RICE UNIVERSITY

Predicting Professional and Technical Performance among Medical Students: Personality, Cognitive Ability, and the Mediating Role of Knowledge

by

Harrison J. Kell

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

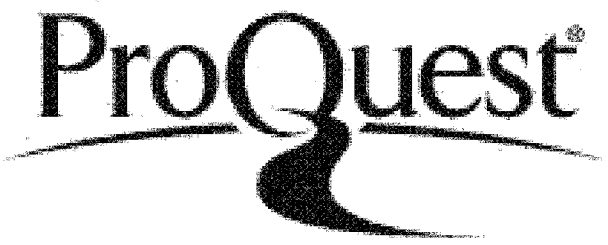
Stephan J. Motowidlo, Herbert S. Autrey Professor, Chair, Psychology

Margaret E. Beier, Associate Professor Psychology

Frederick L. Oswald, Associate Professor Psychology

D. Brent Smith, Associate Professor Management and Psychology

Houston, Texas
April, 2011



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Predicting professional and technical performance among medical students: Personality, cognitive ability, and the mediating role of knowledge

by

Harrison J. Kell

The distinction between technical and contextual performance is widely recognized in the Industrial/Organization Psychology literature (Sackett & Lievens, 2008). Less well-understood are the causal antecedents of performance in these domains and how those antecedents relate to each other. Motowidlo, Borman, and Schmit (1997) proposed that technical performance is determined largely by cognitive ability, which acts through the mediator technical knowledge to influence technical performance. They also proposed that contextual performance is mainly determined by personality traits and that these traits influence contextual performance via the mediating variable contextual knowledge. Although prior research has examined some of the causal antecedents proposed by Motowidlo et al. (1997), no study has examined these four variables simultaneously, in addition to gathering information about performance criteria in the two domains. This study examined these six variables in a sample of medical students. In keeping with the verbiage used in the medical literature, students' contextual knowledge is referred to as professional knowledge and their contextual performance is referred to as professional performance. Medical students (N = 209) beginning their third year at the University of Texas Medical School at Houston completed measures of professional knowledge and the Big Five personality traits and consented to have their MCAT scores (a proxy for cognitive ability) and their first- and second-year course grades (grade point average; a measure

of their technical knowledge) gathered for this investigation. Performance criteria consisted of attending physicians' ratings of students' professional and technical performance during their clinical rotations. Rotations were grouped according to whether they fell into the domain of Primary Care or the Specialties. Notable findings are summarized by a path analytic model. Agreeableness exerted a causal influence on professional knowledge (β = .38) and Primary Care professional performance (β = .14). Extraversion causally affected professional knowledge (β = -.22). Professional knowledge accounted for variance in Primary Care professional (β = .19) and technical performance (β = .22). Openness to experience and conscientiousness influenced technical knowledge (β's -.19 and .25). Cognitive ability was directly related to technical knowledge (β = .43) and Specialties professional (β = -.21) and technical performance (β = -.19). Technical knowledge was related to Primary Care professional (β = .32) and technical performance (β = .42) and also Specialties professional (β = .46) and technical performance (β = .57). Results generally suggest that separate causal paths underlie performance in Primary Care and the Specialties, respectively.

Contents

Contents (continued)

Tables

Tables (continued)

Figures

Appendices

Chapter 1

Introduction

Despite job performance being perhaps the "ultimate criterion" in Industrial/Organizational (I/O) Psychology, it has been conceptually neglected until relatively recently (Austin & Villanova, 1992). Campbell, Dunnette, Lawler, and Weick (1970) provided the first concrete description of job performance: behavior that influences organizational goal accomplishment. Subsequent work has refined this definition (e.g., Borman & Motowidlo, 1993; Campbell, 1990; Motowidlo, 2003) and enhanced understanding of the performance construct.

1.1 Job Performance

Much of the literature emphasizes theory that considers job performance multidimensional. Campbell (1990) developed an eight-factor model of performance that proposes to encompass the latent structure of all jobs. Murphy (1990) divided the performance domain into four dimensions: downtime behaviors, task performance, interpersonal behavior, and destructive behavior. Borman and Motowidlo (1993) offered the most parsimonious multidimensional model of performance, decomposing the construct into two components: task and contextual performance. Task performance encompasses behaviors that either transform raw materials into the organization's products or behaviors that service and maintain the organization's technical core (Borman & Motowidlo, 1993; Motowidlo, 2003). Contextual performance encompasses behaviors that maintain the organizational context in which the technical core operates. The distinction Borman and Motowidlo offered between the largely technical versus largely interpersonal aspects of performance mirrors themes that have consistently appeared in the literature for decades (e.g., Barnard, 1938; Hemphill, 1950; Organ,

1988). These two fundamental facets of performance appear to capture broadly different behavioral patterns that cut across nearly all occupations (Lance, Teachout, & Donnelly, 1992). The multidimensional nature of performance has become fairly well-accepted (Sackett & Lievens, 2008), although some dissent remains (e.g., Viswesvaran, Schmidt, & Ones, 2005).

Causal Antecedents of Job Performance

Various theories addressing the causal antecedents of job performance have been proposed over the past three decades. Prominent in all these models of job performance is the role that knowledge plays in influencing organizationally-relevant behavior.

Schmidt, Hunter, and colleagues (e.g., Hunter, 1983; Schmidt, Hunter, & Outerbridge, 1986) were perhaps the first to highlight the importance of knowledge as it relates to job performance. Using meta-analytic data Hunter (1983) conducted a path analysis linking general cognitive ability, job knowledge, work sample performance, and supervisory ratings of performance. Results showed that cognitive ability affects both knowledge and work sample performance, knowledge influences work sample performance, and knowledge and work sample impact ratings of performance. Cognitive ability is associated with ratings of performance, but its effect is fully mediated by knowledge and work sample performance. Schmidt et al. (1986) confirmed and expanded on these findings, adding experience as an additional predictor. As with cognitive ability, experience only influences performance ratings via its effect on knowledge and work sample performance. They also found that job knowledge is the strongest determinant of work sample performance. Because knowledge mediates the influence of experience and ability on ratings of performance, directly affects these ratings, and also accounts for variance in the only other variable that directly affects these ratings (work sample), it could be argued these results suggest that knowledge is the single most important predictor of performance ratings.

Campbell and colleagues (Campbell, 1990; Campbell, Gasser, & Oswald, 1996; Campbell, McCloy, Oppler, & Sager, 1993) clarified the nature of the relations among the variables studied by Hunter, Schmidt, and associates. Campbell et al. (1996) interpreted work sample performance as a measure of skill and supervisory ratings as a legitimate measure of job performance. Consequently, knowledge and skill fully mediate the effect of cognitive ability on job performance, and job knowledge is again possibly the most important predictor of job performance.

Campbell (1990; Campbell et al., 1996; Campbell et al., 1993) expanded the model presented by Hunter (1983). This theory proposes that knowledge, skill, and motivation are the sole proximal determinants of job performance. These three variables fully mediate the association between cognitive ability and other individual differences (e.g., personality, interests) and job performance. Individual differences interact with learning to shape knowledge, skill, and motivation, which in turn determine job performance.

Motowidlo, Borman, and Schmit (1997) drew on the work of Schmidt and colleagues (1986) and Campbell and colleagues (1993; 1996) to develop a theory of individual differences in task and contextual performance. Their theory argues that the proximal determinants of job performance are knowledge, skill, and habits. Motowidlo et al.'s (1997) work differs from their predecessors' by specifying that the nature of the proximal variables that influence job performance differ across task and contextual domains. Contextual performance is determined by contextual knowledge, skill, and habits while task performance is determined by task-related knowledge, skill, and habits. The variables that impact contextual behavior generally pertain to carrying out effective and ineffective interpersonal actions while the variables that impact task behavior generally pertain to carrying out effective and ineffective actions that directly transform

raw materials into goods and services or maintain an organization's core technical processes. As task-relevant variables pertain to the maintenance of the organization's technical core or the operation of the core itself, task-related variables are essentially technical variables (e.g., technical knowledge, technical performance).

In contrast to prior models of job performance, central to this theory is the idea that two distinctive forms of knowledge and skill exist and give rise to fundamentally different patterns of workplace behavior. A corollary of this idea is that it is expected that contextual knowledge and skill will be more strongly related to contextual performance than task performance and that task-relevant knowledge and skill will be more strongly related to task performance than contextual performance. In sum, job-relevant knowledge and skill have differential consequences in this model of performance.

In keeping with the idea that different types of performance behavior have different proximal antecedents, Motowidlo et al.'s (1997) theory also hypothesizes that different performance domains also have separate distal antecedents. These distal antecedents lead to variability in knowledge, skill, and habits, which in turn influences job performance. The three intervening variables thus have different consequences and different causes. Cognitive ability is hypothesized to affect task-related variables and personality traits are hypothesized to affect contextual variables. Consonant with prior theory and research, the effects of individual differences in cognitive ability and personality on job performance are predicted to be fully mediated by knowledge, skill, and habits.

Causes and Consequences of Knowledge and Skill

Evidence for the differential associations proposed by Motowidlo et al. is relatively sparse (Schmitt, Cortina, Ingerick, & Wiechmann, 2003). Some data demonstrate differential

associations between ability, personality, and the two performance dimensions (e.g., Hattrup, O'Connell, & Wingate, 1998; Hurtz & Donovan, 2000; Kamdar & Van Dyne, 2007; LePine & Van Dyne, 2001; Morgeson, Reider, & Campion, 2005) that support Motowidlo and colleagues' (1997) suppositions. Rarer are studies that have examined knowledge and skill in their investigations of discriminant relations between individual difference and performance variables. No studies to date have reported measuring task or contextual habits. Relevant aspects of the studies that have incorporated measures of knowledge, skill, or both are reviewed below.

Van Scotter and Motowidlo (1996) measured U.S. Air Force mechanics' cognitive ability, personality traits, technical knowledge, task performance, and contextual performance (interpersonal facilitation). Among personality variables they found that only conscientiousness was related to technical knowledge. Cognitive ability was also significantly associated with knowledge. Technical knowledge was significantly related to both task and contextual performance, but the magnitude of the association with task performance significantly exceeded the association between knowledge and contextual performance.

Chan and Schmitt (2002) developed a situational judgment test (SJT) that measured a combination of task and contextual knowledge among entry-level civil service employees. The authors also assessed employees' cognitive ability, personality traits, task performance, and contextual performance. Scores on the SJT were associated with all the Big Five traits (Goldberg, 1990) but were unrelated to cognitive ability. SJT scores were related to both task and contextual performance; the difference between correlation coefficients was not statistically significant.

O'Connell, Hartman, McDaniel, Grubb, and Lawrence's (2007) SJT measured manufacturing employees' contextual knowledge. Data pertaining to employees' cognitive

ability, personality traits, contextual performance, and task performance were also gathered. Cognitive ability, conscientiousness, and agreeableness were associated with SJT score, indicating that more intelligent, conscientious, and agreeable people in this sample tended to have more contextual knowledge. Although O'Connell et al.'s SJT purported to measure contextual knowledge, scores on the device were significantly related to both task and contextual performance, with no evidence for differential relations with performance across the two domains.

Bergman, Donovan, Drasgow, Overton, and Henning (2008) reported the results of a study that included measures of contextual knowledge (customer relations), cognitive ability, personality, task performance, and contextual performance. The study sample consisted of support staff for insurance salespeople. Openness to experience and cognitive ability were related to staff members' knowledge of customer relations, which in turn was related to their performance when interacting with customers. Contextual knowledge was unrelated to staff members' task performance.

Motowidlo, Brownlee, and Schmit (2008) measured the personality traits and contextual knowledge, skill, and performance of retail sales associates. They found that extraversion was significantly associated with participants' knowledge of how to interact with customers effectively but not with their skill in dealing with customers or supervisors' ratings of their performance when interacting with customers. Knowledge exerted a direct influence on skill, which in turn directly affected evaluations of associates' contextual performance.

Motowidlo, Crook, Kell, and Naemi (2009) administered a measure assessing service volunteers' contextual knowledge. They also assessed volunteers' personality traits and contextual performance in three domains: work effort, professionalism, and personal skill.

Conscientiousness and adjustment were significantly associated with volunteers' contextual knowledge. Volunteers' knowledge of effective and ineffective work effort behaviors was related to their performance in the work effort domain, but not associated with their performance in the professionalism or personal skill domains.

In a laboratory setting, Motowidlo, Martin, and Crook (2011) administered a contextual knowledge assessment to undergraduates that asked them to rate how effectively human factors professionals (HFP) behaved when interacting with customers. Participants' contextual skill was measured using a simulation that required them to take the role of an HFP interacting with a customer in nine role play scenarios. Undergraduates also provided self-reports of their personality traits. In this sample, conscientiousness was the only personality variable related to contextual knowledge. Contextual knowledge was significantly related to participants' contextual skill.

Although enough primary studies have been conducted that include measures of personality traits, cognitive ability, and task and contextual performance to warrant a meta-analysis (Hurtz & Donovan, 2000), the preceding review indicates that the intervening variables in Motowidlo et al.'s (1997) theory have been largely neglected. No studies have examined task or contextual habits and only two have included measures of knowledge and skill, both contextual. More surprising is the fact that no investigation has separately assessed both forms of knowledge and both forms of performance. The early results of Hunter (1983) and Schmidt et al. (1986), the interpretation of these results by Campbell et al. (1996), and the recent findings of Motowidlo et al. (2008) suggest that knowledge may be the most important proximal determinant of job performance, as knowledge accounts for variance in skill and the criterion itself. To improve the prediction of task and contextual performance, a logical next step is to

assess task and contextual knowledge and examine whether or not these constructs have differential causes and consequences, as Motowidlo et al.'s (1997) theory predicts. The studies reviewed indicate that the task/contextual distinction – for both knowledge and performance – exists across a wide variety of jobs (e.g., Air Force mechanic, retail associate, service agency volunteer). The apparent generalizability of these constructs implies that the findings of a study that examines them, fairly independent of the job in which those findings are generated, may be generalizable as well. The goal of this study is to investigate the differential consequences and causes of job knowledge in a sample of medical students.

1.2 The Behavioral Content of Contextual and Professional Performance

Contextual Performance

Borman and Motowidlo (1993) divided the performance domain into task and contextual components. They argued that the behaviors comprising task performance can be classified into two types: behaviors that directly transform raw materials into the organization's goods and services and behaviors that maintain the organization's technical processes by replenishing their supply of raw materials, distributing their products, or promoting their continued efficiency and effectiveness. Alternatively, contextual behaviors do not support the organization's technical processes but the work environment in which they occur. There are five categories of contextual behavior: persisting with enthusiasm and extra effort, volunteering to carry out task activities that are not formally part of a worker's own job, helping and cooperating with others, following organizational rules and procedures, and endorsing, supporting, and defending organizational objectives. Included in the category of helping and cooperating with others is assisting or helping

customers. This type of behavior occurs when employees personally distribute an organization's products or services to consumers (Brief & Motowidlo, 1986).

Medical Professionalism and Physicians' Behavior

The specific definition of the term "profession" has not been agreed upon and this study is not intended to contribute to the debate on the subject. The term is used as Evetts (2003) does, referring to professions as "occupations which are service- and knowledge-based and achieved sometimes following years of higher/further education and specified years of vocational training and experience" (p. 397). The education and training that physicians receive is intended to provide them with the knowledge and skills necessary for them to treat their patients effectively. Adequately performing technical behaviors, such as accurately diagnosing diseases and prescribing the correct medications, is necessary but not sufficient for effective patient care; effective patient care requires the satisfactory performance of interpersonal behaviors as well (Stern, Frohna, & Gruppen, 2005). Interpersonal behaviors directed toward patients comprise a large part of the medical professionalism domain.

