11-2009

Personality scale validities increase throughout medical school

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Deniz S. ONES
University of Minnesota - Twin Cities

Stephan DILCHERT
CUNY Bernard M Baruch College

DOI: https://doi.org/10.1037/a0016137

Published in Journal of Applied Psychology,
2009 September, Volume 94, Issue 6, Pages 1514-1535
http://doi.org/10.1037/a0016137

Personality Scale Validities Increase Throughout Medical School

Filip Lievens
Ghent University

Deniz S. Ones
University of Minnesota

Stephan Dilchert
Baruch College, CUNY

Admissions and personnel decisions rely on stable predictor–criterion relationships. The authors studied the validity of Big Five personality factors and their facets for predicting academic performance in medical school across multiple years, investigating whether criterion-related validities change over time. In this longitudinal investigation, an entire European country's 1997 cohort of medical students was studied throughout their medical school career (Year 1, $N = 627$; Year 7, $N = 306$). Over time, extraversion, openness, and conscientiousness factor and facet scale scores showed increases in operational validity for predicting grade point averages. Although there may not be any advantages to being open and extraverted for early academic performance, these traits gain importance for later academic performance when applied practice increasingly plays a part in the curriculum. Conscientiousness, perhaps more than any other personality trait, appears to be an increasing asset for medical students: Operational validities of conscientiousness increased from .18 to .45. In assessing the utility of personality measures, relying on early criteria might underestimate the predictive value of personality variables. Implications for personality measures to predict work performance are discussed.

Keywords: personality validity, longitudinal validation, grades, medical school, professional education

Humphreys (1960) stated five decades ago that "in selection research one should not be satisfied with validation of predictors against the earliest possible criteria" (p. 318). Selection and admissions decisions in organizations rely on stable relationships between predictors and criteria. That is, it is assumed that performance differences between those high and low on various predictor constructs remain relatively stable across the years and that predictor–criterion relationships are of similar strength for individuals of differing tenure. Potential changes in validity coefficients have an impact on the expected utility of selection systems. In personnel selection as well as academic admissions, although we aim to predict criteria of interest (typically performance) over a relatively long duration, the time spans over which criteria are gathered for validation studies most often reflect practical considerations (Sussmann & Robertson, 1986). In predictive studies, the time periods selected for gathering criterion data rarely exceed a year or two—in most cases they are merely a few months. In concurrent studies, criterion scores are often obtained from both newly selected individuals and individuals of varying tenure levels.

Even though there exists a large literature on cognitive ability tests that is directed at whether such tests retain their predictive value in the long run (Barrett, Phillips, & Alexander, 1981; Campbell & Knapp, 2001; Deadrick & Madigan, 1990; F. L. Schmidt, Hunter, Outerbridge, & Goff, 1988), few studies have focused on the long-term predictive validity of noncognitive predictors such as experience, interests, biographical data, assessment center dimension ratings, or personality variables.

In the personality domain, two studies that have examined the relationships between personality and long-term success have focused on income and occupational status (Dodd, Wollowick, & McNamara, 1970; Judge, Higgins, Thoresen, & Barrick, 1999). These studies are extremely valuable in demonstrating the usefulness of personality for predicting long-term *career success*, but by design, their relevance for understanding the validity and utility of personality for predicting *performance* within a given setting (organizational or educational) is limited. To our knowledge, research that has examined the stability of criterion-related validities of personality scales for predicting performance over longer time frames is nonexistent. As will be detailed below, such long-term validation of the stability of personality validities is of key conceptual and practical importance because it would enable theory-driven tests of personality–performance relationships.

The primary objective of this research was to examine the validities of the Big Five personality dimensions (neuroticism, extraversion, openness, agreeableness, and conscientiousness) and their facets longitudinally. We studied the validity of Big Five

Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Ghent, Belgium; Deniz S. Ones, Department of Psychology, University of Minnesota; Stephan Dilchert, Department of Management, Zicklin School of Business, Baruch College, City University of New York.

Order of authorship is arbitrary; all authors contributed equally to this article.

Portions of this research were presented at the annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA, April 2005, and the International Symposium on Personality at Work, Lüneburg, Germany, May 2005.

Correspondence concerning this article should be addressed to Deniz S. Ones, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. E-mail: deniz.s.ones-1@tc.umn.edu

personality factors and their facets for academic performance in all Flemish medical schools across multiple years, investigating whether criterion-related validities change across the entire span of the curriculum. We also explored personality differences between those who leave medical school and those who stay to graduate. Specifically, we examined whether personality determinants of attrition are different early compared with late in medical school. Predicting individuals' success in medical school is of prime importance for two reasons. First, medical education is expensive, and therefore, better selection and admission decisions can help minimize educational costs by decreasing the proportion of dropouts. Second, medical education is professional training to acquire knowledge and skills that will enable individuals to perform their job. These skills are potentially relied upon for careers spanning several decades. Which personality traits predict persistence throughout the various stages of professional education is thus another important question.

Theoretical Background

Research on Longitudinal Changes in Criteria

Although until now the issue of longitudinal validity remained unexamined for personality measures, there is a large literature directed at examining whether cognitive variables retain their predictive value over time. This literature is closely linked to the literature on dynamic criteria. The concept of dynamic criteria first proposed by Ghiselli (1956) refers to "changes in the rank-ordering of individuals in their performance over time" (Barrett, Caldwell, & Alexander, 1985, p. 51) and has been examined in both the educational and work psychology literatures. Humphreys (1960) noted that past behavior is less than perfectly correlated with future behavior and that the "further apart the measures are in the series the lower will be the correlation between them" (p. 315). This observation builds on Guttman's (1955) description of a simplex pattern of correlations.

For longitudinal changes in predictor validity, two primary explanations have been proposed: Either individuals change over time (which would mean that their behavior would change to reflect this change), or the tasks and work being performed change (Alvares & Hulin, 1972). Of these explanations, the changing ability/person explanation has now been largely rejected based on available data. Postdictive validities appear to follow the same patterns of changes as predictive validities (Humphreys & Taber, 1973; Lunneborg & Lunneborg, 1970). As Humphreys and Taber (1973) put it, "The hypothesis that reduced predictive validities and instability of grades is due to change in the broad abilities . . . must be rejected" (p. 181). Similar arguments of stability can be made for personality traits that conceptually reflect stable individual differences. Recent meta-analytic evidence (Fraley & Roberts, 2005; Roberts & DelVecchio, 2000) suggests that rank-order stability is remarkably high (Caspi, Roberts, & Shiner, 2005).

Changes in job or task requirements are likely at the root of "changing patterns of job demands" (Barrett et al., 1985, p. 42), with many authors explaining declines in relationships between increasingly distal criteria with changing tasks and activities (e.g., Alvares & Hulin, 1972; Dunham, 1974; Ghiselli & Haire, 1960; Woodrow, 1938). For academic performance, Humphreys (1960) concluded that "as a matter of fact the subjects would not need to change at all if changes in course content were sufficiently systematic. A gradual shift in emphasis from verbal to quantitative materials, for example, would produce a matrix resembling a simplex" (p. 320). Several variants of the changing tasks model have also been incorporated into more recent theories of skill acquisition in the cognitive domain (e.g., Ackerman, 1987). This literature suggests that the temporal stability of predictor–criterion relationships for cognitive variables differs across types of abilities (general mental ability, psychomotor ability, perceptual ability), settings (educational, work), and types of work (consistent/inconsistent task performance, academic performance, job performance; Keil & Cortina, 2001).

Changes in tasks and work in general can mean that performance requirements may change (Jenkins, 1946) or that the relative importance of performance predictors may vary over time (Deadrick & Madigan, 1990). No investigations on this matter exist for the domain of personality, even though there are conceptual and empirical lines of reasoning that lead us to expect most personality scale validities would increase over time.

There are two main conceptual arguments to support the hypothesis that in predicting performance, validities for personality variables would generally increase over time. First, during initial stages of tenure in a new environment, the "novelty and challenge of the task or job" (Helmreich, Sawin, & Carsrud, 1986, p. 185) tend to reduce naturally occurring motivational differences among individuals. Thus, during the honeymoon phase, personality traits such as achievement orientation, for example, are likely to play less of a role in predicting performance. However, in the long run, honeymoon effects might wear off and motivationally based personality characteristics could therefore show greater predictive validities (Helmreich et al., 1986). This first conceptual argument can be complementarily framed in terms of trait activation theory (Lievens, Chasteen, Day, & Christiansen, 2006; Tett & Burnett, 2003). From a trait activation point of view, the change in the criterion with the increasing emphasis on different behavioral requirements suggests changing performance demands for the expression of various traits over time. That is, the changing demands of the performance situation are likely to influence personality–performance linkages. In the honeymoon stage, the performance demands might also be strong such that full expression of behaviors are constrained, masking the driving force of underlying personality traits. Conversely, when the honeymoon effects wear off and performance demands become weaker or shift into different behavioral domains, only people high on specific personality traits might pick up the relevant situational cues, facilitating the observation of variability in individual differences.

Second, as will be discussed below, in professional education (e.g., medical school), earlier courses (i.e., proximal criteria) typically focus on the acquisition of declarative and procedural knowledge, whereas later courses (i.e., distal criteria) place more emphasis on applications and internship performance. The behavior that needs to be displayed for performance in such courses is more complex, as it relies much more on interpersonal interactions. On the basis of the predictor–criterion matching logic (Campbell, McCloy, Oppler, & Sager, 1993; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990), it might then be expected that different personality factors will be predictive for proximal compared with distal criteria. Due to the changing content of medical courses over time (i.e., from acquisition of knowledge to

performance with mock patients or in the clinic), later grades in medical school may be better predicted by personality traits than earlier grades.

These two conceptual arguments suggest that personality variables can be stronger determinants of increasingly applied performance that occurs later in professional education. From an empirical point of view, Barrett and Alexander (1989) cited several examples of increasing criterion-related validities for noncognitive predictors (e.g., Dodd et al., 1970; Helmreich et al., 1986; Singh, 1978). For noncognitive predictors studied thus far, such as assessment center ratings, criterion-related validities tend to be better in the long run. For example, Hinrichs (1978) found that assessment center ratings predicted organizational level better 8 years postassessment than 1 year postassessment. Similar patterns of results have been reported by Bray and Howard (1983) and more recently by Jansen and Stoop (2001). However, parallel investigations for personality scales are lacking.

Academic Performance in Medical School: A Theoretical Distinction Between Early and Late Performance

In this study, we examine personality determinants of success across seven years, the entire duration of medical education in Belgium. Changing tasks characterize education in general and medical education in particular, providing the possibility that personality determinants of behavior could vary at different points in time.

Performance in educational settings in general can be described broadly as acquisition and applications of declarative and procedural knowledge. Performance in professional education in particular (e.g., law, business administration, pharmacy, medical school) also includes elements of applied practice (e.g., internships). Medical education specifically includes both theoretical coursework and large components of clinical practice (Curry & Makoul, 1998; Dusek & Bates, 2003; Laidlaw, MacLeod, Kaufman, Langille, & Sargeant, 2002; H. G. Schmidt et al., 1996). Conventional medical training can be divided into preclinical and clinical years (Ferguson, Sanders, O'Hehir, & James, 2000, p. 331; Lipton, 1988; Lipton, Huxham, & Hamilton, 1984). Early performance in medical school consists of performance in basic science courses (e.g., gross anatomy, biochemistry, physiology, microbiology) and is assessed by scores on exams that aim to assess primarily declarative knowledge. Later years in medical school are referred to as clinical years, and academic performance during these years is heavily influenced by clerkships during which students are trained with hands-on approaches. Interactions with patients, oral examinations, and ward evaluations typically form the basis of grades (Turner, Helper, & Kriska, 1974). As performance during medical school changes, it can be anticipated that characteristics that are related to working well with others (patients, supervisors, peers) would gain in importance over the years spent in medical school.

Our study's medical school performance criterion (and its changing nature) fits very well with these descriptions. The curriculum of Belgian medical schools as it was operational at the start of this study (in 1997) was a so-called conventional medical curriculum with the following format: In the first year, basic science courses (physics, chemistry, biology, biochemistry) and statistics (mathematics) were taught. In these theoretical courses, medical students were provided with the factual knowledge that served as a foundation for subsequent years. If practical exercises were included, they typically took place in the laboratory (e.g., practical chemistry labs). In terms of teaching format, courses in the first year were lecture-based and given to large groups. Evaluations consisted of written exams, testing whether students acquired the necessary knowledge.

From the second year on, medical courses were taught. In the second and third years, these medical courses focused on the healthy human body (courses such as anatomy, physiology, and histology). In the fourth and fifth years, the emphasis shifted to pathologies and the mechanisms that cause disease (courses such as internal medicine, infectious diseases, pediatrics, surgery, gynecology, and ophthalmology). Another important change in course content was that in the second and third years, clinical experiences were limited to patient contact in introductory courses on patient interviewing, whereas in the fourth and fifth years, multidisciplinary and communication skills courses were included to prepare students for clinical and professional practice. In the sixth and seventh years, several hospital-based clinical clerkships were included. This clerkship program was divided into various rotations (e.g., Children and Youth, Surgery, Primary Care), with 2 to 4 months spent in each unit, and thus introduced applied practice and hands-on experience with patients into the curriculum.

This gradual change in focus from knowledge acquisition to applied practice was also exemplified by the differences in teaching and examination format between the second and seventh years. Specifically, traditional lecture-based teaching methods gradually decreased in importance in favor of small group interactive learning. The latter is exemplified by case-based teaching, clinical case studies, and experiential learning. The examination format reflected this more applied focus in the clinical years with case-based examinations and even clinical practice examinations with "standardized" patients.

