0.20	0.19	0.28
14	0.502	0.496	0.534	0.533	0.543	0.527	0.514	0.25	0.27	0.01	0.05	-0.01	0.08	0.21
15	0.496	0.497	0.486	0.494	0.483	0.490	0.495	0.35	0.36	0.38	0.39	0.41	0.38	0.37
16	0.442	0.440	0.435	0.428	0.436	0.435	0.436	0.29	0.31	0.33	0.32	0.31	0.32	0.32
17	0.449	0.445	0.455	0.454	0.467	0.455	0.451	0.20	0.25	0.28	0.26	0.23	0.25	0.23
18	0.442	0.435	0.528	0.521	0.504	0.497	0.432	0.22	0.26	0.01	0.09	0.18	0.13	0.25
19	0.420	0.435	0.477	0.462	0.459	0.458	0.435	0.36	0.36	0.13	0.23	0.22	0.24	0.32
20	0.486	0.476	0.479	0.469	0.467	0.473	0.482	0.30	0.37	0.37	0.37	0.39	0.37	0.35
Mean	**0.484**	**0.490**	**0.502**	**0.508**	**0.506**	**0.502**	**0.485**	**0.25**	**0.25**	**0.21**	**0.21**	**0.21**	**0.22**	**0.26**
SD	*0.044*	*0.048*	*0.039*	*0.052*	*0.038*	*0.042*	*0.044*	*0.08*	*0.09*	*0.12*	*0.12*	*0.12*	*0.10*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C36

Dataset 5: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Adjustment, Likeability, and Prudence (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.768	6.785	6.788	6.795	6.826	6.798	6.722	0.11	0.08	0.08	0.08	0.08	0.08	0.09
2	6.273	6.175	6.204	6.233	6.313	6.231	6.215	0.07	0.10	0.09	0.08	0.10	0.09	0.07
3	6.141	6.207	6.250	6.228	6.185	6.217	6.167	0.13	0.10	0.06	0.09	0.12	0.09	0.12
4	6.109	6.083	6.065	6.092	6.124	6.091	6.134	0.17	0.17	0.19	0.20	0.20	0.19	0.18
5	5.951	6.097	6.087	6.042	6.102	6.082	5.954	0.09	0.07	0.09	0.10	0.10	0.09	0.10
6	6.268	6.295	6.371	6.378	6.405	6.362	6.337	0.15	0.16	0.15	0.14	0.13	0.14	0.13
7	5.793	5.821	5.854	5.803	5.829	5.827	5.820	0.11	0.06	0.05	0.07	0.08	0.07	0.10
8	6.274	6.280	6.357	6.365	6.271	6.318	6.296	0.08	0.08	0.03	0.02	0.07	0.05	0.07
9	6.345	6.338	6.308	6.298	6.344	6.322	6.337	0.08	0.07	0.08	0.09	0.05	0.07	0.09
10	5.646	5.626	5.634	5.730	5.685	5.669	5.591	0.23	0.20	0.20	0.15	0.16	0.18	0.22
11	5.793	5.749	5.761	5.747	5.751	5.752	5.849	0.05	0.11	0.09	0.11	0.10	0.10	0.03
12	5.812	5.950	6.039	5.954	5.879	5.955	5.856	0.10	0.04	0.03	0.06	0.05	0.05	0.11
13	5.621	5.662	5.740	5.709	5.668	5.695	5.639	0.07	0.07	0.05	0.05	0.06	0.06	0.07
14	5.780	5.949	6.044	6.130	5.983	6.026	5.809	0.14	0.11	0.08	0.04	0.07	0.07	0.11
15	6.246	6.414	6.420	6.497	6.547	6.470	6.367	0.01	-0.01	-0.01	-0.03	-0.02	-0.02	-0.02
16	6.055	6.100	6.098	6.089	6.133	6.105	6.021	0.09	0.08	0.08	0.08	0.04	0.07	0.11
17	6.454	6.459	6.433	6.479	6.496	6.467	6.454	0.14	0.15	0.18	0.14	0.12	0.15	0.17
18	6.557	6.695	6.692	6.702	6.657	6.686	6.560	-0.01	-0.04	-0.04	-0.04	-0.03	-0.04	-0.01
19	6.038	6.082	6.057	6.135	6.086	6.090	6.030	0.17	0.13	0.15	0.11	0.12	0.13	0.17
20	6.575	6.522	6.546	6.536	6.567	6.543	6.608	0.06	0.10	0.07	0.08	0.05	0.07	0.06
Mean	**6.125**	**6.164**	**6.187**	**6.197**	**6.192**	**6.185**	**6.138**	**0.10**	**0.09**	**0.09**	**0.08**	**0.08**	**0.08**	**0.10**
SD	*0.324*	*0.321*	*0.311*	*0.316*	*0.332*	*0.318*	*0.323*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C37

Dataset 6: Predictive performance statistics for the prediction equations developed using the (theoretical) combination of Emotional Orientation, Cognitive Orientation, and Task Orientation (Study 3).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.598	0.612	0.643	0.675	0.682	0.653	0.584	-0.07	-0.06	0.00	0.04	-0.01	-0.01	-0.05
2	0.621	0.616	0.569	0.575	0.621	0.595	0.616	-0.12	-0.07	0.06	0.16	0.11	0.07	-0.05
3	0.577	0.585	0.645	0.586	0.574	0.598	0.586	0.11	0.10	0.13	0.11	0.11	0.11	0.03
4	0.599	0.614	0.530	0.656	0.625	0.606	0.614	-0.19	-0.20	0.16	0.10	0.22	0.07	-0.18
5	0.579	0.617	0.624	0.583	0.584	0.602	0.562	-0.01	-0.09	-0.13	0.10	0.10	-0.01	0.05
6	0.645	0.641	0.617	0.578	0.578	0.603	0.611	-0.21	-0.20	0.01	0.17	0.08	0.01	-0.05
7	0.578	0.772	0.657	0.466	0.543	0.609	0.578	0.17	-0.08	0.18	0.45	0.25	0.20	0.09
8	0.581	0.580	0.595	0.539	0.629	0.586	0.578	-0.02	0.00	0.09	0.17	0.14	0.10	0.05
9	0.569	0.811	0.803	0.667	0.586	0.717	0.585	0.03	0.10	0.17	-0.01	0.18	0.11	0.02
10	0.601	0.610	0.560	0.680	0.636	0.621	0.574	0.03	-0.01	0.22	-0.02	0.04	0.06	0.08
11	0.585	0.582	0.589	0.631	0.617	0.605	0.584	0.08	0.14	0.05	-0.09	-0.09	0.00	0.11
12	0.604	0.621	0.643	0.544	0.546	0.589	0.591	-0.08	0.03	0.06	0.26	0.38	0.18	0.00
13	0.607	0.603	0.586	0.595	0.592	0.594	0.600	0.11	0.17	0.17	0.23	0.25	0.20	0.17
14	0.672	0.681	0.659	0.547	0.557	0.611	0.648	-0.13	-0.12	0.00	0.47	0.46	0.20	-0.02
15	0.553	0.552	0.504	0.614	0.612	0.571	0.542	-0.01	0.03	0.20	-0.04	0.02	0.05	0.04
16	0.602	0.611	0.680	0.706	0.718	0.679	0.618	0.16	0.08	0.05	0.03	0.07	0.06	0.00
17	0.611	0.622	0.620	0.530	0.737	0.627	0.610	0.17	0.15	-0.09	0.29	0.21	0.14	0.02
18	0.657	0.681	0.878	0.727	0.871	0.789	0.647	-0.18	-0.26	0.14	0.01	0.02	-0.02	-0.12
19	0.556	0.553	0.550	0.528	0.547	0.545	0.554	0.08	0.10	0.08	0.30	0.14	0.16	0.13
20	0.665	0.665	0.607	0.633	0.727	0.658	0.635	-0.40	-0.40	-0.03	0.11	-0.01	-0.08	-0.29
Mean	**0.603**	**0.631**	**0.628**	**0.603**	**0.629**	**0.623**	**0.596**	**-0.02**	**-0.03**	**0.08**	**0.14**	**0.13**	**0.08**	**0.00**
SD	*0.034*	*0.065*	*0.087*	*0.068*	*0.082*	*0.055*	*0.029*	*0.15*	*0.15*	*0.10*	*0.15*	*0.14*	*0.08*	*0.11*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C38

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of Extraversion and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	7.173	7.141	7.206	7.141	7.122	7.153	7.193	0.29	0.30	0.29	0.29	0.31	0.30	0.30
2	6.170	6.121	6.214	6.342	6.287	6.241	6.211	0.35	0.34	0.25	0.19	0.21	0.25	0.34
3	6.469	6.523	6.770	6.755	6.676	6.681	6.454	0.32	0.30	0.19	0.20	0.23	0.23	0.32
4	7.836	7.993	8.074	7.997	8.082	8.037	7.865	0.00	-0.02	0.02	0.00	0.01	0.01	0.00
5	6.871	7.003	7.342	7.115	7.175	7.159	6.873	0.30	0.22	0.18	0.22	0.19	0.20	0.30
6	7.261	7.344	7.510	7.704	7.560	7.530	7.298	0.26	0.25	0.20	0.16	0.19	0.20	0.26
7	6.732	7.088	6.819	6.977	6.778	6.916	6.647	0.34	0.23	0.27	0.24	0.30	0.26	0.34
8	6.892	7.116	7.002	6.956	6.996	7.017	7.065	0.26	0.18	0.23	0.22	0.24	0.22	0.23
9	7.175	7.200	7.140	7.137	7.160	7.159	7.135	0.27	0.27	0.32	0.32	0.29	0.30	0.29
10	7.591	7.714	7.682	7.719	7.686	7.700	7.673	0.26	0.14	0.19	0.16	0.18	0.17	0.22
11	7.630	7.594	7.710	7.587	7.590	7.620	7.537	0.42	0.40	0.27	0.33	0.33	0.33	0.41
12	7.031	7.206	7.110	7.347	7.206	7.217	7.099	0.36	0.31	0.36	0.28	0.32	0.31	0.36
13	6.988	6.993	6.981	6.995	7.058	7.006	6.903	0.20	0.20	0.20	0.20	0.19	0.20	0.22
14	6.633	6.601	6.507	6.478	6.482	6.517	6.701	0.30	0.32	0.37	0.38	0.39	0.36	0.34
15	6.864	6.829	6.812	6.798	6.857	6.824	6.879	0.30	0.32	0.36	0.37	0.32	0.34	0.31
16	7.310	7.350	7.302	7.336	7.308	7.324	7.342	0.29	0.27	0.26	0.25	0.26	0.26	0.28
17	7.243	7.224	7.123	7.117	7.145	7.152	7.247	0.31	0.32	0.38	0.38	0.37	0.36	0.33
18	7.199	7.169	7.206	7.212	7.136	7.181	7.201	0.24	0.25	0.24	0.23	0.25	0.24	0.25
19	7.583	7.545	7.547	7.635	7.548	7.569	7.536	0.26	0.26	0.26	0.20	0.24	0.24	0.27
20	7.532	7.738	7.803	7.786	7.822	7.787	7.556	0.24	0.17	0.15	0.18	0.16	0.17	0.23
Mean	**7.109**	**7.175**	**7.193**	**7.207**	**7.184**	**7.189**	**7.121**	**0.28**	**0.25**	**0.25**	**0.24**	**0.25**	**0.25**	**0.28**
SD	*0.411*	*0.431*	*0.442*	*0.428*	*0.436*	*0.429*	*0.409*	*0.08*	*0.09*	*0.08*	*0.09*	*0.08*	*0.08*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C39

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of Extraversion, Openness and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.960	7.040	7.164	7.189	7.069	7.115	6.999	0.34	0.29	0.26	0.23	0.29	0.27	0.31
2	6.059	6.185	6.288	6.477	6.370	6.330	6.084	0.38	0.33	0.25	0.18	0.20	0.24	0.36
3	6.448	6.558	7.054	6.742	6.971	6.831	6.521	0.29	0.27	0.18	0.26	0.20	0.23	0.28
4	7.422	7.767	7.710	7.756	7.927	7.790	7.464	0.08	0.03	0.03	0.02	0.02	0.03	0.07
5	7.261	7.311	7.347	7.467	7.579	7.426	7.229	0.21	0.20	0.18	0.17	0.16	0.18	0.22
6	7.053	7.400	7.512	7.766	7.424	7.525	7.223	0.36	0.24	0.20	0.17	0.25	0.21	0.31
7	6.695	6.878	7.440	6.801	7.180	7.075	6.677	0.26	0.23	0.15	0.25	0.19	0.21	0.26
8	7.051	7.070	7.228	7.185	7.208	7.173	7.237	0.33	0.29	0.25	0.26	0.25	0.26	0.30
9	6.943	6.977	7.069	6.904	7.079	7.007	6.911	0.35	0.34	0.29	0.36	0.29	0.32	0.37
10	7.206	7.396	7.337	7.476	7.263	7.368	7.356	0.37	0.24	0.27	0.26	0.29	0.26	0.32
11	7.324	7.436	7.278	7.477	7.500	7.423	7.351	0.43	0.38	0.40	0.40	0.40	0.39	0.42
12	6.788	6.918	6.870	6.921	7.047	6.939	6.803	0.48	0.43	0.45	0.42	0.37	0.42	0.50
13	6.612	6.696	6.738	6.638	6.881	6.738	6.702	0.31	0.29	0.28	0.28	0.25	0.27	0.29
14	6.351	6.826	6.724	6.573	6.694	6.705	6.885	0.41	0.25	0.29	0.34	0.31	0.30	0.32
15	6.572	6.573	6.497	6.558	6.557	6.546	6.798	0.46	0.45	0.46	0.45	0.46	0.46	0.43
16	6.904	6.902	6.819	6.816	6.888	6.856	6.792	0.37	0.34	0.33	0.34	0.31	0.33	0.34
17	6.947	6.942	7.213	6.965	7.125	7.061	6.985	0.46	0.45	0.32	0.44	0.34	0.39	0.44
18	7.047	7.698	7.698	7.648	7.544	7.647	7.188	0.32	0.08	0.10	0.11	0.12	0.10	0.29
19	7.384	7.462	7.604	7.459	7.784	7.577	7.398	0.31	0.27	0.21	0.28	0.16	0.23	0.30
20	7.534	7.882	8.271	8.470	8.225	8.212	7.616	0.26	0.20	0.14	0.10	0.13	0.14	0.23
Mean	**6.928**	**7.096**	**7.193**	**7.164**	**7.216**	**7.167**	**7.011**	**0.34**	**0.28**	**0.25**	**0.27**	**0.25**	**0.26**	**0.32**
SD	*0.376*	*0.430*	*0.452*	*0.504*	*0.447*	*0.446*	*0.360*	*0.09*	*0.10*	*0.11*	*0.12*	*0.10*	*0.10*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C40

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Extraversion,

Openness *and Conscientiousness (Study 4).*

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.941	7.376	7.549	7.546	7.584	7.514	7.169	0.34	0.26	0.22	0.22	0.20	0.22	0.30
2	6.162	6.210	7.211	6.962	7.010	6.848	6.274	0.29	0.31	0.00	0.07	0.07	0.11	0.26
3	6.447	6.925	6.737	7.042	6.894	6.899	6.574	0.29	0.21	0.28	0.21	0.25	0.24	0.29
4	7.580	7.888	7.848	8.037	7.998	7.943	7.665	0.04	0.01	0.01	0.02	0.04	0.02	0.04
5	7.296	7.540	7.507	8.011	7.864	7.731	7.486	0.21	0.17	0.17	0.09	0.12	0.14	0.18
6	7.026	7.195	7.503	7.345	7.744	7.447	7.235	0.33	0.29	0.21	0.23	0.14	0.22	0.28
7	6.727	6.670	6.954	7.637	7.385	7.161	6.811	0.27	0.30	0.23	0.19	0.18	0.22	0.27
8	7.035	7.013	6.981	7.076	7.249	7.080	6.991	0.33	0.32	0.31	0.27	0.21	0.28	0.31
9	6.938	7.009	7.142	7.128	7.166	7.111	6.966	0.34	0.34	0.31	0.30	0.29	0.31	0.37
10	7.222	7.273	7.307	7.452	7.406	7.360	7.220	0.34	0.30	0.25	0.21	0.21	0.24	0.30
11	7.310	7.224	7.712	7.568	7.307	7.453	7.210	0.44	0.42	0.28	0.32	0.36	0.35	0.44
12	6.903	7.010	8.469	7.953	8.170	7.900	7.023	0.42	0.39	0.12	0.14	0.10	0.19	0.38
13	6.714	7.191	7.153	8.089	7.369	7.451	6.744	0.30	0.24	0.26	0.15	0.22	0.22	0.28
14	6.335	7.544	7.493	7.522	7.447	7.502	6.547	0.41	0.21	0.19	0.18	0.19	0.19	0.38
15	6.537	6.658	6.771	6.774	6.811	6.753	6.590	0.43	0.36	0.32	0.32	0.32	0.33	0.43
16	6.877	7.010	6.853	7.119	7.345	7.082	6.941	0.37	0.31	0.33	0.25	0.21	0.28	0.36
17	7.103	7.004	7.129	7.348	7.282	7.191	7.097	0.38	0.39	0.33	0.18	0.22	0.28	0.38
18	7.035	7.761	8.039	7.448	7.868	7.779	6.978	0.33	0.06	0.04	0.10	0.07	0.07	0.33
19	7.405	7.361	7.484	7.566	7.647	7.515	7.412	0.31	0.30	0.26	0.20	0.18	0.23	0.30
20	7.572	8.115	8.232	8.406	8.059	8.203	7.666	0.26	0.18	0.12	0.08	0.14	0.13	0.24
Mean	**6.958**	**7.199**	**7.404**	**7.501**	**7.480**	**7.396**	**7.030**	**0.32**	**0.27**	**0.21**	**0.19**	**0.19**	**0.21**	**0.31**
SD	*0.380*	*0.432*	*0.467*	*0.418*	*0.372*	*0.377*	*0.364*	*0.09*	*0.10*	*0.10*	*0.08*	*0.08*	*0.08*	*0.09*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C41