The distinction between the technical and professional aspects of physicians' jobs is not identical to the broad distinction between task and contextual performance (Borman & Motowidlo, 1993). Contextual performance is focused on supporting the work environment, while medical professionalism is focused primarily on patient welfare and does not include organizationally-oriented behaviors such as following rules and regulations or defending the organization's objectives. A narrower construct, medical professionalism is not defined in relation to an organization's goals and effective/ineffective professional behaviors can probably best be thought of in terms of their likely impact on patients' welfare. Nonetheless, medical professionalism echoes some of the contextual domain's essential features, especially as both

encompass behaviors that occur when workers directly interact with individuals outside the organization who are seeking its goods and services.

Dividing physicians' performance domain into technical and professional facets, where professional performance is considered roughly analogous to contextual performance, seems justifiable considering the similarity in the behavioral content of these facets and those proposed by Borman and Motowidlo (1993). Further evidence supporting the treatment of professional and contextual performance as essentially identical in this study is presented in the following brief reviews of the relations between personality traits and the two constructs. Although studies of personality antecedents of performance are more common in the I/O Psychology literature than the medical literature, on whole the results of the studies that have been conducted suggest that professional and contextual behavior share the same causal variables, supporting the idea that can be treated as analogous.

Personality and Contextual Performance

Borman and Motowidlo (1993) suggested that personality traits are the most viable antecedents of contextual performance. This has been supported by empirical findings, with the majority of evidence indicating that the Big Five (Goldberg, 1990) traits agreeableness and conscientiousness are most strongly related to contextual performance. Motowidlo and Van Scotter (1994) showed that supervisory ratings of general contextual behavior correlated .31 with dependability and .22 with cooperativeness. Van Scotter and Motowidlo (1996) reported correlations of supervisory ratings of a facet of contextual performance called "interpersonal facilitation" with conscientiousness and agreeableness of .11 and .16, respectively. Borman, Penner, Allen, and Motowidlo's (2001) meta-analysis demonstrated that the mean correlations between agreeableness and conscientiousness and contextual performance are .13 and .19,

respectively, while Hurtz and Donovan (2000) estimated that true-score correlations between these two traits and interpersonal contextual behavior are .20 and .18.

Personality and Professional Performance

Historically, the major predictor of success in medical school has been cognitive ability, with "success" being defined as students' grades during the first two years of their medical education, which centers on classroom learning. This is a consequence of failure rates of 20 to 40% after the first year of medical school that occurred during the early 20th century, the inception of modern medical education in the United States (Barr, 2010). The medical aptitude test (MAT) – the predecessor of the Medical College Admission Test (MCAT) – was developed to measure applicants' cognitive ability and knowledge of basic scientific concepts to reduce the number of drop-outs after the first year of medical school; it was highly successful (Barr, 2010). Making medical students' academic success the criterion of choice caused the relative neglect of examining predictors of students' performance during the last two years of medical school. It is during the second half of medical school that professional performance is most relevant for students.

Third- and fourth-year medical students cycle through clerkship rotations that expose them to various areas of medicine (e.g., Neurology, Psychiatry). These clerkships are often the first time that students consistently interact with hospital patients. Evaluation of their performance is made by attending physicians who supervise students' treatment of patients. Research in the medical literature has consistently shown a relationship between students' performance in their clerkships and various personal characteristics, including personality.

Hojat, Callahan, and Gonnella (2004) investigated differences in global ratings of medical students' clinical competence across six third-year clerkships (e.g., Family Medicine,

Pediatrics, Psychiatry) according to mean group differences in extraversion and neuroticism (Eysenck & Eysenck, 1987). For each clerkship, students' performance was rated on a four-point scale (high honors, excellent, good, and marginal) and scores were averaged across clerkships to derive a total clinical competence score for each student. Students were divided into three groups according to these mean scores: low competence (no high honors ratings), moderate competence (one or two high honors ratings), and high competence (three to six honors ratings). Results indicated that medical students in the moderate and high competence groups were significantly more extraverted than students in the low competence group.

Hojat and colleagues (2002) also examined the association between medical students' empathy and their clinical competence. They found that more empathetic students were more likely to achieve high honors on a global clinical competence rating. Low-scorers on the empathy measure were more likely to be rated as marginally competent in their evaluations across six clerkships (e.g., Psychiatry, Surgery).

Manuel et al. (2005) correlated scores on the Sixteen Personality Factor Questionnaire (Cattell, Cattell, & Cattell, 1993) with medical students' performance during a standardized patient-based clinical skills examination. Medical students' data gathering, physical examination, and communication skills were assessed by faculty members. Physical examination ratings were unrelated to any of the personality traits but data gathering skill was related to warmth (.17) and abstractedness (-.17). Personality was most strongly and consistently related to students' communication skills. Warmth (.21), emotional stability (.14), and perfectionism (.20) were all positively associated with evaluations of students' communication competence, while privateness was negatively related (-.28).

Chibnall and Blaskiewicz (2008) focused on correlations between the Big Five traits and

Psychiatry clerkship performance. A factor analysis of the clerkship clinical evaluation items yielded two factors: knowledge and skill and interpersonal behavior. Interpersonal behavior included items assessing "professionalism in patient care" and "patient rapport" (Chibnall & Blaskiewicz, 2008, p. 202). Conscientiousness correlated .20 with scores on the knowledge and skill factor, while extraversion (.17) and agreeableness (.31) were associated with scores on the interpersonal behavior factor.

In addition to being linked to students' performance during their clerkship rotations, personality has also been linked to students' performance during standardized patient-based clinical skills examinations. Standardized patient examinations entail students interacting with an actor who has been trained to portray a specific medical case in a consistent and reliable manner (Manuel, Borges, & Gerzina, 2005). Faculty physicians observe and rate medical students' competence in several areas while they interact with the standardized patient. Scores on standardized patient exams are negatively associated with the number of communication-related complaints physicians receive (Tamblyn et al., 2007). A significant portion of the variance in scores on both clerkship and standardized patient evaluations is accounted for by how well medical students treat their patients interpersonally (Stern et al., 2005).

In sum, agreeableness and conscientiousness are important predictors of contextual and professional performance. Of the two traits, however, agreeableness appears to be the most strongly related to both contextual behavior and medical professionalism. Agreeableness is characterized by empathy, altruism, and warmth (Goldberg, 1990), all of which have been linked to medical students' interpersonal treatment of patients. This investigation consequently focuses on agreeableness as the primary distal determinant of professional performance.

1.3 Defining Medical Professionalism

The basic tenets of physicians' professional behavior have remained relatively unchanged for thousands of years (Garrison, 1966; Stern et al., 2005). Three examples of recent efforts to define medical professionalism are briefly reviewed to demonstrate that contemporary approaches to the topic converge on the same major themes.

Based on an analysis of the meaning of a profession in general and of the nature of physicians' work in particular, Swick (2000) defined medical professionalism according to "those behaviors by which we – as physicians – demonstrate that we are worthy of the trust bestowed upon us by our patients and the public" (p. 614). He listed nine sets of behaviors that he believed comprise medical professionalism: subordinating own interests to the interests of others; adhering to high ethical and moral standards; responding to societal needs and a social contract with the communities served; demonstrating core humanistic values, including honesty and integrity, caring and compassion, altruism and empathy, respect for others, and trustworthiness; exercising accountability for themselves and for their colleagues; demonstrating a commitment to excellence; demonstrating a commitment to scholarship and to advancing their field; dealing with high levels of complexity and uncertainty; reflecting upon their actions and decisions.

Shortly after the publication of Swick's (2000) article, the American Board of Internal Medicine's (ABIM) "Project Professionalism" (2001) set out six elements as being representative of professionalism: altruism, accountability, excellence, duty, honor and integrity, and respect for others. In 2002 the Association of American Medical Colleges (AAMC) and National Board of Medical Examiners (NBME) invited 25 experts in medical practice and education to a conference that produced perhaps the most definitive model of medical

professionalism to date. Prior to the beginning of the conference, its organizers assembled a list of professionalism categories to serve as a template for discussion among the attending experts. The event's organizers developed the categories from reviews of previous reports that attempted to define medical professionalism, the medical professionalism literature, and publications from various medical schools. During the conference, experts were told to provide examples of specific, observable, and measurable behaviors illustrating professionalism or a lack thereof among medical students. The behavioral examples contributed by experts during the conference were used to develop specific, behavior-based definitions for the professionalism categories that members of the AAMC/NBME had created prior to the conference. These categories and two representative behaviors from each are presented in Table 1.

Table 1

Medical Professionalism Categories and Behavioral Examples Provided by the AAMC/NBME (2002) Conference Report

Honor and Integrity Admits errors. Deals with confidential information discretely and appropriately. **Caring and Compassion** Treats the patient as an individual, taking into account lifestyle, beliefs, personal idiosyncrasies, support system. Communicates bad news with sincerity and compassion. **Respect** Respects patient rights/dignity (privacy/confidentiality, consent); knocks on door, introduces self, drapes patients appropriately, and shows respect for patient privacy needs. Demonstrates tolerance to a range of behaviors and beliefs. **Responsibility and Accountability** Arrives on time. Accountable for deadlines; completes assignments and responsibilities on time.	**Altruism** Offers to help team members who are busy. Contributes to the profession; active in local and national organizations such as the AAMC-Organization of Student Representatives. **Excellence and Scholarship** Is self-critical and able to identify own areas for learning/practice improvement. Has internal focus and direction, setting own goals. **Leadership** Teaches others. Helps build and maintain a culture that facilitates professionalism.

Although many of the definitions and behaviors presented in Table 1 are applicable to physicians' interpersonal interactions with patients, not all are patient-centered (e.g., altruism: active in local medical organizations; excellence and scholarship: is self-critical). Other definitions of medical professionalism (ABIM, 2001; Swick, 2000) also include elements that are perhaps less applicable to physicians' interpersonal treatment of patients than other aspects of their occupation. Constant across the two definitions of professionalism that preceded the list developed by the AAMC/NBME, however, are categories that are perhaps most applicable to physicians' interactions with patients (e.g., altruism, accountability, respect for others; ABIM, 2001; demonstrating caring and compassion, empathy, respect for others, and trustworthiness; Swick, 2000). These professionalism dimensions correspond to three of the AAMC/NBME's categories: Caring and Compassion, Respect, and Responsibility and Accountability. This study focuses on these three elements of medical professionalism because they seem highly relevant to physicians' interactions with patients, they are shared across prominent definitions of professionalism, and they have been defined by specific, concrete, behavioral examples.

1.4 Measurement of Contextual and Professional Knowledge

SJTs have been proposed to be measures of contextual knowledge (Chan & Schmitt, 2005; Schmitt & Chan, 2006). Findings from numerous studies (e.g., Bergman et al., 2008; O'Connell et al., 2007) support this contention and offer evidence for the construct validity of SJTs that is in accordance with the predictions of Campbell (1990) and Motowidlo et al. (1997).

Typical SJTs consist of descriptions of challenging work situations, each accompanied by behavioral alternatives representing actions that a person could potentially carry out in response to the situation. Respondents are often asked to select the response options they believe would be

most and least effective for the situation described, or the response options representing the behaviors they would be most and least likely to carry out in the given situation. Alternatively, test-takers are sometimes asked to rate all of the response options for their effectiveness (e.g., Chan & Schmitt, 2002).

Guided by the rationale that an SJT where test-takers are asked to rate response options for their effectiveness can be simplified such that each item has a single response option, Motowidlo et al. (2009) developed the single-response SJT (SRSJT). The SRSJT consisted of items adapted from critical incidents depicting highly effective and highly ineffective interactions between volunteers and a person in need of help. The critical incidents forming the items were edited such that only information about the situation and the volunteer's behavior were included – all details about the results of the behavior were removed.

In a validity study, volunteers at a service agency completed the SRSJT by rating the effectiveness of each of its items. Information about the volunteers' job performance was obtained from their supervisors. Results showed a correlation of .28 between SRSJT items reflective of work effort and supervisors' ratings of volunteers' work effort on a behavioral summary scale (Borman, 1979).

Although evidence suggests that SJTs are measures of contextual knowledge, an SJT's specific item content likely dictates what type of contextual knowledge it assesses (Motowidlo, Hooper, & Jackson, 2006). An SJT comprised of items depicting situations that entail following organizational rules and procedures is more likely a measure of contextual knowledge of behavior related to organizational policies than an SJT comprised of items depicting situations that entail helping an emotionally distraught or frustrated co-worker (which is likely a measure of contextual knowledge of supportive behavior). Using this line of reasoning, it seems fair to

assume that an SJT comprised of incidents depicting physician-patient interactions would likely address test-takers' knowledge of medical professionalism, which heavily emphasizes interpersonal behavior that promotes patient welfare. Just as contextual knowledge constitutes the knowledge underlying contextual performance, a measure like the one previously described could be argued to assess the knowledge underlying professional performance. In keeping with the labels used by Motowidlo et al. (1997), this knowledge construct can be called professional knowledge.

1.5 The Current Investigation

The overall purpose of this study was to examine the differential causes and consequences of job knowledge. This investigation specifically addresses Motowidlo et al.'s (1997) contention that task and contextual knowledge are distinct constructs and that each is a determinant of different behavioral patterns in the workplace that manifest as task and contextual performance, respectively. It also examines the antecedents of these knowledge constructs in an attempt to determine whether task and contextual knowledge have different antecedents in addition to different effects.

This study was conducted using a sample of medical students and draws on the similarity between medical professionalism and contextual performance. For the purposes of this investigation these two constructs are considered analogous. It is expected that as contextual knowledge is hypothesized to influence contextual performance, professional knowledge will influence professional performance. Similarly, technical knowledge is expected to influence medical students' technical performance. In this investigation, technical knowledge is represented by students' first- and second-year grades. Personality traits are expected to account

for variance in professional knowledge and cognitive ability is expected to account for variance in technical knowledge. Cognitive ability is represented by MCAT score. In sum, the causal paths that give rise to professional and technical performance are expected to be different and separable. Figure 1 summarizes these causal paths.

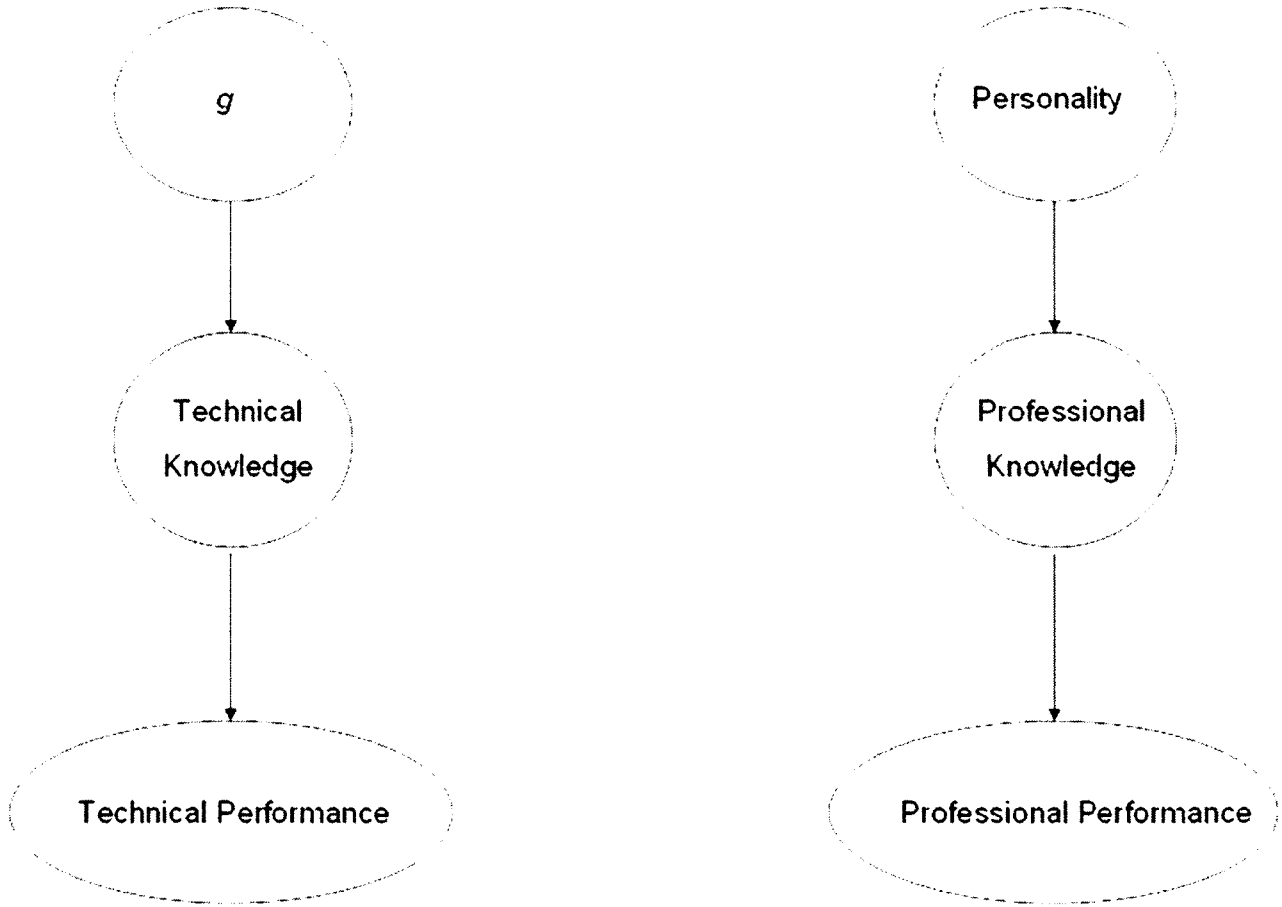


Figure1. Depiction of the differential causes and consequences of job-relevant knowledge.