Thus, the changing nature of our criterion (grade point average [GPA] in medical schools) can be summarized as follows. In the first year, GPA was heavily determined by knowledge acquisition within science courses given in large groups. In the next six years, there was a gradual shift from relatively minor clinical interaction with patients to clinical practice being the primary component of medical education, something that was exemplified by the clerkship program at the end.

We believe that as individuals gain experience and expertise in their fields, validities of occupationally relevant personality traits will increase. Later grades in medical school may be better predicted by personality traits than earlier grades, because personality variables are potentially stronger determinants of increasingly applied performance that occurs later in professional education. Such an increasing validities hypothesis also means that academic performance of groups scoring high and low on various personality traits should diverge throughout medical school. If our general hypothesis of increasing validities is confirmed, the key implication for validation studies is that validation of personality measures against early criteria would underestimate true criterion-related validities and utility of personality scales in such settings.

Attrition and Persistence in Medical School

In addition to predicting early learning success and later interpersonal (practicum) performance among medical school students,

the simple fact of whether students persist through their education is an important criterion of interest. Attrition from medical schools has been the focus of research attention for 140 years. Paget (1869, cited in Gough & Hall, 1975) provided the first summary of base rates of attrition from a medical school in London, in an article titled "What Becomes of Medical Students?" Paget followed 1,000 medical students for a period of 10 years, reporting that only 720 were still in medicine at the end of that time. Gough and Hall (1975) provided an excellent summary of the literature on persistence versus attrition in medical school; readers interested in base rates of medical school attrition and potential causes are directed to that source. Very briefly, it appears that around the world over the past 100 years, medical school attrition rates have ranged approximately between 1.5% and 25%; "an attrition rate of from 5 percent to 10 percent would appear to be 'normal'" (Gough & Hall, 1975, p. 942). In addition, attrition appears to be most closely related to academic performance and cognitive ability (Gough & Hall, 1975). The few studies that have examined personality predictors of medical school attrition have found that dropouts had higher scores on hysteria (Knehr & Kohl, 1959); lower scores on achievement (D. G. Johnson & Hutchins, 1966); higher scores on nurturance and succorance; and lower scores on responsibility, socialization, and communality (Gough & Hall, 1975). However, this research did not distinguish between early and late attrition in students' academic careers. In this study we examine personality variables and attrition across multiple years of medical school.

Theoretical and Empirical Relevance of Big Five Personality Factors and Facets for Performance in Medical School

To date, the literature on the validity of personality scales for prediction of medical school success longitudinally has been rather limited. One relevant study on the topic included only three years of academic performance data and focused on dysfunctional, dark side personality traits (Knights & Kennedy, 2007). The results from that study suggest that "moving away" traits (i.e., paranoid, avoidant, schizoid, borderline, and passive/aggressive) and "moving against" traits (i.e., narcissistic, antisocial, histrionic, and schizotypal) increased in predictive validity across the three years examined (see Knights & Kennedy's, 2007, Table 2). In the current study we investigate validities of normal-range personality traits over longer periods of academic performance. Although our general hypothesis was that personality validities would gain in importance over time, we were open to the idea that this might not be true for all Big Five traits and their facets (see, e.g., Stewart, 1999).

We next present our theoretical considerations of specific linkages between personality factors and facets and early as well as late medical school performance. Where applicable, we note our expectations for medical school attrition as well.

Neuroticism. Neuroticism is the negative pole of emotional stability and describes an individual's tendency to become emotionally upset. Common attributes defining this dimension are anxiety, depression, anger, embarrassment, emotion, worry, fearfulness, instability, and insecurity. Individuals who score high on neuroticism are self-conscious and low on self-esteem and generalized self-efficacy (Judge, Erez, Bono, & Thoresen, 2002). In academic settings, such individuals would be more likely to give up easily and would have problems with approaching difficult tasks. Individuals who are high on neuroticism also are typically vulnerable and employ maladaptive coping strategies in dealing with stressful situations. For example, they become easily disheartened and panicked. In educational settings, one would expect Neuroticism to be negatively related to academic performance and positively related to attrition. Eysenck (1967) pointed out that neurotic individuals tend to score lower on high-stakes tests. This is possibly due to their higher propensity to experience extreme levels of anxiety, including test anxiety, in evaluative settings (e.g., Furnham & Mitchell, 1991; Zeidner, 1995; Zeidner & Matthews, 2000). Previous research has found positive relations between emotional stability and performance in undergraduate classes (e.g., Cattell & Kline, 1977; Goh & Moore, 1978; Lathey, 1991; Sanchez, Rejano, & Rodriguez, 2001; Savage, 1962) as well as in medical school (Barratt & White, 1969).

Medicine is an emotionally demanding field (Marley & Carman, 1999). Students being trained in a medical program need the emotional resources to cope with the general pressures of academic performance and evaluation (i.e., exams) and specific pressures of medical education (e.g., dealing with sick people). The facets that are likely to be useful in the prediction of these performance criteria are anxiety, self-consciousness, and vulnerability. Previous research supports these hypotheses. In a study focusing on medical students, Grover and Smith (1981) found a correlation of –.48 between anxiety and overall GPA. Gough, Bradley, and McDonald (1991) reported an observed correlation of .24 between well-being and clinical performance of anesthesiology residents.

On the one hand, one could expect that for early academic success, emotional stability would be a greater asset, as anxiety is likely to debilitate performance on traditional examinations and, thus, early medical school performance. On the other hand, added stress of applied practice (e.g., patients' lives depending on the treatments prescribed) will also require a stable emotional constitution. Thus, we hypothesize neuroticism validities to remain stable across early and late medical school performance (Hypothesis 1).

Extraversion. Extraversion is defined as a person's capacity for joy and the tendency to seek interpersonal stimulation. Traits relating to sociability, dominance, energy, and positive affect constitute the domain of extraversion. Adjectives such as energetic, active, vigorous, talkative, assertive, fun-loving, gregarious, persuasive, and positive describe individuals high on this trait. Extraverts tend to be socially dominant. They seek situations where they can interact with others. Rolfhus and Ackerman (1999) found that extraverts scored lower on knowledge tests than introverts. The theoretical interpretation offered for this finding is that knowledge acquisition time differs between extraverts and introverts. In learning environments, extraverts "spend more time socializing" (Chamorro-Premuzic & Furnham, 2003, p. 321) rather than devoting their cognitive resources to knowledge acquisition. Extraverts tend to be more distractible, impulsive, and sociable than introverts. Introverts tend to be able to focus better than extraverts on cognitively demanding tasks, have better study skills (Entwistle & Entwistle, 1970), and therefore tend to receive better grades (see, for example, Broadbent, 1958; Sanchez et al., 2001; also see Furneaux, 1957, as cited in De Raad & Schouwenburg, 1996). Perhaps Eysenck (1992) put it best when he concluded that the "extravert socializes, instead of concentrating on his work,

seeks nonacademic outlets (sports, sex) for his energies, and has difficulty in concentrating" (p. 137, cited in De Raad & Schouwenburg, 1996). On the flipside, there is some evidence that extraverts tend to receive higher evaluations in seminar classes (Furnham & Medhurst, 1995), suggesting that they do better in settings that require interpersonal interactions. Extraversion can be expected to be "associated with success when interpersonal skills are required" (Ferguson et al., 2000, p. 324).

On the basis of the foregoing theoretical arguments, we expect that extraversion will be negatively related to early medical school performance. Classes during the preclinical years require concentrated study (e.g., memorization of facts) and individual preparation. During this period, we anticipate extraversion, especially the sociability and excitement-seeking aspects of the trait, to be a liability among medical students. Clinical years of medical education, on the other hand, require greater interactions with both patients and colleagues. Extraverts thrive in social situations in which they can interact with others. Being socially ascendant, affectionate, and warm appears to be more important for later grades that are based on practicums and clerkships. Thus, we expect extraverts to demonstrate better academic performance later in medical school, during clinical years.

In general, then, Hypothesis 2 is that the validities for extraversion would evidence changes from negative to positive throughout medical school. There is one aspect of extraversion that is likely to be equally useful throughout medical school: stamina and energy. Medical school is rigorous and demanding; achieving high grades and performance among other highly able students is likely to require high levels of energy, including the ability to perform well with little sleep. Thus, achievements in medical school are also likely to be related to the energy aspect of extraversion both early and late.

Openness. Openness describes individual differences in imagination, curiosity, originality, broadmindedness, and intellectance. Openness is conceptualized as influencing the breadth and complexity of mental experiences of individuals. Open individuals have wide interests, are imaginative, curious, creative, insightful, and perceived as more intelligent by others. Due to its associations with cognitive ability, particularly vocabulary (Goff & Ackerman, 1992) and divergent thinking (McCrae, 1996), openness has been empirically linked to academic performance (e.g., Blickle, 1996). Its facets of openness to aesthetics and ideas are more closely related to lexical intellect than other facets of the dimension (J. A. Johnson, 1994; Saucier, 1994). Open individuals are also characterized by curiosity and intellectual engagement in cognitive tasks. Curiosity and engagement in intellectual pursuits are requisite attributes in learning environments (Rocklin, 1994), and thus openness is likely to be generally predictive of grades in medical education.

Validities for the openness dimension of the Big Five have usually been poor for predicting performance (Barrick, Mount, & Judge, 2001) and educational success (Hough, 1992). Inconsistent validities have been reported for this personality construct in predicting training performance. A meta-analysis by Barrick and Mount (1991) found a true score correlation of .25 ($N = 2{,}700$, $k = 14$) with training success, whereas Hough (1992) reported a mean observed validity of .02 ($N = 8{,}744$, $k = 35$). Recent studies examining facets of openness have reported more encouraging results (Griffin & Hesketh, 2004).

There are reasons to expect openness validities to increase throughout medical school education. Openness is a trait that is relevant to adapting more adequately to novel and unforeseen changes (LePine, Colquitt, & Erez, 2000). Students are much more likely to encounter new situations during their clinical years. Nonetheless, different aspects of openness may be valuable for predicting academic success in different stages of medical education, including early years. Openness to ideas might be predictive of preclinical year grades, whereas openness to feelings might be more predictive of later grades that are based on interactions with patients. Thus, we hypothesized that the criterion-related validities associated with openness to feelings would increase during medical school (Hypothesis 3).

Agreeableness. Agreeableness is the dimension of the Big Five describing the tendency to help others and behave in prosocial ways. It is a trait that is important in characterizing how people behave in interpersonal interactions. Agreeable individuals are cooperative, nurturing, affectionate, sensitive, caring, altruistic, kind, tender minded, and softhearted. Individuals scoring low on this dimension are described as uncooperative, unfriendly, selfish, hostile, and egocentric. There have been at least two studies that have demonstrated the usefulness of agreeableness-related personality traits in medical education, all focusing on clinical performance as the criterion (Gough et al., 1991; Shen & Comrey, 1997). These studies found that personality traits related to agreeableness are correlated with clinical performance. Of particular importance is the empathy facet of agreeableness. Empathy refers to the "ability to sense what others think and feel" (Gough et al., 1991, p. 993). Gough and colleagues found that empathy predicted clinical performance of anesthesiology residents, and they attributed the personological meaning of this finding to individuals' ability to sense the needs of their patients and medical teams. Thus, we hypothesized validities for agreeableness to increase across the seven years of medical school, as the heaviest component of late performance in medical school is interpersonal in nature (Hypothesis 4).

Conscientiousness. Among the Big Five dimensions of personality, conscientiousness has been most closely related to achievement and success in both educational environments (Blickle, 1996; Busato, Prins, Elshout, & Hamaker, 2000; Costa & McCrae, 1992; De Raad, 1996; De Raad & Schouwenburg, 1996; Goff & Ackerman, 1992; Kling, 2001) and work settings (e.g., Barrick & Mount, 1991; Barrick et al., 2001; Burch & Anderson, 2008; Dudley, Orvis, Lebiecki, & Cortina, 2006; Salgado, 1998). Conscientiousness refers to the cluster of traits relating to achievement striving, prudence, dependability, persistence, order, and impulse control. This dimension of personality is closely aligned with educational achievement through effort and volition (hard work, achievement orientation, and perseverance) as well as the enabling characteristics of being organized and efficient. Individuals high on this personality trait tend to work hard and choose to persist in goal-directed behavior. They also tend to follow rules and norms, plan carefully, and display the ability to delay gratification. Adjectives such as competent, thorough, driven, and work oriented describe individuals high on this trait.

Conscientiousness is a motivational trait: Conscientious individuals not only strive to excel, but they also persist in the face of adversity. Previous work utilizing conscientiousness-related traits has documented its pervasive influence in predicting academic

performance in college (e.g., Wolfe & Johnson, 1995) and graduate studies (e.g., Wiggins, Blackburn, & Hackman, 1969). Indeed, conscientiousness is the Big Five dimension with the most impressive record for predicting academic achievement.

Conscientiousness, and especially achievement-striving aspects of the trait, are expected to predict learning criteria in medical school through motivational effects leading to hard work and persistence. However, one would expect even stronger validities of conscientiousness for grades in medical school where clinical performance is a core component. Interpersonal relationships are facilitated by honesty and dependability. Attention to detail and vigilance are important in diagnosing and treating patients under the supervision of professors. There is some support for these ideas in the medical education literature. Gough et al. (1991) found that socialization and achievement via conformance correlated in the .20s with performance among anesthesiology residents. Thus, we hypothesize validities associated with conscientiousness to be at useful levels throughout medical school but to increase especially during clinical years (Hypothesis 5).

