Predicting Clinical Performance in Medical School: The Contribution of Academic and Non-Academic Characteristics

By

William Poe

A Master's Paper submitted to the faculty of the
University of North Carolina at Chapel Hill in
partial fulfillment of the requirements
for the degree of Master of Public Health
in the Public Health Leadership Program.
Chapel Hill
2012

_____

_____

Date

_____

Date

Abstract

Background: Medical school admissions committees need accurate and precise screening tools to select among well-qualified applicants. Traditional academic performance indicators, such as grade-point average (GPA) and Medical College Admission Test (MCAT) scores, correlate with standard measures of academic success in medical school, such as pre-clinical grades and scores on National Board examinations. In contrast, valid and reliable predictors of clinical performance during clerkships are lacking.

Objective: To compare the power of two admissions tools – one based on previous academic achievement (MCAT scores and GPA) and another based on a content analysis of self-reported biographical information – to predict clerkship performance during the third year of medical school.

Methods: A retrospective cohort study of 109 students at the University of North Carolina at Chapel Hill School of Medicine (UNC-SOM) was performed to evaluate the relationship between various pre-admission tools and clinical performance as assessed by faculty preceptor ratings. The two main predictors were an aggregate academic achievement index derived from MCAT scores and GPA and a "non-academic" index derived by coding self-reported biographical information on the medical school application. Ratings on these two indices were correlated with likert-type preceptor ratings on third year clerkships. I analyzed the data using Pearson's correlation coefficients and multiple linear regression to determine the independent, predictive validity of traditional "academic" variables and "non-academic" variables.

Results: Considered as independent predictors, BCPM and MCAT scores did not correlate with clinical performance ratings on any of the three clerkships. The non-academic achievement index

predicted performance on Pediatrics (standardized $\beta = 0.18$, p=0.09) but did not predict performance in Internal Medicine or Family Medicine. In contrast, the academic achievement index predicted clinical skills ratings in Family Medicine (standardized $\beta = 0.18$, p=0.07) but not performance ratings in the other clerkships. Neither admissions tool reached statistical significance at the 0.05 level. Ceiling effects limited the size of the correlations among predictors and criterion measures.

Conclusion: This study adds to the growing body of literature suggesting that traditional measures of past academic achievement have little, if any, predictive power for clerkship performance during medical school. The findings suggest that self-reported biographical data from student applications may correlate with future performance in a clinical setting. Medical schools should address the challenge of developing more accurate clinical assessment tools and work to find measures beyond academic achievement to select among future physicians.

Acknowledgements

Table of Contents

Introduction

Over the past decade, graduate medical education (GME) and continuing medical education (CME) have reorganized around the six ACGME-endorsed core competencies. Mirroring this trend, medical schools have begun reframing their curricula along the same competency-based model. The proposed rollout of a newly revised MCAT in 2015, with a newly added focus on the social and behavioral determinants of health, manifests that the entire physician-training pipeline – beginning with the selection of medical students – is undergoing realignment upon common standards.

A sea change in the goals, design, and evaluation of an educational system demands an assessment of its raw-material inputs. Two important questions arise: What criteria do we currently use to select students for medical school? More importantly, do these criteria attract and matriculate students who possess the capacity to develop and maintain the core competencies?

To address the first question: admission practices at U.S. medical schools rely heavily upon measures of academic achievement such as standardized test scores and premedical GPA[1][2][3]. Selection committees also evaluate applicants through personal interviews, biographical sketches, personal essays, and letters of reference [4][5].

At present, the evidence is insufficient to answer the second question. The move toward competency-based undergraduate medical education (UME) is a recent shift, and a paucity of research links admissions processes to competency-based outcomes. The majority of medical admissions research focuses on pre-admission variables that correlate with measures of academic achievement. Robust evidence demonstrates that previous academic performance and standardized test scores – most frequently measured by pre-medical GPA and MCAT scores,

respectively – correlate moderately with grades during the preclinical years and scores on subsequent licensure examinations [2][3].

In contrast to their predictive validity for pre-clinical grades and licensure exam scores, pre-medical GPA and standardized test scores correlate inconsistently with clinical performance. Numerous studies have found that MCAT scores and pre-medical GPA do not predict clinical performance in medical school or in residency [6][7,8][9,10].