This investigation expands on two studies reported by Motowidlo, Kell, Martin, Stotts, and Moreno (2011). These studies concern the development and validation of the Opinions about Physicians' Interactions with Patients (OPIP) questionnaire, an SRSJT assessing professional knowledge (a detailed description of the development of this measure is provided in the Method). Initially consisting of 200 items depicting highly effective and highly ineffective physician-patient interactions, the OPIP was completed by 150 third-year University of Texas Medical School at Houston (UTMS-H) medical students in Study 1. Ratings of medical students' performance during their Family and Community Medicine clerkship were obtained from evaluation forms completed by attending physicians who supervised students during the six week rotation period.

Students rated each of the 200 items comprising the OPIP on a 1-7 scale, where 1 (*very ineffective*), 4 (*neither ineffective nor effective*), and 7 (*very effective*). Nurses had previously used the same scale to provide effectiveness ratings for the 200 incidents. Their mean effectiveness ratings were used as norms to judge whether an incident was considered "truly" effective or "truly" ineffective. If the mean effectiveness score was below four (the midpoint of the rating scale) an incident was considered "truly" ineffective and if an incident's mean effectiveness score was above four it was considered "truly" effective. The OPIP consisted of 91 effective items and 109 ineffective items. Students' scores on the OPIP were determined by computing the mean of their ratings across the 91 effective items, reverse-scoring their ratings for the 109 ineffective items and then computing the mean across those items, and finally adding the two scores together to derive a single score representing students' professional knowledge. Scores on the two halves of the measure correlated .29 ($p < .01$) with each other. Reliability estimates treating the OPIP either as a measure of a single construct or a linear composite were

identical: .96.

Clerkship evaluations were available for 145 of Study 1's participants. Principal component analysis (PCA) of the 10-item evaluation forms resulted in a two factor solution, with five items loading on a distinct professional performance factor and five items loading on a distinct technical performance factor. Criteria were developed by summing the five professional scores to derive a professional performance variable and summing the five technical scores to derive a technical performance variable. The two sums correlated .71 with each other, suggesting that attending physicians may have difficulty discriminating between the technical and professional aspects of students' performance, perhaps due to halo effects.

The correlation between OPIP score and professional performance nearly reached the conventional standard of statistical significance ($r = .15$, $p = .07$), while the correlation between OPIP and technical performance did not ($r = .02$, NS). To account for the possibility of halo error attenuating the association between professional knowledge and professional performance, partial correlations between OPIP score and the two performance sums were computed. The partial correlation between OPIP score and the professionalism sum, controlling for the technical sum, was .20 ($p < .05$). The partial correlation between OPIP score and the technical sum, controlling for the professionalism sum, was -.13 (NS). The results of Study 1 indicated that the OPIP is differentially related to the two facets of medical students' performance.

Recognizing that it is unfeasible to use a 200-item measure for applied purposes, Motowidlo et al. (2011) shortened the OPIP to 40 items using empirical keying. The twenty normatively effective items with the highest item-criterion correlations and the twenty normatively ineffective items with the highest item-criterion correlations were selected for inclusion in the shortened scale. Study 2 addressed the construct validity of the OPIP using this

40-item measure.

One-hundred and two undergraduates completed the OPIP under the same instructions given to medical students in Study 1. They also provided self-reports of their Big Five personality traits and six of the 10 values assessed by the Schwartz Values Inventory (Schwartz, 1992). Students' professional skill was assessed using videotaped role play simulations. Undergraduates took the role of a physician interacting with a nurse, patient, or patient's family member across nine one-minute role plays. Six graduate students watched the role play videotapes and rated each on the medical professionalism dimensions of Caring and Compassion, Respect for Patients, and Responsibility and Accountability, in addition to providing ratings for the overall effectiveness of each role play. Because the study required that a total of 918 role plays be evaluated four times, three graduate students rated the exercises performed by half of the participants on the four dimensions, and three different graduate students rated the exercises performed by the second half of the participants on the four dimensions. Total performance scores for each participant were derived by summing the three graduate students' ratings across the four dimensions and nine role play exercises

The OPIP was scored in the same way described for Study 1, with an alpha coefficient of .86 if it was treated as a linear composite and an alpha of .85 if treated as a unidimensional measure. Values and personality traits were unrelated to role play performance. Scores on the OPIP were significantly correlated with ratings of students' role play performance (r = .22). At the outset of Study 2 it was predicted that OPIP score would be positively associated with agreeableness, conscientiousness, and the six values measured. Univariate and multivariate analyses found that, among personality traits, only agreeableness was associated with professional knowledge (r = .34, β = .39). Together, the Big Five traits accounted for 37% of the

variance in students' knowledge of medical professionalism. Among the six values, universalism, benevolence, and achievement were significantly correlated with OPIP score, but when OPIP score was regressed on all six values simultaneously only benevolence was significantly related ($\beta = .39$).

The findings of Study 1 of Motowidlo et al. (2011) are perhaps the first to show that a knowledge construct is associated with the interpersonal aspects of medical students' performance during their medical clerkships. Study 2 provided additional evidence for knowledge being a predictor of skill, while also deepening understanding of the nomological network surrounding professional knowledge. Interpreted according to Motowidlo et al.'s (1997) theory of job performance, the results of these two studies suggest that the OPIP is a measure of professional knowledge, not technical knowledge, and by extension that professional knowledge has different consequences, and possibly different antecedents, than technical knowledge.

The study reported here expands on Motowidlo et al. (2011) by adding data about third-year medical students' cognitive ability, technical knowledge, and performance in all of the clerkships they cycle through during their third year of medical school: Family Medicine, Internal Medicine, Neurology, Obstetrics and Gynaecology (OBGYN), Pediatrics, Psychiatry, and Surgery. As each area of medicine that these clerkships represents entails both technical and interpersonal behaviors on the part of physicians it is expected that the technical/professional distinction found for Family Medicine will be replicated for the other six rotations. Due to this, including performance ratings from multiple samples should provide a more reliable depiction of students' overall technical and professional performance.

Replicating Motowidlo et al.'s (2011) approach for each of the clerkships would result in an impractical number of criterion variables: 14. A theory-driven method was adopted to reduce

the study's criteria to a manageable number of variables. The medical literature locates different areas of medicine within Primary Care or the Specialties. Primary Care physicians practice in the context of families and the community, developing long-term relationships with patients because they are responsible for handling the majority of their patients' health care needs (Starfield, Shi, & Macinko, 2005). Specialty physicians, or Secondary Care physicians, treat patients for specialized problems that Primary Care physicians are not qualified to address. Patients do not have first contact with physicians in the Specialties for their health care needs but are usually referred to them by their Primary Care practitioner (Coulter, 1998). Secondary Care physicians are less likely to develop sustained relationships with their patients because once patients' specialized ailments are cured they no longer require specialized treatment. Of the seven clerkships that UTMS-H students rotate through, Family Medicine, Internal Medicine, and Pediatrics belong to Primary Care and Neurology, OBGYN, Psychiatry, and Surgery belong to the Specialties.

Although Primary Care and the Specialties may differ in the extent to which physicians develop sustained relationships with their patients, both medical domains have clear technical and professional components. Primary Care physicians must be able to carry out technical tasks such as taking patient histories and diagnosing illnesses, just as Specialist physicians must interpersonally interact with patients and their families during difficult times (e.g., Surgery) and may also have to develop long-term relationships with patients whose specialized problems are chronic (e.g., Neurology, Psychiatry). Thus, there were no grounds to expect that the technical/professional distinction differs between the two medical domains or that medical domain type would affect the association between the two performance factors. According to this line of reasoning the 14 criteria that would result from dividing each of the seven clerkship

rotations into technical and professional components can be reduced to an efficient four performance variables: Primary Care – professional, Specialties – professional, Primary Care – technical, , and Specialties – technical.

1.6 Hypotheses

The patterns of expected covariation enumerated in the following hypotheses are summarized in Figure 2.

Schmidt, Hunter, and colleagues (e.g., Hunter, 1983; Schmidt et al., 1986) have consistently demonstrated relations between cognitive ability and task performance, and the theories of Campbell (1990) and Motowidlo et al. (1997) predict that cognitive ability will be positively related to task performance. Borman, White, Pulakos, and Oppler (1991) also found evidence for an effect of conscientiousness on task performance. In this study technical performance is considered equivalent to task performance. Consequently,

Hypothesis 1: MCAT score and conscientiousness will be positively correlated with technical performance.

Numerous studies have reported results indicating that agreeableness is one of the primary predictors of contextual and professional performance (e.g., Borman et al., 2001; Chibnall & Blaskiewicz, 2008; Hurtz & Donovan, 2000; Manuel et al., 2005). Therefore,

Hypothesis 2: Agreeableness will be positively correlated with professional performance.

The work of Schmidt and Hunter (1998) and Borman et al. (1991) has established the link between the acquisition of technical job knowledge, cognitive ability, and conscientiousness. In this study, medical students' grades are considered a measure of their technical knowledge. Therefore,

Hypothesis 3: MCAT score and conscientiousness will be positively correlated with GPA.

McDaniel, Hartman, Whetzel, and Grubb's (2007) meta-analysis showed that agreeableness is one of the personality traits most strongly related to scores on SJTs. Motowidlo and colleagues' (2003; Motowidlo & Beier, 2010; Motowidlo et al., 2006) line of research suggests this finding may be due to the possibility that personality traits can facilitate knowledge acquisition through dispositional fit. They hypothesize that individuals with high standing on a trait (e.g., agreeableness) are more likely to believe that behavioral responses consistent with their own trait (e.g., highly agreeable responses) are the most effective means of handling challenging interpersonal situations. When responses consistent with that trait truly are the most effective means of handling such situations, these individuals possess more knowledge and are thus able to perform better in those situations (Motowidlo, 2003; Motowidlo et al., 2008). Most importantly, Motowidlo et al. (2011) found that of the Big Five traits, only agreeableness was significantly related to scores on the OPIP. Thus, it is expected that,

Hypothesis 4: Agreeableness will be positively correlated with professional knowledge.

According to the models presented by Campbell (1990) and Motowidlo et al. (1997) and evidence provided by Schmidt, Hunter, and colleagues (1986) technical knowledge is a proximal determinant of technical performance and mediates the association between cognitive ability and performance. Thus,

Hypothesis 5: Technical knowledge will be positively correlated with technical performance.

Hypothesis 6: Technical knowledge will mediate[1] the association between MCAT score and technical performance.

Conscientiousness has been hypothesized to exert a direct effect on technical performance even after taking account of mediating variables such as technical knowledge and skill (Campbell et al., 1993; Schmidt & Hunter, 1998). Borman et al. (1991) presented results supporting this hypothesis. Therefore,

Hypothesis 7: Technical knowledge will partially mediate the association between conscientiousness and technical performance.

Motowidlo et al. (2011) found that medical students' professional knowledge is associated with their professional performance. This finding is consonant with Motowidlo and associates' theory of individual differences in job performance, which also specifies that professional knowledge will mediate the effect of personality traits on professional performance. Due to this it is predicted that,

Hypothesis 8: Professional knowledge will be correlated with professional performance.

Hypothesis 9: Professional knowledge will mediate the association between agreeableness and professional performance.

Motowidlo and colleagues' (1997) theory of individual differences in job performance emphasizes that task knowledge should be primarily related to task performance and that contextual knowledge should be primarily related to contextual performance. This study presumes that the causal pathways underlying technical and professional performance do not overlap. Nonetheless, should significant relations be found between professional knowledge and technical performance or technical knowledge and professional performance, it is expected that,

Hypothesis 10: The correlation between technical knowledge and technical performance will be larger than the correlation between professional knowledge and technical

performance. The correlation between professional knowledge and professional performance will be larger than the correlation between technical knowledge and professional performance.

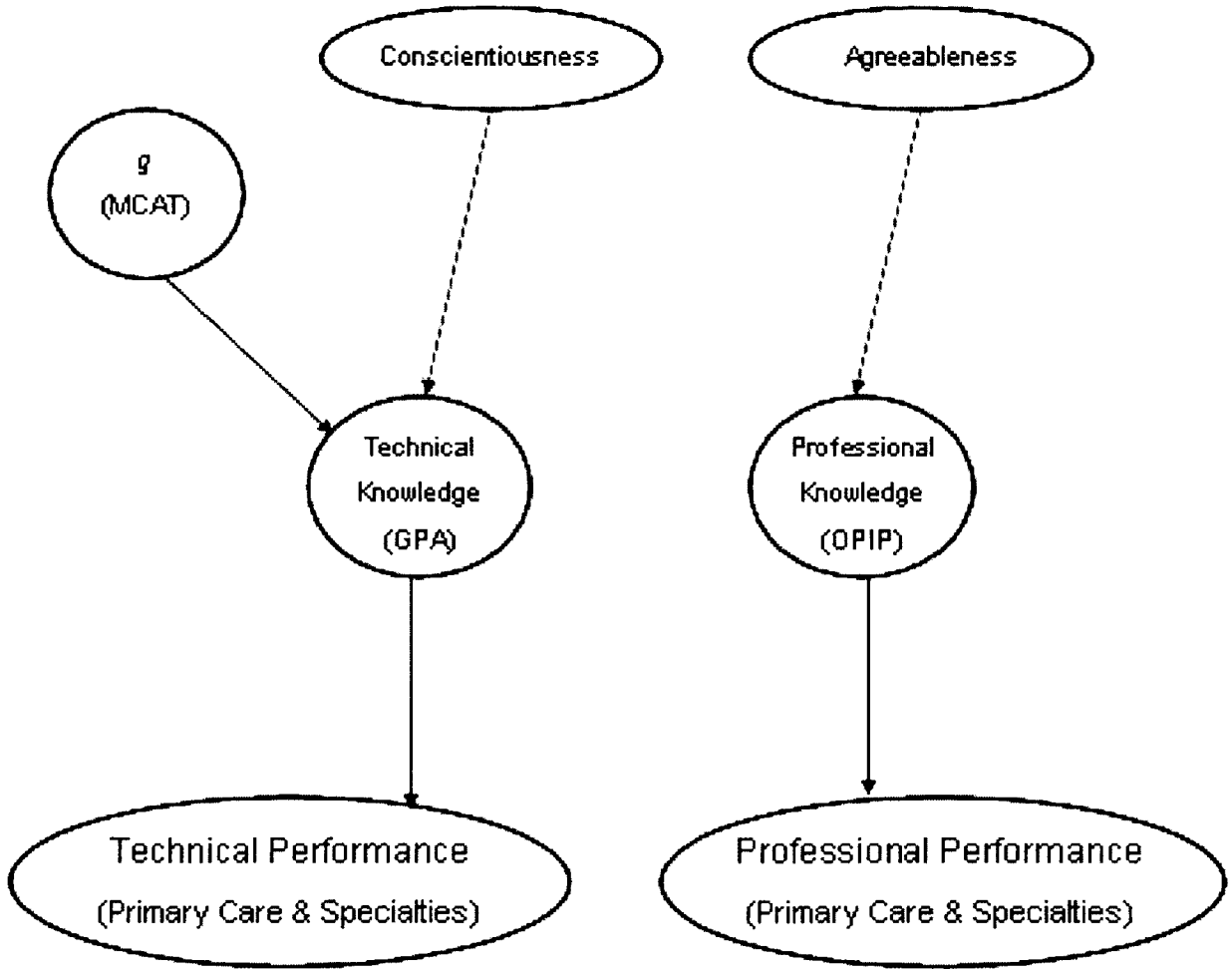


Figure 2. Depiction of the expected pattern of covariation described in Hypotheses 1 through 9.