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.925	7.816	7.666	8.384	8.003	7.967	6.986	0.33	0.20	0.27	0.13	0.18	0.19	0.30
2	6.162	6.766	7.070	7.624	7.728	7.297	6.314	0.29	0.22	0.09	0.16	0.03	0.13	0.29
3	6.442	6.641	7.402	7.475	7.812	7.333	6.593	0.30	0.36	0.31	0.31	0.25	0.31	0.33
4	7.601	7.824	8.563	9.627	9.186	8.800	7.828	0.03	0.12	0.03	0.05	0.01	0.05	0.01
5	7.304	7.126	8.038	8.280	9.541	8.246	7.267	0.21	0.27	0.19	0.16	0.16	0.20	0.24
6	7.168	7.562	8.036	7.604	8.276	7.870	7.245	0.30	0.25	0.18	0.20	0.15	0.20	0.32
7	6.728	6.990	7.578	7.327	7.317	7.303	6.772	0.27	0.27	0.28	0.24	0.25	0.26	0.27
8	7.085	7.501	7.702	7.897	7.260	7.590	7.287	0.32	0.22	0.15	0.20	0.30	0.22	0.28
9	6.940	7.096	7.285	7.682	7.298	7.340	6.976	0.34	0.35	0.31	0.25	0.29	0.30	0.36
10	7.395	7.387	7.489	7.335	7.382	7.398	7.508	0.29	0.31	0.30	0.29	0.24	0.28	0.27
11	7.463	7.724	8.109	7.845	10.844	8.631	7.484	0.38	0.20	0.23	0.22	-0.08	0.14	0.36
12	6.931	6.958	6.644	7.628	8.303	7.383	6.826	0.39	0.33	0.46	0.25	0.16	0.30	0.42
13	6.758	6.788	7.640	7.603	7.304	7.334	6.880	0.29	0.33	0.26	0.21	0.26	0.27	0.32
14	6.384	7.438	6.829	7.265	8.343	7.469	6.608	0.38	0.24	0.35	0.31	0.31	0.30	0.36
15	6.746	6.978	8.037	7.686	7.377	7.520	7.170	0.29	0.29	0.21	0.20	0.20	0.22	0.25
16	6.943	7.187	7.504	7.214	7.964	7.467	6.986	0.34	0.33	0.22	0.31	0.15	0.25	0.33
17	7.114	6.869	7.075	8.395	7.941	7.570	7.169	0.35	0.43	0.33	0.15	0.14	0.26	0.35
18	7.172	7.819	7.662	7.675	7.643	7.700	7.221	0.23	0.00	0.13	0.19	0.20	0.13	0.25
19	7.488	7.310	7.659	7.627	7.675	7.568	7.475	0.26	0.33	0.18	0.23	0.20	0.23	0.26
20	7.607	7.792	8.382	8.316	8.430	8.230	7.505	0.24	0.29	0.12	0.19	0.17	0.19	0.25
Mean	**7.018**	**7.279**	**7.619**	**7.824**	**8.081**	**7.701**	**7.105**	**0.29**	**0.27**	**0.23**	**0.21**	**0.18**	**0.22**	**0.29**
SD	*0.394*	*0.383*	*0.480*	*0.544*	*0.875*	*0.439*	*0.364*	*0.08*	*0.09*	*0.10*	*0.06*	*0.10*	*0.07*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C42

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of Extraversion and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.074	27.227	27.277	27.593	27.232	27.332	27.276	0.25	0.24	0.23	0.23	0.22	0.23	0.24
2	32.719	32.812	32.536	32.359	32.174	32.470	32.581	0.37	0.32	0.37	0.34	0.35	0.34	0.38
3	30.483	31.791	32.071	32.082	32.629	32.144	31.219	0.29	0.16	0.14	0.16	0.13	0.15	0.21
4	30.916	30.725	31.110	31.064	31.734	31.158	31.085	0.23	0.23	0.21	0.23	0.17	0.21	0.22
5	33.033	34.151	34.349	34.498	34.336	34.333	33.823	0.21	0.12	0.10	0.10	0.11	0.11	0.15
6	31.515	32.786	31.153	32.988	33.544	32.618	31.637	0.12	0.12	0.16	0.11	0.11	0.13	0.11
7	33.307	35.110	35.800	35.422	35.093	35.356	34.048	0.20	0.11	0.10	0.07	0.09	0.09	0.16
8	32.345	32.323	32.605	32.587	32.543	32.514	32.299	0.30	0.29	0.28	0.29	0.28	0.28	0.30
9	34.329	34.196	33.324	33.312	33.570	33.601	34.269	0.13	0.15	0.19	0.18	0.17	0.17	0.14
10	34.825	34.693	34.717	34.330	34.527	34.567	34.485	0.17	0.19	0.20	0.23	0.22	0.21	0.18
11	27.425	28.192	27.952	28.049	27.822	28.004	27.214	0.34	0.28	0.29	0.29	0.29	0.29	0.33
12	34.846	34.934	34.883	34.853	34.876	34.887	34.723	0.18	0.18	0.19	0.18	0.19	0.18	0.18
13	32.902	33.174	33.292	33.255	33.345	33.267	32.995	0.24	0.23	0.23	0.23	0.23	0.23	0.23
14	31.433	31.463	31.653	31.981	31.959	31.764	31.208	0.31	0.31	0.27	0.20	0.19	0.24	0.30
15	33.642	33.727	33.353	33.550	33.459	33.522	34.448	0.32	0.31	0.33	0.32	0.32	0.32	0.31
16	28.442	30.273	30.483	30.096	30.496	30.337	29.014	0.22	0.09	0.09	0.11	0.08	0.09	0.18
17	34.523	34.858	34.907	34.753	34.818	34.834	34.564	0.26	0.27	0.27	0.28	0.27	0.27	0.27
18	31.783	32.059	32.002	31.858	31.609	31.882	32.043	0.32	0.28	0.28	0.29	0.30	0.29	0.31
19	30.063	30.712	30.703	30.343	30.603	30.590	30.280	0.25	0.24	0.25	0.26	0.24	0.25	0.25
20	31.022	31.390	31.585	32.018	31.452	31.611	31.813	0.38	0.27	0.26	0.24	0.30	0.27	0.30
Mean	**31.831**	**32.330**	**32.288**	**32.350**	**32.391**	**32.340**	**32.051**	**0.25**	**0.22**	**0.22**	**0.22**	**0.21**	**0.22**	**0.24**
SD	*2.246*	*2.133*	*2.155*	*2.075*	*2.110*	*2.100*	*2.231*	*0.07*	*0.07*	*0.08*	*0.07*	*0.08*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C43

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Extraversion, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.236	27.936	27.330	27.540	27.670	27.619	26.845	0.24	0.23	0.25	0.25	0.24	0.24	0.25
2	32.725	31.755	32.295	31.941	31.971	31.991	32.297	0.37	0.38	0.30	0.35	0.35	0.35	0.38
3	30.905	33.215	32.469	32.921	32.862	32.867	31.489	0.23	0.15	0.16	0.10	0.13	0.13	0.18
4	30.899	30.606	31.864	31.300	31.349	31.280	30.293	0.23	0.27	0.19	0.23	0.22	0.23	0.27
5	32.984	33.863	33.546	33.597	34.025	33.758	32.780	0.19	0.13	0.18	0.16	0.11	0.14	0.18
6	31.596	32.037	31.465	31.606	31.674	31.696	31.285	0.12	0.18	0.22	0.19	0.18	0.19	0.14
7	33.260	34.421	33.946	34.131	33.661	34.040	33.705	0.20	0.19	0.20	0.19	0.21	0.20	0.15
8	32.390	32.869	32.770	33.597	32.491	32.932	32.484	0.29	0.25	0.28	0.26	0.31	0.28	0.29
9	35.213	34.030	34.044	32.911	34.225	33.802	35.015	0.08	0.15	0.14	0.18	0.11	0.14	0.10
10	34.848	33.111	33.283	32.810	33.261	33.116	34.744	0.17	0.29	0.29	0.30	0.29	0.30	0.19
11	27.450	27.151	27.596	27.658	27.575	27.495	27.646	0.31	0.39	0.37	0.35	0.35	0.36	0.34
12	34.985	34.868	35.492	35.260	34.666	35.071	35.445	0.17	0.17	0.15	0.15	0.17	0.16	0.15
13	32.958	32.313	31.712	31.598	32.078	31.925	32.914	0.22	0.27	0.31	0.28	0.28	0.28	0.22
14	31.511	31.239	31.192	31.150	31.125	31.177	31.443	0.28	0.25	0.27	0.27	0.30	0.27	0.27
15	33.648	34.283	33.922	33.179	33.580	33.741	33.067	0.32	0.16	0.24	0.29	0.25	0.24	0.33
16	28.719	29.084	29.242	29.429	28.880	29.159	28.700	0.22	0.18	0.20	0.16	0.18	0.18	0.19
17	34.700	33.436	33.815	33.871	34.231	33.838	34.280	0.25	0.35	0.32	0.33	0.29	0.32	0.27
18	32.008	33.135	33.952	33.122	32.216	33.106	32.740	0.29	0.17	0.17	0.18	0.22	0.18	0.19
19	30.149	29.525	28.927	29.214	29.175	29.210	29.996	0.24	0.27	0.30	0.29	0.30	0.29	0.25
20	31.402	32.306	31.455	32.016	32.249	32.006	31.519	0.31	0.27	0.32	0.29	0.24	0.28	0.28
Mean	**31.979**	**32.059**	**32.016**	**31.943**	**31.948**	**31.991**	**31.934**	**0.24**	**0.24**	**0.24**	**0.24**	**0.24**	**0.24**	**0.23**
SD	*2.261*	*2.134*	*2.185*	*2.041*	*2.078*	*2.086*	*2.287*	*0.07*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C44

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	27.320	28.051	28.149	28.755	28.101	28.264	27.278	0.25	0.23	0.24	0.23	0.23	0.23	0.26
2	32.701	32.774	34.515	34.815	34.660	34.191	32.293	0.37	0.34	0.30	0.29	0.30	0.31	0.40
3	30.961	35.697	35.991	35.267	35.147	35.526	31.420	0.23	0.11	0.06	0.04	0.08	0.07	0.20
4	30.635	32.766	35.257	41.610	33.161	35.698	30.735	0.22	0.08	0.01	-0.04	0.14	0.05	0.21
5	32.874	34.760	39.113	41.995	38.229	38.524	33.261	0.19	0.14	0.07	0.09	0.14	0.11	0.15
6	32.352	32.027	31.297	34.962	34.526	33.203	31.941	0.10	0.11	0.22	0.17	0.20	0.18	0.11
7	33.661	42.022	40.120	39.274	35.602	39.254	33.492	0.18	0.04	0.08	0.08	0.19	0.10	0.16
8	32.652	33.079	33.313	33.199	32.915	33.126	33.040	0.30	0.27	0.28	0.28	0.28	0.28	0.30
9	35.866	39.059	35.370	35.642	35.909	36.495	35.694	0.06	0.05	0.17	0.16	0.15	0.13	0.09
10	35.179	35.982	35.357	35.675	36.474	35.872	35.079	0.15	0.10	0.09	0.10	0.08	0.09	0.11
11	28.791	29.987	31.155	29.435	29.460	30.009	28.386	0.14	0.24	0.20	0.27	0.28	0.25	0.20
12	36.036	35.703	34.406	34.152	35.656	34.979	36.595	0.15	0.13	0.18	0.22	0.19	0.18	0.15
13	33.006	33.292	38.065	32.960	32.381	34.175	33.148	0.22	0.16	0.10	0.22	0.28	0.19	0.25
14	32.254	31.851	31.941	32.576	32.871	32.310	31.382	0.23	0.27	0.25	0.27	0.23	0.25	0.25
15	34.840	37.685	35.448	36.005	36.947	36.521	35.425	0.22	0.12	0.16	0.19	0.13	0.15	0.20
16	29.260	31.136	33.302	32.206	30.612	31.814	29.362	0.18	0.11	0.04	0.09	0.15	0.10	0.18
17	35.120	36.440	35.782	37.721	36.028	36.493	35.062	0.20	0.12	0.17	0.08	0.15	0.13	0.20
18	32.973	34.664	32.709	32.429	33.874	33.419	32.320	0.22	0.12	0.25	0.27	0.25	0.22	0.26
19	30.356	31.659	31.553	33.629	32.883	32.431	31.216	0.23	0.23	0.28	0.18	0.22	0.23	0.24
20	31.713	31.659	31.034	32.793	32.327	31.953	31.773	0.23	0.31	0.34	0.28	0.29	0.30	0.24
Mean	**32.428**	**34.015**	**34.194**	**34.755**	**33.888**	**34.213**	**32.445**	**0.20**	**0.16**	**0.18**	**0.17**	**0.20**	**0.18**	**0.21**
SD	*2.318*	*3.192*	*2.878*	*3.348*	*2.493*	*2.651*	*2.378*	*0.06*	*0.08*	*0.09*	*0.09*	*0.07*	*0.08*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C45

Dataset 3: Predictive performance statistics for the prediction equations developed using the combination of Agreeableness and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.035	1.045	1.052	1.039	1.043	1.044	1.033	0.18	0.16	0.12	0.20	0.17	0.16	0.18
2	1.087	1.122	1.123	1.116	1.118	1.120	1.089	0.25	0.12	0.12	0.17	0.17	0.14	0.25
3	1.097	1.114	1.127	1.116	1.120	1.119	1.101	0.15	0.14	0.11	0.14	0.11	0.12	0.17
4	1.014	0.990	1.002	0.994	1.007	0.998	1.002	0.30	0.28	0.26	0.28	0.23	0.26	0.33
5	1.077	1.080	1.081	1.080	1.078	1.080	1.073	0.03	0.05	0.06	0.06	0.07	0.06	0.06
6	0.958	0.980	0.968	0.955	0.956	0.965	0.955	0.13	0.11	0.15	0.17	0.18	0.15	0.14
7	1.103	1.101	1.094	1.100	1.096	1.098	1.101	0.26	0.25	0.26	0.25	0.26	0.25	0.26
8	1.073	1.087	1.076	1.079	1.079	1.080	1.079	0.12	0.08	0.11	0.09	0.08	0.09	0.12
9	1.049	1.063	1.062	1.063	1.063	1.063	1.054	0.10	0.07	0.07	0.03	0.03	0.05	0.08
10	1.052	1.055	1.056	1.054	1.050	1.054	1.056	0.18	0.12	0.14	0.14	0.16	0.14	0.16
11	1.029	1.025	1.023	1.022	1.017	1.022	1.026	0.15	0.15	0.14	0.16	0.15	0.15	0.16
12	1.056	1.054	1.052	1.057	1.060	1.056	1.059	0.11	0.12	0.15	0.16	0.15	0.15	0.12
13	1.130	1.142	1.142	1.137	1.142	1.141	1.129	0.15	0.11	0.09	0.06	0.08	0.09	0.14
14	1.129	1.145	1.166	1.149	1.149	1.152	1.124	0.06	0.09	0.04	0.06	0.05	0.06	0.07
15	1.130	1.132	1.149	1.134	1.156	1.143	1.135	0.12	0.11	0.07	0.09	0.06	0.08	0.12
16	1.021	1.111	1.126	1.120	1.118	1.119	1.059	0.17	-0.03	-0.05	-0.05	-0.05	-0.04	0.05
17	0.962	0.995	1.105	1.054	1.055	1.052	0.957	0.15	0.10	0.03	0.05	0.04	0.06	0.15
18	1.068	1.095	1.084	1.097	1.084	1.090	1.075	0.17	0.11	0.16	0.12	0.13	0.13	0.14
19	1.085	1.101	1.096	1.110	1.110	1.104	1.071	0.12	0.03	0.03	0.02	0.01	0.02	0.09
20	1.131	1.140	1.176	1.159	1.166	1.160	1.161	0.00	-0.02	-0.05	-0.05	-0.05	-0.04	-0.01
Mean	**1.064**	**1.079**	**1.088**	**1.082**	**1.083**	**1.083**	**1.067**	**0.15**	**0.11**	**0.10**	**0.11**	**0.10**	**0.10**	**0.14**
SD	*0.050*	*0.050*	*0.053*	*0.052*	*0.053*	*0.051*	*0.053*	*0.07*	*0.07*	*0.08*	*0.09*	*0.08*	*0.08*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C46