Why should we be interested in identifying predictors of clinical performance during medical school? One reason is that the six core competencies – medical knowledge, patient care, professionalism, interpersonal and communication skills, practice-based learning, and systems-based practice – are designed to be assessed in a clinical context. If these are the criteria by which we aim to judge success in medical school, it follows that the outcomes of admissions-related research should include competency-based, clinical assessments.

Another reason to search for better predictors is that the ultimate goal of the admissions process is to select applicants with the potential to practice effective clinical medicine. The performance of doctors in practice involves both intellectual knowledge and a variety of professional and interpersonal qualities such as communication skills, empathy, compassion, and integrity. As board scores and preclinical grades during medical school primarily measure intellectual knowledge, they provide only intermediate, partial surrogates for the full range of abilities necessary to practice clinical medicine. In order to complement these measures, the effectiveness of admissions criteria should also be assessed over a longer time horizon that includes clinically based performance endpoints beyond medical school.

It is intuitive that clinical performance within medical school would correlate more robustly with long-term clinical behavior than would performance on a series of fact-based

written examinations. Thus, it is not surprising that many studies have found that clinical performance in medical school predicts performance in residency [11] [12] [13-15]. Evidence also supports a relationship between clinical performance and long-term physician behavior. Numerous studies have linked unprofessional behavior in a clinical setting during medical school to subsequent disciplinary action by medical boards [16] [17] [18]. In the same studies, MCAT scores and undergraduate GPA were not associated with disciplinary action. These findings suggest that correlates of clinical performance at the UME level could serve as valid and reliable predictors of long-term competence and success.

Altogether, this chain of evidence manifests the value of searching for valid and reliable predictors of clinical performance at the level of UME. In turn, the medical school admissions process could benefit greatly from tools that identify those applicants most likely to succeed in a competency-based, clinical environment. To this end, I conducted a pilot study at the UNC School of Medicine (UNC-SOM) to evaluate the predictive validity of "non-academic" components of the medical school application for clinical performance as assessed by the ACGME competencies. In this study "non-academic" refers to applicant qualities apart from MCAT scores and undergraduate GPA, as found in applicants' biographical sketches of activities. I conducted a first test of their predictive validity of these criteria by comparing their association with third year clinical performance to that of traditional admissions benchmarks such as MCAT scores and pre-medical GPA.

This paper begins by reviewing the evidence on predictors of clinical performance: can we anticipate which applicants will excel in the clinical curriculum during the 3rd and 4th years of medical school?. I then present the methods, results, and conclusions from this initial pilot test of a novel method for predicting competency-based clinical performance.

Evidence Of Predictors of Clinical Performance in Medical School

I conducted a systematic review of the literature to determine what is known about predictors of clinical performance in medical school (further details of the review are in Appendix A, including the search strategy and a critical appraisal of the articles). Recent, well-conducted reviews address the relationship of academic, or 'cognitive' predictors (e.g. tests scores and pre-med GPA) and clinical performance[2,3]. Therefore, my review focused on studies assessing portions of the application other than test scores and pre-medical GPA – namely personal interviews, biographical sketches, personal essays, and letters of reference. Personality measures and learning styles were not addressed in this review for two reasons: recent, good quality reviews deal with these variables [7,19], and Medical schools do not routinely use these measures in evaluating and selecting students.

I reviewed and evaluated the quality of 18 articles published from 1978 to 2011. Numerous methodological problems complicate analysis of the studies I included. One difficulty is that using populations of only admitted students means that ceiling effects restrict the range of predictor variables. By the same token, clinical performance measurements are frequently skewed, limiting the range of the criterion measure. Ceiling effects not only weaken the correlation among predictors and outcomes but also make it difficult to extrapolate the results of a particular study to a population of potential medical school students – the ultimate goal of medical admissions research.

Another challenge is that the literature does not use a common definition of clinical performance. Some studies defined this variable in terms of relatively subjective measures such as preceptor clinical ratings, whereas others used more objective measures such as standardized

test scores (i.e. clerkship shelf exams). This heterogeneity in clinical performance measures makes it difficult to compare the results between studies. The lack of a common standard of assessing clinical performance in the literature is not surprising and reflects a challenge in medical education that the competency system is designed to address.