We also expect conscientiousness to relate negatively to academic attrition; conscientious individuals may be expected to have the self-discipline to persist in medical school. Empirical evidence supporting our hypothesis is reported by D. G. Johnson and Hutchins (1966), who found that achievement orientation was related to persistence in medical school. Also, Gough and Hall (1975) found that responsibility and socialization (which assess primarily conscientiousness, but also agreeableness and emotional stability; see Hough & Ones, 2001), were negatively related to medical school attrition.

Method

Sample and Procedure

In 1997, 785 students entered medical studies across all six Flemish universities in Belgium, having successfully passed a cognitive ability-based admission exam, consisting of a general mental ability and situational judgment test (see Lievens, Buyse, & Sackett, 2005, for details on tests and the selection procedure). Students were selected based on their standing on an overall score composite, which was mostly determined by the general mental ability test (the observed correlation of the general mental ability test with the overall score was .78 in this sample). The general mental ability test, the situational judgment test, and the overall score composite correlated only negligibly with the focal (personality-based) predictors employed in this study (the strongest positive correlation was .09 with openness; the strongest negative correlation was −.13 with agreeableness).

During classes at the beginning of the first academic year, students were asked to participate in this study by completing a personality inventory and granting the researchers access to their academic records throughout their medical school career. Participants were informed of the purpose of the study, were assured that results were to be used only for research purposes, and were guaranteed confidentiality of the information provided. Feedback on personality test scores was given to those individuals who were interested. For this purpose, student identification numbers, which were also used to match test scores to criterion data collected in later stages of the study, were employed.

Of the 785 students initially enrolled across all six universities, 627 agreed to participate, yielding a total response rate of 79.9% (ranging from 67.8% to 84.2% across universities). Of those students participating, 403 were female (64.3%) and 224 male (35.7%); the mean age of participants was 18 years 3 months (SD = 10 months) at the beginning of the study. A portion of the data used in this research (from three preclinical years of the seven-year curriculum) was previously reported by Lievens, Coetsier, De Fruyt, and De Maeseneer (2002). However, Lievens et al. did not study changes in criterion-related validity over time. The later GPA criterion available for the present article had not been collected at that point. Further, Lievens and colleagues focused on differences between medical students and other students in terms of their personality characteristics; this issue is not a focus in the present article beyond its influence on predictor score range restriction and enhancement (see discussion below).

The mean personality scale scores of the 627 medical students who agreed to participate in this study did not differ notably from the population of Belgian college students (normative N = 1,560); the average absolute magnitude of group differences across Big Five factors and the 30 facet scales of the NEO Personality Inventory-Revised (NEO PI-R; Costa & McCrae, 1992) was .07 standard deviation units.[1] With regard to variability in personality scale scores, this sample was between 6% less variable and 9% more variable than the population of Belgian college students, depending on the personality scale under investigation. These findings of relatively small differences are intuitively appealing when we consider that the personality inventory was administered at the beginning of the first academic year for research purposes only, and individuals' scores were not used in any selection decisions. As such, we concluded that there was no bias in either mean levels or variability of personality scores in this sample of medical students when compared with the general student population.

One strength of this sample is that it consisted entirely of medical students (an entire country's cohort). This afforded us a greater extent of curriculum standardization than is typically found in studies involving students across several academic disciplines. By using this sample, we held constant a number of potentially contaminating influences that operate on reliabilities of grades and on observed validities in the prediction of academic success (e.g., differences in course difficulty across fields; see discussion below).

Measures

Personality measure. Participants completed the authorized Flemish translation (Hoekstra, Ormel, & De Fruyt, 1996) of the NEO PI-R (Costa & McCrae, 1992). The NEO PI-R is a 240-item personality inventory assessing the Big Five dimensions of neuroticism, extraversion, openness, agreeableness, and conscientiousness, as well as 6 specific facets per factor (8 items per facet). The item response scale ranges from 1 (*strongly agree*) to 5 (*strongly disagree*). In this sample, a principal components analysis, followed by varimax-rotation, yielded five factors that adequately described the data, with eigenvalues ranging from 1.71 to 5.75 (60.6% of variance explained). All but 1 of the 30 facets

[1] We thank Filip De Fruyt for generously sharing the normative data of this large sample of Belgian college students.

measured by the NEO PI-R displayed their primary loadings on the factor they were purported to measure. The only exception was observed in the impulsiveness facet of neuroticism, which was found to load primarily on the factor extraversion. This deviance in factor structure when compared with the original U.S. version of the test is one that is often observed in analyses of Dutch/Flemish data sets (De Fruyt & Mervielde, 1998).

Criterion measures. Participants' attrition status and GPA were obtained from university records at the end of each academic year for seven consecutive years (i.e., throughout the whole curriculum). We refer to the earlier section in our literature review for a year-by-year description of the criterion. GPA served as a measure of academic performance in medical school. In Belgium, GPA is measured on a scale from 0 to 20, with higher scores indicating better grades. In order to circumvent potential distortion effects caused by differences in harsh or lenient grading policies among universities, individuals' GPAs were standardized within each of the six universities from which data were obtained.

These data were available for 608 students at the end of the first year, for 405 students at the end of the second year, for 353 students at the end of the third year, for 339 students at the end of the fourth year, for 334 students at the end of the fifth year, for 307 students at the end of the sixth year, and for 306 students at the end of the seventh and final year in medical school. The reduction in the number of participants was due to attrition of students across years primarily as a result of individuals failing final year exams and eventually dropping out of medical school.

These longitudinal data enabled us to compare the relative predictive validity of personality traits over time, although, as previously discussed, attrition among students needed to be taken into account as it could potentially lead to restriction of range in the variables examined. Comparisons of validities for the GPA criterion across seven years constitute our primary examination. However, we also summarize mean differences between persisters and dropouts in terms of personality characteristics.

Analyses

Correlations were computed between Big Five factor and facet scores, as assessed by the NEO PI-R, and students' GPA, separately by year. Means and standard deviations were computed for students' personality test scores as well as for GPA at the end of each academic year, to determine the extent of potential range restriction on the predictor and criterion variables that was due to student attrition (see below for more details on artifact corrections).

To examine differences between those who left medical school and those who stayed in terms of Big Five personality factors and facets, we computed standardized effect sizes (*d* values) comparing leavers and persisters. An effect size is computed by expressing the differences in the means of the two personality scale scores in pooled standard deviation units. In this study, positive *d* values indicate higher mean scores for persisters, and negative *d* values indicate higher mean scores for dropouts. Although effect sizes can theoretically range between positive and negative infinity, given a normal distribution, 95.44% of all effect sizes are found between 2.00 and –2.00. Effect sizes of about .20 in magnitude are small, around .50 are medium, and above .80 are large (Cohen, 1977). In this research, effect sizes close to zero would indicate the equivalence of the means of the two groups being compared.

Our choice of standardized effect sizes (*d* values) to compare persisters and leavers was based on two major considerations. First, different personality scales (e.g., Big Five factor scores versus facet scale scores) do not have the same scoring. When effect sizes are used to convert observed score differences to standardized difference scores, the resulting value is independent of the original units of measurement. Second, and more important for this research, *d* values indicate the magnitudes of the differences between two groups being compared, irrespective of sample sizes of individuals in each group. This was crucial for the interpretability of our findings as the degree of attrition was greatest between the first and the second year, with fewer and fewer students leaving medical school in subsequent years. In computing 90% confidence intervals around the *d* values, unequal sizes of the groups were taken into account in estimating sampling errors (Hedges & Olkin, 1985). Due to the smaller number of dropouts in later years, we report these analyses only until Year 4, where meaningful comparisons can still be made.

Corrections for range restriction on the criterion. In longitudinal validation designs, attrition creates range restriction on the criterion, typically performance. For instance, in work settings, poor performers are fired or laid off, and good performers are promoted (Bass, 1962). Thus, performance measures assess only those individuals remaining at a particular point in time. As Sturman and Trevor (2001) pointed out, "Research predicting individual job performance over time that excludes leavers may suffer from nonrandom mortality, which threatens the internal validity of such studies"; thus, "It is important that analyses not be biased by excluding an often sizable, and conceptually important, portion of the workforce" (p. 695). In educational settings, poor performance is the leading cause of academic dropouts (Humphreys, 1968). As Humphreys put it, "A good deal of the change in *N*s overall, however, is due to academic selection. Academic dropouts decrease the range of talent and attenuate correlation coefficients" (p. 375). Examinations of criteria over time must take attrition and turnover influences into account (Sturman & Trevor, 2001) because they can work to produce spurious effects such as declining validities due to range restriction (Lin & Humphreys, 1977).

Thus, in assessing the predictive value of a variable, it is desirable to have an estimate of what the predictor–criterion relationship would have been had there not been any range restriction (i.e., had no one dropped out). Conclusions about the validity and usefulness of selection and admissions instruments need to apply to the whole pool of applicants and not only to those who remain in the organization or who stay to graduate. In this study, as the demands on students increase over the course of seven academic years, many individuals leave medical school because they fail to meet learning criteria (i.e., they fail end-of-year exams). As academic attrition results in a reduced number of (presumably high-performing) students over the course of seven years, the range of scores on the GPA criterion among each year's sample should be notably reduced and thus attenuate observed correlations. We were interested in the predictability of academic performance *for all students* and not only for those who persisted to graduation. We therefore followed Humphreys's (1968) recommendations and statistically corrected for range restriction in the criterion.

In longitudinal prediction studies such as this one, it is crucial to disentangle range restriction effects from other influences resulting in reductions in criterion variability. In addition to true changes in performance (e.g., individuals become more homogeneous in their performance as a result of their training), one artifactual influence that reduces criterion score variability is scale compression. If medical students in later years mostly receive a limited set of high grades, a ceiling effect on grades can produce reduced variability in GPAs that is substantively quite distinct from the effects of direct range restriction on the criterion. GPA scale compression influences variability in criterion scores but is a separate phenomenon from criterion range restriction caused by attrition. Although such scale compression artifactually depresses observed validity, it should not be corrected for like range restriction, especially in instances where it cannot be distinguished from true changes in performance variability, as is the case in the present study.

In making artifact corrections, our goal was to be as accurate as possible. Thus, we sought to *not* include the influence of non-attrition-related reductions in criteria (i.e., scale compression or true variability reductions) in our range restriction corrections but rather adjust range restriction correction factors (*u* values) to provide more conservative (i.e., lower) estimates of operational validities. Thus, we examined whether scale compression and/or true reductions in criterion variability affected GPA variability across years. To this end, we computed means and standard deviations on GPA separately for those groups of students who were still enrolled each year across all preceding years separately, by university. For example, we selected all individuals still enrolled in the seventh year and computed standard deviations on GPA for only this group for Years 1 to 7.

These analyses indicated that those groups of students who persisted to go on to later years were much less variable in their performance even in early years, compared with the entire group of individuals entering medical school. In addition, there were temporal declines in the academic performance variability of persisters. For example, when we selected only those students who persisted to the seventh year and computed GPA standard deviations for this select group of students separately for each year in medical school, we found the respective standard deviations to be 1.65 for Year 1, 1.83 for Year 2, 1.64 for Year 3, 1.65 for Year 4, 1.33 for Year 5, 1.07 for Year 6, and 0.96 for Year 7. Because this group consisted only of those students persisting to the seventh year, attrition can be ruled out as a cause of declining standard deviations.

The existence of scale compression effects and/or true reductions in variability prevented the use of traditional range restriction information in corrections for attenuation due to restriction caused by attrition. Normally, we would have obtained a range restriction correction (*u*) value (cf. Hunter & Schmidt, 2004; Sackett & Yang, 2000) by dividing the standard deviation of the restricted group (those students still enrolled in the year under investigation) by the standard deviation of the unrestricted group (all students enrolled in the first year of studies).[2]

However, since there was evidence of non-attrition-related reductions in variability in our data, we had to devise and apply an alternate strategy in order not to overestimate operational validities by applying range restriction corrections. Our aim was to isolate attrition-based range restriction effects from scale compression and/or true variability reduction effects. We first computed a scale compression/true variability reduction index for each year by dividing the standard deviation for that year's GPA (using all students in that year) by the standard deviation of GPA for the same students (a given year's persisters) in Year 1. For example, the factor for Year 3 was computed by dividing the standard deviation of GPA in Year 3 by the standard deviation of GPA for exactly the same students (Year 3 persisters) in Year 1.

To answer the question "For each year, what would be the standard deviation of grades if academic attrition were not operating?", we multiplied the standard deviation of grades in Year 1 (before any range restriction on the criterion occurred) by each year's scale adjustment factor. This provided unrestricted group GPA standard deviations for each year, unhampered by effects that also reduce variability but are not due to attrition (e.g., scale compression and increased performance homogeneity). These new estimates of unrestricted group GPA variability were notably smaller than those we would have obtained looking at observed Year 1 scores.

We then computed range restriction correction (*u*) values by dividing the GPA standard deviations of those who persisted to each given year by the new unrestricted group GPA standard deviation (as described earlier, after taking GPA scale reduced variability effects into account). For example, for Year 3, the *u* value was obtained by dividing the Year 3 GPA standard deviation by the GPA standard deviation for all students in Year 1 multiplied by the scale variability reduction factor for Year 3. Such a procedure focuses only on the effects of attrition on the criterion and does not correct for reduced variability in criterion scores due to other influences (e.g., scale compression or true changes in performance). This procedure was repeated for all years to obtain a unique, attrition-based *u* value for each year. These more conservative *u* values were utilized in range restriction corrections. No corrections for range restriction were applied to Year 1 validities, as attrition based on academic performance was not evidenced until the end of that year.