1.7 Research Questions

The preceding hypotheses are supplemented by two research questions. These inquiries have been framed as questions rather than hypotheses due to their exploratory nature.

1. Will any personality traits, other than those hypothesized, be related to professional knowledge, technical knowledge, or the four criterion variables?

2. When all predictors, distal and proximal, are simultaneously included in a regression equation which variables will be related to each of the four performance variables?

Chapter 2

Method

Participants

The study's participants were 215 students beginning their third-year at UTMS-H. The author recruited participants by announcing to the assembled third-year class at the end of its morning orientation session that an opportunity was available for its members to participate in a research study examining medical professionalism. After explaining the details of the investigation to the students the author indicated that those interested could participate by going to the UTMS-H lounge from 11:30 am to 3:30 pm and obtaining the relevant materials from study personnel.

Six participants' data were entirely dropped from the analysis. Four participants were transfer students from dental school and did not have MCAT scores or first-year grades. Two additional students' data were excluded because their OPIP scores fell more than three standard deviations below the mean OPIP score and were treated as outliers. After these exclusions the study's total sample size was 209 students (114 males, 95 females). Additional cases of missing data were handled on a measure-by-measure basis and are described accordingly below.

Raters

Medical students' performance in each of the seven clerkship rotations was judged by an attending physician. Students approach the attending physician they would like to be their evaluator on the first day of their clerkships. The attending physician shadows the medical student two days a week while the medical student interacts with patients. On these days the physician observes all student-patient interactions during the eight-hour work period. Clerkship rotations are six weeks long. At the end of the clerkship rotation the attending physician

completes a form evaluating the student's performance during that rotation. Performance ratings for each of the seven clerkships are therefore based on approximately 12 days (96 hours) of observation of the medical student's behavior by the attending physician. The performance forms are nearly identical across rotations and are described below.

Materials

Big-Five 5 Broad Domains (Goldberg, 1999). Participants' Big Five personality traits were assessed using the 50-item Big-Five 5 Broad Domains questionnaire derived from the International Personality Item Pool (IPIP). This questionnaire is a broad-bandwidth, public domain measure available on the Internet. This measure consists of a number of brief statements, with "I" added prior to each to aid ease of responding. Subjects were asked to rate how well each statement describes themselves using a seven-point Likert-type scale, ranging from 1 (*very inaccurate*) to 7 (*very accurate*). When corrected for attenuation due to unreliability of both scales the correlation between the Big-Five 5 Broad Domains and Goldberg's (1992) 100 Big Five factor markers is .81 (Goldberg, 1999). Reliabilities for this study were: adjustment (.87), agreeableness (.80), conscientiousness (.83), extraversion (.89), and openness (.80).

Participants' scores for each of the Big Five traits were excluded from the analysis if they were missing more than two items (one-fifth of the total scale) for each of the five traits. If participants were missing two items or less from each trait scale their average score for that trait was calculated using the total number of items they did complete as the denominator. Using this criterion for exclusion 208 participants had complete personality data (one participant accounted for all the missing personality data).

Opinions about Physicians' Interactions with Patients (OPIP; Motowidlo et al., 2011). The following description of the development of the OPIP is adapted from Motowidlo et al.

(2011). The OPIP is an SRSJT consisting of 200 examples of particularly effective or ineffective physician-patient interactions. The instrument was developed by asking 50 nurses to describe occasions when they saw a physician interact with a patient in a way that struck them as especially effective or ineffective. They provided 426 brief descriptions of physician-patient interactions. Then, 30 other nurses sorted each incident into one of the AAMC/NBME (2002) dimensions of professionalism relevant to patient interaction (Caring and Compassion, Respect, or Responsibility and Accountability) and also rated each incident on a Likert-type scale for its level of effectiveness, ranging from 1 (*very ineffective*) to 4 (*neither ineffective nor effective*) to 7 (*very effective*) . Two-hundred incidents that more than half the nurses sorted into the same dimension and that had mean effectiveness ratings at the extremes of the effectiveness scale (i.e., 5 or more or 3 or less) were identified and selected for inclusion in the questionnaire. Incidents with ratings at the extreme ends of the effectiveness continuum were chosen as SJT response options that are either highly effective or ineffective tend to demonstrate the greatest validity (Waugh & Russell, 2006). This procedure also ensured that the incidents would be good examples of the professionalism categories defined by the AAMC/NBME. Representative items are included in Appendix A. Motowidlo et al. (2011) subsequently shortened the OPIP to 40 items by selecting the 20 normatively effective items with the highest item-criterion correlations and the 20 normatively ineffective items with the highest item-criterion correlations. The 40-item OPIP was administered for the purposes of this study. The OPIP was scored using the same method used by Motowidlo et al. (2011), described in the Current Investigation.

If participants were missing data for more than four of either the "effective" or "ineffective" items their OPIP scores were excluded from the analysis. Students missing four or fewer data points for either half of the scale had their averages for that half of the scale computed

with the number of items they did complete as the denominator. Using this strategy, the OPIP data set had no missing total scores.

Medical College Admission Test (MCAT; Association of American Medical Colleges, 1998). The MCAT is a multiple-choice test used to select individuals to attend medical school. The MCAT consists of three sections (biological sciences, physical sciences, verbal reasoning) and assesses applicants' knowledge of entry-level science concepts, capacity for scientific problem solving, and analytical thinking ability (Koenig, Sireci, & Wiley, 1998). Although the MCAT assesses test-takers' knowledge of specific topics it is also considered a measure of cognitive ability (Shen & Comrey, 1997). Meara and Sireci's (2000) analysis of the dimensionality of MCAT test scores concluded they consist of two lower order factors representing knowledge of science materials and verbal reasoning, respectively, and a higher order factor representing cognitive ability. Consequently, the MCAT can be considered a measure of fluid and crystallized intelligence (Cattell, 1971).

Each subsection is scored from 0 to 15 points. Point values for each subtest are summed to derive a total score for the MCAT. Total MCAT score is used to assess students' admissibility to medical school and in studies examining the predictive validity of the test (Wiley & Koenig, 1996). There were no missing data for MCAT score.

Scores on the biological and physical sciences sections correlated .46 ($p < .01$), while scores on neither of these subtests correlated significantly with scores on the verbal reasoning section (r's .11 and .26, respectively). This finding reflects prior research showing that MCAT scores are comprised of separate factors representing scientific knowledge and verbal reasoning ability (Meara & Sireci, 2000). Supporting this conclusion is the fact that the alpha coefficient for all three scores was .52, while the alpha coefficient of a scale comprised solely of scores on

the natural sciences tests was .63. According to Nunnally and Bernstein (1993), neither coefficient meets the standard acceptable for basic research, .70. Combined MCAT score is used when making decisions about whether applicants are admitted to medical school. To reflect the reality of what occurs in medical school admissions practices, despite the lower reliability obtained when the three scores are treated as a single measure, the three items were summed to form a single score for the MCAT. Despite this decision the low reliability of the combined MCAT score in this study should lead to caution when interpreting its association with other variables, as half its variance is attributable to error rather than true score variance.

Grade point average (GPA). UTMS-H students take 19 courses during their first two years of medical school. During their first year students complete: Biochemistry, Clinical Applications, Developmental Anatomy, Gross Anatomy, Histology and Cell Biology, Introduction to Clinical Medicine, Immunology, Microbiology, Neuroscience, and Physiology. During their second year students complete: Behavioral Sciences, Ethics and Professionalism, Fundamentals of Clinical Medicine, Genetics, Integrative Clinical Experience/Problem-Based Learning (ICE/PBL), Pathology, Pharmacology, Physical Diagnosis, and Reproductive Biology.

Students can receive five possible grades for their performance in each course: honors, high pass, pass, marginal performance, and fail. Alphabetical grades were converted to numbers (5 = honors to 1 = fail). Correlations among course grades ranged from .19 to .77, with an average correlation of .49. Despite the range in the correlations among grades, examination of the correlation matrix and an exploratory PCA suggested that students' grades load on a single factor. Due to this, the mean of students' scores across the 19 courses was computed to form a single GPA variable. Cronbach's alpha for GPA was .95.

Exclusion criteria for GPA were similar to the criteria used for the Big Five traits. If participants were missing scores for four or more classes (more than one-fifth of the total scale) they were excluded from the analyses. If students were missing scores for three classes or less their GPA was computed using the total number of courses they did have data for as the denominator. Using these criteria all 209 participants had complete GPA data.

Performance appraisal. UTMS-H medical students cycle through seven clerkship rotations during their third-year of medical school: Family Medicine, Internal Medicine, Neurology, OBGYN, Pediatrics, Psychiatry, and Surgery. At the end of each six week rotation their performance is evaluated by an attending physician on nine dimensions: assessment and organizational skills; clinical problem solving; technical skills; knowledge in subject area; relationships with patients; professional relationships; educational attitudes; initiative and interest; attendance and dependability. All evaluations are made using a five-point Likert-type scale. Some rotations feature slight variation in the item content of their evaluations. For example, Internal Medicine replaces the item "technical skills" with "history, physical exam, and presentation skills." Appendix B presents the evaluation form for Family Medicine, which deviates the most from the other evaluations. In addition to including all the standard items listed, it contains two additional items: "mastery of principles of family practice" and "overall clinical performance." The final variable for this rotation was not included in any analyses, as none of the other clerkship evaluations include a summary variable addressing overall clerkship performance.

For each of the four criterion variables if participants were missing a score for more than one of the clerkship evaluations their data were excluded from the analyses. For individuals missing one score or less each of the criterion variables was computed using the available

clerkship evaluation scores. Using this standard for exclusion 207 participants had complete criterion data for Primary Care (professional and technical performance) and 208 participants had complete criterion data for the Specialties (professional and technical performance).

Procedure

Participants received the Big-Five 5 Broad Domains questionnaire, the OPIP, a demographic form asking them to indicate their gender, and consent and receipt forms. After reading and signing the consent form students completed the surveys in the UTMS-H lounge and returned them to study personnel. Average completion time was 20 minutes. Students were paid $50 and signed the receipt form to verify that they had been reimbursed for their participation.

As students agreed to when signing the consent forms, the author obtained their grades for the 19 courses they completed during their first and second years of medical school, their MCAT scores, and the evaluations for their seven clerkship rotations from the Office of Student Affairs. Personality and OPIP data were gathered in July, 2009. The final data set was assembled in November, 2010, as it takes one year for third-year students to complete their clinical rotations.

Chapter 3

Results

The pervasive issue of the criterion problem (Austin & Villanova, 1992) suggests that criteria be identified and developed prior to addressing their predictors. Consequently, results are presented first for the criterion variables, then for their proximal predictors, and then for their distal predictors.

3.1 Criterion Development

Guided by the results of Motowidlo et al. (2011) and prior research in the I/O Psychology and medical professionalism literatures (e.g., Ramsey, Wenrich, Carline, Inui, Larson, & LoGerfo, 1993) it was expected that PCAs of the clerkship evaluations for each of the seven rotations would yield two factors, one representing professional performance, the other representing technical performance. Exploratory PCAs were conducted for each of the seven clerkship evaluations. Each of the analyses was restricted to produce a two component, varimax-rotated solution. Results are presented in Tables 2 through 8. Following each exploratory PCA a confirmatory factor analysis (CFA) was performed for each of the clerkship evaluations, allowing for the assessment of how well the results of the PCAs fit the data.

Principal Component and Confirmatory Factor Analyses of Clerkship Rotations

Family Medicine. A single factor emerged with an eigenvalue greater than 1.00 (7.09), with a second approaching 1.00 (.80). The next largest factor was .41. Together the first two components accounted for 78.83% of the total variance in Family Medicine clerkship evaluations. A CFA specifying that the five items the PCA indicated loaded on the professional factor loaded on a single latent factor and that the five items the PCA indicated loaded on the

technical factor loaded on a second latent factor showed adequate fit (comparative fit index [CFI] = .94, root-mean-square error of approximation [RMSEA] = .09, standardized root-mean-square residual [SRMR] = .06).

Table 2

Results for Family Medicine Clerkship Evaluation Principal Component Analysis (N = 190)

	Professional Factor	Technical Factor
Attendance & Dependability	.719	.496
Educational Attitudes	.754	.495
Initiative & Interest	.645	.556
Professional Relations	.850	.361
Relationships with Patients	.867	.266
Assessment & Organizational Skills	.289	.809
Clinical Problem Solving	.403	.767
Knowledge in Subject Area	.359	.835
Mastery of Principles of Family Practice	.471	.786
Technical Skills	.382	.778

Internal Medicine. A single factor emerged with an eigenvalue greater than 1.00 (6.63), with a second approaching 1.00 (.95). The next largest factor was .35. Together these two components accounted for 84.20% of the total variance in Internal Medicine clerkship evaluations. A CFA consistent with the results of the PCA showed adequate fit (CFI = .94, RMSEA = .15, SRMR = .05).

Table 3

Results for Internal Medicine Clerkship Evaluation Principal Component Analysis (N = 205)

	Professional Factor	Technical Factor
Attendance & Dependability	.850	.335
Educational Attitudes	.768	.528
Initiative & Interest	.767	.512
Professional Relations	.894	.300
Relationships with Patients	.836	.331
Assessment & Organizational Skills	.371	.826
Clinical Problem Solving	.361	.855
Knowledge in Subject Area	.296	.848

Pediatrics. A single factor emerged from the PCA with an eigenvalue greater than 1.00 (5.29), with a second approaching 1.00 (.85). The next largest factor was .69. Together these two components accounted for 68.25% of the total variance in Pediatrics clerkship evaluations. CFA results showed good fit (CFI = .97, RMSEA = .08, SRMR = .03).

Table 4

Results for Pediatrics Clerkship Evaluation Principal Component Analysis (N = 206)

	Professional Factor	Technical Factor
Attendance & Dependability	.835	.287
Educational Attitudes	.727	.466
Initiative & Interest	.798	.339
Professional Relations	.820	.285
Relationships with Patients	.812	.235
Assessment & Organizational Skills	.446	.672
Clinical Problem Solving	.430	.638
Knowledge in Subject Area	.241	.684
Technical Skills	.160	.777

Neurology. The PCA generated two factors with eigenvalues greater than 1.00 (6.42 and 1.02). The next largest factor was .43. Together these two components accounted for 82.67% of the total variance in Neurology clerkship evaluations. CFA results showed adequate fit for the two-factor model (CFI = .91, RMSEA = .15, SRMR = .06).