Dataset 3: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.044	1.054	1.042	1.045	1.043	1.046	1.046	0.15	0.14	0.16	0.13	0.16	0.15	0.16
2	1.084	1.133	1.140	1.126	1.124	1.131	1.094	0.25	0.08	0.07	0.12	0.12	0.10	0.20
3	1.099	1.146	1.123	1.132	1.133	1.133	1.103	0.14	0.03	0.08	0.06	0.06	0.06	0.13
4	1.012	1.017	1.013	1.059	1.064	1.038	1.022	0.31	0.23	0.13	0.06	0.10	0.13	0.29
5	1.076	1.078	1.085	1.096	1.084	1.086	1.073	0.03	0.05	0.05	0.07	0.07	0.06	0.05
6	0.958	0.971	0.955	0.975	0.987	0.972	0.961	0.13	0.11	0.16	0.14	0.12	0.13	0.13
7	1.105	1.095	1.096	1.099	1.107	1.099	1.100	0.25	0.26	0.25	0.28	0.20	0.25	0.25
8	1.073	1.073	1.062	1.069	1.061	1.066	1.076	0.12	0.12	0.14	0.12	0.15	0.13	0.11
9	1.049	1.064	1.064	1.062	1.056	1.061	1.045	0.09	0.08	0.07	0.04	0.07	0.06	0.07
10	1.052	1.051	1.037	1.046	1.043	1.044	1.050	0.17	0.16	0.20	0.17	0.17	0.18	0.18
11	1.029	1.038	1.031	1.038	1.037	1.036	1.025	0.15	0.11	0.12	0.14	0.14	0.13	0.15
12	1.068	1.066	1.061	1.069	1.061	1.064	1.069	0.06	0.09	0.13	0.13	0.17	0.13	0.09
13	1.132	1.137	1.207	1.144	1.171	1.165	1.126	0.11	0.10	0.06	0.10	0.05	0.07	0.12
14	1.142	1.178	1.167	1.173	1.157	1.169	1.149	0.01	0.01	0.03	0.02	0.04	0.03	0.02
15	1.154	1.165	1.182	1.185	1.198	1.183	1.159	0.05	0.06	0.05	0.06	0.04	0.05	0.06
16	1.023	1.114	1.119	1.126	1.128	1.122	1.038	0.15	0.06	-0.01	0.01	-0.02	0.01	0.06
17	0.964	1.009	0.997	0.988	0.993	0.997	0.979	0.14	0.06	0.09	0.07	0.09	0.08	0.08
18	1.068	1.090	1.090	1.080	1.105	1.091	1.074	0.17	0.12	0.10	0.14	0.06	0.11	0.13
19	1.084	1.120	1.097	1.120	1.124	1.115	1.068	0.12	0.01	0.07	0.04	0.02	0.03	0.11
20	1.137	1.186	1.161	1.172	1.176	1.174	1.166	-0.01	-0.01	-0.02	-0.04	-0.03	-0.02	-0.03
Mean	**1.068**	**1.089**	**1.086**	**1.090**	**1.093**	**1.090**	**1.071**	**0.13**	**0.09**	**0.10**	**0.09**	**0.09**	**0.09**	**0.12**
SD	*0.053*	*0.057*	*0.064*	*0.057*	*0.058*	*0.058*	*0.053*	*0.08*	*0.07*	*0.06*	*0.07*	*0.06*	*0.06*	*0.07*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C47

Dataset 3: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	1.040	1.042	1.045	1.042	1.032	1.040	1.033	0.18	0.18	0.14	0.19	0.20	0.17	0.19
2	1.087	1.163	1.180	1.199	1.199	1.185	1.146	0.13	0.02	0.02	0.03	0.01	0.02	0.05
3	1.099	1.149	1.223	1.173	1.181	1.182	1.113	0.17	0.01	0.03	0.02	0.02	0.02	0.11
4	1.015	1.053	1.091	1.061	1.063	1.067	1.021	0.20	0.08	0.04	0.05	0.03	0.05	0.13
5	1.082	1.085	1.112	1.099	1.101	1.099	1.086	0.07	0.06	0.04	0.05	0.05	0.05	0.06
6	0.950	0.963	0.990	0.953	1.001	0.977	0.952	0.16	0.19	0.14	0.18	0.16	0.17	0.17
7	1.101	1.105	1.106	1.127	1.124	1.115	1.113	0.26	0.21	0.22	0.19	0.20	0.20	0.20
8	1.069	1.071	1.099	1.087	1.110	1.092	1.057	0.15	0.15	0.08	0.06	0.06	0.09	0.16
9	1.078	1.086	1.088	1.084	1.096	1.089	1.077	0.02	-0.02	-0.03	0.01	-0.05	-0.02	0.02
10	1.054	1.051	1.039	1.046	1.061	1.049	1.063	0.16	0.16	0.22	0.17	0.13	0.17	0.12
11	1.068	1.105	1.091	1.088	1.109	1.098	1.071	-0.01	0.01	0.04	0.05	0.03	0.03	0.00
12	1.075	1.085	1.098	1.103	1.107	1.098	1.084	0.10	0.07	0.07	0.07	0.06	0.07	0.08
13	1.122	1.143	1.126	1.156	1.142	1.142	1.124	0.14	0.11	0.16	0.04	0.14	0.11	0.12
14	1.143	1.172	1.186	1.184	1.197	1.185	1.160	0.02	0.01	-0.01	-0.03	-0.03	-0.02	0.00
15	1.155	1.174	1.179	1.196	1.242	1.198	1.155	0.00	-0.01	0.04	0.02	-0.01	0.01	0.00
16	1.062	1.119	1.161	1.143	1.128	1.138	1.076	0.06	0.04	-0.01	-0.02	-0.03	0.00	0.02
17	0.986	1.026	1.043	1.049	1.067	1.046	1.023	0.09	0.07	0.00	-0.02	0.00	0.01	0.04
18	1.078	1.090	1.100	1.098	1.101	1.097	1.096	0.15	0.10	0.08	0.08	0.07	0.08	0.11
19	1.075	1.079	1.167	1.157	1.175	1.145	1.084	0.13	0.06	-0.01	0.01	0.01	0.02	0.09
20	1.135	1.207	1.218	1.205	1.270	1.225	1.156	0.02	-0.02	-0.02	-0.02	0.00	-0.01	0.01
Mean	**1.074**	**1.098**	**1.117**	**1.113**	**1.125**	**1.113**	**1.085**	**0.11**	**0.07**	**0.06**	**0.06**	**0.05**	**0.06**	**0.08**
SD	*0.049*	*0.057*	*0.061*	*0.064*	*0.067*	*0.061*	*0.051*	*0.07*	*0.07*	*0.08*	*0.07*	*0.07*	*0.07*	*0.06*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C48

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of Openness and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.477	0.486	0.491	0.493	0.483	0.488	0.477	0.11	0.09	0.09	0.07	0.11	0.09	0.10
2	0.441	0.445	0.446	0.452	0.451	0.448	0.446	0.16	0.15	0.18	0.18	0.17	0.17	0.15
3	0.546	0.551	0.546	0.549	0.548	0.548	0.541	0.09	0.05	0.06	0.05	0.05	0.05	0.08
4	0.496	0.492	0.494	0.495	0.496	0.494	0.497	0.28	0.29	0.29	0.26	0.24	0.27	0.28
5	0.482	0.489	0.491	0.516	0.497	0.499	0.485	0.26	0.25	0.25	0.20	0.22	0.23	0.24
6	0.460	0.467	0.461	0.496	0.490	0.478	0.457	0.26	0.23	0.25	0.10	0.08	0.16	0.25
7	0.509	0.521	0.533	0.528	0.531	0.528	0.509	0.19	0.19	0.21	0.17	0.20	0.19	0.19
8	0.534	0.534	0.530	0.533	0.533	0.533	0.534	0.30	0.30	0.33	0.30	0.30	0.31	0.31
9	0.580	0.583	0.584	0.581	0.582	0.582	0.576	0.25	0.24	0.25	0.27	0.29	0.26	0.27
10	0.449	0.448	0.462	0.463	0.461	0.458	0.444	0.20	0.19	0.12	0.13	0.12	0.14	0.17
11	0.537	0.552	0.564	0.557	0.553	0.557	0.536	0.31	0.23	0.12	0.17	0.19	0.18	0.29
12	0.448	0.461	0.455	0.455	0.453	0.456	0.452	0.31	0.28	0.28	0.28	0.29	0.28	0.29
13	0.465	0.453	0.457	0.455	0.455	0.455	0.459	0.24	0.22	0.18	0.20	0.21	0.20	0.25
14	0.502	0.498	0.496	0.497	0.496	0.497	0.506	0.26	0.25	0.27	0.27	0.27	0.27	0.26
15	0.502	0.504	0.490	0.488	0.493	0.494	0.493	0.33	0.31	0.27	0.27	0.26	0.28	0.33
16	0.435	0.442	0.448	0.439	0.461	0.448	0.428	0.30	0.20	0.19	0.20	0.16	0.19	0.28
17	0.450	0.454	0.519	0.508	0.470	0.488	0.449	0.20	0.13	0.00	0.05	0.10	0.07	0.18
18	0.442	0.460	0.475	0.475	0.480	0.472	0.456	0.23	0.15	0.10	0.09	0.08	0.11	0.17
19	0.409	0.413	0.407	0.458	0.469	0.437	0.415	0.38	0.37	0.41	0.19	0.17	0.28	0.37
20	0.483	0.485	0.482	0.483	0.482	0.483	0.484	0.35	0.34	0.36	0.35	0.34	0.35	0.34
Mean	**0.482**	**0.487**	**0.492**	**0.496**	**0.494**	**0.492**	**0.482**	**0.25**	**0.22**	**0.21**	**0.19**	**0.19**	**0.20**	**0.24**
SD	*0.042*	*0.042*	*0.043*	*0.038*	*0.036*	*0.039*	*0.041*	*0.07*	*0.08*	*0.10*	*0.09*	*0.08*	*0.08*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C49

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.472	0.509	0.515	0.491	0.496	0.503	0.466	0.16	0.21	0.20	0.19	0.19	0.20	0.19
2	0.444	0.446	0.454	0.461	0.467	0.457	0.437	0.19	0.22	0.19	0.20	0.18	0.20	0.21
3	0.549	0.536	0.541	0.529	0.540	0.536	0.536	0.06	0.16	0.16	0.20	0.14	0.17	0.14
4	0.494	0.481	0.493	0.492	0.490	0.489	0.490	0.33	0.37	0.29	0.29	0.29	0.31	0.37
5	0.487	0.498	0.514	0.518	0.486	0.504	0.484	0.30	0.24	0.26	0.26	0.29	0.27	0.31
6	0.490	0.516	0.547	0.553	0.565	0.545	0.503	0.16	0.10	0.04	0.12	0.07	0.08	0.12
7	0.505	0.514	0.513	0.509	0.504	0.510	0.501	0.21	0.19	0.20	0.22	0.23	0.21	0.22
8	0.533	0.524	0.518	0.512	0.514	0.517	0.531	0.33	0.40	0.40	0.41	0.42	0.41	0.35
9	0.579	0.563	0.564	0.561	0.561	0.562	0.574	0.28	0.33	0.33	0.36	0.34	0.34	0.32
10	0.472	0.474	0.492	0.488	0.488	0.485	0.468	0.19	0.14	0.11	0.11	0.12	0.12	0.16
11	0.537	0.564	0.610	0.553	0.564	0.573	0.544	0.31	0.09	-0.08	0.19	0.15	0.09	0.31
12	0.445	0.461	0.467	0.476	0.485	0.472	0.446	0.35	0.25	0.12	0.15	0.10	0.15	0.37
13	0.476	0.462	0.481	0.475	0.471	0.472	0.466	0.22	0.27	0.16	0.19	0.20	0.20	0.26
14	0.504	0.509	0.503	0.503	0.506	0.505	0.510	0.28	0.26	0.28	0.28	0.27	0.27	0.30
15	0.500	0.487	0.488	0.485	0.497	0.489	0.489	0.39	0.36	0.42	0.42	0.37	0.39	0.42
16	0.432	0.437	0.433	0.444	0.457	0.443	0.423	0.35	0.28	0.31	0.26	0.26	0.27	0.32
17	0.446	0.446	0.498	0.523	0.504	0.493	0.441	0.25	0.16	0.09	0.01	0.04	0.07	0.20
18	0.444	0.453	0.459	0.495	0.485	0.473	0.447	0.27	0.19	0.18	0.16	0.22	0.19	0.25
19	0.459	0.459	0.482	0.515	0.465	0.480	0.451	0.17	0.21	0.17	0.11	0.20	0.17	0.18
20	0.500	0.519	0.514	0.527	0.517	0.519	0.495	0.26	0.16	0.20	0.16	0.21	0.18	0.28
Mean	**0.488**	**0.493**	**0.504**	**0.506**	**0.503**	**0.501**	**0.485**	**0.25**	**0.23**	**0.20**	**0.21**	**0.21**	**0.21**	**0.26**
SD	*0.038*	*0.037*	*0.039*	*0.030*	*0.032*	*0.033*	*0.039*	*0.08*	*0.08*	*0.12*	*0.10*	*0.10*	*0.09*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C50

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of Extraversion, Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.474	0.509	0.535	0.540	0.522	0.527	0.468	0.16	0.14	0.12	0.10	0.10	0.11	0.17
2	0.445	0.450	0.447	0.463	0.478	0.460	0.442	0.20	0.18	0.22	0.16	0.12	0.17	0.21
3	0.551	0.567	0.620	0.568	0.581	0.584	0.535	0.06	0.08	0.02	0.14	0.09	0.08	0.19
4	0.496	0.486	0.616	0.595	0.544	0.560	0.496	0.32	0.35	0.17	0.18	0.18	0.22	0.31
5	0.487	0.507	0.538	0.523	0.507	0.519	0.481	0.30	0.32	0.29	0.29	0.31	0.30	0.31
6	0.496	0.506	0.557	0.552	0.601	0.554	0.492	0.15	0.11	-0.09	-0.12	0.00	-0.03	0.16
7	0.518	0.601	0.578	0.612	0.592	0.596	0.520	0.16	0.18	0.23	0.19	0.19	0.20	0.16
8	0.533	0.522	0.558	0.517	0.525	0.531	0.526	0.33	0.43	0.25	0.44	0.37	0.37	0.39
9	0.579	0.584	0.554	0.559	0.553	0.562	0.574	0.28	0.26	0.41	0.40	0.41	0.37	0.36
10	0.473	0.494	0.530	0.511	0.534	0.517	0.466	0.19	0.12	0.11	0.06	0.05	0.08	0.17
11	0.537	0.563	0.679	0.625	0.646	0.628	0.540	0.31	0.06	-0.13	-0.05	0.09	-0.01	0.28
12	0.446	0.461	0.513	0.531	0.503	0.502	0.444	0.36	0.25	0.14	0.10	0.13	0.16	0.37
13	0.480	0.549	0.625	0.561	0.677	0.603	0.463	0.21	0.17	0.02	0.09	-0.01	0.07	0.26
14	0.504	0.505	0.521	0.514	0.509	0.512	0.501	0.28	0.25	0.24	0.27	0.29	0.26	0.29
15	0.505	0.483	0.480	0.506	0.527	0.499	0.492	0.35	0.38	0.43	0.25	0.16	0.30	0.38
16	0.440	0.466	0.518	0.472	0.458	0.479	0.432	0.33	0.26	0.14	0.30	0.34	0.26	0.30
17	0.446	0.447	0.444	0.471	0.473	0.459	0.432	0.25	0.19	0.23	0.20	0.17	0.20	0.24
18	0.451	0.471	0.505	0.535	0.539	0.512	0.454	0.23	0.16	0.15	0.11	0.10	0.13	0.24
19	0.458	0.493	0.526	0.491	0.494	0.501	0.480	0.18	0.17	0.21	0.28	0.26	0.23	0.12
20	0.502	0.523	0.513	0.503	0.516	0.514	0.498	0.26	0.16	0.17	0.27	0.21	0.20	0.23
Mean	**0.491**	**0.509**	**0.543**	**0.532**	**0.539**	**0.531**	**0.487**	**0.25**	**0.21**	**0.17**	**0.18**	**0.18**	**0.18**	**0.26**
SD	*0.038*	*0.043*	*0.057*	*0.044*	*0.055*	*0.045*	*0.038*	*0.08*	*0.10*	*0.14*	*0.13*	*0.12*	*0.11*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C51