Descriptive statistics for predictor and criterion variables are presented in Appendix A. As indicated in Appendix A, the standard deviation of GPA was 3.46 in Year 1, whereas it was 0.99 in Year 7 (based on GPA raw score units). As discussed earlier, this reduction in variability was due to (a) academic attrition, (b) true performance changes, and (c) scale compression. The effects of attrition-based range restriction on predictor and criterion variability are summarized in Appendix B. For the GPA criterion, the *u* values reported indicate that the greatest effect of attrition on variability in grades occurs in Year 2 (the year when the greatest number of academic dropouts occurs—203 of 608 original students dropping out). The *u* value of .53 indicates that those students who made it to Year 2 are 47% less variable on GPA than would be the case if all Year 1 students were allowed to advance

[2] This procedure was followed by the only study we could locate that addressed and dealt with the issue of range restriction in criterion scores due to student attrition over time (Humphreys, 1968). In reporting corrected correlations between ability and grades, Humphreys compared the standard deviation of grades for each group for which validity was to be estimated with that of the most unrestricted group (students in the first semester). This correction was appropriate because there was no evidence of scale compression in Humphreys's study.

to Year 2 (recall that this takes into account only attrition-based range restriction). As such, the *u* values presented in Appendix B do not confound scale compression effects/true reductions in criterion variability with attrition-based range restriction effects. Appendix C presents observed correlations between Big Five personality factors and facets and GPA, by year. The range restriction and range enhancement *u* values reported in Appendix B were used in range restriction corrections of observed correlations presented in Appendix C.

Corrections for range restriction on the predictor. Results of the personality assessment were not used in admission decisions or in selecting individuals to proceed to subsequent years throughout their medical school career. However, we were sensitive to the fact that student attrition across the seven years may result in less variability in students' personality scores in later years, as can be hypothesized by an attraction–selection–attrition model (B. Schneider, 1987). Thus, we also corrected observed correlations for range restriction on the predictor variable, as we were interested in what the validity of the predictor (administered at time of admission) would be for the entire applicant pool. In the case of personality test scores, this sometimes resulted in correction for range enhancement in scores rather than corrections for range restriction. In applying range restriction/enhancement corrections, we relied on the improved approximation provided by Alexander, Carson, Alliger, and Carr (1987).

Corrections for unreliability. No corrections were applied for attenuation due to unreliability in predictor scores, as we were interested in the operational validity of personality scales in predicting GPA. Also, unreliability in the criterion was not taken into account. In this study, we did not want to rely on assumptions or meta-analytic estimates of criterion reliability, despite the fact that good and stable estimates would be available in the literature on which we could base our corrections. Thus, our estimates of operational validities underestimate the real operational validity of personality in predicting medical school GPA to the extent that GPA is an unreliably measured criterion of academic performance. If one were to assume a level of unreliability in the GPA criterion commensurate with that typically observed in educational settings (the meta-analysis by Kuncel, Hezlett, & Ones, 2001, reports a mean reliability of .83 for graduate GPA), or if employing the value of .75 used for such corrections in previous studies (Lievens et al., 2005), all validities would be 9.7% (or 15.5%) larger than the values reported here, which are not corrected for attenuation due to unreliability.

Multiple regression analyses. Multiple regression analyses were conducted on the basis of correlation matrices obtained after correcting all observed criterion-related validities for effects of range restriction and enhancement caused by academic attrition. For each year, GPA was regressed on all Big Five factors entered into the regression simultaneously, in order to obtain an estimate of the amount of variance in academic performance that is explained by individual differences in personality, and also to investigate the relative value of each of the Big Five factors in the prediction.

Results

Operational Validity of Big Five Factor and Facet Scores

Table 1 presents operational validities of Big Five personality factors and facets for grade point average by year. The results show that operational validities of personality for predicting grades tend to increase for most Big Five factor and facet scales over the seven years of medical school education.

For predicting GPA in Year 1, few personality factor and facet scales displayed operational validities that exceeded .10. These scales were extraversion (−.11), gregariousness facet of extraversion (−.13), and excitement-seeking facet of extraversion (−.16); ideas facet of openness (.12); conscientiousness (.18), as well as competence (.16), achievement striving (.14), self-discipline (.22), and deliberation (.14) facets of conscientiousness. By Year 7, operational validities for the Big Five factors of extraversion, openness to experience, agreeableness, and conscientiousness were substantially higher. Their validities for predicting academic performance in the seventh year of medical school were .31, .30, .17, and .45, respectively. These values represent increases in operational validity of .28 for openness, .16 for agreeableness, and .27 for conscientiousness. The validity of extraversion initially was –.11 and increased to .31 by Year 7. Only a very modest increase in validity was observed for the Big Five factor of neuroticism, as assessed by the NEO PI-R. Operational validity (not corrected for unreliability in the criterion) increased from a negligible value of $\rho = .03$ in the first year (higher scores on neuroticism indicating higher GPA) to a high of −.07 in the seventh year (lower scores in neuroticism indicating higher GPA).

Early on, only 7 of 24 facet scales of extraversion, openness, agreeableness, and conscientiousness exceeded an operational validity of .10. By Year 7, 19 of the 24 facet scales had substantial validities, ranging between .17 and .47. That is, with only a few exceptions, virtually all facets of these four dimensions of the Big Five factors predicted medical school performance with good efficiency. Average increases in operational validities were .18 across extraversion facets, .15 across openness facets, .10 across agreeableness facets, and .21 across conscientiousness facets. We next discuss patterns of increasing operational validities for each of the Big Five.

Operational Validities for Neuroticism

In general, validities for neuroticism remained relatively unchanged for predicting grades, supporting Hypothesis 1. Academic performance differences between those high versus low on neuroticism remained relatively constant across the years, but this Big Five dimension had virtually negligible predictive value to begin with ($r = .03$ to $-.07$). Nontrivial validity increases, however, were found for two facets of neuroticism. Validities for self-consciousness increased by .10 validity points; validity for the vulnerability facet increased by .11 validity points. According to Costa and McCrae (1992), self-conscious individuals are "uncomfortable around others, sensitive to ridicule, and prone to feelings of inferiority," and vulnerable individuals are "unable to cope with stress, becoming dependent, hopeless and panicked" (p. 16). As medical students move to dealing with patients in applied practice during their education, it appears that personality characteristics relating to being comfortable around others, as well as having the ability to cope with stress, gain in validity for predicting medical school performance.

Table 1
Operational Validities (ρ) of Big Five Factors and Facets for Predicting Grade Point Average Across Years

Variable	Year 1 (N = 608)	Year 2 (N = 405)	Year 3 (N = 353)	Year 4 (N = 339)	Year 5 (N = 334)	Year 6 (N = 307)	Year 7 (N = 306)
Neuroticism	.03	−.01	−.01	.02	−.05	−.09	−.07
Anxiety	.06	.10	.02	.08	−.01	−.04	.02
Angry Hostility	.03	−.08	.08	.05	−.06	.01	−.01
Depression	.04	.01	.03	.01	−.04	.00	.00
Self-Consciousness	.04	.01	−.04	.03	−.08	−.19	−.14
Impulsiveness	−.05	−.09	−.06	−.05	.07	.00	−.07
Vulnerability	.01	−.02	−.08	−.02	−.12	−.17	−.12
Extraversion	−.11	−.01	.03	.14	.25	.25	.31
Warmth	−.08	.07	.05	.18	.26	.26	.33
Gregariousness	−.13	−.09	−.06	.04	.20	.05	.18
Assertiveness	.01	.10	.24	.28	.29	.36	.44
Activity	−.01	.03	.07	.16	.14	.15	.29
Excitement-Seeking	−.16	−.18	−.16	−.13	−.07	−.02	−.11
Positive Emotions	−.07	.07	.00	.05	.23	.25	.17
Openness	.02	.18	.25	.28	.42	.35	.30
Fantasy	−.01	.08	.01	.03	.24	.18	−.02
Aesthetics	.04	.17	.32	.34	.40	.40	.34
Feelings	.01	.13	.15	.34	.38	.30	.28
Actions	−.09	−.05	.00	−.01	.08	.14	.19
Ideas	.12	.22	.36	.30	.34	.29	.31
Values	−.06	.12	.06	.01	.26	.08	.08
Agreeableness	−.01	.09	−.03	.01	.17	.06	.17
Trust	−.02	.11	.03	.03	.22	.06	.17
Straightforwardness	.06	.14	.08	.09	.23	.09	.23
Altruism	−.09	.04	−.01	.02	.11	.19	.27
Compliance	.05	.18	−.10	−.12	.02	−.06	.03
Modesty	−.06	−.12	−.14	−.05	−.04	−.06	−.08
Tender-Mindedness	−.02	.03	.01	.10	.20	.05	.13
Conscientiousness	.18	.41	.33	.26	.38	.43	.45
Competence	.16	.25	.30	.19	.36	.42	.42
Order	.08	.22	.11	.02	.17	.24	.24
Dutifulness	.09	.26	.15	.22	.21	.21	.23
Achievement Striving	.14	.34	.31	.35	.41	.48	.44
Self-Discipline	.22	.46	.38	.34	.40	.43	.47
Deliberation	.14	.29	.22	−.01	.16	.16	.30

Note. N = sample size in a given year; ρ = operational validity, corrected for range restriction and range enhancement on predictor and criterion measures. See Appendix B for u values used in corrections.

Operational Validities for Extraversion

Of all the Big Five dimensions, operational validities showed the greatest change for extraversion and its facets. The change in validity between Year 1 and Year 7 was from −.11 to .31 for the overall extraversion factor scale. The mean validity for extraversion facets was −.08 in Year 1 and .25 in Year 7. Largest gains in validity were found for assertiveness and warmth facets. Being socially ascendant, affectionate, and warm appears to be important for later grades that are based on practicums and internships.

Note that extraversion scores were negatively associated with academic performance early in medical school (e.g., operational validity for Year 1 was –.11). Extraverted individuals tended to obtain lower GPAs. However, later on in medical school, extraversion scores were positively and more strongly related to the criterion (validity in Year 7 was .31). Thus, Hypothesis 2 of increasing validities for extraversion received support. Extraversion and many of its facets predicted later grades in medical school better than earlier grades. As activities with significant interpersonal interactions became part of the official curriculum, extraverted individuals performed better.

Early in medical school, extraversion appears to be a liability. Particularly excitement-seeking and gregariousness facet scales predict academic performance negatively (ρs of −.16 and −.13, respectively). Furthermore, the excitement-seeking facet of extraversion was consistently negatively correlated with GPA throughout medical school. One explanation for the negative correlation between excitement-seeking and academic success might be the potential negative influence that excitement-seeking behavior can have on study habits of individuals; high scorers in this domain are considerably more prone to engage in social interactions and activities that take time away from their studies. Personality tendencies such as craving excitement and stimulation, as well as a preference for the company of others, detract from concentrating on academic work.

Operational Validities for Openness

Open individuals are characterized by curiosity about different life domains and unconventionality. In our data, openness was not predictive of academic performance in the first year of medical school (ρ = .02). However, the validity of openness increased to

the .30 to .40 range by the fifth to seventh year of medical school. Note that validities of openness to aesthetics, feelings, and ideas are higher than the validities of other openness facet scales throughout the seven years. According to Costa and McCrae (1992), individuals who score high on openness to aesthetics appreciate art and beauty; individuals who score high on openness to feelings evaluate emotion as an important part of life; and individuals who score high on openness to ideas are intellectually curious. Although it is relatively easy to understand why being intellectually curious and open to feelings would result in better academic performance, particularly in those classes that involve patients, it is not clear to us why appreciating art and beauty would translate into better grades in medical school. We can formulate no convincing explanation for the pattern in validity of openness to aesthetics.

We noted that the largest gains in validity were observed for the openness to aesthetics, feelings, actions, and ideas facet scales. In general, Hypothesis 3 of increasing validities for the domain of openness to experience received support. But it may be worthwhile to point out that the overall openness and the openness to aesthetics, feelings, and ideas operational validities suggest an asymptote to a level of about .30. The relationship between year in medical school and the validity of openness scales can be characterized as a monotonically increasing, negatively accelerated curve.

Operational Validities for Agreeableness

The increase in operational validity was modest for the Big Five factor of agreeableness (total increase of .16 correlational points across seven years). It could be argued that grades partially reflect likeability in later academic years. Our Hypothesis 4 of increasing validities also received support for altruism, straightforwardness, trust, and tender-mindedness facet scales of agreeableness. Seventh-year validities for these facets ranged between .13 and .27. These facets of agreeableness are descriptive of individuals who are willing to help those in need, are sincere, are well intentioned, and who show sympathy and concern for others. It is not difficult to see how these qualities would be beneficial in interacting with patients during practicums and internships and thus get reflected in later medical school GPA. The modesty facet of agreeableness, it is interesting to note, displayed a consistent small, negative validity over time.

Operational Validities for Conscientiousness

Conscientiousness is the Big Five dimension that has most consistently been found to predict training performance and learning in educational settings (Barrick & Mount, 1991; Barrick et al., 2001; Hough, 1992; Salgado, 1997). Accordingly, validities for this Big Five dimension were strong. The results for conscientiousness also suggest that academic performance differences between high- and low-scoring groups on this trait diverge throughout medical school. That is, greater differences in academic performance are found between students high and low on conscientiousness later in medical school. By Year 7, the validity for the overall dimension was .45 and validities across all six facets ranged between .23 and .47. The mean increase in facet validities was .21, lending support to Hypothesis 5.

In general, the more proactive conscientiousness traits (self-discipline, achievement striving, and competence) appeared to better predict medical school performance than inhibitory and regulatory conscientiousness traits (order, deliberation, and dutifulness). Largest long-term validities were found for the overall conscientiousness factor and self-discipline, achievement striving, and competence facets (ρs of .45, .47, .44, and .42, respectively, for Year 7 GPA). Individuals who start and finish tasks in a timely manner "despite boredom and distractions," who "work hard to achieve their goals" (Costa & McCrae, 1992, p. 18), and those who are prudent do well early in medical school, but they do even better later, compared with individuals lower on these traits.