The literature also lacks a common approach to the measurement of predictor variables. For instance, among studies that examined the predictive value of interviews, interview formats ranged from unstructured to semi-structured, and each study used a different technique to derive interview ratings. Similar measurement variance characterized other predictor variables in the included studies; these inconsistencies make it difficult to compare the correlations between predictors and clinical performance across studies. Furthermore, many of the included studies had small sample sizes and nearly all were conducted at a single medical school, both of which result in poor generalizability.

Despite these limitations, we can find meaningful trends, depending on the type of predictors and the criterion variables used to measure clinical performance. I identified three general categories of predictor variables: Interviews, narrative portions of the application (letters of reference, personal essays, and the biographical list of activities), and educational background and life experience (e.g. undergraduate major or military experience). I use these categories to organize the rest of the review.

Overall, the evidence for the predictive validity of interview ratings for clinical performance is inconsistent. Studies using structured interviews tended to find reasonable correlations (0.20-0.60) [20] [21] [22] with clinical performance; unstructured interviews, not surprisingly, are less well correlated [23,24] or not significantly correlated at all[25-27]. The ability of interviews to predict clinical performance also depends on the operational definition of "clinical

performance." Specifically, interviews consistently predict clinical performance as measured by preceptor clinical ratings [20,21,23,24] but this correlation is attenuated or disappears altogether when the clinical performance metric includes standardized test scores [21,23,25-27]. These findings suggest that standardized tests of clinical knowledge and preceptor ratings evaluate distinct but interrelated aptitudes or skill sets. The observation that MCAT scores and pre-medical grades correlate relatively well with NBME shelf exam scores, whereas interview ratings do not, provides empirical support for this theory.

Further evidence to support the predictive validity of interviews comes from studies using proxy measures of clinical performance including dean's letter ratings or residency match rankings [22,28,29]. Two studies found small but significant correlations between interview ratings and dean's letter ratings (r = 0.22-0.37) [28,29]. A third study found a significant relationship between interview scores and family medicine residency rankings for senior medical students. [22] In contrast to the waning predictive validity of academic measures over the course of medical school, these results suggest that interviews may capture stable, long-term information relevant to an individual's clinical performance.

Objective structured clinical exams (OSCE's) theoretically provide a more direct and accurate means of measuring clinical proficiency. Few studies have used OSCE scores as an outcome, but the trend is that structured interviews correlate moderately with OSCE scores [21] whereas unstructured interviews do not [22,30]. Research demonstrates that both OSCE scores and preceptor ratings positively correlate with residency director ratings [31]. It is therefore interesting that interviews correlate differentially with preceptor ratings and OSCE scores. OSCEs are more standardized and lend themselves to discrete (often correct vs. incorrect) responses whereas personal interviews tend to be more open-ended and permit an array of equally meritorious

answers. In addition, OSCEs test a limited range of clinical skills, but preceptors might be assessing a wider range of abilities. One hypothesis to explain the different findings is that the method of assessment (standardized vs. open-ended) influences the relationship between predictor and outcome; this hypothesis would predict that unstructured interviews do not correlate with OSCE scores.

Information derived from the narrative components of the application – the biographical sketch, letters of reference, and the personal essay – does not consistently predict clinical performance[21,22,32,33]. Some studies show promising results [22,32,33] while others have found no significant association between these variables and clinical performance. [21] The literature offers no guide to whether some narrative components are better predictors than are others, and the inconsistent findings once again impel us to develop studies using both both comparable predictor variables and clinical performance measures.

The robust pre-medical curriculum and emphasis on basic science during the preclinical years – Abraham Flexner's legacy – naturally leads to interest in whether educational background influences medical school performance. Do medical students who majored in social sciences or humanities perform as well as do those who studied more traditional pre-medical natural sciences? For clinical performance, the answer is yes. The small collection of studies that address this question [34-36] do not demonstrate significant or meaningful differences in clinical performance among students of different undergraduate disciplines. These results are consistent with other studies showing no association between undergraduate course of study and preclinical grades or licensure examination scores [37-39]