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.473	0.521	0.499	0.585	0.607	0.553	0.471	0.14	0.05	0.18	0.16	0.06	0.11	0.16
2	0.442	0.467	0.497	0.532	0.542	0.509	0.437	0.21	0.17	0.17	0.18	0.14	0.17	0.19
3	0.546	0.582	0.584	0.579	0.569	0.578	0.552	0.09	0.01	0.15	0.16	0.16	0.12	0.14
4	0.490	0.503	0.540	0.544	0.740	0.582	0.489	0.35	0.27	0.14	0.17	0.18	0.19	0.34
5	0.479	0.484	0.555	0.686	0.522	0.562	0.471	0.31	0.35	0.15	0.08	0.26	0.21	0.34
6	0.494	0.488	0.588	0.608	0.564	0.562	0.485	0.16	0.19	-0.01	0.11	0.10	0.10	0.19
7	0.537	0.560	0.616	0.622	0.609	0.602	0.522	0.10	0.16	0.12	0.21	0.13	0.16	0.17
8	0.528	0.542	0.529	0.707	0.537	0.579	0.524	0.36	0.25	0.36	0.18	0.26	0.26	0.39
9	0.580	0.579	0.559	0.533	0.583	0.564	0.575	0.31	0.31	0.35	0.43	0.32	0.35	0.37
10	0.474	0.492	0.536	0.522	0.544	0.523	0.467	0.22	0.14	-0.07	-0.01	-0.08	-0.01	0.18
11	0.542	0.581	0.603	0.597	0.614	0.599	0.562	0.28	0.05	0.10	0.05	0.10	0.08	0.25
12	0.440	0.463	0.548	0.501	0.456	0.492	0.433	0.39	0.24	-0.13	0.22	0.28	0.15	0.43
13	0.473	0.505	0.548	0.612	0.544	0.552	0.462	0.24	0.11	0.16	0.10	0.22	0.15	0.29
14	0.503	0.506	0.632	0.620	0.572	0.582	0.498	0.28	0.27	0.15	0.04	0.10	0.14	0.31
15	0.502	0.492	0.484	0.519	0.490	0.496	0.497	0.33	0.32	0.41	0.35	0.38	0.36	0.34
16	0.435	0.470	0.471	0.453	0.515	0.477	0.432	0.32	0.24	0.28	0.36	0.26	0.29	0.32
17	0.443	0.443	0.557	0.442	0.450	0.473	0.440	0.25	0.22	0.07	0.27	0.25	0.20	0.27
18	0.443	0.459	0.545	0.509	0.515	0.507	0.444	0.25	0.19	0.07	0.18	0.16	0.15	0.26
19	0.461	0.520	0.533	0.587	0.543	0.546	0.448	0.17	0.01	0.19	0.11	0.14	0.11	0.18
20	0.504	0.529	0.500	0.559	0.602	0.548	0.502	0.22	0.13	0.24	0.18	0.06	0.15	0.23
Mean	**0.490**	**0.509**	**0.546**	**0.566**	**0.556**	**0.544**	**0.486**	**0.25**	**0.18**	**0.15**	**0.18**	**0.17**	**0.17**	**0.27**
SD	*0.040*	*0.041*	*0.042*	*0.067*	*0.062*	*0.039*	*0.043*	*0.08*	*0.10*	*0.13*	*0.11*	*0.10*	*0.09*	*0.08*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C52

Dataset 5: Predictive performance statistics for the prediction equations developed using the combination of Ambition and Prudence (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.684	6.787	6.794	6.758	6.712	6.763	6.693	0.17	0.09	0.09	0.10	0.13	0.10	0.14
2	6.184	6.195	6.190	6.198	6.222	6.201	6.195	0.11	0.12	0.13	0.10	0.10	0.11	0.11
3	6.144	6.165	6.173	6.167	6.181	6.172	6.159	0.15	0.16	0.15	0.15	0.14	0.15	0.14
4	6.212	6.230	6.229	6.206	6.215	6.220	6.158	0.12	0.16	0.15	0.15	0.15	0.15	0.14
5	5.929	5.897	5.887	5.907	5.904	5.899	5.931	0.11	0.15	0.16	0.15	0.15	0.15	0.13
6	6.252	6.318	6.291	6.252	6.271	6.283	6.260	0.17	0.16	0.16	0.17	0.16	0.16	0.17
7	5.872	5.964	5.976	5.974	5.963	5.969	5.876	0.08	0.03	0.03	0.04	0.03	0.03	0.06
8	6.276	6.265	6.253	6.247	6.258	6.256	6.302	0.08	0.13	0.12	0.13	0.12	0.12	0.08
9	6.373	6.375	6.361	6.391	6.377	6.376	6.330	0.08	0.11	0.12	0.10	0.11	0.11	0.13
10	5.669	5.696	5.714	5.714	5.698	5.706	5.658	0.17	0.14	0.13	0.13	0.13	0.13	0.16
11	5.721	5.762	5.758	5.782	5.780	5.770	5.740	0.17	0.16	0.17	0.14	0.14	0.15	0.18
12	5.918	5.962	5.963	5.918	5.941	5.946	5.857	0.06	0.08	0.08	0.09	0.09	0.09	0.10
13	5.637	5.686	5.726	5.714	5.709	5.708	5.658	0.09	0.10	0.10	0.10	0.10	0.10	0.09
14	5.747	5.831	5.843	5.828	5.824	5.831	5.715	0.16	0.14	0.13	0.14	0.14	0.14	0.14
15	6.174	6.172	6.180	6.163	6.167	6.171	6.166	0.10	0.13	0.13	0.14	0.14	0.13	0.12
16	5.961	5.996	6.017	5.991	6.036	6.010	5.979	0.15	0.15	0.14	0.15	0.12	0.14	0.17
17	6.420	6.443	6.463	6.454	6.455	6.454	6.437	0.19	0.21	0.19	0.19	0.20	0.20	0.19
18	6.395	6.484	6.476	6.465	6.447	6.468	6.395	0.05	0.06	0.06	0.05	0.05	0.05	0.06
19	6.083	6.119	6.117	6.104	6.109	6.112	6.101	0.14	0.11	0.11	0.12	0.12	0.11	0.13
20	6.519	6.484	6.518	6.548	6.490	6.510	6.525	0.10	0.15	0.12	0.11	0.13	0.13	0.12
Mean	**6.108**	**6.142**	**6.146**	**6.139**	**6.138**	**6.141**	**6.107**	**0.12**	**0.13**	**0.12**	**0.12**	**0.12**	**0.12**	**0.13**
SD	*0.288*	*0.287*	*0.284*	*0.282*	*0.273*	*0.281*	*0.291*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C53

Dataset 5: Predictive performance statistics for the prediction equations developed using the combination of Adjustment, Ambition, and Prudence (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.722	6.777	6.785	6.892	6.829	6.821	6.695	0.14	0.08	0.09	0.03	0.06	0.06	0.13
2	6.167	6.143	6.174	6.143	6.171	6.158	6.158	0.11	0.13	0.09	0.11	0.11	0.11	0.12
3	6.133	6.156	6.176	6.181	6.165	6.170	6.166	0.15	0.14	0.12	0.12	0.14	0.13	0.14
4	6.197	6.193	6.153	6.174	6.241	6.190	6.285	0.12	0.17	0.17	0.16	0.15	0.16	0.13
5	5.942	5.915	5.905	5.913	5.894	5.907	5.914	0.08	0.13	0.14	0.12	0.14	0.13	0.10
6	6.241	6.272	6.263	6.339	6.383	6.314	6.228	0.17	0.15	0.14	0.12	0.12	0.13	0.17
7	5.870	6.013	5.958	5.912	5.948	5.958	5.922	0.08	-0.01	0.03	0.04	0.03	0.02	0.06
8	6.331	6.327	6.243	6.378	6.401	6.337	6.254	0.05	0.05	0.08	0.06	0.04	0.06	0.07
9	6.373	6.363	6.341	6.414	6.391	6.377	6.379	0.08	0.11	0.11	0.06	0.08	0.09	0.09
10	5.660	5.724	5.724	5.733	5.741	5.730	5.661	0.15	0.15	0.15	0.13	0.13	0.14	0.16
11	5.775	5.844	5.864	5.828	5.842	5.845	5.817	0.09	0.08	0.07	0.10	0.09	0.08	0.06
12	5.926	6.111	6.075	6.152	6.192	6.132	5.937	0.05	-0.04	-0.02	-0.03	-0.02	-0.03	0.04
13	5.729	5.802	5.821	5.868	5.863	5.839	5.742	0.04	0.03	0.03	0.00	0.03	0.02	0.02
14	5.741	5.863	5.833	5.857	5.906	5.865	5.819	0.14	0.12	0.13	0.10	0.10	0.11	0.15
15	6.167	6.191	6.193	6.215	6.192	6.198	6.174	0.09	0.10	0.09	0.10	0.10	0.10	0.11
16	6.010	6.079	6.109	6.092	6.090	6.093	6.029	0.12	0.11	0.08	0.11	0.10	0.10	0.14
17	6.445	6.482	6.504	6.461	6.468	6.479	6.463	0.17	0.14	0.10	0.18	0.14	0.14	0.16
18	6.461	6.583	6.562	6.538	6.578	6.565	6.452	0.02	0.01	0.01	0.01	0.01	0.01	0.02
19	6.059	6.046	6.093	6.069	6.073	6.070	6.054	0.13	0.13	0.09	0.10	0.10	0.11	0.13
20	6.521	6.499	6.559	6.548	6.518	6.531	6.541	0.10	0.14	0.11	0.11	0.13	0.12	0.11
Mean	**6.124**	**6.169**	**6.167**	**6.185**	**6.194**	**6.179**	**6.134**	**0.10**	**0.10**	**0.09**	**0.09**	**0.09**	**0.09**	**0.11**
SD	*0.287*	*0.271*	*0.274*	*0.288*	*0.278*	*0.277*	*0.276*	*0.04*	*0.06*	*0.05*	*0.05*	*0.05*	*0.05*	*0.04*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C54

Dataset 5: Predictive performance statistics for the prediction equations developed using the combination of Adjustment, Ambition, Likeability, and Prudence (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.744	6.732	6.802	6.762	6.737	6.758	6.685	0.12	0.12	0.08	0.13	0.12	0.11	0.13
2	6.227	6.241	6.200	6.169	6.274	6.221	6.171	0.09	0.06	0.12	0.13	0.11	0.11	0.10
3	6.127	6.207	6.259	6.226	6.210	6.226	6.160	0.15	0.11	0.09	0.11	0.12	0.11	0.16
4	6.191	6.168	6.225	6.197	6.204	6.199	6.167	0.12	0.16	0.16	0.16	0.19	0.17	0.17
5	5.942	5.975	6.054	6.037	6.095	6.040	5.922	0.08	0.11	0.10	0.11	0.12	0.11	0.11
6	6.253	6.281	6.275	6.462	6.483	6.375	6.274	0.17	0.16	0.18	0.13	0.13	0.15	0.17
7	5.879	5.985	5.992	6.059	6.030	6.016	5.874	0.07	0.03	0.05	0.06	0.06	0.05	0.08
8	6.329	6.280	6.325	6.496	6.446	6.387	6.308	0.05	0.10	0.09	0.01	0.06	0.06	0.09
9	6.404	6.404	6.413	6.381	6.377	6.394	6.330	0.05	0.05	0.07	0.07	0.08	0.07	0.07
10	5.655	5.755	5.736	5.796	5.768	5.764	5.653	0.16	0.15	0.16	0.17	0.16	0.16	0.15
11	5.787	5.842	5.814	5.811	5.959	5.857	5.796	0.07	0.03	0.07	0.08	0.03	0.05	0.09
12	5.934	6.051	5.991	6.056	6.061	6.040	5.916	0.05	0.06	0.06	0.01	0.01	0.03	0.06
13	5.725	5.735	5.751	5.739	5.753	5.744	5.742	0.04	0.04	0.05	0.05	0.05	0.05	0.05
14	5.791	6.295	6.291	6.550	6.344	6.370	5.956	0.12	0.07	0.09	0.04	0.04	0.06	0.08
15	6.237	6.424	6.481	6.378	6.410	6.423	6.313	0.04	0.02	0.02	0.05	0.05	0.04	0.04
16	6.048	6.028	6.108	6.140	6.051	6.082	6.023	0.11	0.13	0.08	0.07	0.11	0.10	0.12
17	6.456	6.543	6.536	6.532	6.558	6.542	6.406	0.14	0.11	0.13	0.09	0.08	0.10	0.21
18	6.527	6.660	6.681	6.688	6.695	6.681	6.637	0.00	0.04	0.01	0.04	0.06	0.04	0.01
19	6.055	6.034	6.096	6.280	6.072	6.120	6.000	0.13	0.13	0.10	0.13	0.14	0.13	0.15
20	6.555	6.569	6.798	6.615	6.681	6.666	6.591	0.08	0.10	-0.01	0.09	0.09	0.07	0.08
Mean	**6.143**	**6.210**	**6.241**	**6.269**	**6.260**	**6.245**	**6.146**	**0.09**	**0.09**	**0.08**	**0.09**	**0.09**	**0.09**	**0.11**
SD	*0.295*	*0.281*	*0.307*	*0.291*	*0.283*	*0.286*	*0.290*	*0.05*	*0.04*	*0.05*	*0.04*	*0.05*	*0.04*	*0.05*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C55

Dataset 5: Predictive performance statistics for the prediction equations developed using the combination of Adjustment, Ambition, Intellectance, Likeability, and Prudence (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.735	6.806	6.773	6.681	6.785	6.761	6.714	0.13	0.09	0.11	0.15	0.13	0.12	0.12
2	6.316	6.249	6.220	6.265	6.360	6.274	6.222	0.06	0.10	0.08	0.11	0.09	0.09	0.10
3	6.117	6.167	6.173	6.142	6.186	6.167	6.158	0.15	0.13	0.15	0.17	0.16	0.15	0.15
4	6.193	6.419	6.224	6.304	6.357	6.326	6.273	0.12	0.14	0.14	0.18	0.17	0.16	0.13
5	5.994	6.369	6.506	6.309	6.335	6.380	6.102	0.04	-0.03	-0.09	-0.04	-0.02	-0.04	0.02
6	6.332	6.592	6.657	6.873	6.848	6.742	6.350	0.13	0.02	0.06	0.00	0.02	0.02	0.12
7	5.880	5.976	6.097	6.159	6.048	6.070	5.948	0.07	0.07	0.04	0.08	0.06	0.06	0.07
8	6.336	6.378	6.342	6.445	6.595	6.440	6.357	0.05	0.03	0.07	0.02	0.01	0.03	0.06
9	6.421	6.362	6.367	6.400	6.395	6.381	6.381	0.04	0.06	0.05	0.04	0.05	0.05	0.05
10	5.663	5.774	5.673	5.708	5.827	5.745	5.630	0.16	0.12	0.15	0.14	0.11	0.13	0.17
11	5.786	5.833	5.821	5.890	5.858	5.850	5.782	0.08	0.08	0.10	0.06	0.07	0.08	0.08
12	5.948	5.979	6.022	6.147	6.112	6.065	5.946	0.05	0.03	0.06	-0.02	0.04	0.03	0.05
13	5.736	5.941	5.719	5.844	5.922	5.857	5.736	0.04	0.04	0.08	0.09	0.05	0.06	0.05
14	5.923	6.393	6.418	6.655	6.789	6.564	6.082	0.05	-0.01	-0.01	-0.02	-0.02	-0.02	0.02
15	6.247	6.414	6.385	6.545	6.395	6.435	6.285	0.04	0.04	0.06	0.00	0.03	0.03	0.05
16	6.054	6.254	6.138	6.069	6.090	6.138	6.082	0.11	0.02	0.09	0.13	0.10	0.09	0.11
17	6.463	6.461	6.505	6.448	6.478	6.473	6.432	0.12	0.17	0.14	0.13	0.12	0.14	0.12
18	6.529	6.584	6.586	6.536	6.590	6.574	6.458	0.00	0.03	0.02	0.06	0.06	0.04	0.04
19	6.055	6.126	6.146	6.144	6.255	6.168	6.006	0.13	0.12	0.10	0.14	0.15	0.13	0.16
20	6.545	6.659	6.820	6.860	6.832	6.793	6.543	0.09	0.07	0.05	0.07	0.11	0.08	0.10
Mean	**6.164**	**6.287**	**6.280**	**6.321**	**6.353**	**6.310**	**6.174**	**0.08**	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**	**0.09**
SD	*0.289*	*0.275*	*0.313*	*0.312*	*0.312*	*0.296*	*0.274*	*0.04*	*0.05*	*0.06*	*0.07*	*0.06*	*0.05*	*0.04*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C56