The validities for the Big Five personality domains reported earlier are impressive for noncognitive predictors, especially in the light of findings that typically suggest declining validities in longitudinal investigations. We were encouraged by comments received during the peer review process of this article to conduct additional analyses to firmly illustrate that the effects reported here are not simply due to the artifact corrections applied, but are true effects of rising validity. Recall that we did not correct for unreliability in the predictor or the criterion, and we also did not correct for reduced variability in criterion scores due to scale compression or increasing performance homogeneity, and thus already provided very conservative (downwardly biased) artifact corrections. But because the (conservative) corrections we applied are relatively uncommon, we wanted to further reinforce the fact that they are not the source of the pattern of rising validities but only serve to estimate operational validities more accurately.

First, we checked on the potential impact that the adjustment we made to range restriction (*u)* values (in order not to overcorrect given evidence of criterion scale compression and/or true reduced performance variability in later years) had on conclusions drawn from range restriction corrected correlations. When the analyses reported in this article were repeated without the adjustment of *u* values for the effect of criterion variability reductions observed in later years, the exact same pattern of results was replicated, yet operational validity estimates were much higher. To illustrate this, we computed the correlations between the operational validity results we reported earlier and those obtained without taking criterion variability reductions in later years into account. For the Big Five traits, the correlations of these two respective sets of validities across the seven years were .98, .98, .92, .96, and .89 for neuroticism, extraversion, openness, agreeableness, and conscientiousness, respectively.

Second, following the suggestion of one of the anonymous reviewers, we also analyzed the subsample of students who persevered until Year 7 separately ($n = 297$ after listwise deletion on all variables). This sample is not influenced by range restriction on the criterion, as there was no attrition for this group. However, results indicate that the variability in criterion scores was reduced across years for this subsample as well, due to either scale compression (which we again did not correct for) or true reductions in performance variability. The means and standard deviations of GPA in all seven years for this subsample only are reported in Appendix D. In addition, we provide observed correlations between all personality scales and the GPA criteria. These data show patterns of increasing validities for the Big Five factors similar to those observed among operational validities for the entire sample. In sum, regardless of the analytic approaches taken, validities for

personality variables, particularly conscientiousness and extraversion, increased over the duration of medical school education.

Multiple Regression Results

We also conducted multiple regression analyses to determine the overall predictive validity of the Big Five personality dimensions for academic performance in each year of medical school. The results of the regression analyses are reported in Table 2. The regression results confirm some of the earlier results from zero-order correlations. Conscientiousness starts out as a moderate predictor of medical school performance and strengthens in its predictive value considerably for later phases of medical school. Openness initially contributes little to the prediction of medical school grades but over time contributes consistently, and nontrivially, to prediction of this criterion. Extraversion starts out as a negative predictor of performance but eventually turns into a positive predictor. The regression analysis for Year 1 indicates that conscientiousness (positively) and extraversion (negatively) most strongly predict performance.

The adjusted multiple correlation (R) for all Big Five factors was .22. Performance in Year 7 is most strongly predicted by conscientiousness, extraversion, and openness (all with positive regression weights); adjusted R for Big Five factors was .56. The contribution of neuroticism to the prediction of grades is trivial in each year. In some years, agreeableness appears to emerge as a suppressor variable. For example, in Years 3 and 4, the zero-order correlations between agreeableness and GPA are virtually zero (.01 and −.03). However, the sizable standardized regression weights for agreeableness suggest a suppressor role for this variable. Inspection of the R^2 shows that effects are fairly homogeneous within three stages of performance (i.e., Year 1, Years 2–4, and Years 5–7). This is consistent with our description of how the criterion gradually changes through the years. At first, there is a heavy emphasis on the acquisition of declarative knowledge, with limited patient contact. In the second stage, clinical courses and patient management problems are also part of the curriculum. Finally, students actually deal with patients in various clerkships.

Two overarching conclusions are warranted based on the results discussed above. First, the predictive validities of the Big Five personality factors increase over time. Second, the validity of an optimal linear combination of the Big Five predicts long-term medical school success with an overall validity of adjusted R = .56.

Attrition: Comparing Persisters and Leavers

Table 3 presents the standardized mean differences between persisters and leavers in terms of Big Five dimension and facet scores as well as in terms of GPA. The table is organized to present sample sizes of persisters (n_1) and leavers (n_2), d values, and 90% confidence intervals associated with each d value by academic year in medical school. For example, Year 2 data contrast those who stayed to the second year with those who left prior to the start of the second year (i.e., those who left during or at the end of Year 1). The numbers of leavers after the third year of medical school are too small to warrant discussion. The sample sizes of leavers for Years 1, 2, 3, 4, 5, and 6 are 203, 55, 22, 11, 19, and 1, respectively. Thus, we report, discuss, and interpret personality differences only for Year 2 persisters versus Year 1 leavers, Year 3 persisters versus Year 2 leavers, and Year 4 persisters versus Year 3 leavers.

The largest difference between leavers and persisters was in terms of GPA for those who persevered to Year 2 compared with those leaving during or at the end of Year 1 (d = .96). The GPA of the two groups differed by almost a standard deviation, confirming that early on the primary reason for attrition is poor academic performance. It is interesting that differences between persisters and leavers on GPA were much more modest and in the opposite direction for years after the first year. That is, in subsequent years the GPAs of leavers were slightly higher than those of persisters. As such, after the first year in medical school, poor grades do not appear to be the primary cause for attrition.

Personality differences between persisters and leavers were negligible to modest. Across the years, few effect sizes were greater than .15, with associated 90% confidence interval not excluding zero. Focusing on differences between medical students in Year 2 and those who left prior to Year 2, there were small differences in terms of conscientiousness (d = .18) and conscientiousness facet scales of achievement and self-discipline (ds = .20 and .22, respectively). Medical students persisting to Year 2 were somewhat higher in terms of conscientiousness. They were also less gregarious (d = –.19), and less open to actions and values (ds = –.24 for both). Personality differences between medical students in

Table 2
Results of Regression Analyses for Big Five Factors

	Year 1 (N = 608)			Year 2 (N = 405)			Year 3 (N = 353)			Year 4 (N = 339)			Year 5 (N = 334)			Year 6 (N = 307)			Year 7 (N = 306)		
Variable	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β	Adj. R	Adj. R^2	β
Model	.22	.05		.46	.21		.44	.20		.40	.16		.56	.32		.57	.32		.56	.32	
Neuroticism			.06			.11			.09			.12			.09			.07			.11
Extraversion			−.13			−.09			−.05			.08			.11			.14			.22
Openness			.08			.21			.31			.28			.38			.33			.21
Agreeableness			−.06			−.06			−.19			−.13			−.04			−.16			−.03
Conscientiousness			.22			.45			.39			.31			.39			.46			.46

Note. The range restriction corrected matrices served as input for the regression analyses. N = sample size in a given year; Adj. R = adjusted multiple correlation; Adj. R^2 = adjusted squared multiple correlation; β = standardized regression coefficient.

Table 3
Predicting Attrition: Comparison of Persisters and Leavers

Variable	Year 2 persisters − leavers					Year 3 persisters − leavers					Year 4 persisters − leavers				
	n_1	n_2	d	90% CI	SD_{per}/SD_{leav}	n_1	n_2	d	90% CI	SD_{per}/SD_{leav}	n_1	n_2	d	90% CI	SD_{per}/SD_{leav}
GPA	608	203	0.96	0.82, 1.10	1.14	405	55	−0.19	−0.42, 0.05	2.13	353	22	−0.37	−0.73, −0.01	1.86
Neuroticism	608	203	0.13	−0.01, 0.26	0.89	405	55	0.05	−0.19, 0.28	1.29	353	22	−0.11	−0.47, 0.25	1.27
Anxiety	608	203	0.15	0.01, 0.28	0.97	405	55	0.19	−0.05, 0.43	1.21	353	22	−0.26	−0.62, 0.10	1.03
Angry Hostility	608	203	0.15	0.02, 0.29	0.96	405	55	−0.15	−0.39, 0.08	1.02	353	22	0.17	−0.19, 0.53	1.08
Depression	608	203	0.11	−0.03, 0.24	0.96	405	55	0.07	−0.17, 0.30	1.25	353	22	−0.16	−0.52, 0.20	1.08
Self-Consciousness	608	203	0.10	−0.04, 0.23	0.96	405	55	0.13	−0.11, 0.36	1.24	353	22	0.17	−0.19, 0.53	0.95
Impulsiveness	608	203	−0.02	−0.15, 0.11	1.02	405	55	−0.14	−0.38, 0.09	1.14	353	22	−0.16	−0.52, 0.20	1.00
Vulnerability	608	203	0.06	−0.07, 0.19	0.89	405	55	0.08	−0.16, 0.31	1.34	353	22	−0.19	−0.55, 0.17	1.01
Extraversion	608	203	−0.14	−0.27, −0.01	1.01	405	55	−0.08	−0.31, 0.16	1.13	353	22	−0.06	−0.42, 0.30	0.78
Warmth	608	203	−0.13	−0.27, 0.00	0.99	405	55	0.06	−0.18, 0.29	1.20	353	22	−0.17	−0.53, 0.19	0.78
Gregariousness	608	203	−0.19	−0.32, −0.06	1.07	405	55	−0.04	−0.28, 0.19	1.17	353	22	0.09	−0.27, 0.45	0.79
Assertiveness	608	203	−0.02	−0.15, 0.11	0.93	405	55	−0.25	−0.49, −0.01	1.06	353	22	−0.29	−0.65, 0.07	1.00
Activity	608	203	0.00	−0.13, 0.14	1.06	405	55	−0.04	−0.27, 0.20	1.05	353	22	−0.16	−0.52, 0.20	0.76
Excitement-Seeking	608	203	−0.15	−0.29, −0.02	1.20	405	55	−0.08	−0.32, 0.15	1.00	353	22	0.46	0.10, 0.82	0.77
Positive Emotions	608	203	−0.08	−0.21, 0.05	1.01	405	55	0.06	−0.17, 0.30	1.15	353	22	−0.20	−0.56, 0.16	0.80
Openness	608	203	−0.09	−0.23, 0.04	1.00	405	55	0.15	−0.08, 0.39	1.16	353	22	−0.29	−0.65, 0.07	0.98
Fantasy	608	203	−0.01	−0.14, 0.12	0.99	405	55	0.30	0.06, 0.53	1.08	353	22	−0.11	−0.47, 0.25	0.74
Aesthetics	608	203	0.05	−0.09, 0.18	0.98	405	55	−0.14	−0.37, 0.10	1.14	353	22	−0.23	−0.59, 0.13	1.16
Feelings	608	203	−0.05	−0.18, 0.09	0.90	405	55	0.13	−0.11, 0.36	1.00	353	22	−0.13	−0.49, 0.23	0.80
Actions	608	203	−0.24	−0.38, −0.11	0.99	405	55	−0.06	−0.30, 0.17	1.13	353	22	−0.38	−0.74, −0.02	1.01
Ideas	608	203	0.03	−0.11, 0.16	1.04	405	55	0.06	−0.17, 0.30	1.11	353	22	−0.18	−0.55, 0.18	1.08
Values	608	203	−0.24	−0.37, −0.10	0.97	405	55	0.38	0.14, 0.61	1.11	353	22	−0.11	−0.47, 0.25	0.88
Agreeableness	608	203	−0.06	−0.19, 0.07	0.97	405	55	0.27	0.03, 0.50	0.94	353	22	−0.11	−0.47, 0.25	1.17
Trust	608	203	−0.12	−0.25, 0.01	0.94	405	55	0.25	0.01, 0.49	1.00	353	22	0.04	−0.32, 0.40	1.18
Straightforwardness	608	203	0.00	−0.13, 0.14	0.90	405	55	0.14	−0.09, 0.38	0.93	353	22	−0.25	−0.61, 0.11	1.06
Altruism	608	203	−0.11	−0.24, 0.03	0.95	405	55	0.18	−0.05, 0.42	1.22	353	22	−0.22	−0.58, 0.14	0.76
Compliance	608	203	−0.02	−0.15, 0.11	1.01	405	55	0.22	−0.02, 0.45	0.91	353	22	−0.01	−0.37, 0.35	0.90
Modesty	608	203	0.00	−0.14, 0.13	1.06	405	55	0.01	−0.23, 0.25	0.95	353	22	0.00	−0.36, 0.36	1.10
Tender-Mindedness	608	203	−0.03	−0.16, 0.10	1.06	405	55	0.37	0.13, 0.60	0.94	353	22	−0.02	−0.38, 0.35	0.98
Conscientiousness	608	203	0.18	0.05, 0.31	0.90	405	55	0.23	0.00, 0.47	1.05	353	22	0.05	−0.31, 0.41	0.96
Competence	608	203	0.13	−0.01, 0.26	0.95	405	55	0.09	−0.15, 0.33	1.12	353	22	0.16	−0.20, 0.52	0.98
Order	608	203	0.07	−0.06, 0.20	0.99	405	55	−0.04	−0.28, 0.19	1.03	353	22	−0.08	−0.44, 0.28	0.92
Dutifulness	608	203	0.10	−0.03, 0.23	0.99	405	55	0.28	0.05, 0.52	1.05	353	22	−0.10	−0.46, 0.26	0.90
Achievement Striving	608	203	0.20	0.07, 0.34	0.98	405	55	0.16	−0.08, 0.39	1.06	353	22	0.05	−0.31, 0.41	0.72
Self-Discipline	608	203	0.22	0.09, 0.36	0.92	405	55	0.32	0.08, 0.56	1.01	353	22	−0.02	−0.38, 0.34	1.01
Deliberation	608	203	0.08	−0.05, 0.22	0.92	405	55	0.21	−0.03, 0.44	0.92	353	22	0.21	−0.15, 0.57	0.99

Note. Year 6–Year 7 comparison was not computed, as only one individual left after Year 6. Results of analyses for Years 5 and 6 are available from Stephan Dilchert on request. n_1 = sample size for persisters; n_2 = sample size for leavers; d = Cohen's d value, mean difference in standard deviation units; CI = confidence interval; SD_{per} = standard deviation of group that persisted to a given year; SD_{leav} = standard deviation for group that left before a given year.