Table 5

Results for Neurology Clerkship Evaluation Principal Component Analysis (N = 198)

	Professional Factor	Technical Factor
Attendance & Dependability	.834	.386
Educational Attitudes	.811	.410
Initiative & Interest	.751	.480
Professional Relations	.877	.324
Relationships with Patients	.861	.258
Assessment & Organizational Skills	.377	.829
Clinical Problem Solving	.386	.838
Knowledge in Subject Area	.308	.822

Psychiatry. Two factors emerged with eigenvalues greater than or equal to 1.00 (5.01 and 1.00). The next largest factor was .69. Together these two components accounted for 66.85% of the total variance in Psychiatry clerkship evaluations. A CFA consistent with the PCA results showed adequate fit (CFI = .94, RMSEA = .11, SRMR = .05).

Table 6

Results for Psychiatry Clerkship Evaluation Principal Component Analysis (N = 204)

	Professional Factor	Technical Factor
Attendance & Dependability	.833	.286
Educational Attitudes	.850	.256
Initiative & Interest	.611	.494
Professional Relations	.669	.402
Relationships with Patients	.864	.222
Assessment & Organizational Skills	.339	.634
Clinical Problem Solving	.206	.762
Knowledge in Subject Area	.287	.735

OBGYN. The PCA produced two factors with eigenvalues greater than 1.00 (4.94 and 1.14). The next largest factor was .68. Together these two components accounted for 67.55% of the total variance in OBGYN clerkship evaluations. The subsequent CFA showed good fit for the specified model (CFI = .96, RMSEA = .08, SRMR = .05).

Table 7

Results for OBGYN Clerkship Evaluation Principal Component Analysis (N = 203)

	Professional Factor	Technical Factor
Attendance & Dependability	.796	.171
Educational Attitudes	.794	.251
Initiative & Interest	.704	.198
Professional Relations	.749	.251
Relationships with Patients	.668	.388
Assessment & Organizational Skills	.318	.839
Clinical Problem Solving	.403	.767
Knowledge in Subject Area	.278	.839
Technical Skills	.463	.618

Surgery. A single factor emerged with an eigenvalue greater than 1.00 (6.54), with a second approaching 1.00 (.69). The next largest factor was .40. Together these two components accounted for 80.30% of the total variance in Surgery clerkship evaluations. A CFA showed adequate fit (CFI = .96, RMSEA = .11, SRMR = .03).

Table 8

Results for Surgery Clerkship Evaluation Principal Component Analysis (N = 203)

	Professional Factor	Technical Factor
Attendance & Dependability	.785	.457
Educational Attitudes	.747	.532
Initiative & Interest	.731	.496
Professional Relations	.844	.355
Relationships with Patients	.856	.295
Assessment & Organizational Skills	.518	.712
Clinical Problem Solving	.316	.865
Knowledge in Subject Area	.355	.811
Technical Skills	.431	.727

Summary. Interpreting PCA results is subjective and accepting factors based solely on whether their eigenvalues exceed 1.00 is not recommended (Nunnally & Bernstein, 1993). Interpretation of the results as being indicative of professional and technical performance factors was based on the size of the eigenvalues produced by the PCA, inspection of scree plots, and the decline of the eigenvalue for each successive factor extracted. Except in the cases of Pediatrics and Surgery these decision criteria very clearly suggest the presence of professional and technical factors for each rotation. The results were not so clear-cut for Pediatrics, with a drop of only .16 in the eigenvalues between the second and third factors (small compared to the decreases for the second and third factors for the other evaluations), and Surgery, with the second factor of .69 representing a much larger drop from the size of first eigenvalue than for the other clerkships. Despite the somewhat ambivalent findings for Pediatrics and Surgery, however, the overall pattern of results suggests separate professional and technical factors for clerkship performance. The same items consistently loaded on separate factors for all seven evaluations. Follow-up CFAs testing two-factor solutions for each clerkship rotation uniformly produced adequate to good fit statistics.

Computation of Composite Scores as Criterion Variables

Professional and technical performance. Professional and technical performance variables were developed for each of the seven rotations by summing the items that loaded on the professional and technical factors, respectively, resulting in 14 composite variables. This was done to determine the internal consistency of ratings of students' technical and professional

Table 9

Correlations among Professional and Technical Scores for Seven Clerkship Evaluations (N = 190-207)

	M	*SD*	1	2	3	4	5	6	7	8	9	10
1. Family Medicine (prof.)	21.91	2.15	.93									
2. Family Medicine (tech.)	20.54	2.28	.81	.93								
3. Internal Medicine (prof.)	23.57	2.11	.18	.19	.95							
4. Internal Medicine (tech.)	17.27	2.15	.17	.23	.76	.93						
5. Pediatrics (prof.)	21.29	2.57	.27	.28	.22	.22	.92					
6. Pediatrics (tech.)	15.22	1.80	.19	.22	.23	.22	.72	.78				
7. Neurology (prof.)	23.13	1.77	.13	.07	.13	.20	.15	.11	.95			
8. Neurology (tech.)	17.23	1.58	.22	.22	.21	.28	.28	.22	.73	.95		
9. Psychiatry (prof.)	23.72	2.00	.12	.11	.22	.18	.11	.10	.12	.14	.89	
10. Psychiatry (tech.)	17.64	1.82	.12	.13	.14	.21	.15	.13	.19	.15	.67	.79
11. OBGYN (prof.)	23.79	1.86	.17	.20	.23	.29	.23	.17	.05	.09	.19	.20
12. OBGYN (tech.)	17.64	2.26	.14	.22	.16	.21	.21	.22	.06	.11	.26	.23
13. Surgery (prof.)	20.95	3.40	.16	.18	.22	.19	.23	.19	.25	.36	.01	.04
14. Surgery (tech.)	15.26	2.66	.17	.24	.14	.17	.21	.19	.17	.35	.10	.15

Note. Prof. = Professional performance. Tech. = Technical performance.
All correlations above .13 significant at $p < .05$. All correlations above .18 significant at $p < .01$.

Table 9 (continued).

Correlations among Professional and Technical Scores for Seven Clerkship Evaluations (N = 190-207)

	11	12	13	14
1. Family Medicine (prof.)				
2. Family Medicine (tech.)				
3. Internal Medicine (prof.)				
4. Internal Medicine (tech.)				
5. Pediatrics (prof.)				
6. Pediatrics (tech.)				
7. Neurology (prof.)				
8. Neurology (tech.)				
9. Psychiatry (prof.)				
10. Psychiatry (tech.)				
11. OBGYN (prof.)	.84			
12. OBGYN (tech.)	.64	.87		
13. Surgery (prof.)	.08	.22	.90	
14. Surgery (tech.)	.08	.24	.82	.94

Note. Prof. = Professional performance. Tech. = Technical performance.
All correlations above .13 significant at $p < .05$. All correlations above .18 significant at $p < .01$.

performance for each clerkship and to assess the intercorrelations of the professional and technical factors within and between each clerkship rotation.

Alpha coefficients for the professional and technical performance components were acceptable, ranging from .78 to .95, and are displayed on the diagonal in Table 9. Uncorrected correlations between the professional and technical performance facets of each clerkship were large, ranging from .67 (Psychiatry) to .82 (Surgery). Correlations of these magnitudes approach or exceed acceptable estimates of internal consistency (Nunnally & Bernstein, 1993), suggesting that attending physicians may have been unable to discriminate between the technical and professional aspects of medical students' performance despite the amount of time the physicians spent supervising them. The potential presence of halo error calls into question the distinction between professional and technical performance taken as a given at the outset of this investigation.

Between clerkship rotation correlations across the professional performance subscores were small to moderate, ranging from .01 (Surgery and Psychiatry) to .27 (Family Medicine and Pediatrics), with an average intercorrelation of .17. Covariation among the technical subscores was higher, ranging from .11 to .35, with an average intercorrelation of .21. These results suggest that students' performance, even within the same domain (i.e., professional or technical), was not highly consistent across the seven clerkship rotations. One possible explanation for this finding is that students may differ in how interested they are in the content of each clerkship rotation. Students are likely to be more motivated to acquire knowledge and perform well in clerkships they are interested in than those they are not (Ackerman, 1996). The possibility that raters may have been differentially rigorous across the clerkships, attenuating correlations between students' performance appraisals, has been ruled unlikely by a subject matter expert. The small

correlations of technical and professional performance sums between rotations, coupled with the large correlations between technical and professional performance sums within-clerkship rotation, cast further doubt on the distinction between the two performance domains in this investigation.

Correlations between the 14 professional and technical sums and all of the predictor variables were computed to attempt to determine whether the two performance domains could be considered separable. If the predictor variables were systematically differentially related to at least some of the professional and technical performance variables it could be argued that there was some empirical support in these data for distinguishing between technical and professional performance. Examination of the correlation coefficients presented in Table 10 suggests little to no discernible pattern between the predictors and criterion variables. While conscientiousness is significantly related to four technical performance sums versus two professional performance sums and OPIP score is related to two professional sums versus one technical sum, these results are scant evidence for what are presumed to be fundamentally different performance domains (Motowidlo et al., 1997).

These data are at odds with previous theoretical (e.g., Borman & Motowidlo, 1993) and empirical (e.g., Hurtz & Donovan, 2000; Motowidlo & Van Scotter, 1994) work that supports the distinction between the two performance domains. Although finding large correlations between the technical and professional aspects of performance within-clerkship is not surprising in light of Motowidlo et al.'s (2011) results, the small between-clerkship correlations for the two performance facets are surprising considering task and contextual performance are largely accepted as distinct domains (Sackett & Lievens, 2008). More curious is that these correlational findings contradict the results of the PCAs and CFAs of all of the clerkship evaluations. It is

difficult to consider the results of these analyses to be anomalies as they replicate prior findings both among UTMS-H medical students (Motowidlo et al., 2011) and among peer evaluations of practicing neurologists and dermatologists (Ramsey et al., 1993).

Two courses of action seemed equally reasonable. Following the correlational results presented in Tables 9 and 10 it could be assumed that the distinction between professional and technical performance cannot be supported in these data. In this scenario the most appropriate next step would be to collapse across the two domains to form seven criterion variables, each representing overall performance for one of the clerkship rotations. Alternatively, giving priority to previous empirical results and the findings of the PCAs and CFAs conducted in this study leads to the conclusion that the distinction between technical and professional performance is theoretically sound, even if the data imply that attending physicians could not discriminate between professional and technical performance during the clerkship evaluation process. Taking into account the fact that collapsing across the performance domains would require the revision of many of this study's hypotheses, in addition to yielding a cumbersome number of criterion variables, the author decided to proceed with the original analyses as planned.

Primary Care and Specialties. For the sake of efficiency the 14 professional and technical performance sums were combined according to whether they belonged to Primary Care or the Specialties. The sum computed from the Family Medicine, Internal Medicine, and Pediatrics scores formed a Primary Care variable and the sum computed from the Neurology, OBGYN, Psychiatry, and Surgery scores formed a Specialties variable. The distinction between professional and technical performance was retained, resulting in four final criterion variables: Primary Care - professional performance (15 items), Primary Care - technical performance (13

Table 10

Correlations between Professional and Technical Criteria and Predictors (N = 190-207)

	Adj.	Agr.	Con.	Ext.	Open.	MCAT	OPIP	GPA
1. Family Medicine (prof.)	-.09	.14	.03	.03	.03	-.07	.21**	.17*
2. Internal Medicine (prof.)	.05	.04	.13	.03	-.11	.04	.02	.26**
3. Pediatrics (prof.)	.06	.28**	.09	.09	-.12	-.10	.21**	.24**
4. Neurology (prof.)	.02	.08	.13	.17*	.10	-.02	-.06	.20**
5. OBGYN (prof.)	.06	.12	.20**	.00	-.13	-.09	.03	.18**
6. Psychiatry (prof.)	-.20**	.05	-.03	-.07	-.02	-.05	.02	.15*
7. Surgery (prof.)	-.02	-.04	.15*	.15*	-.02	-.06	-.08	.31**
8. Family Medicine (tech.)	.00	.03	.08	-.04	-.01	-.06	.22**	.19**
9. Internal Medicine (tech.)	.07	.05	.16*	-.01	-.04	.10	.09	.38**
10. Pediatrics (tech.)	.06	.14*	.18*	.02	.01	.00	.12	.29**
11. Neurology (tech.)	.00	.08	.12	.18**	.13	-.05	-.08	.27**
12. OBGYN (tech.)	.08	.02	.17*	-.05	-.16*	.06	-.02	.37**
13. Psychiatry (tech.)	-.06	.09	.08	-.10	.05	-.03	.06	.22**
14. Surgery (tech.)	-.01	-.05	.15*	.09	-.01	-.01	-.09	.39**

Note. Adj. = Adjustment. Agr. = Agreeableness. Con. = Conscientiousness. Ext. = Extraversion. Open. = Openness. Prof. = Professional performance. Tech. = Technical performance.
* $p < .05$. ** $p < .01$

items), Specialties - professional performance (24 items), and Specialties - technical performance (20 items).

Internal consistency estimates were satisfactory for each of the four criteria: Primary Care - professional (.88), Primary Care - technical (.85), Specialties - professional (.84), and Specialties - technical (.88). Correlations between the technical and professional subscores for Primary Care (.76) and Specialties (.81) approached but did not exceed the reliabilities for these variables. The average intercorrelation of the four final criterion scores was .54. These results may speak to the continuing influence of halo error affecting physicians' ratings of medical students' performance.

3.2 Proximal Predictors

OPIP

Cronbach's alpha for the 20 normatively effective items was .79 and the reliability coefficient for the 20 normatively ineffective items was .84. After reverse-scoring the ineffective items the two halves of the instrument correlated .44 ($p < .01$) with each other. Treating the OPIP as a linear composite resulted in a reliability estimate of .85 (Nunnally & Berstein, 1993, pp. 268-269). Treating the OPIP as a measure of a unidimensional construct and computing its reliability accordingly resulted in a nearly identical alpha estimate, .87. Following Motowidlo et al. (2011) the mean score for the (reverse-scored) ineffective items was added to the mean score for the effective items. Larger scores indicate greater capacity to differentiate professional from unprofessional actions and, thus, greater professional knowledge.

GPA

Descriptive statistics for GPA are presented in Table 11. Associations between GPA and the other variables examined are discussed in the context of the Hypotheses and Research Questions.

3.3 Distal Predictors

MCAT

Descriptive statistics for the MCAT are presented in Table 11. Associations between the MCAT and the other variables examined are discussed in the context of the Hypotheses and Research Questions.

Personality

Descriptive statistics for the Big Five personality traits are presented in Table 11. Associations between the traits and the other variables examined are discussed in the context of the Hypotheses and Research Questions.

Table 11

Correlations between Predictors and Criteria (N = 207-209)

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Adjustment	4.69	1.10	.87											
2. Agreeableness	5.98	.67	.02	.80										
3. Conscientiousness	5.44	.90	.30**	.14*	.83									
4. Extraversion	4.56	1.11	.08	.31**	.06	.89								
5. Openness	5.42	.74	.12	.14*	.19**	.16*	.80							
6. MCAT	30.06	3.23	.12	-.21**	-.14	-.13	.22**	.52						
7. OPIP Score	4.80	.79	.08	.31**	.12	-.11	.04	-.09	.85					
8. GPA	3.96	.55	.09	-.11	.16*	-.09	-.05	.35**	.02	.95				
9. Primary Care (prof.)	22.28	1.60	.03	.23**	.12	.07	-.10	-.08	.21**	.31**	.88			
10. Primary Care (tech.)	17.59	1.50	.05	.11	.20**	-.02	-.01	.05	.20**	.43**	.76**	.85		
11. Specialties (prof.)	22.89	1.36	-.06	.09	.19**	.14*	-.06	-.10	-.05	.39**	.43**	.39**	.84	
12. Specialties (tech.)	16.94	1.36	.00	.05	.19**	.06	-.03	-.02	-.05	.50**	.38**	.46**	.81**	.88

Note. Prof. = Professional performance. Tech. = Technical performance.
* $p < .05$. ** $p < .01$.