Dataset 5: Predictive performance statistics for the prediction equations developed using the combination of Adjustment, Ambition, Sociability, Intellectance, Likeability and Prudence (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	6.766	6.746	6.784	6.725	6.758	6.753	6.706	0.11	0.11	0.10	0.13	0.13	0.12	0.12
2	6.315	6.277	6.264	6.179	6.218	6.234	6.162	0.06	0.04	0.08	0.11	0.11	0.08	0.11
3	6.117	6.136	6.119	6.247	6.210	6.178	6.141	0.15	0.14	0.13	0.09	0.09	0.11	0.13
4	6.218	6.331	6.409	6.402	6.383	6.381	6.226	0.11	0.11	0.08	0.10	0.13	0.11	0.13
5	6.009	6.490	6.458	6.406	6.314	6.417	6.067	0.03	-0.08	-0.03	0.01	0.00	-0.03	0.03
6	6.386	6.444	6.603	7.137	6.744	6.732	6.510	0.09	0.07	0.06	-0.02	0.02	0.03	0.03
7	5.889	6.188	6.235	6.325	6.489	6.309	5.936	0.06	-0.06	-0.02	-0.04	-0.11	-0.06	0.03
8	6.358	6.404	6.393	6.466	6.521	6.446	6.303	0.04	0.05	0.04	0.05	0.02	0.04	0.07
9	6.482	6.445	6.269	6.379	6.375	6.367	6.414	0.00	0.02	0.08	0.04	0.05	0.05	0.04
10	5.665	5.798	5.820	5.988	6.038	5.911	5.674	0.16	0.09	0.11	0.06	0.11	0.09	0.15
11	5.790	5.908	5.966	5.917	5.863	5.914	5.775	0.07	0.02	-0.02	0.07	0.03	0.03	0.06
12	5.948	6.014	5.940	6.075	5.985	6.003	5.956	0.05	0.02	0.07	0.04	0.07	0.05	0.04
13	5.771	5.849	5.810	5.880	5.861	5.850	5.756	0.03	0.07	0.06	0.07	0.12	0.08	0.06
14	5.960	6.281	6.645	6.454	6.504	6.471	6.224	0.02	-0.04	-0.07	-0.06	-0.07	-0.06	-0.02
15	6.253	6.346	6.411	6.630	6.358	6.436	6.316	0.03	0.01	0.02	-0.02	0.06	0.02	0.03
16	6.065	6.010	6.185	6.067	6.046	6.077	6.085	0.10	0.14	0.03	0.13	0.15	0.11	0.10
17	6.478	6.646	6.653	6.608	6.583	6.622	6.481	0.11	0.04	0.09	0.07	0.07	0.07	0.10
18	6.558	6.646	6.556	6.772	6.646	6.655	6.544	-0.02	-0.01	-0.04	0.01	0.03	0.00	0.01
19	6.065	6.133	6.078	6.096	6.222	6.132	5.981	0.13	0.12	0.12	0.16	0.12	0.13	0.16
20	6.553	6.714	6.916	7.008	6.829	6.867	6.546	0.08	0.06	0.05	0.05	0.04	0.05	0.10
Mean	**6.182**	**6.290**	**6.326**	**6.388**	**6.347**	**6.338**	**6.190**	**0.07**	**0.05**	**0.05**	**0.05**	**0.06**	**0.05**	**0.07**
SD	*0.295*	*0.278*	*0.306*	*0.340*	*0.285*	*0.291*	*0.283*	*0.05*	*0.06*	*0.06*	*0.06*	*0.07*	*0.06*	*0.05*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C57

Dataset 6: Predictive performance statistics for the prediction equations developed using the combination of Emotional Orientation and Cognitive Orientation (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.601	0.600	0.563	0.577	0.585	0.581	0.600	-0.08	-0.06	0.07	0.02	-0.01	0.01	-0.06
2	0.624	0.614	0.607	0.606	0.612	0.610	0.619	-0.12	0.01	0.02	0.04	0.00	0.02	-0.07
3	0.578	0.586	0.608	0.567	0.549	0.578	0.587	0.27	0.14	0.08	0.29	0.37	0.22	0.26
4	0.599	0.599	0.545	0.565	0.557	0.566	0.597	-0.19	-0.15	0.10	0.15	0.20	0.08	-0.11
5	0.575	0.579	0.589	0.590	0.580	0.585	0.571	0.04	0.01	0.02	0.19	0.20	0.10	0.08
6	0.638	0.633	0.581	0.543	0.561	0.579	0.629	-0.21	-0.18	0.16	0.34	0.19	0.13	-0.12
7	0.588	0.638	0.551	0.569	0.597	0.589	0.573	0.31	0.16	0.29	0.29	0.29	0.26	0.39
8	0.580	0.593	0.603	0.546	0.555	0.574	0.570	-0.01	-0.08	-0.01	0.10	0.09	0.03	0.18
9	0.559	0.792	0.606	0.643	0.662	0.676	0.554	0.09	0.10	0.00	0.00	0.20	0.07	0.08
10	0.602	0.622	0.570	0.671	0.634	0.624	0.609	0.02	0.09	0.18	0.11	0.11	0.12	0.00
11	0.588	0.581	0.591	0.572	0.597	0.585	0.593	0.07	0.17	0.09	0.23	0.06	0.14	0.03
12	0.603	0.627	0.609	0.577	0.541	0.588	0.591	-0.07	0.05	-0.06	0.14	0.26	0.10	-0.01
13	0.607	0.600	0.596	0.577	0.577	0.587	0.586	0.13	0.15	0.25	0.31	0.29	0.25	0.20
14	0.667	0.660	0.599	0.574	0.596	0.607	0.646	-0.14	-0.10	0.32	0.40	0.31	0.23	-0.12
15	0.554	0.550	0.520	0.583	0.599	0.563	0.565	0.03	0.07	0.17	0.02	-0.02	0.06	-0.02
16	0.598	0.611	0.567	0.574	0.567	0.580	0.599	0.21	0.08	0.27	0.30	0.35	0.25	0.08
17	0.617	0.624	0.618	0.610	0.654	0.626	0.612	0.25	0.14	-0.01	0.10	-0.08	0.04	0.12
18	0.633	0.632	0.608	0.567	0.565	0.593	0.626	-0.02	-0.02	0.12	0.25	0.26	0.15	0.04
19	0.554	0.564	0.522	0.503	0.506	0.524	0.557	0.09	0.03	0.45	0.50	0.48	0.36	0.23
20	0.615	0.614	0.589	0.580	0.558	0.586	0.602	-0.40	-0.33	0.18	0.21	0.32	0.10	0.20
Mean	**0.599**	**0.616**	**0.582**	**0.580**	**0.583**	**0.590**	**0.594**	**0.01**	**0.01**	**0.14**	**0.20**	**0.19**	**0.14**	**0.07**
SD	*0.028*	*0.048*	*0.029*	*0.034*	*0.037*	*0.029*	*0.024*	*0.17*	*0.13*	*0.13*	*0.13*	*0.15*	*0.09*	*0.14*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C58

Dataset 6: Predictive performance statistics for the prediction equations developed using the combination of Emotional Orientation, Cognitive Orientation, Interpersonal Orientation, and Task Orientation (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.584	0.600	0.615	0.718	0.923	0.714	0.572	-0.02	-0.04	-0.03	-0.26	-0.15	-0.12	0.09
2	0.616	0.609	0.630	0.751	0.729	0.680	0.615	-0.09	-0.03	-0.13	-0.19	-0.08	-0.11	-0.09
3	0.572	0.585	0.655	0.780	0.880	0.725	0.602	0.15	0.10	-0.20	-0.21	-0.24	-0.14	-0.09
4	0.599	0.605	0.532	0.788	0.669	0.649	0.599	-0.18	-0.18	0.16	-0.08	0.24	0.04	-0.18
5	0.583	0.594	0.679	0.620	0.574	0.617	0.598	-0.04	-0.09	-0.14	0.12	0.23	0.03	-0.03
6	0.640	0.712	0.648	0.689	0.917	0.741	0.620	-0.19	-0.09	-0.03	-0.24	-0.20	-0.14	-0.07
7	0.585	0.819	0.612	0.816	0.863	0.777	0.575	0.09	-0.12	-0.17	-0.25	-0.03	-0.14	0.04
8	0.583	0.581	0.628	0.684	0.775	0.667	0.599	-0.04	-0.02	-0.03	-0.02	0.05	0.00	-0.16
9	0.569	0.617	0.633	0.784	0.743	0.694	0.584	0.01	-0.09	0.01	-0.16	0.11	-0.03	-0.07
10	0.592	0.628	0.584	0.653	0.646	0.628	0.593	0.02	0.09	0.13	-0.07	-0.06	0.02	0.06
11	0.571	0.621	0.519	0.547	0.693	0.595	0.583	0.10	0.12	0.17	0.13	-0.22	0.05	0.08
12	0.599	0.643	0.669	0.687	0.738	0.684	0.588	-0.12	0.05	-0.05	0.20	0.13	0.08	-0.03
13	0.606	0.577	0.608	0.586	0.964	0.684	0.604	0.13	0.19	0.14	0.30	0.20	0.21	0.19
14	0.675	0.688	0.718	0.908	0.688	0.750	0.643	-0.14	-0.18	-0.18	-0.06	0.09	-0.08	-0.13
15	0.541	0.551	0.495	0.647	0.689	0.595	0.555	0.03	0.02	0.19	-0.10	0.06	0.04	-0.05
16	0.588	0.627	0.657	0.850	1.491	0.906	0.617	0.20	0.08	0.08	0.14	-0.22	0.02	0.01
17	0.602	0.817	0.790	0.754	1.134	0.874	0.597	0.07	0.10	0.12	-0.01	-0.03	0.05	0.15
18	0.652	0.700	0.748	0.811	0.801	0.765	0.640	-0.16	-0.25	0.10	-0.13	-0.22	-0.13	-0.10
19	0.560	0.567	0.640	0.585	0.571	0.591	0.570	0.03	0.00	-0.22	0.06	0.12	-0.01	0.10
20	0.669	0.669	0.886	0.623	1.078	0.814	0.664	-0.41	-0.42	-0.19	0.14	-0.08	-0.14	-0.33
Mean	**0.599**	**0.640**	**0.647**	**0.714**	**0.828**	**0.708**	**0.601**	**-0.03**	**-0.04**	**-0.01**	**-0.03**	**-0.01**	**-0.02**	**-0.03**
SD	*0.035*	*0.073*	*0.088*	*0.095*	*0.213*	*0.087*	*0.026*	*0.14*	*0.14*	*0.14*	*0.16*	*0.16*	*0.09*	*0.12*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C59

Dataset 6: Predictive performance statistics for the prediction equations developed using the combination of Emotional Orientation, Social Orientation, Cognitive Orientation, Interpersonal Orientation, and Task Orientation (Study 4).

	MAE							Cross-validity coefficient						
Partition	LR	H1	H2	H3	H4	ANN1	ANN2	LR	H1	H2	H3	H4	ANN1	ANN2
1	0.585	0.593	0.595	0.740	1.125	0.763	0.593	-0.04	-0.03	-0.01	-0.14	-0.31	-0.12	-0.08
2	0.617	0.608	0.594	0.776	1.290	0.817	0.621	-0.09	-0.02	-0.09	-0.16	0.05	-0.05	-0.08
3	0.578	0.594	0.628	0.995	0.946	0.791	0.609	0.09	0.10	-0.18	-0.20	-0.19	-0.12	-0.10
4	0.598	0.611	0.614	0.974	0.774	0.743	0.595	-0.18	-0.18	0.06	-0.07	0.05	-0.04	-0.15
5	0.606	0.617	0.649	0.631	1.085	0.746	0.595	-0.10	-0.13	-0.09	0.15	0.14	0.02	-0.09
6	0.640	0.808	0.656	0.923	1.297	0.921	0.626	-0.19	-0.12	-0.08	-0.12	0.04	-0.07	-0.09
7	0.593	0.606	0.591	0.959	1.210	0.842	0.601	0.04	-0.11	-0.01	0.01	0.05	-0.02	-0.04
8	0.584	0.613	0.589	0.684	1.286	0.793	0.581	-0.05	-0.21	-0.01	0.04	-0.39	-0.14	0.01
9	0.578	0.613	0.702	0.904	1.902	1.030	0.565	-0.03	-0.07	-0.14	-0.01	0.00	-0.06	-0.06
10	0.606	0.613	0.608	0.791	1.188	0.800	0.614	-0.03	-0.03	0.06	-0.28	-0.31	-0.14	0.01
11	0.589	0.774	0.703	0.895	1.342	0.929	0.595	-0.02	-0.28	-0.24	-0.37	-0.35	-0.31	-0.02
12	0.606	0.608	0.668	0.942	1.075	0.823	0.599	-0.16	-0.09	-0.32	0.10	0.01	-0.08	-0.08
13	0.608	0.682	0.647	0.656	0.618	0.651	0.613	0.08	0.10	0.07	0.15	0.29	0.15	0.13
14	0.686	0.679	0.769	0.886	0.883	0.804	0.670	-0.19	-0.43	-0.15	-0.12	0.15	-0.13	-0.17
15	0.542	0.574	0.620	0.776	0.803	0.693	0.540	0.00	-0.15	-0.24	0.07	0.03	-0.07	-0.02
16	0.605	0.615	0.832	1.006	2.460	1.228	0.636	0.09	0.03	-0.06	-0.05	-0.25	-0.08	-0.11
17	0.602	0.845	0.704	0.761	1.436	0.937	0.599	0.07	0.10	0.09	-0.24	0.21	0.04	0.19
18	0.654	0.705	0.713	0.863	0.884	0.791	0.642	-0.16	-0.24	-0.02	0.04	0.05	-0.04	-0.04
19	0.595	0.598	0.573	0.694	1.235	0.775	0.600	-0.19	-0.19	-0.07	-0.04	-0.09	-0.10	-0.21
20	0.689	0.659	0.749	0.947	1.026	0.845	0.670	-0.45	-0.32	-0.17	-0.02	0.19	-0.08	-0.39
Mean	**0.608**	**0.651**	**0.660**	**0.840**	**1.193**	**0.836**	**0.608**	**-0.07**	**-0.11**	**-0.08**	**-0.06**	**-0.03**	**-0.07**	**-0.07**
SD	*0.035*	*0.075*	*0.068*	*0.117*	*0.399*	*0.124*	*0.030*	*0.13*	*0.14*	*0.11*	*0.14*	*0.20*	*0.09*	*0.12*

Note: LR = linear regression equations, H1 to H4 = neural networks developed without early stopping, ANN1 = mean of H1 to H4, ANN2 = early stopping neural network committees.

Table C60

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Neuroticism facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	7.415	7.526	7.288	0.18	0.21	0.19
2	6.401	7.671	6.388	0.17	0.17	0.18
3	6.439	6.702	6.400	0.25	0.27	0.28
4	7.490	8.492	7.494	-0.02	0.19	-0.01
5	6.604	7.579	6.689	0.31	0.18	0.32
6	7.377	7.449	7.341	0.14	0.26	0.17
7	6.696	10.723	7.278	0.25	-0.13	0.13
8	7.142	7.579	7.183	0.10	0.17	0.15
9	7.137	7.937	7.153	0.24	0.18	0.21
10	7.517	8.650	7.531	0.18	0.12	0.22
11	7.853	8.442	7.912	0.17	0.11	0.16
12	7.152	7.744	7.256	0.29	0.26	0.24
13	6.657	7.259	6.615	0.24	0.22	0.27
14	6.783	7.541	6.792	0.10	0.05	0.10
15	6.941	8.379	6.953	0.18	0.04	0.20
16	7.362	7.477	7.125	0.33	0.23	0.36
17	7.240	8.176	7.181	0.30	0.17	0.31
18	7.325	8.116	7.505	0.15	0.05	0.10
19	7.971	9.002	8.116	0.08	-0.08	0.03
20	7.824	7.998	7.776	0.08	0.10	0.08
Mean	**7.166**	**8.022**	**7.199**	**0.19**	**0.14**	**0.19**
SD	*0.462*	*0.835*	*0.467*	*0.09*	*0.11*	*0.10*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C61

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Extraversion facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	7.324	7.921	7.278	0.10	-0.02	0.14
2	6.438	7.378	6.311	0.20	0.22	0.26
3	6.716	9.294	6.775	0.15	0.01	0.14
4	8.139	9.153	8.050	-0.08	-0.06	-0.08
5	6.996	8.554	7.053	0.19	0.09	0.14
6	7.572	8.367	7.517	0.16	0.07	0.17
7	6.973	7.292	6.932	0.18	0.19	0.19
8	7.114	8.288	7.464	0.17	0.05	0.13
9	7.347	8.784	7.783	0.12	-0.16	-0.11
10	7.532	8.316	7.615	0.23	0.00	0.10
11	7.643	8.379	7.612	0.37	0.04	0.37
12	7.523	8.295	7.480	0.15	0.05	0.12
13	7.425	8.205	7.430	0.10	0.00	0.10
14	7.054	6.996	7.070	0.03	0.04	0.04
15	7.024	8.043	6.948	0.23	0.12	0.27
16	7.417	7.712	7.410	0.30	0.11	0.21
17	7.410	8.034	7.330	0.20	0.05	0.21
18	7.514	7.753	7.432	0.03	0.07	0.08
19	7.847	9.380	7.867	0.18	0.08	0.18
20	7.416	8.324	7.481	0.19	0.09	0.15
Mean	**7.321**	**8.223**	**7.342**	**0.16**	**0.05**	**0.14**
SD	*0.387*	*0.631*	*0.403*	*0.10*	*0.08*	*0.11*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C62

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Openness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	7.518	8.149	7.554	0.09	0.13	0.09
2	6.429	7.613	6.290	0.23	0.17	0.25
3	6.701	6.770	6.525	0.11	0.35	0.18
4	6.931	7.273	6.769	0.23	0.23	0.25
5	7.187	8.358	7.277	0.21	0.12	0.21
6	7.881	8.363	7.888	0.05	0.06	0.05
7	7.086	8.632	7.077	0.15	0.10	0.15
8	6.610	7.204	6.837	0.35	0.29	0.33
9	7.773	8.070	7.752	-0.08	0.04	-0.04
10	7.308	7.336	7.161	0.45	0.32	0.38
11	8.284	7.761	8.269	0.05	0.14	0.05
12	7.360	7.257	7.233	0.25	0.22	0.25
13	6.520	7.066	6.457	0.28	0.24	0.34
14	6.718	6.928	6.633	0.25	0.25	0.25
15	7.028	7.373	7.000	0.24	0.24	0.26
16	7.228	7.289	7.246	0.32	0.12	0.24
17	7.422	7.459	7.384	0.22	0.19	0.21
18	7.500	8.776	7.738	0.19	0.07	0.11
19	7.634	8.433	7.738	0.22	-0.04	0.16
20	7.518	7.666	7.503	0.11	0.12	0.12
Mean	**7.232**	**7.689**	**7.216**	**0.20**	**0.17**	**0.19**
SD	*0.487*	*0.595*	*0.527*	*0.12*	*0.10*	*0.11*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C63

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Agreeableness facets (Study 5).