Year 3 and those who left prior to Year 3 were about the same magnitude. There were modest differences in conscientiousness (d = .23) and agreeableness (d = .27). Several conscientiousness and agreeableness facet scales also distinguished between persisters and leavers, although the associated effect sizes were modest. These results suggest that even by Year 2, characteristics that relate to interpersonal interactions, such as agreeableness, relate to attrition. Contrary to our expectations, neuroticism was not related to attrition. While some modest group differences were found between Year 4 persisters and dropouts, the group of dropouts at that point was too small to conclude with certainty that the observed effect sizes would replicate in other settings.

In sum, the first year in medical school appears to act as an academic-achievement-driven weed-out mechanism. Personality characteristics associated with this effect are conscientiousness and openness to experience. Later attrition in medical school appears not to be primarily academic performance based. Personality characteristics associated with agreeableness, particularly trust and tender-mindedness, differentiate persisters from leavers. Conceptually then, early attrition (Year 1 to 2) is explained by poor grades, low conscientiousness, and high openness. Later attrition (Year 2 to 3) is associated with lower conscientiousness as well as agreeableness.

Discussion

The long-term predictive validity of personality traits is rarely studied and reported. Previous research has examined the relationships between personality variables and extrinsic and intrinsic career success (Judge et al., 1999), but similar long-term investigations have not been carried out for academic performance. As such, our study fills a void in this area. We examined the longi-

tudinal validities of the Big Five personality dimensions and their facets for predicting academic performance as well as attrition in medical school.

Results across seven years indicated that validities for extraversion, openness, agreeableness, and conscientiousness factor and facet scales increased in predictive potency. Largest gains in validity were obtained for extraversion, openness, and conscientiousness. Virtually constant, negligible validities were found for neuroticism. Conscientiousness, perhaps more so than any other personality trait, is an increasing asset for medical students during their education. Further, the results of this study suggest that although there may be no advantages to being open to experience and extraverted in terms of predicting early academic performance, these traits increasingly gain in importance later in an individual's academic experience in medical school and in applied settings. More generally, it is noticeable that predictors related to "getting ahead" (cf. Hogan & Holland, 2003) are predictive in the early years (conscientiousness), whereas both getting ahead and "getting along" predictors are important in later years. For example, consider the validities for the facets warmth, feelings, and altruism (conceptually matched to getting along criteria) and assertiveness, activity, competence, achievement striving, and self-discipline (related to getting ahead criteria). The former start out as negligible (or even negative) performance predictors and obtain good validities in later years, whereas the latter set of facets start out as good predictors and predict performance even better in later years.

The increasing validities observed for personality scales seem to be related to changes in the GPA criterion. Early GPA in professional schools reflects interpersonal behaviors only to a modest degree. Later GPA reflects successful applications of knowledge in practice where interpersonal behaviors play a key role (as in internship performance, interacting with patients, and so forth).[3] Framing the nature of this criterion change in socioanalytic theory (Hogan & Holland, 2003), the getting ahead dimension seems to be more important than the getting along dimension in the first year. Between the second and sixth years, there occurs a gradual change. Although the getting ahead dimension stays important (eligibility for specialty studies is primarily contingent upon the grades obtained in later years), the importance of clinical and interpersonal work with actual patients in the later years shows that the getting along dimension is becoming increasingly prevalent.

There are implications of such a criterion change for prediction. For example, previous research has demonstrated that personality characteristics such as empathy are more useful in predicting clinical performance than cognitive ability (Gough & Hall, 1975; Shen & Comrey, 1997; Spiegel, Smolen, & Hopfensperger, 1986; Turner et al., 1974). In a related field, dentistry, Chamberlain, Catano, and Cunningham (2005) suggested that

> certain academic related competencies, such as time management, task organization, reading comprehension, memorization, test taking, and concentration, may be more critical to success during the first year of dental school. As students progress through dental school, the curriculum begins to include clinical interactions with patients and behavioural skills such as verbal communication and empathy become more valuable. (p. 3)

Smithers, Catano, and Cunningham (2004) also demonstrated that different predictors were associated with performance in academic and clinical courses.

In addition to changes in the nature of the criterion, increasing validities can also be interpreted in the light of trait-activation theory (see earlier discussion). Both explanations are viable and even complementary. We know for a fact that the nature of the GPA criterion changed over time, shifting emphasis from an assessment of declarative knowledge acquisition to performance in applied courses and internships. The fact that our specific hypotheses were supported only further strengthens the former explanation. The latter, trait-activation-based explanation also seems viable. Within professional education, performance settings become less and less structured as students venture into more complex, real-life settings (e.g., simulation-based evaluations or even long-term internships). It is likely that these weaker situations leave more room for individual differences in personality to determine student behavior and thus predict performance variability.

It is intriguing to compare our results for personality with the results typically obtained in the ability domain. In predicting academic performance (i.e., undergraduate, graduate, and professional school grades), cognitive ability tests have shown declining validities over time (Humphreys, 1968; Humphreys & Taber, 1973; Lin & Humphreys, 1977). Highest validities are reported for academic performance in early semesters and years, and gradually declining validities are observed thereafter. The patterns found in the ability domain are in direct contrast to those reported for personality variables in this study. As F. L. Schmidt et al. (1988) noted, the different nature of GPA criteria in different years may also account for the declining validities of cognitive ability tests.

As noted earlier, students who participated in this study also completed cognitive ability measures as part of the admissions process. The validation results for these tests have been reported elsewhere (Lievens, 2004; Lievens et al., 2005; Lievens & Sackett, 2007) and will be summarized only briefly here. To make results comparable, we used normative information from our applicant pool to correct observed correlations between the general mental ability test and the GPA criterion for restriction of range in students' predictor scores. We also conducted the same analyses for the score composite used to make admissions decisions (general mental ability and situational judgment test). The level of range restriction was very consistent across the seven years; u values ranged from .81 to .82 (for the general mental ability test alone) and from .66 to .67 for the overall score composite. Parallel to the analyses reported for personality, we also corrected for restriction of range in GPA using the u values reported in Appendix B. Operational validities for the general mental ability test and the score composite varied across years (no systematic increase or decrease was observed) but were generally in the .30 range (ρ = .33 for the general mental ability test and .39 for the overall score composite in Year 7). Thus, at the end of the curriculum, the predictive power of personality (see results for conscientiousness

[3] During the review process, the question of whether changes in validity over time were linear or curvilinear in nature was also raised. The present dataset, despite its longitudinal design, does not lend itself to investigating this question (curve estimation would be based on only 7 data points). However, while the trends of effect sizes across the seven years in our data can be described as linear, it is conceivable that curvilinear effects could be observed in other investigations, depending on potential changes in criterion measures used to validate the predictors in question.

as well as the Big Five factors as a set) was at least on par with that of the cognitive ability measures.

Implications for Practice

Our results of personality factors being predictive for more distal, interpersonal criteria have immediate implications for broadening medical school admissions criteria and for making admissions decisions to professional education in general. These findings provide an important message for universities seeking to diversify their admissions procedures. Performance in professional education (e.g., law, business administration, pharmacy, medical school) over the years becomes less reliant on the acquisition of declarative knowledge and incorporates more strongly interpersonal and motivational qualities. Admission to professional education cannot be based on only one type of predictor or on maximizing one type of criterion. As our data clearly show, early attrition is based mostly on academic performance in a structured classroom setting. Cognitive ability will always be the best predictor of knowledge acquisition and must be an integral part in selecting students who will be able to master the required material. However, to better predict and ensure the long-term success of individuals in professional education, noncognitive characteristics must be assessed such that those selected persevere and make the most of their education. Thus, our findings of personality validities increasing over time are particularly relevant to contexts wherein high-end cognitive ability-based preselection of candidates has already taken place.

Implications for Validation Research

In assessing the utility of personality measures for academic performance, relying on early grades in validation is likely to underestimate the predictive value of this noncognitive predictor domain. Our results suggest that the true value of personality traits may be higher than estimated by the values typically observed in concurrent and most predictive validation studies. A previous quantitative summary has reported negligible validities for most Big Five factors in educational settings (Hough, 1992). There were two exceptions: Emotional stability and the achievement facet of conscientiousness predicted grades with mean observed validities of .20 and .29, respectively. Our findings of higher and more pervasive validities for all Big Five factors (except neuroticism) and many specific facets stand in contrast. We suspect an underestimation of validities due to cross-sectional or short-term validation designs reported in previous quantitative summaries of operational validities of personality scales in predicting academic success. Our results highlight the importance of examining validity longitudinally in educational contexts. Similarly, criteria used in validating personality measures in occupational settings should capture contributions of workers not just during the initial months they spend on the job (i.e., the honeymoon period) but during a longer time span or even their entire tenure with the organization. It can be expected that we will obtain a more accurate estimate of the predictive value of personality for job performance when validation studies are designed so that noncognitive predictors have room to distinguish among performance of individuals once they are well beyond the honeymoon phase of their organizational tenure.

There are also implications for meta-analytic research. Whereas meta-analyses have traditionally emphasized validity generalization across various local settings and samples, our study shows that validity generalization across different time periods might be as important in order to obtain an accurate estimate of the predictive validity of personality.

Directions for Future Research

Our research examined the stability of criterion-related validities of personality measures for medical school performance. Changing criteria in medical school were likely at the root of the increasing validities we observed. Similarly intriguing would be the exploration of temporal changes in the relationships between personality measures and job performance. Clearly, our findings need to be replicated in other settings. Nonetheless, there are conceptual reasons as well as promising empirical support from a small sample study to expect that similar conclusions might ensue in the work domain. Theoretically, when tasks are well learned and employees are in a "maintenance" job stage (Murphy, 1989), one might expect personality determinants of performance to gain predictive value. We believe it is important to also investigate whether validity coefficients of personality traits exhibit similar or differential increases for different performance dimensions (task performance versus contextual performance) over time. The only study to investigate these matters (Thoresen, Bradley, Bliese, & Thoresen, 2004) has not focused on behavior and performance but on a criterion not entirely under the control of employees being measured. Thoresen and colleagues examined the relationship between a Big Five personality inventory and sales, and found different personality traits to be predictive of sales for maintenance and transitional job stages. For employees at a maintenance job stage, conscientiousness predicted sales and extraversion and conscientiousness predicted sales growth ($N = 99$). For employees at a transitional job stage, agreeableness and openness predicted both sales and sales growth ($N = 48$).

Another intriguing possibility is potential nonlinearity in effects of personality over time. Conceptually, longitudinal changes in magnitudes of criterion-related validities of personality variables may be expected when changes in tasks and situational demands in work settings are experienced. Such changes might create various configurations of curvilinear relationships between time and personality scale validities. Future research, if based on many more temporal data points and conducted over longer periods of time, will have the opportunity to start exploring such effects.

So far, there is a scant and fragmented literature documenting increases in validity of personality-based measures (assessment center dimensions, competencies, etc.) using longitudinal designs. Hinrichs (1978) reported that aggressiveness measured during assessment center exercises predicted organizational level better after 8 years on the job than earlier. Other studies (Dodd et al., 1970; Helmreich et al., 1986; Singh, 1978) all reported higher long-term validities of specific personality scales for job performance. Perhaps most relevant to the results reported in this research, Jansen and Stoop (2001) found that the assessment center dimension of interpersonal effectiveness showed validity only after a number of years on the job. Similarly, in a study of 98 executives, Russell (2001) reported that "people-oriented" competencies were more predictive of later performance than "resource problem-solving oriented" competencies. Thus, there are indications that personality-based measures are likely to show increases in predictive validities over time in work settings as well.

Conclusions

Despite the large amount of evidence supporting their applied value in both work and educational settings (Barrick et al., 2001; Burch & Anderson, 2008; De Raad & Schouwenburg, 1996; Ferguson et al., 2000; Gough, 1964; Hough, 1992; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ones, Dilchert, Viswesvaran, & Judge, 2007; Salgado, 2003), personality variables are sometimes criticized for having modest predictive validities. A variety of reasons have been suggested. Among these are poor predictor–criterion matching in validation studies (Hough, 1992, 1998), various arguments around bandwidth-fidelity trade-off (Hogan & Holland, 2003; Ones & Viswesvaran, 1996; R. J. Schneider, Hough, & Dunnette, 1996), and predictive inefficiencies of non-Big Five scales (Hurtz & Donovan, 2000). Another key reason for seemingly modest validities of Big Five personality scales may be that validation studies are often conducted with only limited time frames. Few studies have focused on the long-term validity of personality scales in predicting success in occupational and educational settings. This study showed that in medical education, validation of personality measures against early criteria underestimates their true validity and utility. We would like to suggest the possibility that the true value of personality traits is amplified in the long run in occupational settings as well. Large-scale investigations of Big Five traits and their facets in predicting long-term job performance are called for.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Carr, L. (1987). Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Educational and Psychological Measurement, 47,* 309–315.

Alvares, K. M., & Hulin, C. L. (1972). Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. *Human Factors, 14,* 295–308.

Barratt, E. S., & White, R. (1969). Impulsiveness and anxiety related to medical students' performance and attitudes. *Journal of Medical Education, 44,* 604–607.