Study Rationale

My review underscores several gaps in the literature. I address the areas for future research in detail in Appendix A, but several in particular provided the rationale for the present study. One salient issue is that the most promising predictive admissions tools for clinical performance (e.g. the MMI) tend to be the most time and resource intensive. Unfortunately, admissions committee participation is frequently voluntary, and time and effort is already stretched at many U.S. medical schools (and perhaps especially so at public institutions). Most schools do not have the personnel to conduct a content analysis of each applicant's personal statements, and there will likely never be a standardized, multiple-choice test for the interpersonal characteristics desirable in future physicians. Hence the need for research that pilots efficient and scalable admissions screening tools for clinical performance. All applicants at U.S. medical schools are required to submit a common application through the American Medical College Application Service (AMCAS). The AMCAS application contains both academic data (MCAT scores, pre-medical GPA) and non-academic, or narrative, information including a personal essay and a biographical sketch of experiences. Therefore, a predictive admissions tool derived from the systematic, efficient evaluation of the AMCAS application could provide a standardized, scalable method of comparing the anticipated clinical performance of applicants. Moreover, this type of common assessment would not require a fundamental alteration of the current admissions process.

Few studies of predictors of clinical performance at U.S. medical schools lead us to turn to the results of studies conducted in Australia, Canada, or Europe, but are they relevant to American medical education, given notable differences in each country's general educational

and medical training systems?. Studies designed to predict clinical performance in an explicitly American contextwould help address this need.

Finally, I did not identify any predictive studies that assessed clinical performance among U.S students using competency-based metrics, despite the obvious need for evidence linking admissions criteria to policy-related outcomes.

Considering these gaps in evidence, I developed a rating scheme for the narrative portions of the AMCAS application and tested its ability to predict clinical performance. The study compares the predictive validity of this tool to MCAT scores and GPA, the traditional predictors of medical school performance. The UNC School of Medicine, one of the nation's leading public medical institutions, is an ideal setting in which to pilot a generalizable clinical performance prediction tool among U.S. students.

Methods

Participants

The UNC SOM received approximately 4,500 applications during the 2007-2008 admissions cycle. Of these applicants, nearly 500 were invited to complete a supplemental application and on-site interview. Of those who interviewed, 210 were accepted and 160 matriculated in the fall of 2008. The selection pools for both dual degree (M.D./Ph.D) and out-of-state candidates are systematically different than the in-state applicant pool. Admissions processes differ for the former two groups compared with in-state students. Therefore, only in-state, traditional M.D candidates (e.g. students seeking a four-year degree) were included in this initial pilot study. The goal of the study was to predict clinical performance; therefore, after receiving IRB approval from the UNC IRB, and with the help of staff in the UNC SOM's Office

of Medical Education who served as data custodians, I had a final database that included students beginning the M.D. program in August, 2008 who completed the first three years of the program consecutively, performing their 3rd year clinical rotations in the 2010-2011 academic year. Thus, those who chose to take a leave of absence *after* their third year in order to pursue a research fellowship, an alternative degree such as an M.P.H. (which about 20% of the UNC medical school class does) or for other reasons were included.. Interview data were missing for 17 students; therefore these students were excluded during analysis. The total remaining study population was 92 students, whose admissions and clinical performance data were linked with a unique identifier held by the data custodian, and not known to the study team.

Predictor/Admissions Variables

Academic Achievement: MCAT& Undergraduate Grade Point Average: The Medical College Admissions Test (MCAT) is a standardized test consisting of multiple-choice questions in 3 subsection: Biological Science (BS), Physical Science (PS) and Verbal Reasoning (VR). Each subsection is scored on a scale of 1-15; the total MCAT score is the overall score out of 45. The AMCAS application subdivides undergraduate GPA into two cumulative averages – one for courses taken in biology, chemistry, physics, and math (BCPM), and one based on courses taken in all other subjects (AO). For this study I used BCPM, since it is used in practice by the UNC admissions committee as a way to compare GPAs among applicants of different undergraduate concentrations or Majors. Both the total MCAT score and the BCPM were converted into single digit scores using pre-determined cutoffs (appendix x) on a scale of 0-3. These two scores were added to form a composite academic achievement variable, which ranged from 0-6. This composite score was treated as continuous variables in multiple regression analysis.