Partition	MAE LR	MAE ANN1	MAE ANN2	Cross-validity coefficient LR	Cross-validity coefficient ANN1	Cross-validity coefficient ANN2
1	7.752	8.489	7.965	-0.11	0.00	-0.15
2	6.871	7.990	6.978	-0.01	-0.06	-0.05
3	6.788	8.007	6.727	-0.02	-0.09	0.02
4	7.765	8.239	7.748	-0.15	-0.06	-0.15
5	7.432	7.945	7.424	-0.04	0.01	-0.06
6	7.786	10.571	8.457	0.00	-0.12	-0.16
7	7.485	10.598	7.694	-0.04	-0.06	-0.04
8	7.420	7.272	7.341	-0.01	0.08	0.00
9	7.895	8.464	8.036	-0.02	-0.10	-0.09
10	8.383	9.868	8.334	-0.11	0.02	-0.06
11	8.118	11.314	8.199	0.01	-0.20	-0.11
12	7.833	8.061	7.963	-0.07	-0.06	-0.18
13	7.311	7.913	7.294	-0.06	-0.09	-0.06
14	6.801	9.014	6.973	0.11	-0.02	0.15
15	7.287	8.307	7.303	-0.04	-0.08	-0.02
16	7.698	9.652	7.969	0.08	-0.13	0.01
17	7.715	8.378	7.782	0.08	-0.11	0.04
18	7.585	10.901	7.792	-0.08	-0.07	-0.19
19	8.017	9.123	8.080	0.07	-0.19	0.03
20	7.956	9.166	8.017	-0.10	0.00	-0.11
Mean	**7.595**	**8.964**	**7.704**	**-0.03**	**-0.07**	**-0.06**
SD	*0.430*	*1.155*	*0.478*	*0.07*	*0.07*	*0.09*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C64

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Conscientiousness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	8.012	8.916	8.174	0.07	0.02	0.06
2	6.503	7.023	6.505	0.21	-0.02	0.18
3	6.547	8.406	6.724	0.26	-0.08	0.19
4	7.895	8.496	7.912	-0.08	-0.04	-0.05
5	7.239	9.502	7.248	0.15	-0.11	0.13
6	7.290	10.154	7.866	0.25	-0.16	0.08
7	6.810	7.788	6.893	0.26	0.00	0.23
8	6.980	7.542	6.838	0.18	0.02	0.17
9	7.372	9.581	7.565	0.20	-0.11	0.15
10	7.955	8.105	7.720	0.10	0.08	0.13
11	7.791	8.500	7.888	0.23	0.06	0.16
12	7.037	9.897	7.258	0.27	-0.10	0.14
13	7.174	8.391	7.264	0.11	0.02	0.11
14	7.312	8.835	7.361	-0.05	-0.10	-0.08
15	6.957	9.619	6.891	0.22	0.08	0.25
16	7.451	9.149	7.620	0.17	0.10	0.13
17	7.429	7.426	7.362	0.14	0.08	0.18
18	7.605	8.502	7.583	0.11	0.05	0.11
19	7.663	8.238	7.735	0.24	0.09	0.21
20	7.474	7.955	7.453	0.22	0.11	0.24
Mean	**7.325**	**8.601**	**7.393**	**0.16**	**0.00**	**0.14**
SD	*0.433*	*0.856*	*0.447*	*0.10*	*0.08*	*0.09*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C65

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the eighteen facets from the three theoretically relevant factors Neuroticism, Openness, and Conscientiousness (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	8.103	15.038	7.903	0.22	-0.01	0.25
2	5.992	15.744	6.000	0.38	0.05	0.39
3	6.536	15.473	6.239	0.26	0.08	0.34
4	7.304	13.728	7.143	0.20	0.11	0.21
5	7.206	15.563	7.507	0.26	0.21	0.18
6	7.845	16.546	8.040	0.16	0.16	0.12
7	6.183	15.243	6.768	0.36	0.20	0.29
8	6.529	17.148	6.676	0.34	0.02	0.30
9	7.776	19.245	7.575	0.18	-0.03	0.16
10	7.080	13.736	7.204	0.33	0.20	0.32
11	7.496	13.522	7.675	0.32	0.05	0.29
12	6.765	17.744	7.027	0.43	0.17	0.38
13	6.541	16.375	6.203	0.37	0.17	0.40
14	6.586	15.763	6.958	0.22	0.18	0.15
15	6.570	17.364	6.757	0.40	0.20	0.36
16	6.907	19.735	6.930	0.37	-0.02	0.33
17	7.413	16.919	7.729	0.27	0.05	0.20
18	7.426	16.342	7.596	0.25	-0.03	0.20
19	8.029	17.337	8.012	0.18	-0.06	0.16
20	7.267	16.515	7.543	0.29	0.08	0.28
Mean	**7.078**	**16.254**	**7.174**	**0.29**	**0.09**	**0.27**
SD	*0.608*	*1.644*	*0.609*	*0.08*	*0.09*	*0.09*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C66

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the six Neuroticism facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	27.166	30.148	27.245	0.18	0.04	0.21
2	34.398	35.139	34.056	0.17	0.25	0.20
3	33.268	33.694	33.163	0.04	0.12	0.05
4	30.712	32.498	31.538	0.20	0.23	0.17
5	33.751	35.144	33.037	0.04	0.04	0.08
6	31.878	33.733	31.727	0.05	-0.01	0.04
7	34.499	34.513	34.461	0.03	0.21	0.06
8	34.074	36.326	33.773	0.07	0.13	0.04
9	34.855	37.579	34.766	-0.01	-0.03	-0.08
10	34.945	34.890	34.119	0.06	0.13	0.09
11	28.954	31.887	28.707	0.13	0.24	0.19
12	36.294	38.034	36.266	0.05	0.05	0.06
13	33.568	34.850	33.210	0.10	0.03	0.07
14	33.111	33.025	33.874	0.10	0.19	0.13
15	35.991	44.101	36.439	-0.02	0.00	-0.05
16	30.016	29.649	29.603	0.07	0.14	0.08
17	35.449	38.239	36.265	0.11	0.01	0.05
18	32.123	33.019	33.714	0.24	0.23	0.09
19	32.169	33.398	31.508	0.00	0.19	0.05
20	34.024	36.980	33.786	-0.01	0.00	0.00
Mean	**33.062**	**34.842**	**33.063**	**0.08**	**0.11**	**0.08**
SD	*2.380*	*3.216*	*2.432*	*0.07*	*0.10*	*0.08*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C67

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the six Extraversion facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	25.784	27.256	25.669	0.22	0.22	0.24
2	32.311	33.459	32.450	0.40	0.26	0.36
3	30.864	33.553	31.654	0.29	0.05	0.05
4	31.251	45.752	31.726	0.22	0.00	0.18
5	35.311	40.510	34.919	0.10	-0.01	0.08
6	30.941	32.070	30.961	0.16	0.17	0.13
7	35.343	38.348	36.371	0.15	-0.08	0.04
8	33.981	33.810	33.628	0.21	0.24	0.25
9	33.832	48.353	34.304	0.10	0.06	0.11
10	34.034	35.147	33.729	0.13	0.13	0.13
11	26.915	29.635	28.100	0.37	0.16	0.24
12	34.148	38.483	35.564	0.11	0.01	0.07
13	33.696	33.993	33.216	0.18	0.20	0.19
14	31.943	36.660	32.367	0.37	0.00	0.26
15	33.212	37.092	33.545	0.33	0.11	0.22
16	28.598	29.566	28.886	0.25	0.03	0.15
17	33.773	36.576	33.913	0.34	0.15	0.32
18	31.828	34.559	32.395	0.29	0.16	0.23
19	31.188	32.399	30.542	0.22	0.12	0.22
20	32.939	34.289	33.600	0.19	0.08	0.08
Mean	**32.095**	**35.576**	**32.377**	**0.23**	**0.10**	**0.18**
SD	*2.577*	*5.088*	*2.589*	*0.10*	*0.09*	*0.09*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C68

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the six Openness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	26.635	29.682	27.386	0.17	0.05	0.11
2	34.347	36.840	34.744	0.23	0.12	0.19
3	32.688	33.705	32.720	0.11	0.08	0.11
4	30.180	37.318	31.086	0.22	0.08	0.16
5	32.343	36.063	33.186	0.19	0.12	0.11
6	33.909	35.669	34.954	-0.11	0.01	-0.10
7	34.716	37.223	35.017	0.06	0.07	0.06
8	32.318	35.616	32.732	0.33	0.01	0.27
9	33.953	40.947	34.129	0.06	-0.02	0.04
10	36.864	40.109	36.741	0.08	-0.04	0.05
11	28.978	31.111	29.961	0.15	0.04	0.10
12	36.657	38.902	37.267	0.07	0.05	0.08
13	31.277	35.017	31.922	0.27	0.14	0.24
14	34.359	39.182	34.044	0.11	0.06	0.10
15	33.974	37.578	34.414	0.12	0.05	0.10
16	29.958	33.463	30.200	0.05	0.01	0.03
17	33.958	34.329	33.844	0.21	0.18	0.26
18	33.817	37.235	34.396	0.12	0.11	0.11
19	29.952	33.642	29.857	0.20	0.14	0.22
20	32.952	35.090	33.532	0.13	0.16	0.11
Mean	**32.692**	**35.936**	**33.107**	**0.14**	**0.07**	**0.12**
SD	*2.559*	*2.844*	*2.446*	*0.10*	*0.06*	*0.09*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C69

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the six Agreeableness facets (Study 5).

Partition	MAE LR	MAE ANN1	MAE ANN2	Cross-validity coefficient LR	Cross-validity coefficient ANN1	Cross-validity coefficient ANN2
1	28.282	35.428	29.598	0.03	-0.13	0.05
2	33.838	36.058	33.375	0.15	0.16	0.20
3	32.332	34.956	33.094	0.03	-0.06	-0.03
4	31.774	34.331	33.196	0.07	0.12	-0.03
5	32.717	37.821	33.903	0.12	-0.02	0.06
6	31.407	33.204	31.803	0.00	0.06	0.01
7	34.309	37.200	35.541	0.07	0.03	-0.01
8	33.986	37.783	33.871	-0.01	-0.09	0.02
9	32.074	38.454	32.813	0.17	0.01	0.04
10	34.426	36.933	34.869	0.04	0.23	-0.02
11	28.053	28.669	28.158	0.11	0.15	0.10
12	35.581	41.715	36.050	0.06	-0.09	0.01
13	32.907	34.580	33.920	0.17	0.07	0.05
14	36.585	36.322	36.669	-0.14	-0.04	-0.19
15	37.647	41.254	37.481	-0.01	0.07	-0.04
16	28.637	31.061	28.680	0.13	-0.04	0.08
17	36.422	36.223	35.998	0.07	0.12	0.08
18	33.128	37.736	33.439	0.04	-0.04	0.00
19	29.935	33.885	30.468	0.16	-0.03	0.08
20	32.757	38.326	32.933	0.16	0.03	0.15
Mean	**32.840**	**36.097**	**33.293**	**0.07**	**0.03**	**0.03**
SD	*2.699*	*3.079*	*2.559*	*0.08*	*0.10*	*0.08*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C70

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the six Conscientiousness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	28.801	30.322	28.368	0.12	0.14	0.15
2	33.117	40.631	34.222	0.31	0.07	0.20
3	30.750	32.948	31.851	0.25	0.11	0.16
4	30.449	31.498	30.836	0.24	0.14	0.18
5	33.260	36.758	34.220	0.18	0.07	0.11
6	30.819	33.682	30.737	0.15	0.01	0.12
7	34.204	36.715	33.481	0.14	0.10	0.15
8	33.108	34.772	33.283	0.23	0.18	0.22
9	35.168	39.245	35.048	0.11	0.09	0.12
10	37.432	38.488	36.858	0.08	0.01	0.09
11	28.618	31.208	28.967	0.21	0.07	0.15
12	35.572	34.103	35.566	0.12	0.20	0.15
13	34.114	35.086	34.155	0.11	0.09	0.11
14	32.073	36.365	32.036	0.24	0.10	0.23
15	33.737	36.402	34.051	0.25	0.19	0.23
16	29.819	32.742	29.672	0.15	0.16	0.18
17	35.434	41.705	35.382	0.15	-0.01	0.15
18	32.672	35.283	32.279	0.24	0.14	0.22
19	30.491	33.502	31.029	0.20	0.13	0.20
20	32.706	36.229	32.066	0.15	0.16	0.20
Mean	**32.617**	**35.384**	**32.705**	**0.18**	**0.11**	**0.16**
SD	*2.383*	*3.061*	*2.311*	*0.06*	*0.06*	*0.04*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C71

Dataset 2: Predictive performance statistics for the prediction equations developed using the combination of the twenty-four facets from the four theoretically relevant factors Neuroticism, Extraversion, Openness, and Conscientiousness (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	27.942	55.513	26.182	0.25	0.16	0.31
2	33.369	64.988	32.635	0.34	0.16	0.39
3	31.376	60.994	31.767	0.25	0.06	0.22
4	31.281	63.982	30.233	0.34	0.23	0.37
5	35.196	67.763	34.262	0.20	0.12	0.23
6	32.500	61.108	32.483	0.08	0.02	0.06
7	36.013	57.808	36.078	0.11	0.22	0.12
8	34.098	54.648	33.351	0.23	0.15	0.26
9	34.137	56.990	33.705	0.19	0.16	0.21
10	37.631	63.744	38.053	0.09	0.00	0.08
11	30.245	73.856	28.644	0.21	-0.03	0.23
12	37.167	52.745	36.379	0.15	0.07	0.12
13	35.745	70.652	33.967	0.16	0.05	0.16
14	32.189	63.690	31.151	0.22	0.13	0.24
15	36.208	55.859	35.710	0.15	0.08	0.09
16	31.457	56.486	30.028	0.11	0.30	0.11
17	34.061	58.717	35.394	0.32	0.11	0.24
18	34.726	57.562	34.999	0.23	0.08	0.19
19	31.546	66.252	30.070	0.25	0.09	0.29
20	36.096	39.577	35.184	0.02	0.28	0.02
Mean	**33.649**	**60.147**	**33.014**	**0.19**	**0.12**	**0.20**
SD	*2.540*	*7.414*	*2.967*	*0.09*	*0.09*	*0.10*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C72

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of the six Neuroticism facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.501	0.690	0.492	0.06	-0.17	0.08
2	0.474	0.542	0.470	0.07	-0.07	0.08
3	0.564	0.591	0.547	-0.01	0.08	0.04
4	0.499	0.672	0.503	0.15	-0.10	0.06
5	0.495	1.017	0.488	0.17	0.27	0.17
6	0.506	0.743	0.515	0.10	-0.08	0.06
7	0.547	0.837	0.539	0.01	-0.10	0.00
8	0.553	0.640	0.553	0.14	-0.08	0.13
9	0.587	1.096	0.598	0.29	-0.14	0.23
10	0.474	0.497	0.471	0.05	-0.06	0.07
11	0.553	0.716	0.558	0.21	0.07	0.21
12	0.448	0.821	0.447	0.32	-0.19	0.25
13	0.462	0.896	0.472	0.28	-0.03	0.22
14	0.508	1.035	0.532	0.22	-0.14	0.13
15	0.500	0.653	0.500	0.28	0.14	0.27
16	0.479	0.599	0.486	0.14	-0.09	0.10
17	0.436	1.019	0.442	0.24	0.23	0.22
18	0.430	0.545	0.433	0.11	-0.01	0.12
19	0.427	0.745	0.441	0.18	-0.31	0.03
20	0.523	0.895	0.525	0.00	-0.22	0.00
Mean	**0.498**	**0.762**	**0.501**	**0.15**	**-0.05**	**0.12**
SD	*0.046*	*0.182*	*0.045*	*0.10*	*0.15*	*0.08*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C73

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of the six Extraversion facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.490	0.598	0.482	0.07	0.18	0.11
2	0.472	0.642	0.460	0.15	0.04	0.13
3	0.537	0.725	0.544	0.02	-0.13	0.01
4	0.528	0.712	0.528	0.08	-0.15	0.06
5	0.503	0.568	0.505	0.19	0.19	0.17
6	0.485	0.629	0.489	0.16	0.19	0.15
7	0.562	0.919	0.559	0.09	-0.04	0.08
8	0.552	0.644	0.547	0.27	0.16	0.27
9	0.606	0.724	0.602	0.21	0.06	0.22
10	0.509	0.571	0.517	-0.08	0.04	-0.09
11	0.554	0.685	0.574	0.24	0.02	0.14
12	0.457	0.748	0.461	0.39	-0.03	0.36
13	0.475	0.640	0.479	0.13	-0.16	0.05
14	0.512	0.643	0.521	0.19	-0.11	0.13
15	0.536	0.591	0.531	0.11	0.11	0.10
16	0.473	0.640	0.475	0.14	0.12	0.13
17	0.441	0.531	0.437	0.30	0.19	0.28
18	0.452	0.618	0.446	0.13	-0.05	0.14
19	0.474	0.780	0.467	0.04	-0.07	0.04
20	0.574	0.883	0.561	-0.01	0.22	-0.02
Mean	**0.510**	**0.674**	**0.509**	**0.14**	**0.04**	**0.12**
SD	*0.045*	*0.101*	*0.046*	*0.11*	*0.13*	*0.11*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C74