Barrett, G. V., & Alexander, R. A. (1989). Rejoinder to Austin, Humphreys, and Hulin: Critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology, 42,* 597–612.

Barrett, G. V., Caldwell, M. S., & Alexander, R. A. (1985). The concept of dynamic criteria: A critical reanalysis. *Personnel Psychology, 38,* 41–56.

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology, 66,* 1–6.

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., & 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.

Bass, B. M. (1962). Mood changes during training laboratory. *Journal of Applied Psychology, 46,* 361–364.

Blickle, G. (1996). Personality traits, learning strategies, and performance. *European Journal of Personality, 10,* 337–352.

Bray, D. W., & Howard, A. (1983). The AT&T longitudinal study of managers. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 112–146). New York: Guilford Press.

Broadbent, D. E. (1958). *Perception and communication.* Elmsford, NY: Pergamon Press.

Burch, G. S. J., & Anderson, N. (2008). Personality as a predictor of work-related behavior and performance: Recent advances and directions for future research. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 23, pp. 261–305). Hoboken, NJ: Wiley.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29,* 1057–1068.

Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Erlbaum.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco: Jossey-Bass.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453–484.

Cattell, R. B., & Kline, P. (1977). *The scientific analysis of personality and motivation.* Oxford, England: Academic Press.

Chamberlain, T. C., Catano, V. M., & Cunningham, D. P. (2005). Personality as a predictor of professional behavior in dental school: Comparisons with dental practitioners. *Journal of Dental Education, 69,* 1222–1237.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37,* 319–338.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* San Diego: Academic Press.

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

Curry, R. H., & Makoul, G. (1998). The evolution of courses in professional skills and perspectives for medical students. *Academic Medicine, 73,* 10–13.

Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology, 43,* 717–744.

De Fruyt, F., & Mervielde, I. (1998). The assessment of the Big Five in the Dutch language domain. *Psychologica Belgica, 38,* 1–22.

De Raad, B. (1996). Personality traits in learning and education. *European Journal of Personality, 10,* 185–200.

De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality, 10,* 303–336.

Dodd, W. E., Wollowick, H. B., & McNamara, W. J. (1970). Task difficulty as a moderator of long-range prediction. *Journal of Applied Psychology, 54,* 265–270.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40–57.

Dunham, R. B. (1974). Ability-skill relationships: An empirical explanation of change over time. *Organizational Behavior and Human Performance, 12,* 372–382.

Dusek, T., & Bates, T. (2003). Analysis of European medical schools' teaching programs. *Croatian Medical Journal, 44,* 26–31.

Entwistle, N. J., & Entwistle, D. (1970). The relationships between personality, study methods and academic performance. *British Journal of Educational Psychology, 40,* 132–143.

Eysenck, H. J. (1967). Personality patterns in various groups of businessmen. *Occupational Psychology, 41,* 249–250.

Eysenck, H. J. (1992). Personality and education: The influence of extra-

version, neuroticism and psychoticism. *Zeitschrift für Pädagogische Psychologie, 6,* 133–144.

Ferguson, E., Sanders, A., O'Hehir, F., & James, D. (2000). Predictive validity of personal statements and the role of the five-factor model of personality in relation to medical training. *Journal of Occupational and Organizational Psychology, 73,* 321–344.

Fraley, R., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112,* 60–74.

Furnham, A., & Medhurst, S. (1995). Personality correlates of academic seminar behaviour: A study of four instruments. *Personality and Individual Differences, 19,* 197–208.

Furnham, A., & Mitchell, J. (1991). Personality, needs, social skills and academic achievement: A longitudinal study. *Personality and Individual Differences, 12,* 1067–1073.

Ghiselli, E. E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology, 40,* 1–4.

Ghiselli, E. E., & Haire, M. (1960). The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology, 13,* 225–231.

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

Goh, D. S., & Moore, C. (1978). Personality and academic achievement in three educational levels. *Psychological Reports, 43,* 71–79.

Gough, H. G. (1964). Academic achievement in high school as predicted from the California Psychological Inventory. *Journal of Educational Psychology, 55,* 174–180.

Gough, H. G., Bradley, P., & McDonald, J. S. (1991). Performance of residents in anesthesiology as related to measures of personality and interests. *Psychological Reports, 68,* 979–994.

Gough, H. G., & Hall, W. B. (1975). An attempt to predict graduation from medical school. *Journal of Medical Education, 50,* 940–950.

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.

Grover, P. L., & Smith, D. U. (1981). Academic anxiety, locus of control, and achievement in medical school. *Journal of Medical Education, 56,* 727–736.

Guttman, L. (1955). A generalized simplex for factor analysis. *Psychometrika, 20,* 173–192.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.

Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986). The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. *Journal of Applied Psychology, 71,* 185–188.

Hinrichs, J. R. (1978). An eight-year follow-up of a management assessment center. *Journal of Applied Psychology, 63,* 596–601.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *NEO persoonlijkheidsvragenlijsten NEO PI-R en NEO-FFI.* Lisse, The Netherlands: Swets & Zeitlinger.

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

Hough, L. M. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance, 5,* 139–155.

Hough, L. M. (1998). The millennium for personality psychology: New horizons or good ole daze. *Applied Psychology: An International Review, 47,* 233–261.

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.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology: Vol. 1. Personnel psychology* (pp. 233–277). London: Sage.

Humphreys, L. G. (1960). Investigations of the simplex. *Psychometrika, 25,* 313–323.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59,* 375–380.

Humphreys, L. G., & Taber, T. (1973). Postdiction study of the Graduate Record Examination and eight semesters of college grades. *Journal of Educational Measurement, 10,* 179–184.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis.* Thousand Oaks, CA: Sage.

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

Jansen, P. G. W., & Stoop, B. A. M. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology, 86,* 741–753.

Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology, 10,* 93–98.

Johnson, D. G., & Hutchins, E. B. (1966). Doctor or dropout? A study of medical student attrition. *Journal of Medical Education, 41,* 1099–1269.

Johnson, J. A. (1994). Clarification of Factor Five with the help of the AB5C model. *European Journal of Personality, 8,* 311–334.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology, 83,* 693–710.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52,* 621–652.

Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin, 127,* 673–697.

Kling, K. C. (2001, July). *The role of personality, academic ability, and gender in predicting academic achievement.* Paper presented at the meeting of the International Society for the Study of Individual Differences, Edinburgh.

Knehr, C. A., & Kohl, R. N. (1959). MMPI screening of entering medical students. *Journal of Psychology: Interdisciplinary and Applied, 47,* 297–304.

Knights, J. A., & Kennedy, B. J. (2007). Medical school selection: Impact of dysfunctional tendencies on academic performance. *Medical Education, 41,* 362–368.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181.

Laidlaw, T. S., MacLeod, H., Kaufman, D. M., Langille, D. B., & Sargeant, J. (2002). Implementing a communication skills programme in medical school: Needs assessment and programme change. *Medical Education, 36,* 115–124.

Lathey, J. W. (1991). Temperament style as a predictor of academic achievement in early adolescence. *Journal of Psychological Type, 22,* 52–58.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology, 53,* 563–593.

Lievens, F. (2004). Longitudinal study of the validity of different cognitive ability tests in a student admission context. *Applied H. R. M. Research, 9,* 27–30.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90,* 442–452.

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006).

Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91,* 247–258.

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.

Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology, 92,* 1043–1055.

Lin, P.-C., & Humphreys, L. G. (1977). Predictions of academic performance in graduate and professional school. *Applied Psychological Measurement, 1,* 249–257.

Lipton, A. (1988). Predictors of success in a cohort of medical students. *Medical Education, 18,* 203–210.

Lipton, A., Huxham, G., & Hamilton, D. (1984). School results as predictors of medical school achievement. *Medical Education, 22,* 381–388.

Lunneborg, C. E., & Lunneborg, P. W. (1970). Relations between aptitude changes and academic success during college. *Journal of Educational Psychology, 61,* 169–173.

Marley, J., & Carman, I. (1999). Selecting medical students: A case report of the need for change. *Medical Education, 33,* 455–459.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin, 120,* 323–337.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335–354.

Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance, 2,* 183–200.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60,* 995–1027.

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

Paget, J. (1869). What becomes of medical students? *St. Bartholomew's Hospital Report, 5,* 238–242.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3–25.

Rocklin, T. (1994). Relation between typical intellectual engagement and openness: Comment on Goff and Ackerman (1992). *Journal of Educational Psychology, 86,* 145–149.

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91,* 511–526.

Russell, C. J. (2001). A longitudinal study of top-level executive performance. *Journal of Applied Psychology, 86,* 560–573.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85,* 112–118.

Salgado, J. F. (1997). The Five Factor Model of personality and job performance in the European Community. *Journal of Applied Psychology, 82,* 30–43.

Salgado, J. F. (1998). Big Five personality dimensions and job performance in Army and civil occupations: A European perspective. *Human Performance, 11,* 271–288.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76,* 323–346.

Sanchez, M. M., Rejano, E. I., & Rodriguez, Y. T. (2001). Personality and academic productivity in the university student. *Social Behavior and Personality, 29,* 299–305.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63,* 506–516.

Savage, R. D. (1962). Personality factors and academic performance. *British Journal of Educational Psychology, 32,* 251–253.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology, 73,* 46–57.

Schmidt, H. G., Machiels-Bongaerts, M., Hermans, H., ten Cate, T. J., Venekamp, R., & Boshuizen, H. P. (1996). The development of diagnostic competence: Comparison of a problem-based, an integrated, and a conventional medical curriculum. *Academic Medicine, 71,* 658–664.

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437–453.

Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior, 17,* 639–655.

Shen, H., & Comrey, A. L. (1997). Predicting medical students' academic performances by their cognitive abilities and personality characteristics. *Academic Medicine, 72,* 781–786.

Singh, S. (1978). Achievement motivation and entrepreneurial success: A follow-up study. *Journal of Research in Personality, 12,* 500–503.

Smithers, S., Catano, V. M., & Cunningham, D. P. (2004). What predicts performance in Canadian dental schools? *Journal of Dental Education, 68,* 598–613.

Spiegel, D. A., Smolen, R. C., & Hopfensperger, K. A. (1986). Medical student stress and clerkship performance. *Journal of Medical Education, 61,* 929–931.

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.

Sturman, M. C., & Trevor, C. O. (2001). The implications of linking the dynamic performance and turnover literatures. *Journal of Applied Psychology, 86,* 684–696.

Sussmann, M., & Robertson, D. U. (1986). The validity of validity: An analysis of validation study designs. *Journal of Applied Psychology, 71,* 461–468.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500–517.

Thoresen, C. J., Bradley, J. C., Bliese, P. D., & Thoresen, J. D. (2004). The Big Five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology, 89,* 835–853.

Turner, E. V., Helper, M. M., & Kriska, S. D. (1974). Predictors of clinical performance. *Journal of Medical Education, 49,* 338–342.

Wiggins, N., Blackburn, M., & Hackman, J. R. (1969). Prediction of first-year graduate success in psychology: Peer ratings. *Journal of Educational Research, 63,* 81–85.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement, 55,* 177–185.

Woodrow, H. (1938). The effect of practice on groups of different initial ability. *Journal of Educational Psychology, 29,* 268–278.

Zeidner, M. (1995). Personality trait correlates of intelligence. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 299–319). New York: Plenum Press.

Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 581–610). New York: Cambridge University Press.

(*Appendixes follow*)