3.4 Tests of Hypotheses

Hypothesis 1

Hypothesis 1 predicted that MCAT score and conscientiousness would be positively correlated with technical performance in both medical domains. This hypothesis was partially supported (Table 11). MCAT score correlated .05 (NS) with technical performance in Primary Care and -.02 (NS) with technical performance in the Specialties. Conscientiousness was associated with Specialties technical performance ($r = .19, p < .01$) but not Primary Care technical performance ($r = .12$, NS). Together the two predictors accounted for 5% ($p < .05$) of the variance in technical performance in Primary Care and 4% ($p < .01$) of the variance in technical performance in the Specialties.

Hypothesis 2

Agreeableness was expected to be positively correlated with professional performance. As shown in Table 11, agreeableness was significantly associated with professional performance in Primary Care ($r = .23, p < .01$) but not in the Specialties ($r = .09$, NS). Hypothesis 2 received partial support.

Hypothesis 3

Hypothesis 3 stated that conscientiousness and MCAT score would be positively associated with technical knowledge (GPA). This hypothesis was fully supported, with conscientiousness ($r = .16, p < .05$) and MCAT score ($r = .35, p < .01$) being significantly positively correlated with GPA.

Regressing GPA on the two predictors produced results similar to those found at the zero-order level. When simultaneously included in a regression equation conscientiousness ($\beta = .21, p$

< .01) and MCAT score ($\beta = .38, p < .01$) were positively associated with technical knowledge. Together, the two predictors accounted for 17% of the variance in GPA.

Hypothesis 4

This hypothesis stipulated that agreeableness would be positively correlated with professional knowledge (OPIP score). The zero-order correlation between agreeableness and OPIP score was .31 ($p < .01$) and agreeableness accounted for a moderate proportion of the variance in professional knowledge ($R^2 = .10$). Hypothesis 4 was supported

Hypothesis 5

Hypothesis 5 predicted that technical knowledge would be positively correlated with technical performance across medical domains. This supposition was supported. GPA was correlated with technical performance in Primary Care ($r = .43, p < .01$) and the Specialties ($r = .50, p < .01$). Technical knowledge accounted for a large proportion of the variance in technical performance in both Primary Care ($R^2 = .18$) and the Specialties ($R^2 = .25$).

Hypothesis 6

Hypothesis 6 anticipated that technical knowledge would mediate the association between MCAT score and technical performance in both medical domains. The extent of mediation was not specified because of the variables theorized to fully mediate the relationship between basic traits and performance (e.g., knowledge, skill, habits; Campbell, 1990; Motowidlo et al., 1997), only knowledge was measured in this investigation.

According to the frequently used causal steps method to test for mediation (Baron & Kenny, 1986) this hypothesis cannot be tested because the association between MCAT score and technical performance is not statistically significant in either medical domain. MacKinnon and Fairchild (2009), however, argued that a significant direct association between an independent

variable and a dependent variable is not a prerequisite for mediation. One method for testing for mediation when there is not direct effect of an independent variable on a dependent variable is the product of coefficients approach (Hayes, 2009). The product of coefficients approach specifies that the indirect effect of an independent variable *X* on dependent variable *Y* is carried through the proposed mediatior *M*. The significance of the indirect effect can be tested by multiplying the coefficients for paths *a* (from predicting *M* from *X*) and *b* (from predicting *Y* from *X* and *M*), dividing the product of *ab* by its standard error, and comparing this ratio to the standard normal distribution (MacKinnon, Lockwood, & Williams, 2004).

The most commonly used application of the product of coefficients approach is the Sobel test (Sobel, 1982). A weakness of the Sobel test, however, is its assumption that the sampling distribution of the indirect effect is normal when it frequently is not (Stone & Sobel, 1990). An alternative to the Sobel test is bootstrapping (Hayes, 2009; MacKinnon & Fairchild, 2009; MacKinnon et al., 2004). Bootstrapping treats a study's sample as a representation of the entire population and repeatedly resamples from it with replacement. Coefficients *a*, *b*, and their product are then estimated from the resampled data *k* number of times (where *k* is specified by the researcher). Inferences about the magnitude of the indirect effect in the true population can be drawn from the resampled distribution by developing a percentile-based bootstrap confidence interval. Unlike the Sobel test, bootstrapping does not assume that the indirect effect's sampling distribution is normal. If the 95% confidence interval for an indirect effect does not include zero it is equivalent to finding that an indirect is significant at the $p < .05$ level. All mediation tests were conducted using Preacher and Hayes' (2004) macro for SPSS with supplementary information provided by hierarchical regressions (Baron & Kenny, 1986). For all bootstrap tests *k* was specified as 5000.

Bootstrap tests of the indirect effect of MCAT on technical performance in Primary Care (95% CI [.05, .11]) and the Specialties (95% CI [.06, .12]) were statistically significant. Hierarchical regression results (Table 12) indicated that MCAT score likely acts as a suppressor variable in the Specialties. In this domain the direct effect of MCAT on technical performance is -.02 (NS) but increases to -.21 ($p < .01$) when GPA is added at Step 2. MCAT score is strongly correlated with GPA ($r = .35$), which in turn is strongly correlated with Specialties technical performance ($r = .50$). This pattern of relations meets the definition of a classical suppressor variable, which is characterized by "a lack of association with the criterion and high intercorrelation with one or more predictors" (Cascio & Aguinis, 2005, p. 313). As MCAT and GPA are strongly correlated they have a large amount of overlapping variance, which is controlled for when both variables are included in the same regression equation. Due to the positive correlation between GPA and MCAT, and GPA's positive correlation with technical performance, controlling for this shared variance essentially removes variance in MCAT score that is positively related to technical performance. Much of the remaining variance in MCAT score is negatively associated with performance, as GPA is not negatively related to performance. Leaving largely only the variance in MCAT negatively related to performance exaggerates the association between MCAT score and technical performance, leading to the significant negative effect of MCAT on Specialties technical performance. This association is essentially based on a very weak negative correlation between MCAT and technical performance. Since finding such a weak association could be due to chance many additional studies supporting a negative association between MCAT score and any type of clerkship performance in medical school would have to occur before the hypothesis that MCAT and Specialties technical performance are inversely related could be seriously evaluated.

MCAT score acting as a suppressor variable in the Specialties makes evaluation of Hypothesis 6 difficult. The most conservative approach is to treat the unpredicted finding of MCAT as a suppressor as indicative of lack of support for the hypothesis in the Specialties. Bootstrapping results indicated the indirect effect of MCAT score on Primary Care technical performance was significant, however. Hypothesis 6 was thus partially supported.

Table 12

Hierarchical Regressions of Primary Care and Specialties Technical Performance Scores on MCAT and GPA

	Standardized Betas	
Step	Primary Care	Specialties
First		
R^2	.00	.00
F	.47	.10
β		
A. MCAT	.05	-.02
Second		
R^2 change	.20	.29
F change	49.92**	84.17**
β		
A. MCAT	-.12	-.21**
B. GPA	.47**	.58**
Total Equation		
R^2	.20	.30
F	25.25**	42.12**
df	206	207

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

Hypothesis 7

It was predicted that technical knowledge would partially mediate the association between conscientiousness and technical performance. Partial mediation was specified because Campbell (1990) hypothesized that motivation (i.e., conscientiousness) exerts a direct effect on job performance beyond knowledge and skill. The data supported this hypothesis in Primary Care but not in the Specialties (Table 13). The test of the indirect effect was significant in Primary Care (95% CI [.02, .22]). The direct path between conscientiousness and technical performance in Primary Care was .20 ($p < .01$) and decreased to .13 ($p < .05$) when GPA was included in the regression, indicating partial mediation.

The mediating effect of technical knowledge was also significant for technical performance in the Specialties (95% CI [.02, .22]). The direct path between conscientiousness and technical performance in the Specialties was .19 ($p < .01$) but dropped to .12 ($p = .06$) when GPA was included, however, suggesting full rather than partial mediation. Hypothesis 7 was partially supported.

Table 13

Hierarchical Regressions of Primary Care and Specialties Technical Performance Scores on Conscientiousness and GPA

	Standardized Betas	
Step	Primary Care	Specialties
First		
R^2	.04	.04
F	8.74**	7.96**
β		
A. Conscientiousness	.20**	.19**
Second		
R^2 change	.17	.23
F change	42.45**	62.11**
β		
A. Conscientiousness	.13*	.12
B. GPA	.41**	.48**
Total Equation		
R^2	.21	.26
F	26.48**	36.22**
df	205	206

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

Hypothesis 8

OPIP score was expected to positively correlate with professional performance. Professional knowledge correlated .21 ($p < .01$) with Primary Care professional performance and accounted for 4% of its variance. Professional knowledge correlated -.05 (NS) with professional performance in the Specialties. Hypothesis 8 received partial support.

Hypothesis 9

Hypothesis 9 stipulated that professional knowledge would mediate the association between agreeableness and professional performance across medical domains. Results of bootstrap tests supported this hypothesis in Primary Care (95% CI [.02, .24]), indicating that the drop in the effect of agreeableness on Primary Care professional performance when OPIP score is added to the regression equation ($\beta = .23$ to $\beta = .19$; Table 14) is statistically significant. Bootstrap results did not support this hypothesis in the Specialties (95% CI [-.17, .03]). Hypothesis 9 was partially supported.

Table 14

Hierarchical Regressions of Primary Care and Specialties Professional Performance Scores on Agreeableness and OPIP

	Standardized Betas	
Step	Primary Care	Specialties
First		
R^2	.05	.01
F	11.76**	1.49
β		
A. Agreeableness	.23**	.09
Second		
R^2 change	.02	.01
F change	4.47*	1.29
β		
A. Agreeableness	.19**	.11
B. OPIP	.15*	-.08
Total Equation		
R^2	.08	.01
F	8.22**	1.39
df	205	206

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

Hypothesis 10

Hypothesis 10 predicted that the correlation between technical knowledge and technical performance would be larger than the correlation between professional knowledge and technical performance and that the correlation between professional knowledge and professional performance would be larger than the correlation between technical knowledge and professional performance. As these correlations were drawn from the same sample this hypothesis was

evaluated using the Hotelling-Williams test (Hotelling, 1940; Williams, 1959). In both Primary Care ($t(205) = 2.53, p < .01$) and the Specialties ($t(206) = 6.42, p < .01$) the correlation between technical knowledge and technical performance significantly exceeded the correlation between professional knowledge and technical performance.

The correlations between technical knowledge and professional knowledge did not significantly differ in Primary Care ($t(205) = 1.87$, NS). In the Specialties the correlations were significantly different ($t(206) = 4.08, p < .01$), but in the opposite of the direction predicted. In the Specialties, the association between GPA and professional performance ($r = .39$) was larger than the association between OPIP and professional performance ($r = -.05$). Hypothesis 10 was partially supported.

3.5 Research Questions

Research Question 1

The first research question was directed toward examining any unhypothesized associations between the Big Five traits and the proximal predictors and criterion variables. At the zero-order level, extraversion was related to professional performance in the Specialties ($r = .14, p < .05$) and conscientiousness was correlated with professional performance in the Specialties ($r = .19, p < .01$).

Regressing the proximal predictors on the Big Five traits (Table 15) showed an unexpected, negative association between extraversion and OPIP score ($\beta = -.24, p < .01$). No other significant, unhypothesized associations were found with either professional or technical knowledge. Regression of each of the criteria on the Big Five traits (Table 16, Step 1) resulted in a significant positive association between extraversion and professional performance in the

Specialties ($\beta = .15, p < .05$), mirroring the univariate results, and a significant negative association between openness and professional performance in Primary Care ($\beta = -.16, p < .05$).

Table 15

Regression of Proximal Predictors on the Big Five Personality Traits

	Standardized Betas	
Step	OPIP	GPA
β		
A. Adjustment	.08	.05
B. Agreeableness	.38**	-.11
C. Conscientiousness	.05	.18*
D. Extraversion	-.24**	-.06
E. Openness	.01	-.06
Total Equation		
R^2	.16	.05
F	7.47**	2.31*
df	207	207

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

Research Question 2

The second research question was directed at examining which variables are significantly related to the four performance criteria when all the distal and proximal predictors are entered into a regression equation simultaneously. Results of regressing the criterion variables on all eight of the predictor variables are presented in Table 16 (Step 2).

When all predictors were taken into account, OPIP ($\beta = .14, p < .05$), GPA ($\beta = .37, p < .01$), and agreeableness ($\beta = .20, p < .01$) were significantly related to professional performance in Primary Care, and OPIP ($\beta = .15, p < .05$) and GPA ($\beta = .45, p < .01$) were significantly

related to technical performance in Primary Care. Significant predictors of professional performance in the Specialties when all variables were accounted for were MCAT (β = -.20,

Table 16

Hierarchical Regression of Criterion Variables on Predictor Variables

	Standardized Betas			
Step	P.C. (prof.)	P.C. (tech.)	Spec. (prof.)	Spec. (tech.)
First				
R^2	.09	.05	.08	.05
F	3.81**	2.29*	3.57**	2.13
β				
A. Adjustment	.00	-.01	-.13	-.07
B. Agreeableness	.23**	.11	.03	.02
C. Conscientiousness	.12	.20**	.24**	.22**
D. Extraversion	.02	-.06	.15*	.06
E. Openness	-.16*	-.05	-.11	-.08
Second				
R^2 change	.13	.19	.18	.27
F change	11.23**	16.98**	15.50**	25.60**
β				
A. Adjustment	-.01	-.03	-.11	-.05
B. Agreeableness	.20**	.09	.07	.08
C. Conscientiousness	.02	.10	.12	.09
D. Extraversion	.07	.00	.13	.05
E. Openness	-.11	-.01	-.03	.01
F. MCAT	-.13	-.06	-.20**	-.18**
G. OPIP	.14*	.15*	-.09	-.11
H. GPA	.37**	.45**	.46**	.57**
Total Equation				
R^2	.22	.25	.26	.32
F	6.96**	8.14**	8.53**	11.42**
df	205	205	206	206

Note. P.C. = Primary Care. Spec. = Specialties. Prof. = Professional performance. Tech. = Technical performance.
* $p < .05$. ** $p < .01$.

$p < .01$) and GPA ($\beta = .46, p < .01$), and for technical performance MCAT ($\beta = -.18, p < .01$) and GPA ($\beta = .57, p < .01$). Inclusion of the eight variables resulted in strong effect sizes for R^2's for all four criteria, with the least amount of variance accounted for being for Primary Care professional performance ($R^2 = .22, p < .01$) and the greatest amount being for technical performance in the Specialties ($R^2 = .32, p < .32$).

3.6 Dichotomous Scoring

As a measure of knowledge, the OPIP should be less susceptible to faking than self-report measures of personality. It is still possible that some OPIP test-takers may seek to distort their responses in order to obtain higher scores, however. This can be accomplished by test-takers rating items they believe to be effective as "very effective" and items they believe to be ineffective as "very ineffective." Because the OPIP is scored by adding the means computed for the effective and (reverse-scored) ineffective items, in most cases this approach will result in a higher score than if test-takers answered the questions according to how effective or ineffective they truly believe they are.

A strategy to protect against potential response distortion by OPIP test-takers involves re-scoring the test dichotomously. This scoring scheme awards individuals one point each time they rate a truly effective item as "slightly," "somewhat," or "very effective," and awards them zero points each time they rate a truly effective time as "neither ineffective nor effective," "slightly," "somewhat," or "very ineffective." The scoring strategy is the same for ineffective items, except it rewards test-takers who correctly rate truly ineffective items as belonging anywhere on the ineffective, rather than effective, continuum. Point totals for the effective and ineffective items are then summed to determine a final score.