Dataset 1: Predictive performance statistics for the prediction equations developed using the combination of the six Openness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.504	0.676	0.495	0.00	-0.19	-0.02
2	0.453	0.438	0.441	0.14	0.21	0.20
3	0.537	0.640	0.532	0.10	0.03	0.10
4	0.504	0.516	0.507	0.28	0.21	0.23
5	0.515	0.634	0.505	0.21	0.13	0.21
6	0.474	0.497	0.482	0.14	0.12	0.10
7	0.513	0.563	0.514	0.13	0.03	0.09
8	0.511	0.548	0.522	0.25	0.18	0.21
9	0.605	0.693	0.599	0.15	0.11	0.15
10	0.458	1.014	0.457	0.22	-0.15	0.20
11	0.587	0.669	0.596	0.15	-0.01	0.07
12	0.489	0.862	0.479	0.18	-0.13	0.18
13	0.472	0.529	0.460	0.25	0.14	0.26
14	0.525	0.552	0.516	0.11	0.16	0.16
15	0.510	0.538	0.503	0.23	0.14	0.21
16	0.457	0.472	0.465	0.22	0.20	0.18
17	0.427	0.532	0.435	0.37	0.24	0.31
18	0.470	0.511	0.477	0.13	0.08	0.12
19	0.427	0.597	0.422	0.25	0.10	0.23
20	0.490	0.527	0.502	0.26	0.18	0.17
Mean	**0.496**	**0.600**	**0.496**	**0.19**	**0.09**	**0.17**
SD	*0.046*	*0.137*	*0.046*	*0.08*	*0.13*	*0.08*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C75

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of the six Agreeableness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.503	0.597	0.503	0.06	0.01	0.08
2	0.466	0.505	0.470	0.02	0.16	0.03
3	0.545	0.672	0.529	-0.14	-0.15	-0.05
4	0.526	0.526	0.523	0.13	0.42	0.03
5	0.523	0.714	0.522	0.04	0.25	0.10
6	0.536	0.723	0.538	-0.15	-0.20	-0.07
7	0.518	0.753	0.531	0.17	-0.03	0.13
8	0.549	0.561	0.557	0.11	0.13	0.08
9	0.640	0.890	0.620	-0.09	-0.07	0.08
10	0.482	0.530	0.481	0.13	0.13	0.19
11	0.578	0.894	0.578	0.04	0.03	0.12
12	0.497	0.734	0.488	-0.07	0.19	0.04
13	0.509	0.548	0.500	0.01	0.17	0.04
14	0.560	0.587	0.549	-0.02	0.11	0.05
15	0.554	0.652	0.533	-0.02	-0.15	0.20
16	0.517	0.532	0.504	-0.05	0.05	-0.03
17	0.445	0.636	0.441	0.19	0.08	0.30
18	0.449	0.545	0.441	0.19	0.09	0.26
19	0.478	0.557	0.482	-0.12	0.04	-0.01
20	0.530	0.669	0.542	0.00	0.02	-0.01
Mean	**0.520**	**0.641**	**0.517**	**0.02**	**0.06**	**0.08**
SD	*0.046*	*0.115*	*0.044*	*0.11*	*0.15*	*0.10*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C76

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of the six Conscientiousness facets (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.460	0.500	0.455	0.24	0.08	0.26
2	0.432	0.527	0.438	0.28	0.14	0.24
3	0.524	0.535	0.518	0.17	0.11	0.19
4	0.523	0.694	0.515	0.12	-0.29	0.14
5	0.475	0.576	0.492	0.30	0.09	0.22
6	0.488	0.577	0.490	0.20	0.04	0.18
7	0.519	0.560	0.524	0.18	0.14	0.17
8	0.526	0.568	0.531	0.37	0.23	0.37
9	0.557	0.632	0.568	0.35	0.05	0.29
10	0.459	0.562	0.461	0.19	0.06	0.18
11	0.535	0.574	0.546	0.34	0.17	0.31
12	0.434	0.535	0.445	0.39	0.05	0.33
13	0.461	0.484	0.455	0.30	0.15	0.31
14	0.501	0.528	0.498	0.18	0.09	0.19
15	0.485	0.604	0.492	0.37	0.12	0.30
16	0.439	0.489	0.439	0.29	0.08	0.30
17	0.475	0.562	0.466	0.15	-0.03	0.14
18	0.456	0.584	0.467	0.22	-0.07	0.20
19	0.442	0.500	0.453	0.23	0.13	0.22
20	0.482	0.553	0.491	0.36	-0.06	0.38
Mean	**0.484**	**0.557**	**0.487**	**0.26**	**0.06**	**0.25**
SD	*0.037*	*0.050*	*0.037*	*0.08*	*0.11*	*0.07*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Table C77

Dataset 4: Predictive performance statistics for the prediction equations developed using the combination of the eighteen facets from the three theoretically relevant factors Neuroticism, Extraversion, and Conscientiousness (Study 5).

	MAE			Cross-validity coefficient		
Partition	LR	ANN1	ANN2	LR	ANN1	ANN2
1	0.479	0.890	0.464	0.20	0.13	0.24
2	0.475	1.231	0.464	0.17	-0.20	0.13
3	0.558	1.081	0.543	0.04	-0.04	0.09
4	0.545	0.832	0.515	0.00	-0.04	0.10
5	0.502	0.845	0.490	0.15	0.15	0.21
6	0.514	0.989	0.501	0.22	0.01	0.19
7	0.559	0.807	0.533	0.12	0.15	0.13
8	0.532	1.092	0.532	0.34	-0.15	0.25
9	0.555	1.121	0.566	0.38	0.04	0.34
10	0.493	0.978	0.484	0.06	0.05	0.00
11	0.537	1.155	0.529	0.28	0.01	0.34
12	0.443	1.241	0.418	0.37	-0.06	0.46
13	0.451	0.981	0.455	0.30	-0.06	0.28
14	0.524	0.937	0.528	0.16	-0.06	0.12
15	0.484	1.024	0.490	0.30	0.02	0.25
16	0.456	0.847	0.441	0.27	-0.06	0.27
17	0.483	1.084	0.449	0.13	0.12	0.22
18	0.438	0.884	0.441	0.25	0.26	0.25
19	0.484	0.814	0.477	0.14	0.19	0.12
20	0.556	1.079	0.544	0.04	-0.08	0.09
Mean	**0.503**	**0.996**	**0.493**	**0.20**	**0.02**	**0.20**
SD	*0.041*	*0.137*	*0.042*	*0.11*	*0.12*	*0.11*

Note: LR = linear regression equations, ANN1 = neural networks developed without early stopping (H4 hidden units), ANN2 = early stopping neural network committees.

Appendix D: Corrected t-Test Results

Table D1.

Study 1: Corrected t-test comparing MAE of linear regression and neural networks.

Dataset and predictor	ψ	SD	SE	t
1. University students				
Neuroticism	0.001	0.111	0.083	0.01
Extraversion	-0.004	0.105	0.078	-0.05
Openness	-0.033	0.036	0.027	-1.22
Agreeableness	-0.043	0.062	0.046	-0.92
Conscientiousness	-0.031	0.098	0.073	-0.43
2. Police recruits				
Neuroticism	-0.214	0.446	0.330	-0.65
Extraversion	-0.038	0.630	0.466	-0.08
Openness	-0.379	0.528	0.390	-0.97
Agreeableness	-0.352	0.381	0.282	-1.25
Conscientiousness	-0.131	0.256	0.190	-0.69
3. Flight attendants				
Neuroticism	-0.006	0.007	0.005	-1.06
Extraversion	-0.005	0.010	0.008	-0.63
Openness	-0.006	0.009	0.007	-0.84
Agreeableness	-0.014	0.015	0.011	-1.22
Conscientiousness	-0.001	0.008	0.006	-0.24
4. Managers				
Neuroticism	-0.006	0.011	0.008	-0.76
Extraversion	-0.001	0.009	0.007	-0.22
Openness	-0.007	0.012	0.009	-0.76
Agreeableness	-0.005	0.010	0.008	-0.72
Conscientiousness	-0.003	0.004	0.003	-1.04
5. Bus drivers				
Adjustment	-0.009	0.018	0.014	-0.64
Ambition	-0.012	0.020	0.015	-0.81
Sociability	-0.030	0.065	0.048	-0.63
Intellectance	0.055	0.068	0.051	1.09
Likeability	-0.016	0.036	0.027	-0.61
Prudence	-0.013	0.037	0.028	-0.46
6. Professionals				
Emotional Orientation	-0.008	0.026	0.019	-0.43
Social Orientation	-0.004	0.006	0.004	-0.92
Cognitive Orientation	-0.010	0.014	0.011	-0.97
Interpersonal Orientation	-0.042	0.060	0.044	-0.94
Task Orientation	0.000	0.010	0.007	0.07

Note: ψ = linear regression MAE – neural network MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D2

Study 1: Corrected t-test comparing cross-validity coefficients (CVR) of linear regression and neural networks.

Dataset and predictor	ψ	SD	SE	t
1. University students				
Neuroticism	-0.04	0.08	0.06	-0.69
Extraversion	0.01	0.05	0.04	0.22
Openness	0.06	0.08	0.06	0.98
Agreeableness	0.01	0.05	0.04	0.23
Conscientiousness	0.02	0.05	0.04	0.61
2. Police recruits				
Neuroticism	0.04	0.07	0.05	0.73
Extraversion	0.04	0.11	0.08	0.56
Openness	0.02	0.04	0.03	0.64
Agreeableness	0.04	0.07	0.05	0.88
Conscientiousness	0.01	0.02	0.01	0.80
3. Flight attendants				
Neuroticism	0.04	0.04	0.03	1.23
Extraversion	0.02	0.03	0.02	0.91
Openness	0.03	0.04	0.03	0.81
Agreeableness	0.04	0.04	0.03	1.33
Conscientiousness	0.01	0.03	0.02	0.56
4. Managers				
Neuroticism	0.05	0.09	0.07	0.80
Extraversion	-0.04	0.06	0.04	-0.84
Openness	-0.05	0.06	0.04	-1.16
Agreeableness	0.11	0.15	0.11	0.92
Conscientiousness	0.01	0.01	0.01	1.42
5. Bus drivers				
Adjustment	0.02	0.03	0.02	0.97
Ambition	0.01	0.02	0.01	0.75
Sociability	0.03	0.06	0.04	0.73
Intellectance	-0.15	0.04	0.03	-4.63**
Likeability	0.03	0.06	0.04	0.58
Prudence	0.00	0.03	0.03	-0.14
6. Professionals				
Emotional Orientation	-0.03	0.11	0.08	-0.40
Social Orientation	0.04	0.07	0.06	0.71
Cognitive Orientation	0.09	0.16	0.12	0.76
Interpersonal Orientation	0.01	0.15	0.11	0.07
Task Orientation	-0.03	0.05	0.04	-0.68

Note: ψ = linear regression CVR – neural network CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.
** $p < .01$

Table D3

Study 2. Corrected t-test comparing MAE of linear regression and early stopping neural networks.

Dataset and predictor	ψ	SD	SE	t
<u>1. University students</u>				
Neuroticism	0.016	0.065	0.048	0.33
Extraversion	0.014	0.065	0.048	0.29
Openness	-0.009	0.045	0.034	-0.26
Agreeableness	0.002	0.034	0.025	0.06
Conscientiousness	-0.005	0.068	0.051	-0.09
<u>2. Police recruits</u>				
Neuroticism	-0.031	0.242	0.179	-0.17
Extraversion	-0.069	0.309	0.229	-0.30
Openness	0.004	0.226	0.167	0.03
Agreeableness	-0.131	0.235	0.174	-0.75
Conscientiousness	-0.053	0.178	0.132	-0.40
<u>3. Flight attendants</u>				
Neuroticism	-0.003	0.005	0.004	-0.68
Extraversion	-0.001	0.008	0.006	-0.12
Openness	-0.002	0.007	0.005	-0.50
Agreeableness	-0.006	0.005	0.004	-1.61
Conscientiousness	-0.004	0.006	0.005	-0.88
<u>4. Managers</u>				
Neuroticism	-0.002	0.004	0.003	-0.66
Extraversion	0.001	0.004	0.003	0.44
Openness	-0.001	0.006	0.005	-0.18
Agreeableness	-0.001	0.005	0.003	-0.22
Conscientiousness	-0.001	0.003	0.002	-0.54
<u>5. Bus drivers</u>				
Adjustment	-0.008	0.036	0.026	-0.30
Ambition	-0.008	0.043	0.032	-0.27
Sociability	-0.022	0.034	0.025	-0.86
Intellectance	0.027	0.050	0.037	0.73
Likeability	-0.005	0.022	0.017	-0.32
Prudence	-0.012	0.044	0.033	-0.37
<u>6. Professionals</u>				
Emotional Orientation	-0.002	0.009	0.007	-0.25
Social Orientation	0.001	0.008	0.006	0.09
Cognitive Orientation	-0.003	0.006	0.004	-0.66
Interpersonal Orientation	-0.007	0.013	0.010	-0.73
Task Orientation	-0.002	0.006	0.005	-0.41

Note: ψ = linear regression MAE – early stopping neural network MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D4

Study 2: Corrected t-test comparing cross-validity coefficients (CVR) of linear regression and early stopping neural networks.

Dataset and predictor	ψ	SD	SE	t
1. University students				
Neuroticism	-0.03	0.04	0.03	-1.00
Extraversion	-0.01	0.03	0.02	-0.24
Openness	0.01	0.02	0.01	0.96
Agreeableness	0.00	0.04	0.03	-0.11
Conscientiousness	0.01	0.02	0.01	0.44
2. Police recruits				
Neuroticism	0.01	0.03	0.02	0.47
Extraversion	0.03	0.05	0.04	0.72
Openness	0.01	0.01	0.01	0.94
Agreeableness	0.02	0.05	0.04	0.50
Conscientiousness	0.00	0.01	0.01	0.75
3. Flight attendants				
Neuroticism	0.02	0.02	0.02	0.99
Extraversion	0.01	0.01	0.01	0.85
Openness	0.01	0.02	0.02	0.74
Agreeableness	0.01	0.02	0.02	0.94
Conscientiousness	0.00	0.01	0.01	0.57
4. Managers				
Neuroticism	0.01	0.02	0.02	0.58
Extraversion	-0.04	0.03	0.02	-1.60
Openness	-0.01	0.06	0.04	-0.17
Agreeableness	0.07	0.15	0.11	0.59
Conscientiousness	0.00	0.00	0.00	1.03
5. Bus drivers				
Adjustment	0.01	0.01	0.01	0.68
Ambition	0.00	0.01	0.01	0.53
Sociability	0.03	0.05	0.04	0.72
Intellectance	-0.09	0.06	0.04	-2.06
Likeability	0.05	0.12	0.09	0.61
Prudence	-0.01	0.02	0.01	-0.57
6. Professionals				
Emotional Orientation	-0.03	0.16	0.12	-0.29
Social Orientation	0.04	0.08	0.06	0.74
Cognitive Orientation	0.04	0.09	0.07	0.66
Interpersonal Orientation	-0.04	0.22	0.16	-0.27
Task Orientation	0.03	0.06	0.04	0.65

Note: ψ = linear regression CVR – early stopping neural network CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D5

Study 3. Corrected t-test comparing MAE of linear regression and neural networks developed without early stopping (ANN1).

Dataset	ψ	SD	SE	t
1. University students	-0.154	0.142	0.106	-1.45
2. Police recruits	-0.439	1.141	0.844	-0.52
3. Flight attendants	-0.060	0.035	0.026	-2.32*
4. Managers	-0.018	0.023	0.017	-1.03
5. Bus drivers	-0.061	0.082	0.062	-0.98
6. Professionals	-0.020	0.051	0.038	-0.52

Note: ψ = linear regression MAE – ANN1 MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.
* $p < .05$

Table D6

Study 3. Corrected t-test comparing cross-validity coefficients (CVR) of linear regression and neural networks developed without early stopping (ANN1).

Dataset	ψ	SD	SE	t
1. University students	0.05	0.04	0.03	1.75
2. Police recruits	-0.01	0.05	0.04	-0.34
3. Flight attendants	0.08	0.06	0.04	1.88
4. Managers	0.04	0.09	0.06	0.56
5. Bus drivers	0.02	0.03	0.02	0.80
6. Professionals	-0.10	0.12	0.09	-1.13

Note: ψ = linear regression CVR – ANN1 CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D7

Study 3. Corrected t-test comparing MAE of linear regression an early stopping neural networks (ANN2).