Appendix A

Descriptive Statistics for Predictor and Criterion Variables

Variable	Year 1 (N = 608)		Year 2 (N = 405)		Year 3 (N = 353)		Year 4 (N = 339)		Year 5 (N = 334)		Year 6 (N = 307)		Year 7 (N = 306)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
GPA	11.32	3.46	13.05	2.49	13.78	1.87	14.29	1.88	14.96	1.40	15.29	1.16	15.50	0.99
Neuroticism	138.10	22.15	139.04	21.19	138.97	21.90	138.83	21.99	139.01	22.05	138.75	22.06	138.69	22.08
Anxiety	24.50	5.74	24.78	5.66	24.88	5.81	24.78	5.78	24.81	5.82	24.72	5.79	24.72	5.80
Angry Hostility	20.94	4.68	21.18	4.61	21.04	4.63	21.12	4.64	21.17	4.63	21.09	4.60	21.06	4.58
Depression	23.88	5.32	24.06	5.23	24.06	5.38	24.00	5.36	23.99	5.39	23.93	5.35	23.92	5.36
Self-Consciousness	23.80	5.00	23.97	4.93	24.02	5.05	24.06	5.02	24.12	5.01	24.19	5.04	24.18	5.04
Impulsiveness	25.42	4.69	25.39	4.72	25.28	4.78	25.24	4.76	25.22	4.80	25.22	4.88	25.21	4.88
Vulnerability	19.57	4.89	19.67	4.68	19.68	4.84	19.63	4.80	19.70	4.89	19.60	4.93	19.61	4.93
Extraversion	166.08	19.91	165.15	19.97	164.96	20.29	164.99	19.83	164.79	20.49	164.63	20.63	164.55	20.61
Warmth	29.63	4.29	29.43	4.27	29.49	4.36	29.44	4.24	29.36	4.35	29.43	4.44	29.43	4.45
Gregariousness	28.50	5.16	28.17	5.26	28.14	5.36	28.19	5.25	28.11	5.37	28.14	5.38	28.13	5.39
Assertiveness	23.65	5.37	23.62	5.24	23.43	5.28	23.37	5.28	23.52	5.25	23.41	5.28	23.37	5.25
Activity	25.99	4.26	26.00	4.34	25.97	4.38	25.95	4.29	25.96	4.34	25.87	4.38	25.85	4.37
Excitement-Seeking	27.46	4.83	27.21	5.08	27.15	5.08	27.29	4.95	27.21	4.98	27.19	4.99	27.19	5.00
Positive Emotions	30.85	5.03	30.71	5.04	30.77	5.14	30.76	5.02	30.63	5.19	30.60	5.24	30.58	5.25
Openness	169.49	18.73	168.91	18.72	169.22	19.13	168.86	18.98	168.81	19.17	169.50	19.50	169.49	19.53
Fantasy	27.97	5.43	27.96	5.42	28.16	5.45	28.10	5.30	28.07	5.33	28.21	5.40	28.21	5.41
Aesthetics	28.99	5.72	29.07	5.69	28.91	5.82	28.84	5.86	28.85	5.90	28.95	5.96	28.96	5.97
Feelings	30.06	4.54	29.99	4.38	30.06	4.41	30.04	4.31	30.02	4.37	30.02	4.41	30.01	4.41
Actions	24.60	4.20	24.26	4.17	24.21	4.27	24.12	4.26	24.13	4.30	24.20	4.32	24.20	4.33
Ideas	27.76	5.35	27.81	5.43	27.85	5.52	27.76	5.51	27.82	5.46	28.02	5.52	28.02	5.53
Values	30.11	3.75	29.82	3.70	30.03	3.72	29.99	3.67	29.92	3.80	30.08	3.82	30.09	3.82
Agreeableness	170.04	19.47	169.64	19.27	170.41	19.07	170.13	19.34	169.51	19.41	169.91	19.42	170.00	19.38
Trust	28.17	4.57	27.99	4.46	28.16	4.45	28.14	4.49	28.05	4.51	28.16	4.50	28.18	4.50
Straightforwardness	27.81	5.48	27.82	5.29	27.92	5.24	27.82	5.25	27.69	5.20	27.79	5.19	27.81	5.19
Altruism	30.25	3.84	30.11	3.77	30.22	3.85	30.14	3.75	30.02	3.83	29.99	3.82	29.98	3.83
Compliance	24.07	4.62	24.04	4.64	24.19	4.58	24.15	4.57	24.03	4.51	24.05	4.51	24.08	4.48
Modesty	29.02	4.78	29.01	4.87	29.05	4.84	29.03	4.87	28.99	4.87	29.06	4.91	29.08	4.91
Tender-Mindedness	30.72	4.05	30.68	4.13	30.88	4.06	30.85	4.07	30.74	4.21	30.86	4.21	30.88	4.22
Conscientiousness	166.31	20.06	167.52	19.27	168.15	19.38	168.33	19.26	167.93	19.38	167.59	19.71	167.63	19.73
Competence	28.19	3.55	28.34	3.49	28.39	3.54	28.44	3.51	28.40	3.55	28.38	3.56	28.39	3.56
Order	25.56	4.97	25.68	4.95	25.65	5.00	25.67	4.99	25.64	4.95	25.36	4.91	25.36	4.92
Dutifulness	30.29	4.34	30.43	4.33	30.61	4.34	30.58	4.28	30.49	4.31	30.51	4.38	30.51	4.39
Achievement Striving	29.20	4.52	29.51	4.48	29.59	4.52	29.60	4.41	29.53	4.50	29.51	4.54	29.49	4.54
Self-Discipline	27.75	4.69	28.10	4.54	28.30	4.51	28.32	4.50	28.27	4.51	28.28	4.59	28.27	4.60
Deliberation	25.32	4.91	25.46	4.77	25.62	4.72	25.72	4.69	25.61	4.70	25.56	4.73	25.60	4.67

Appendix B

Effects of Range Restriction and Enhancement on Predictor and Criterion Variables

Variable	*u*						
	Year 1 (N = 608)	Year 2 (N = 405)	Year 3 (N = 353)	Year 4 (N = 339)	Year 5 (N = 334)	Year 6 (N = 307)	Year 7 (N = 306)
Grade point average	1.00	0.53	0.52	0.42	0.45	0.43	0.47
Neuroticism	1.00	0.96	0.99	0.99	1.00	1.00	1.00
Anxiety	1.00	0.99	1.01	1.01	1.01	1.01	1.01
Angry Hostility	1.00	0.99	0.99	0.99	0.99	0.98	0.98
Depression	1.00	0.98	1.01	1.01	1.01	1.01	1.01
Self-Consciousness	1.00	0.98	1.01	1.00	1.00	1.01	1.01
Impulsiveness	1.00	1.01	1.02	1.02	1.02	1.04	1.04
Vulnerability	1.00	0.96	0.99	0.98	1.00	1.01	1.01
Extraversion	1.00	1.00	1.02	1.00	1.03	1.04	1.04
Warmth	1.00	0.99	1.02	0.99	1.01	1.04	1.04
Gregariousness	1.00	1.02	1.04	1.02	1.04	1.04	1.04
Assertiveness	1.00	0.98	0.98	0.98	0.98	0.98	0.98
Activity	1.00	1.02	1.03	1.01	1.02	1.03	1.03
Excitement-Seeking	1.00	1.05	1.05	1.02	1.03	1.03	1.03
Positive Emotions	1.00	1.00	1.02	1.00	1.03	1.04	1.04
Openness	1.00	1.00	1.02	1.01	1.02	1.04	1.04
Fantasy	1.00	1.00	1.00	0.98	0.98	0.99	1.00
Aesthetics	1.00	0.99	1.02	1.02	1.03	1.04	1.04
Feelings	1.00	0.97	0.97	0.95	0.96	0.97	0.97
Actions	1.00	0.99	1.02	1.01	1.02	1.03	1.03
Ideas	1.00	1.01	1.03	1.03	1.02	1.03	1.03
Values	1.00	0.98	0.99	0.98	1.01	1.02	1.02
Agreeableness	1.00	0.99	0.98	0.99	1.00	1.00	1.00
Trust	1.00	0.98	0.97	0.98	0.99	0.98	0.98
Straightforwardness	1.00	0.97	0.96	0.96	0.95	0.95	0.95
Altruism	1.00	0.98	1.00	0.98	1.00	1.00	1.00
Compliance	1.00	1.00	0.99	0.99	0.98	0.97	0.97
Modesty	1.00	1.02	1.01	1.02	1.02	1.03	1.03
Tender-Mindedness	1.00	1.02	1.00	1.00	1.04	1.04	1.04
Conscientiousness	1.00	0.96	0.97	0.96	0.97	0.98	0.98
Competence	1.00	0.98	0.99	0.99	1.00	1.00	1.00
Order	1.00	1.00	1.01	1.00	1.00	0.99	0.99
Dutifulness	1.00	1.00	1.00	0.99	0.99	1.01	1.01
Achievement Striving	1.00	0.99	1.00	0.98	1.00	1.00	1.00
Self-Discipline	1.00	0.97	0.96	0.96	0.96	0.98	0.98
Deliberation	1.00	0.97	0.96	0.95	0.96	0.96	0.95

Note. u = values used in range restriction/range enhancement corrections, computed by dividing the standard deviation of the restricted group by the standard deviation of the unrestricted group. For personality scales, for each year, the restricted group consisted of those students still enrolled; the unrestricted group consisted of all students who were enrolled in the 1st year of data collection. For grade point average (GPA) in each year, the restricted group's standard deviation was the previous year's standard deviation for those students persisting to that year; the unrestricted group consisted of all students enrolled in the previous year. Thus, u values reported for GPA are those where scale compression effects have been removed.

(*Appendixes continue*)

Appendix C

Observed Correlations (r_{obs}) Between Big Five Factors and Facets and Grade Point Average, by Year

Variable	r_{obs}						
	Year 1 (N = 608)	Year 2 (N = 405)	Year 3 (N = 353)	Year 4 (N = 339)	Year 5 (N = 334)	Year 6 (N = 307)	Year 7 (N = 306)
Neuroticism	.03	−.01	.00	.01	−.02	−.04	−.03
Anxiety	.06	.05	.01	.03	.00	−.02	.01
Angry Hostility	.03	−.04	.04	.02	−.03	.00	.00
Depression	.04	.00	.01	.00	−.02	.00	.00
Self-Consciousness	.04	.00	−.02	.01	−.04	−.08	−.07
Impulsiveness	−.05	−.05	−.03	−.02	.03	.00	−.03
Vulnerability	.01	−.01	−.04	−.01	−.05	−.08	−.06
Extraversion	−.11	.00	.02	.06	.12	.12	.16
Warmth	−.08	.04	.03	.08	.13	.12	.17
Gregariousness	−.13	−.05	−.03	.02	.10	.02	.09
Assertiveness	.01	.05	.13	.13	.13	.17	.24
Activity	−.01	.02	.04	.07	.07	.07	.15
Excitement-Seeking	−.16	−.10	−.09	−.05	−.03	−.01	−.06
Positive Emotions	−.07	.04	.00	.02	.11	.12	.08
Openness	.02	.10	.14	.13	.22	.17	.16
Fantasy	−.01	.04	.00	.01	.11	.08	−.01
Aesthetics	.04	.09	.18	.16	.21	.20	.18
Feelings	.01	.07	.07	.15	.19	.14	.14
Actions	−.09	−.03	.00	.00	.04	.06	.09
Ideas	.12	.13	.21	.14	.17	.14	.17
Values	−.06	.06	.03	.00	.12	.03	.04
Agreeableness	−.01	.05	−.02	.01	.08	.02	.08
Trust	−.02	.06	.02	.01	.10	.03	.08
Straightforwardness	.06	.07	.04	.04	.10	.04	.11
Altruism	−.09	.02	−.01	.01	.05	.09	.14
Compliance	.05	.10	−.05	−.05	.01	−.03	.01
Modesty	−.06	−.07	−.07	−.02	−.02	−.03	−.04
Tender-Mindedness	−.02	.02	.01	.04	.09	.02	.07
Conscientiousness	.18	.24	.18	.11	.18	.21	.25
Competence	.16	.14	.17	.08	.18	.21	.23
Order	.08	.12	.06	.01	.08	.11	.11
Dutifulness	.09	.14	.08	.09	.10	.10	.12
Achievement Striving	.14	.20	.17	.16	.21	.25	.24
Self-Discipline	.22	.28	.21	.15	.20	.22	.26
Deliberation	.14	.16	.12	.00	.07	.07	.15

Note. N = sample size in a given year.

Appendix D

Descriptive Statistics and Prediction of 7-Year Persisters' Grade Point Average Across Years

Variable	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
M_{GPA}	0.00	0.02	0.21	0.47	0.84	1.00	1.09
SD	1.00	1.11	1.00	1.00	0.81	0.65	0.58
Neuroticism	−.04	−.04	−.03	−.03	−.02	−.03	−.04
Anxiety	.02	.03	.02	.02	.01	−.02	.00
Angry Hostility	−.04	−.01	−.01	−.03	−.03	.04	−.01
Depression	−.02	.00	−.02	−.03	−.01	−.01	−.01
Self-Consciousness	−.07	−.09	−.06	−.04	−.06	−.06	−.07
Impulsiveness	−.04	−.04	−.05	−.03	.03	.01	−.04
Vulnerability	−.03	−.05	−.04	−.01	−.02	−.08	−.07
Extraversion	−.05	.05	.02	.10	.16	.13	.17
Warmth	−.05	.04	.03	.09	.15	.11	.18
Gregariousness	−.09	−.03	−.03	.04	.09	.03	.09
Assertiveness	.13	.20	.14	.13	.21	.23	.24
Activity	−.02	.07	.02	.08	.08	.09	.15
Excitement-Seeking	−.11	−.09	−.10	−.03	−.01	−.03	−.06
Positive Emotions	−.09	.01	.03	.08	.14	.12	.09
Openness	.13	.11	.14	.14	.20	.17	.16
Fantasy	−.02	.01	.02	.04	.10	.08	−.01
Aesthetics	.12	.15	.18	.18	.20	.19	.18
Feelings	.04	.06	.07	.16	.18	.16	.15
Actions	.04	.00	.00	.03	.05	.06	.09
Ideas	.26	.18	.22	.13	.16	.15	.17
Values	.05	−.01	.02	.00	.07	−.01	.03
Agreeableness	−.01	.00	.00	.06	.07	−.02	.08
Trust	.03	.02	.04	.04	.11	−.01	.08
Straightforwardness	.07	.05	.05	.08	.08	.00	.11
Altruism	−.04	.01	.02	.05	.10	.08	.14
Compliance	.07	.07	−.02	.03	.01	−.07	.02
Modesty	−.10	−.10	−.09	−.01	−.05	−.04	−.04
Tender-Mindedness	−.10	−.06	.00	.06	.06	−.03	.06
Conscientiousness	.16	.25	.21	.16	.21	.20	.26
Competence	.14	.18	.17	.12	.19	.20	.23
Order	.09	.14	.08	.05	.10	.11	.12
Dutifulness	.02	.10	.10	.12	.12	.09	.12
Achievement Striving	.14	.24	.19	.17	.25	.28	.26
Self-Discipline	.23	.31	.24	.22	.22	.23	.27
Deliberation	.10	.15	.15	.03	.07	.01	.15

Note. M_{GPA} = mean grade point average (GPA) for 7-year persisters (z score in terms of 1st-year GPA for that group); *SD* = standard deviation of GPA for the same group of persisters in standard score form. The remaining values are correlations between personality scales and GPA, corrected only for restriction and enhancement of range in personality scores (for this sample of 7-year persisters, there was no attrition and thus no need to correct for range restriction in GPA). $N = 297$, based on complete listwise deletion on all variables.

Received June 16, 2008
Revision received February 20, 2009
Accepted March 26, 2009 ■