This approach is equivalent to test-takers distorting their responses to every item on the OPIP, the most extreme form of faking possible. When scored normally the OPIP tests the extent to which individuals can discriminate between incidents of varying levels of effectiveness. When scored dichotomously the OPIP tests whether individuals can simply discriminate between effective and ineffective incidents. Table 17 shows the association between OPIP score and the other variables in this study when the OPIP is scored using an interval scale and using a dichotomous scale.

Associations between scores on the two versions of the OPIP and other study variables differ trivially. Most importantly, the criterion-related validities of the two versions of the test are nearly identical. These results suggest that dichotomous scoring may be a viable means of scoring the OPIP to defend against response distortion, as it does not appear to affect the measure's validity in a meaningful way.

Table 17

Correlation Coefficients for the OPIP Using Interval and Dichotomous Scoring Methods

	OPIP (I.S.)	OPIP (D.S.)
Adjustment	.08	.07
Agreeableness	.31**	.20*
Conscientiousness	.12	.07
Extraversion	-.11	-.15*
Openness	.04	.06
MCAT	-.09	.01
Primary Care (prof.)	.21**	.21**
Primary Care (tech.)	.20**	.22**
Specialties (prof.)	-.05	-.03
Specialties (tech.)	-.05	-.03

Note. I.S. = Interval scoring. D.S. = Dichotomous scoring. Prof. = Professional performance. Tech. = Technical performance.
* $p < .05$. ** $p < .01$.

3.7 Path Analysis

Figure 2 depicts the pattern of covariation that was expected among the variables in this study. This figure did not constitute a causal model, only a summary of Hypotheses 1 through 9. After testing this study's initial hypotheses and evaluating its research questions, however, the results were informative enough to lead to the formation of hypotheses about the causal relations among the variables in this investigation. It is important to note that these hypotheses cannot be considered entirely a priori as they were formulated after the examination and evaluation of many of the relations among this study's variables. The hypotheses were not entirely post hoc, however, as the causal model developed and tested was not a result of exploratory theory trimming (Pedhazur, 1982) but informed by the analyses already carried out, along with Motowidlo et al.'s (1997) theory of job performance. Consequently, it seems reasonable to assume that the model developed was amenable to significance-testing using a path analysis. The model tested is described below.

1. Agreeableness exerts causal influence on professional knowledge and professional performance in Primary Care.
2. Extraversion exerts causal influence on professional knowledge.
3. Openness exerts causal influence on technical knowledge.
4. Conscientiousness exerts causal influence on technical knowledge.
5. MCAT exerts causal influence on technical knowledge, professional performance in the Specialties, and technical performance in the Specialties.
6. Technical knowledge exerts causal influence on professional and technical performance in Primary Care and the Specialties.

7. Professional knowledge exerts causal influence on professional performance and technical performance in Primary Care.

The model tested is presented in Figure 3. Path coefficients are standardized beta weights. A test of the model showed excellent fit (χ^2 = .17, CFI = .99, RMSEA = .04, SRMR = .05). Despite these encouraging findings, however, the model should be interpreted with caution as it was not specified a priori and it is the first time a model like this has been proposed for clinical performance for medical students or physicians.

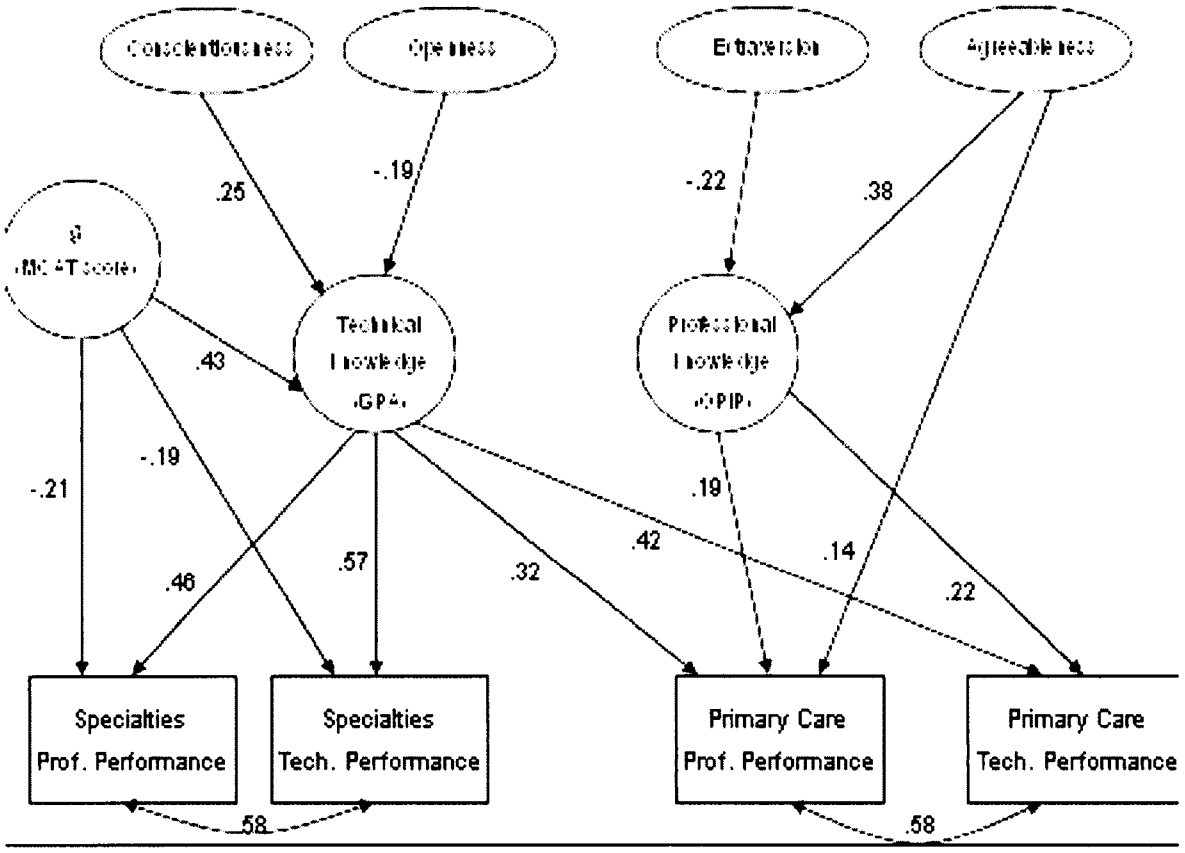


Figure 3. Path analysis results.

Chapter 4

Discussion

The overall purpose of this study was to examine the differential consequences and causes of job knowledge. This study specifically investigated professional and technical knowledge among medical students, the relations between these constructs and technical and professional performance, and relations between these knowledge constructs and cognitive ability and the Big Five personality traits. Despite difficulty distinguishing between the two performance domains due to possible halo error, the expectation that evidence for separate causal chains for different performance domains would emerge was met, although not as predicted.

4.1 Theoretical Implications

Path analytic results of this study's data indicated that the direct determinants of professional performance in Primary Care are professional and technical knowledge and agreeableness. Distal determinants of professional performance in Primary Care are agreeableness and extraversion, which influence professional knowledge. More agreeable people appear to possess more professional knowledge while more extraverted people appear to possess less professional knowledge. The effect of agreeableness on Primary Care professional performance is not mediated by knowledge. Although this finding is contrary to expectations, it is possible that agreeableness acts on professional performance in this domain through skill or habits, rather than knowledge (Motowidlo et al., 1997). Inclusion of measures of these constructs in subsequent studies may shed further light on the mechanism(s) through which agreeableness impacts professional performance in Primary Care.

Professional performance in the Specialties is directly influenced only by cognitive ability (MCAT score) and technical knowledge (GPA). Distal antecedents of Specialties

professional performance are likely conscientiousness and openness to experience. Conscientiousness is positively related to students' GPA and openness to experience is negatively related to students' GPA.

Technical performance in Primary Care is directly affected by only professional and technical knowledge. Technical performance in the Specialties is directly affected by MCAT score and technical knowledge. As with agreeableness and Primary Care professional performance, the influence of MCAT on Primary Care technical performance may be mediated by variables other than knowledge, such as technical habits or skill.

When examined in terms of technical and professional performance, these results do not suggest that job knowledge has differential effects. Professional knowledge is related to performance (professional and technical) only in Primary Care while technical knowledge is related to performance in all four of the domains studied. More clear is that these two types of knowledge have different antecedents. Extraversion and agreeableness account for variance in professional knowledge but are wholly unrelated to technical knowledge. MCAT score, openness, and conscientiousness account for variance in technical knowledge but are wholly unrelated to professional knowledge. The antecedents of these two types of knowledge appear to be completely independent of each other.

Examination of Figure 3 indicates that largely separate causal chains underlie performance in different domains, but that in this study the domains of Primary Care and the Specialties are better differentiated than the professional and technical performance domains. Performance across the three Primary Care rotations is influenced by professional knowledge, technical knowledge, agreeableness, and extraversion. Performance across the four Specialties rotations is influenced by technical knowledge, MCAT score, conscientiousness, and openness.

The only predictor that these two areas share is technical knowledge. This finding may be somewhat unique to the medical profession, as some of the courses contributing to GPA are designed to prepare medical students for their clinical experiences (e.g., Ethics and Professionalism, Fundamentals of Clinical Medicine). Even classes designed only to enhance students' technical knowledge may also improve their professional knowledge if instructors use anecdotes or case studies pertaining to interactions with patients in the course of their teaching. A measure of technical knowledge completely devoid of any professional content would presumably be unrelated to students' professional performance and would perhaps also be less strongly related to performance in Primary Care.

The stronger contrast between performance in Primary Care versus the Specialties – as opposed to technical versus professional performance – in this study could be due to several reasons. First, attending physicians apparently had difficulty differentiating between different aspects of students' performance. If professional and technical performance are not adequately separated in the data, it may obscure the differential relations that job knowledge has with each criterion. Second, attending physicians in Primary Care may be more patient-oriented than attending physicians in the Specialties, causing them to weigh the professional aspects of students' performance more heavily than Secondary Care physicians. If Primary Care and Specialties attending physicians have different conceptions of what effective performance constitutes it could explain the distinction between performance in Primary Care versus performance in the Specialties. If this distinction truly is due to differential weighting of the professional and technical elements of performance across the two domains, one way of interpreting this study's results is to consider performance in Primary Care a proxy for professional performance and performance in the Specialties a proxy for technical performance.

The nomological networks for these two performance variables are almost identical to those expected for technical and professional performance at the study's outset and clearly demonstrate that job knowledge has both different antecedents and different consequences.

4.2 Practical Implications

The results of this study have important applied as well as theoretical implications. Findings suggest that medical school admissions boards should account for personality-related variables when selecting students as their current procedures target only half of the relevant performance domain. Although the need to take into account medical school applicants' personality traits is paid lip service by admissions committees (Barr, 2010), no serious effort has been undertaken in the United States to incorporate personality tests into medical school admissions. One likely reason for this is the transparency of self-report personality measures and concern about applicants faking their responses to these assessments (Hough, Eaton, Dunnete, Kamp, & McCloy, 1990). An instrument that measures professional knowledge, like the OPIP, may represent a solution to this dilemma by serving as a proxy for personality testing. Unlike personality measures the OPIP is not transparent and it is seemingly highly resistant to faking, especially when dichotomously scored. OPIP scores are also unrelated to MCAT scores, which is one of the major criteria that medical schools use to select students. The inclusion of a predictor that is unrelated to one of the major selection tools currently used to admit medical students should improve selection efforts (Cascio & Aguinis, 2005).

The finding that the OPIP is related to personality traits and professional performance, but unrelated to MCAT score, GPA, and technical performance, supports its construct validity as a measure of professional knowledge. In turn, this supports the construct validity of SJTs in-general as measures of contextual knowledge (Motowidlo & Beier, 2010; Schmitt & Chan,

2006). This study's results also suggest that SRSJTs are viable measures of contextual knowledge and capable of predicting the interpersonal aspects of job performance.

4.3 Limitations and Future Directions

The findings described should be interpreted cautiously due to the limitations of this investigation. First, MCAT score was treated as a proxy for cognitive ability when analyses of the MCAT suggest that it is a measure of cognitive ability, verbal reasoning, and basic science knowledge (Meara & Sireci, 2000). It is possible that cognitive ability was significantly related to professional knowledge in this study's sample but that this relationship was attenuated due to the fact that MCAT is partially a measure of basic science knowledge. This may be a plausible explanation because the measure of students' advanced scientific knowledge (GPA) was uncorrelated with OPIP, the measure of professional knowledge. Thus, because MCAT is a measure of both cognitive ability and scientific knowledge it is possible that variance in MCAT score attributable to scientific knowledge is obscuring the relationship between the variance in MCAT score attributable to cognitive ability and variance in professional knowledge. Future studies may want to incorporate a purer measure of cognitive ability in order to obtain a more definitive result regarding the relationship between ability and professional knowledge.

Second, attending physicians' ratings seemed to be contaminated by halo error due to the high correlation between the technical and professional aspects of students' performance ratings across all rotations. Attending physicians may not be able to adequately distinguish between these two facets of students' performance because the behaviors that distinguish them may occur simultaneously. For example, if a student is treating a patient very well interpersonally while taking the patient's history but makes an error during the history-taking process the attending physician may not consider the error particularly important because of how well the patient was

being treated otherwise. By the same token, if a student is somewhat abrupt with a patient but is able to diagnosis a very obscure illness the attending physician may consider the student's deficiency in treating the patient to be less important due to the student's skill in diagnosing a rare disease. Because physicians do not fill-out evaluations everyday they shadow the medical students they are forced to rely on their memories of students' behavior over a six week period, which possibly further obscures the relation between technical and professional performance.

Difficulty obtaining accurate ratings of physicians' and medical students' performance in clinical settings is not restricted to this study. Carline, Paauw, Thiede, and Ramsey (1992) conducted a generalizability study in an attempt to determine how many observations would be necessary to obtain reliable ratings of medical students' performance during their Internal Medicine clerkship rotation. The investigators used a nine-item rating scale almost identical to the performance measures used in this study and defined a single observation as an entire 12-week clerkship period. Carline and colleagues (1992) determined that 14 observations would be necessary to obtain reliable (.80) ratings of students' technical performance and that between 14 and 27 observations would be needed to obtain reliable ratings of students' professional performance. Wenrich, Carline, Giles, and Ramsey (1993) reported the results of a similar study, but with nurses being responsible for rating the performance of internists. Although they found that fewer observations were necessary to obtain reliable ratings than Carline et al. (1993), Wenrich and colleagues (1993) still came to the conclusion that 10 to 15 ratings would be necessary to obtain reliable estimates of internists' professional performance.

Due to the very large number of observations that may be necessary to obtain accurate estimates of medical students' clinical behavior a useful alternative may be to evaluate their performance during standardized patient examinations. These examinations have been found to

be highly reliable (Colliver & Williams, 1993; Tamblyn, Klass, Schabl, & Kopelow, 1991), as opposed to ratings in the clinical setting. Nonetheless, scores on standardized patient exams could potentially be considered a measure of medical students' skill rather than a substitute for ratings of their actual behavior during their clerkships. The best approach likely entails gathering evaluations of students' clerkship performance and their scores on standardized patient examinations.