Dataset	ψ	SD	SE	t
1. University students	0.001	0.112	0.083	0.01
2. Police recruits	0.210	0.442	0.327	0.64
3. Flight attendants	-0.016	0.026	0.019	-0.81
4. Managers	-0.001	0.008	0.006	-0.13
5. Bus drivers	-0.013	0.043	0.033	-0.41
6. Professionals	0.007	0.015	0.011	0.67

Note: ψ = linear regression MAE – ANN2 MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D8

Study 3. Corrected t-test comparing cross-validity coefficients (CVR) of linear regression an early stopping neural networks (ANN2).

Dataset	ψ	SD	SE	t
1. University students	0.01	0.02	0.02	0.48
2. Police recruits	-0.01	0.03	0.02	-0.47
3. Flight attendants	0.01	0.04	0.03	0.51
4. Managers	-0.01	0.03	0.02	-0.39
5. Bus drivers	0.00	0.02	0.01	0.33
6. Professionals	-0.03	0.08	0.06	-0.41

Note: ψ = linear regression CVR – ANN2 CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D9

Study 5. Corrected t-test comparing MAE of linear regression and neural networks developed without early stopping (ANN1).

Dataset and predictors	ψ	SD	SE	t
<u>1. University students</u>				
Neuroticism facets	-0.856	0.848	0.630	-1.36
Extraversion facets	-0.902	0.578	0.430	-2.10*
Openness facets	-0.457	0.524	0.390	-1.17
Agreeableness facets	-1.369	1.047	0.778	-1.76
Conscientiousness facets	-1.276	0.907	0.675	-1.89
Theoretical combination	-9.176	1.737	1.292	-7.10**
<u>2. Police recruits</u>				
Neuroticism facets	-1.780	1.854	1.372	-1.30
Extraversion facets	-3.481	4.048	2.995	-1.16
Openness facets	-3.244	1.664	1.231	-2.63*
Agreeableness facets	-3.257	2.102	1.555	-2.09*
Conscientiousness facets	-2.767	1.924	1.423	-1.94
Theoretical combination	-26.497	8.456	6.256	-4.24**
<u>4. Managers</u>				
Neuroticism facets	-0.264	0.180	0.134	-1.97
Extraversion facets	-0.165	0.088	0.065	-2.52*
Openness facets	-0.104	0.136	0.102	-1.02
Agreeableness facets	-0.121	0.093	0.069	-1.74
Conscientiousness facets	-0.073	0.041	0.030	-2.42*
Theoretical combination	-0.492	0.142	0.106	-4.66**

Note: ψ = linear regression MAE – ANN1 MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.
* $p < .05$, ** $p < .01$

Table D10

Study 5. Corrected t-test comparing cross-validity coefficients (CVR) of linear regression and neural networks developed without early stopping (ANN1).

Dataset and predictors	ψ	SD	SE	t
1. University students				
Neuroticism facets	0.05	0.12	0.09	0.51
Extraversion facets	0.11	0.10	0.07	1.47
Openness facets	0.03	0.11	0.08	0.37
Agreeableness facets	0.04	0.12	0.09	0.45
Conscientiousness facets	0.16	0.13	0.09	1.75
Theoretical combination	0.20	0.10	0.08	2.66*
2. Police recruits				
Neuroticism facets	-0.03	0.08	0.06	-0.49
Extraversion facets	0.13	0.11	0.08	1.59
Openness facets	0.07	0.09	0.06	1.09
Agreeableness facets	0.05	0.11	0.08	0.57
Conscientiousness facets	0.07	0.08	0.06	1.31
Theoretical combination	0.07	0.13	0.09	0.78
4. Managers				
Neuroticism facets	0.20	0.17	0.12	1.62
Extraversion facets	0.10	0.16	0.12	0.87
Openness facets	0.10	0.10	0.08	1.27
Agreeableness facets	-0.04	0.14	0.10	-0.42
Conscientiousness facets	0.20	0.11	0.08	2.47*
Theoretical combination	0.18	0.17	0.13	1.37

Note: ψ = linear regression CVR – ANN1 CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.
* $p < .05$

Table D11

Study 5. Corrected t-test comparing MAE of linear regression an early stopping neural networks (ANN2).

Dataset and predictors	ψ	SD	SE	t
<u>1. University students</u>				
Neuroticism facets	-0.032	0.159	0.118	-0.27
Extraversion facets	-0.021	0.141	0.105	-0.20
Openness facets	0.015	0.115	0.086	0.18
Agreeableness facets	-0.109	0.167	0.124	-0.88
Conscientiousness facets	-0.068	0.167	0.124	-0.55
Theoretical combination	-0.096	0.242	0.180	-0.54
<u>2. Police recruits</u>				
Neuroticism facets	0.000	0.606	0.448	0.00
Extraversion facets	-0.282	0.569	0.421	-0.67
Openness facets	-0.415	0.398	0.295	-1.41
Agreeableness facets	-0.453	0.570	0.422	-1.07
Conscientiousness facets	-0.088	0.540	0.400	-0.22
Theoretical combination	0.635	0.833	0.616	1.03
<u>4. Managers</u>				
Neuroticism facets	-0.002	0.009	0.007	-0.33
Extraversion facets	0.000	0.008	0.006	0.04
Openness facets	0.001	0.009	0.006	0.13
Agreeableness facets	0.004	0.010	0.007	0.51
Conscientiousness facets	-0.004	0.008	0.006	-0.61
Theoretical combination	0.010	0.012	0.009	1.11

Note: ψ = linear regression MAE – ANN2 MAE. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Table D12

Study 5. Corrected t-test comparing cross-validity coefficients (CVR) of linear regression an early stopping neural networks (ANN2).

Dataset and predictors	ψ	SD	SE	t
<u>1. University students</u>				
Neuroticism facets	0.00	0.04	0.03	0.05
Extraversion facets	0.02	0.07	0.05	0.40
Openness facets	0.00	0.04	0.03	0.15
Agreeableness facets	0.03	0.06	0.04	0.76
Conscientiousness facets	0.03	0.05	0.04	0.71
Theoretical combination	0.02	0.04	0.03	0.80
<u>2. Police recruits</u>				
Neuroticism facets	0.00	0.05	0.04	0.06
Extraversion facets	0.05	0.07	0.05	1.07
Openness facets	0.02	0.03	0.02	1.00
Agreeableness facets	0.04	0.05	0.04	1.15
Conscientiousness facets	0.02	0.04	0.03	0.58
Theoretical combination	0.00	0.04	0.03	-0.07
<u>4. Managers</u>				
Neuroticism facets	0.03	0.05	0.04	0.72
Extraversion facets	0.02	0.03	0.02	0.83
Openness facets	0.02	0.04	0.03	0.70
Agreeableness facets	-0.06	0.07	0.06	-1.01
Conscientiousness facets	0.02	0.03	0.02	0.71
Theoretical combination	-0.01	0.05	0.04	-0.20

Note: ψ = linear regression CVR – ANN2 CVR. The standard error (SE) is obtained by applying the adjustment outlined in Appendix B. Degrees of freedom = 19 for all comparisons.

Appendix E: Polynomial Regression Results

Table E1

Dataset 1: Polynomial regression results.

Predictor and step	R^2	ΔR^2	df	t
<u>Neuroticism</u>				
Step 1: Linear	.003	.003	225	0.78
Step 2: Quadratic	.014	.011	224	-1.57
Step 3: Cubic	.015	.001	223	0.53
<u>Extraversion</u>				
Step 1: Linear	.017	.017	225	-1.96
Step 2: Quadratic	.024	.007	224	-1.30
Step 3: Cubic	.024	.000	223	-0.18
<u>Openness</u>				
Step 1: Linear	.015	.015	225	1.82
Step 2: Quadratic	.015	.000	224	0.33
Step 3: Cubic	.015	.000	223	0.21
<u>Agreeableness</u>				
Step 1: Linear	.001	.001	225	0.49
Step 2: Quadratic	.002	.001	224	0.36
Step 3: Cubic	.002	.000	223	-0.24
<u>Conscientiousness</u>				
Step 1: Linear	.040	.040	225	3.06**
Step 2: Quadratic	.044	.004	224	0.99
Step 3: Cubic	.048	.004	223	-0.99

** $p < .01$

Table E2

Dataset 2: Polynomial regression results.

Predictor and step	R^2	ΔR^2	df	t
Neuroticism				
Step 1: Linear	.026	.026	284	-2.74**
Step 2: Quadratic	.027	.001	283	-0.56
Step 3: Cubic	.028	.001	282	0.68
Extraversion				
Step 1: Linear	.026	.026	284	2.74**
Step 2: Quadratic	.032	.006	283	1.32
Step 3: Cubic	.044	.012	282	-1.94
Openness				
Step 1: Linear	.011	.011	284	1.75
Step 2: Quadratic	.011	.000	283	0.36
Step 3: Cubic	.019	.008	282	-1.49
Agreeableness				
Step 1: Linear	.012	.012	284	1.85
Step 2: Quadratic	.013	.001	283	-0.54
Step 3: Cubic	.015	.002	282	-0.71
Conscientiousness				
Step 1: Linear	.075	.075	284	4.79**
Step 2: Quadratic	.075	.000	283	0.03
Step 3: Cubic	.076	.001	282	-0.68

** $p < .01$

Table E3

Dataset 3: Polynomial regression results.

Predictor and step	R^2	ΔR^2	df	t
Neuroticism				
Step 1: Linear	.014	.014	303	-2.09*
Step 2: Quadratic	.014	.000	302	-0.23
Step 3: Cubic	.015	.001	301	0.58
Extraversion				
Step 1: Linear	.023	.023	303	2.68**
Step 2: Quadratic	.025	.002	302	0.69
Step 3: Cubic	.025	.000	301	0.31
Openness				
Step 1: Linear	.022	.022	303	2.62**
Step 2: Quadratic	.024	.002	302	0.68
Step 3: Cubic	.025	.001	301	-0.77
Agreeableness				
Step 1: Linear	.030	.030	303	3.04**
Step 2: Quadratic	.030	.000	302	-0.25
Step 3: Cubic	.036	.006	301	-1.39
Conscientiousness				
Step 1: Linear	.013	.013	303	1.98*
Step 2: Quadratic	.015	.002	302	0.74
Step 3: Cubic	.015	.000	301	0.38

* $p < .05$, ** $p < .01$

Table E4

Dataset 4: Polynomial regression results.

Predictor and step	R^2	ΔR^2	df	t
Neuroticism				
Step 1: Linear	.039	.039	177	-2.69**
Step 2: Quadratic	.039	.000	176	-0.11
Step 3: Cubic	.042	.003	175	-0.66
Extraversion				
Step 1: Linear	.011	.011	177	1.43
Step 2: Quadratic	.028	.017	176	1.74
Step 3: Cubic	.032	.004	175	-0.87
Openness				
Step 1: Linear	.000	.000	177	0.28
Step 2: Quadratic	.005	.005	176	0.87
Step 3: Cubic	.009	.004	175	0.85
Agreeableness				
Step 1: Linear	.006	.006	177	-1.07
Step 2: Quadratic	.011	.005	176	0.87
Step 3: Cubic	.011	.000	175	-0.34
Conscientiousness				
Step 1: Linear	.076	.076	177	3.82**
Step 2: Quadratic	.076	.000	176	0.17
Step 3: Cubic	.077	.001	175	0.39

** $p < .01$

Table E5

Dataset 5: Polynomial regression results.

Predictor and step	R^2	ΔR^2	df	t
<u>Adjustment</u>				
Step 1: Linear	.007	.007	484	1.80
Step 2: Quadratic	.007	.000	483	0.62
Step 3: Cubic	.007	.000	482	-0.09
<u>Ambition</u>				
Step 1: Linear	.011	.011	484	2.33*
Step 2: Quadratic	.012	.001	483	-0.58
Step 3: Cubic	.012	.000	482	-0.38
<u>Sociability</u>				
Step 1: Linear	.000	.000	484	-0.19
Step 2: Quadratic	.000	.000	483	0.35
Step 3: Cubic	.000	.000	482	-0.11
<u>Intellectance</u>				
Step 1: Linear	.000	.000	484	0.08
Step 2: Quadratic	.005	.005	483	-1.56
Step 3: Cubic	.020	.015	482	2.75**
<u>Likeability</u>				
Step 1: Linear	.009	.009	484	2.12*
Step 2: Quadratic	.011	.002	483	0.96
Step 3: Cubic	.013	.002	482	0.87
<u>Prudence</u>				
Step 1: Linear	.021	.021	484	3.19**
Step 2: Quadratic	.026	.005	483	-1.68
Step 3: Cubic	.027	.001	482	0.58

* $p < .05$, ** $p < .01$

Table E6

Dataset 6: Polynomial regression results.

	R^2	ΔR^2	df	t
Emotional Orientation				
Step 1: Linear	.007	.007	118	0.92
Step 2: Quadratic	.021	.014	117	1.28
Step 3: Cubic	.021	.000	116	-0.09
Social Orientation				
Step 1: Linear	.005	.005	118	-0.77
Step 2: Quadratic	.005	.000	117	-0.14
Step 3: Cubic	.005	.000	116	0.11
Cognitive Orientation				
Step 1: Linear	.016	.016	118	-1.37
Step 2: Quadratic	.016	.000	117	-0.29
Step 3: Cubic	.038	.022	116	-1.61
Interpersonal Orientation				
Step 1: Linear	.012	.012	118	-1.20
Step 2: Quadratic	.017	.005	117	-0.73
Step 3: Cubic	.017	.000	116	0.18
Task Orientation				
Step 1: Linear	.000	.000	118	0.04
Step 2: Quadratic	.004	.004	117	0.65
Step 3: Cubic	.004	.000	116	-0.28

Appendix F: Moderated Multiple Regression Results

Table F1.

Dataset 1. Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
N x O	.020	.000	223	-0.24
N x C	.057	.000	223	0.28
O x C	.053	.000	223	-0.02
N x O x C	.080	.004	219	1.04

Note. For each product term, only the results associated with the last step of the analysis are provided.
N = Neuroticism, O = Openness, C = Conscientiousness.

Table F2.

Dataset 2. Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
N x E	.049	.008	282	-1.56
N x O	.037	.005	282	1.23
N x C	.076	.001	282	0.45
E x O	.035	.007	282	-1.44
E x C	.082	.000	282	0.05
O x C	.081	.000	282	0.36
N x E x O	.076	.010	278	1.73
N x E x C	.115	.021	278	2.53*
N x O x C	.102	.014	278	2.12*
E x O x C	.115	.030	278	-2.69**
N x E x O x C	.169	.001	270	-0.48

Note. For each product term, only the results associated with the last step of the analysis are provided.
N = Neuroticism, E = Extraversion, O = Openness, C = Conscientiousness.
* $p < .05$, ** $p < .01$

Table F3.

Dataset 3: Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
N x E	.027	.000	301	0.05
N x O	.032	.000	301	0.15
N x A	.031	.000	301	0.03
N x C	.020	.002	301	-0.78
E x O	.032	.001	301	0.47
E x A	.036	.001	301	0.35
E x C	.029	.002	301	0.72
O x A	.036	.000	301	0.29
O x C	.030	.000	301	-0.33
A x C	.036	.005	301	1.30
N x E x O	.038	.001	297	0.58
N x E x A	.047	.011	297	1.85
N x E x C	.037	.003	297	1.03
N x O x A	.043	.005	297	1.24
N x O x C	.038	.001	297	0.54
N x A x C	.040	.001	297	0.61
E x O x A	.043	.002	297	-0.78
E x O x C	.041	.001	297	-0.45
E x A x C	.048	.002	297	-0.79
O x A x C	.049	.001	297	-0.53
N x E x O x A	.059	.002	289	0.47
N x E x O x C	.058	.001	289	-0.63
N x E x A x C	.056	.000	289	-0.11
N x O x A x C	.071	.007	289	-1.45
E x O x A x C	.073	.005	289	-1.24
N x E x O x A x C	.113	.001	273	0.08

Note. For each product term, only the results associated with the last step of the analysis are provided.
N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness.

Table F4.

Dataset 4: Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
N x E	.063	.020	175	-1.91
N x C	.085	.000	175	0.10
E x C	.080	.001	175	-0.09
N x E x C	.119	.008	171	1.23

Note. For each product term, only the results associated with the last step of the analysis are provided.
N = Neuroticism, E = Extraversion, C = Conscientiousness.

Table F5.

Dataset 5: Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
Adj. x Lik.	.016	.005	482	1.49
Adj. x Pru.	.021	.000	482	-0.41
Lik. x Pru.	.022	.000	482	-0.16
Adj. x Lik. x Pru.	.030	.001	478	-0.81

Note. For each product term, only the results associated with the last step of the analysis are provided.
Adj. = Adjustment, Lik. = Likeability, Pru. = Prudence.

Table F6.

Dataset 6: Moderated multiple regression results.

Product term	R^2	ΔR^2	df	t
EO x CO	.062	.043	116	-2.30*
EO x TO	.055	.047	116	-2.40*
CO x TO	.018	.001	116	0.28
EO x CO x TO	.118	.026	112	-1.79

Note. For each product term, only the results associated with the last step of the analysis are provided.
EO = Emotional Orientation, CO = Cognitive Orientation, TO = Task Orientation.
* $p < .05$