Future research should seek to further explore the nomological network of professional knowledge by incorporating measures of other individual differences beyond the Big Five traits. Numerous studies in the medical literature indicate that such constructs as empathy (Hojat et al., 2002) and moral reasoning (Sheehan, Husted, Candee, Cook, & Bargen, 1980) are associated with how medical students and physicians behave professionally. It might be interesting to examine how these variables are causally related to both knowledge and professional performance. Although this study affirms the importance of professional knowledge in predicting professional performance, results also showed that agreeableness is directly related to both professional knowledge and professional performance. It is unclear where empathy and especially moral reasoning fall in the nomological network. Evidence suggests that moral reasoning can be increased through exposure to instruction pertaining to morality and ethics (Bunch, 2005; Self, Olivarez, & Baldwin, 1998). Consequently, technical knowledge as measured by medical students' grades may be an antecedent of moral reasoning. Including a measure of moral reasoning in a study similar to this one would allow for the examination of the possibility that moral reasoning is responsible for – and mediates – the association between technical knowledge and students' professional performance.

Bibliography

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227-257.

American Board of Internal Medicine. (2001). *Project professionalism.* Philadelphia, PA: Author.

Association of American Medical Colleges and National Board of Medical Examiners (2002). *Embedding professionalism in medical education: Assessment as a tool for implementation.* Baltimore, MD: Author.

Association of American Medical Colleges. (1998). *MCAT interpretive manual.* Washington, DC: Author.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*, 836-874.

Barnard, C. I. (1938). *The functions of the executive.* Cambridge, MA: Harvard University Press.

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

Barr, D. A. (2010). *Questioning the premedical paradigm.* Baltimore, MD: The Johns Hopkins University Press.

Bergman, M. E., Donovan, M. A., Drasgow, F., Overton, R. C., & Henning, J. B. (2008). Test of Motowidlo et al.'s (1997) theory of individual differences in task and contextual performance. *Human Performance, 21*, 227-253.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64,* 410-421.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

Brief, A. P. & Motowidlo, S. J. (1986). Prosocial organizational behaviors. *Academy of Management Journal, 11,* 710-725.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76,* 863-872.

Bunch, W. H. (2005). Changing moral reasoning in seminary students. *Journal of Moral Education, 34*, 363-370.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.

Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258–299). San Francisco: Jossey-Bass.

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.

Carline, J. D., Paauw, D. S., Thiede, K. W., & Ramsey, P. G. (1992). Factors affecting the

reliability of ratings of students' clinical skills in a medicine clerkship. *Journal of General Internal Medicine, 7*, 506-510.

Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* New York: Houghton Mifflin.

Cattell, R. B., Cattell, A. K., & Cattell, H. E. (1993). *Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15,* 233-254.

Chan, D., & Schmitt, N. (2005). Situational judgment tests. In N. Anderson, A. Evers, & O. Voskuijil (Eds.), *Blackwell handbook of selection* (pp. 219-242). Oxford, England: Blackwell.

Chibnall, J. T., & Blaskiewicz, R. J. (2008). Do clinical evaluations in a psychiatry clerkship favor students with positive personality characteristics? *Academic Psychiatry, 32,* 199-205.

Colliver, J. A., & Williams, R. G. (1993). Technical issues: Test application. *Academic Medicine, 68*, 454-460.

Coulter, A. (1998). Managing demand at the interface between primary and secondary care. *British Medical Journal, 316*, 1974-1976.

Evetts, J. (2003). The sociological analysis of professionalism: Occupational change in the modern world. *International Sociology, 18,* 395-415.

Eysenck, H. J., & Eysenck, S. B. G. (1987). *Manual of the Eysenck Personality Questionnaire,*

juniors and adults. Kent, England: Hodder & Stoughton.

Garrison, F. H. (1966). *History of medicine*. Philadelphia: W. B. Saunders Company.

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

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

Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11,* 305-319.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76,* 408-420.

Hemphill, J. K. (1950). Relations between the size of a group and the behavior of "superior" leaders. *Journal of Social Psychology, 32,* 11-22.

Hojat, M., Callahan, C. A., & Gonnella, J. S. (2004). Students' personality and ratings of clinical competence in medical school clerkships: A longitudinal study. *Psychology, Health, & Medicine, 9,* 247-252.

Hojat, M., Gonnella, J. S., Mangione, S., Nasca, T. J., Veloski1, J. J., Erdmann, J. B. et al. (2002). Empathy in medical students as related to academic performance, clinical competence and gender. *Medical Education, 36,* 522-527.

Hotelling, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. *Annals of Mathematical Statistics, 11,* 271-283.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-

related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581-595.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.

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

Kamdar, D., & Van Dyne, L. (2007). The joint effects of personality and workplace social exchange relationships in predicting task performance and citizenship performance. *Journal of Applied Psychology, 92,* 1286-1298.

Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine, 73,* 1095-1106.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology, 77,* 437-452.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86,* 326-336.

MacKinnon, D. P., & Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science, 18,* 16-20.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39,* 99-128.

Manuel, R. S., Borges, N. J., & Gerzina, H. A. (2005). Personality and clinical skills: Any correlation? *Academic Medicine, 80,* S30–S33.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60,* 63-91.

Meara, K., & Sireci, S. G. (2000). *Appraising the dimensionality of the Medical College Admission Test.* Washington, DC: American Association of Medical Colleges.

Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology, 58,* 583-611.

Motowidlo, S. J. (2003). Job performance. In W. C. Borman, D. R. Ilgen, and R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 39-53). New York: Wiley.

Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology, 95,* 321-333.

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

Motowidlo, S. J., Brownlee, A. L., & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment, 16,* 272-281.

Motowidlo, S. J., Crook, A. E., Kell, H. J., & Naemi, B. (2009). Measuring procedural knowledge more simply with a single-response situational judgment test. *Journal of*

Business and Psychology, 24, 281-288.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). A theoretical basis for situational judgment tests. In J. Weekley & R. Ployhart (Eds.), *Situational judgment tests* (pp. 57-81). San Francisco: Jossey-Bass.

Motowidlo, S. J., Kell, H. J., Martin, M. P., Stotts, A. L., & Moreno, C. A. (2011). *Medical students' knowledge about medical professionalism predicts their professionalism performance.* Manuscript submitted for publication.

Motowidlo, S. J., Martin, M. P., & Crook, A. E. (2011). *Relations between personality, knowledge, and behavior in professional service encounters.* Manuscript submitted for publication.

Murphy, K. R. (1990). Job performance and productivity. In K. R. Murphy & F. Saal (Eds.), *Psychology in organizations: Integrating science and practice* (pp. 157-176). Hillsdale, NJ: Erlbaum.

Nunnally, J. C., & Bernstein, I. H. (1993). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, Inc.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment, 15,* 19-29.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome.* Lexington, MA: Lexington Books.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart, & Winston.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects

in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717-731.

Ramsey, P. G., Wenrich, M. D., Carline, J. D., Inui, T. S., Larson, E. B., & LoGerfo, J. P. (1993). Use of peer ratings to evaluate physician performance. *The Journal of the American Medical Association, 269*, 1655-1660.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Reviews of Psychology, 59,* 419-450.

Schmidt, F. L., & Hunter, J. E. (1998). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science,3,* 89-92.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact pf job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432-439.

Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. Weekley & R. Ployhart (Eds.), *Situational judgment tests.* (pp. 135-155). San Francisco: Jossey-Bass.

Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. In W. C. Borman, D. R. Ilgen, and R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 77-105). New York: Wiley.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). New York: Academic Press.

Self, D. J., Olivarez, M., & Baldwin, D. C. (1998). The amount of small-group case-study

discussion required to improve moral reasoning skills of medical students. *Academic Medicine, 73*, 521-523.

Sheehan, T. J., Husted, S. D. R., Candee, D., Cook, C. D., & Bargen, M. (1980). Moral judgment as a predictor of clinical performance. *Evaluation and the Health Professions, 3*, 393-404.

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

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290-312). San Francisco: Jossey-Bass.

Starfield, B., Shi, L., & Macinko, J. (2005). Contribution of Primary Care to health systems and health. *The Milbank Quarterly, 83*, 457-502.

Stern, D. T., Frohna, A. Z., & Gruppen, L. D. (2005). The prediction of professional behavior. *Medical Education, 39,* 75-82.

Stone, C. A., & Sobel, M. E. (1990). The robustness of total indirect effects in covariance structure models estimated with maximum likelihood. *Psychometrika, 55*, 337-352.

Swick, H. M. (2000). Toward a normative definition of medical professionalism. *Academic Medicine, 75*, 612-616.

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D.,... Hanley, J. A. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *The Journal of the American Medical Association, 298*, 993-1001.

Tamblyn, R., Klass, D. J., Schabl, G. K., & Kopelow, M. L. (1991). Sources of unreliability and

bias in standardized-patient rating. *Teaching and Learning in Medicine, 3*, 74-85.

Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology, 81,* 525-531.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90,* 108-131.

Waugh, G. W., & Russell, T. L. (2006, April). *The effects of content and empirical parameters on the predictive validity of a situational judgment test.* Poster session presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Wenrich, M. D., Carline, J. D., Giles, L. M., & Ramsey, P. G. (1993). Ratings of the performances of practicing internists by hospital-based registered nurses. *Academic Medicine, 68*, 680-687.

Wiley, A., & Koenig, J. A. (1996). The validity of the Medical College Admission Test for predicting performance in the first two years of medical school. *Academic Medicine, 71*, S83-S85.

Williams, E. (1959). *Regression analysis*. New York: John Wiley.

Appendix A

Sample Items from the Opinions about Physicians' Interactions with Patients Questionnaire

Effective Items:

1. A patient found dead at home was taken to the ER. The physician spent 25 minutes with the family explaining what efforts were made to revive the patient after he was admitted to the ER and answering their questions about the patient's pain and final moments.
2. Although not on-call, the physician came to visit one of his patients in the intensive care unit late one night to answer any questions she might have.

Ineffective Items:

1. When a patient's nurse neglected to bring an advanced knee-rehabilitation machine into the room as this physician had asked, the physician slammed the nurse up against the wall in front of the patient and reprimanded her.
2. On the day after getting a check-up, a patient called the physician's office with some questions. When the receptionist told the physician that the patient was on the phone, he said, "Is that her again?" loudly enough for the patient to overhear him on the phone.

(Motowidlo et al., 2011)

Appendix B

Family and Community Medicine Clerkship Evaluation Form

University of Texas
Yr3 Family & Community Medicine

Report on student::

FAMILY MEDICINE – THIRD YEAR CLERKSHIP FINAL EVALUATION FORM

INSTRUCTIONS: In the context of the student's level of training, please indicate your assessment of the student's performance by checking the appropriate box. Please use the boxes provided.

Clinical Knowledge and Skills

	Unsatisfactory, Needs work on acquiring, recording, and analyzing the patient data base.	Has basic patient data. Needs work on organization, assessment or case presentations.	Patient data complete and concise. Satisfactory organization, assessment, and case presentations.	Patient data and assessment are above average. Good case presentations.	Patient data base and assessment are outstanding. Excellent case presentations.	**N**	**Mean**
Assessment and Organizational Skills	0	0	0	0	0		
	Has difficulty identifying the key problems. Demonstrates little independence. Uses time inefficiently.	Identifies major problems but cannot set priorities. Somewhat inefficient.	Identifies major problems. Adequate utilization of lab and other parameters. Efficient.	Identifies major problems. Above average grasp of information. Efficient use of lab and other services.	Identifies major and minor problems in perspective. Superior grasp of information. Very efficient use of lab and services.	**N**	**Mean**
Clinical Problem Solving	0	0	0	0	0		
	Unable to demonstrate basic skills of interview/PE/	Minimal level of basic skills. Needs work on interviews/	Satisfactory basic skills appropriate to clerkship. Steady improvement.	Demonstrates above average mastery of basic skills. Performs above average clerkship level.	Demonstrates superior mastery of basic skills. Performs far in advance of clerkship level.	**N**	**Mean**
Technical Skills	0	0	0	0	0		
	Shows inadequate knowledge of medical principles and patho-physiology related to the patient's problems.	Shows a minimal amount of knowledge related to the patient's problems.	Shows adequate comprehension of basic pathophysiology and relates them to the patient's problems.	Shows above average comprehension of basic medical principles relating to the patient's problems.	Shows superior knowledge of the basic medical principles relating to the patient's problems.	**N**	**Mean**
Knowledge in Subject Area	0	0	0	0	0		

Interpersonal Relationships

	Often discourteous and/or non-empathetic with patients. Puts personal convenience above patient's needs.	Fair rapport, occasionally discourteous if patient is hostile.	Generally good rapport with patients. Generally empathetic.	Good rapport with patients. Empathetic.	Consistently courteous and empathetic. Gives patient's needs priority, even with unpleasant or hostile patients.	**N**	**Mean**

Relationships with Patients	0	0	0	0	0		
	Behavior interferes with satisfactory performance. Discourteous to nurses and/or residents. Hostile or uncooperative.	Occasionally discourteous or uncooperative. Sometimes does not work with others.	Cooperative and courteous with staff, other students, nurses.	Works well with others. Consistently courteous.	Works very well with others. Consistently courteous. Has admiration and respect of coworkers.	**N**	**Mean**
Professional Relationships	0	0	0	0	0		
	Is often sullen, hostile, and argumentative. Unresponsive to suggestions. Reacts poorly to criticism.	Responsive to questions but does not volunteer. Rarely contributes to discussions.	Good participation. Contributes to discussion. Accepts criticism well.	Above average participation. Actively contributes to discussions. Active learner.	Excellent participation. Eager to learn and be evaluated. Stimulates the learning process.	**N**	**Mean**
Educational Attitudes	0	0	0	0	0		
Personal/Professional Characteristics							
	Not well motivated. Avoids "doing" when possible. Appears disinterested. Never volunteers.	Accepts average load of work. Rarely volunteers or actively participates.	Does all work expected. Often volunteers.	Works hard. Regularly volunteers. Interested in learning.	Works exceptionally hard. Active leader/participant. Seeks new learning experiences.	**N**	**Mean**
Initiative and Interest	0	0	0	0	0		
	Consistently absent or late to conferences and/or patient rounds. Not prepared for didactic or patient care activities.	Occasionally late or absent. Sometimes unprepared for conferences or rounds.	Attends all teaching conferences and rounds. Is prepared and prompt. Provides appropriate patient care.	Prompt and prepared for scheduled conferences and rounds. Occasionally assumes added responsibilities for patient care.	Consistently prompt and prepared at scheduled conferences/rounds. Assumes added responsibilities for patient care.	**N**	**Mean**
Attendance and Dependability	0	0	0	0	0		
	Consistently fails to demonstrate an understanding of the concepts or principle of Family Practice.	Seldom demonstrates an understanding of the concepts or principles of Family Practice.	Demonstrates an average mastery of the concepts and principles of Family Practice.	Demonstrate an above average mastery of the concepts of Family Practice.	Consistently demonstrated an excellent mastery of the concepts and principles of Family Practice.	**N**	**Mean**
Mastery of Principles of Family Practice	0	0	0	0	0		
	SO FAR BELOW EXPECTATIONS THAT STUDENT HAS FAILED	BELOW EXPECTATIONS – REMEDIATION NEEDED	STUDENT MEETS ALL EXPECTATIONS	EXCEEDS EXPECTATIONS IN MOST RESPECTS	GREATLY EXCEEDS EXPECTATIONS IN VIRTUALLY ALL RESPECTS (TOP 10%)	**N**	**Mean**
Overall Clinical Performance	0	0	0	0	0		

Please comment below on the overall performance of the student, indicating specific areas of strength or weakness in the student's knowledge or performance. Comments are mandatory.

Comments:

Footnote

[1] The extent of mediation is not specified for Hypotheses 6 and 9 because this investigation only assesses medical students' knowledge. Several theories of job performance (e.g., Campbell, 1990; Motowidlo et al., 1997; Schmidt & Hunter, 1998) specify that knowledge in addition to variables such as skill and habits fully mediate the association between basic individual differences such as personality and cognitive ability and performance. Because these theories do not make predictions about the extent to which knowledge alone mediates relations between distal predictors of job performance and performance itself Hypotheses 6 and 9 do not address the magnitude of the mediated effect.