Other Achievement: Non-academic Variables: The American Medical College Application Service (AMCAS) includes a section in which applicants may list and describe up to 15 activities relevant to their motivation, qualification, and capacity to become an effective physician. The study authors developed a rating methodology for this portion of the application containing six dimensions: Motivation for and exposure to clinical medicine, service orientation, research experience, maturity, capacity for leadership and teamwork, cultural awareness and appreciation of diversity, and evidence of dedication or commitment to activities. The first three dimensions were rated on a scale of 0-3 (with 0 being the lowest score and 3 being the highest), whereas the latter four each were collapsed into a binary scale (0 being average or poor evidence of the particular quality or attribute and 1 being the strong presence of, or excellence in, the quality or attribute.) Interview ratings for each student were converted into single digit-scores on a scale of 0-3 and added to the total application rating. Therefore the composite "non-academic' rating ranged from 0-16. The study author rated the biographical sketch. The composite numerical score was analyzed as a continuous variable and compared with the composite academic achievement variable for its ability to predict clinical performance ratings.

Demographic Variables and Potential Confounders: Potential confounders or variables known to influence medical school performance [7] including age, race, and sex, were also collected. Socioeconomic data was unavailable for the majority of the study population, therefore this variable was not included in the analysis. Age was treated as a continuous variable; race and sex were coded categorically. These variables were treated as covariates in multiple regression analysis.

Clinical Performance Measures

UNC students complete 8 mandatory clerkships (Adult Ambulatory Medicine, Family Medicine, Internal Medicine, Obstetrics and Gynecology, Neurology, Pediatrics, Surgery, and Psychiatry) over a 12-month period during their third year. In each clerkship, residents and attending physicians evaluate the individual learner's clinical performance. These evaluations include quantitative ratings and qualitative summary comments in numerous competency-aligned domains such as medical knowledge, patient care, interpersonal and communication skills, and professionalism. Each domain contains subcategories. For instance, within the category of clinical skills, preceptors are asked to rate an individual's ability to elicit an appropriate history, perform an appropriate physical exam, develop and critically evaluate a differential diagnosis, among other skills. In internal medicine and pediatrics, students may receive multiple clinical evaluations. For these clerkships, multiple ratings within each subdomain were averaged. These subdomain averages were then averaged across domains to create composite scores for Clinical Skills, Medical Knowledge, and Professionalism. Figure 1 presents the conceptual methodology for deriving these summary statistics for the Internal Medicine clerkship. These composite scores were treated as continuous outcomes.

This study was conducted prior to the rollout of UNC's common clinical assessment form; therefore, not all clerkships had realigned their assessments with the core competencies during the period in which the study cohort completed their clinical rotations. As a result, I used performance evaluations from the clerkships that did have competency-based clinical assessments: Internal Medicine, Family Medicine, and Pediatrics. These clerkships were also chosen because they comprise UNC's core primary care clinical curriculum. The evaluations for specific clerkships differ in terms of the number and type of rating items in each dimension as

well as in the numerical scale – performance in Internal medicine is evaluated on a scale of 1-5, Family medicine is evaluated on a scale of 1-8, and Pediatrics are rated on a scale of 1-9. Family medicine professionalism was rated as binary and had no range. Therefore this variable was dropped from analysis.

Data

The Office of Medical Education at UNC maintains records of student AMCAS application information for internal purposes. Data on all predictor variables were stripped of any unique identifiers and extracted from extant databases into excel files; no data were obtained directly from the online AMCAS portal. UNC uses a commercial product (one45 software) to record clinical performance ratings. These one45 quantitative ratings, without qualitative comments, were stripped of all unique identifiers and exported into excel files. All excel files were linked by a single 3 digit random ID, held in the sole possession of the data custodian for the project. Study investigators remained blinded to the clinical ratings until the AMCAS-derived predictor scores were completed.

Statistical Analysis

All statistical analyses were performed using STATA version 10.1 statistical software (Stata Corporation, College Station, TX). I calculated descriptive statistics (mean, standard deviation, range, percentages) for predictor variables and clinical ratings.

I examined the relationship between admissions measures and clinical performance ratings using Pearson's correlation coefficients. I used Spearman's rank correlation for variables that did not fit a normal distribution. Demographic predictors of performance were analyzed with parametric (ANOVA or t-tests) or their corresponding non-parametric (Kruskal-Wallis or

Wilcoxen rank sum) tests. I used multiple linear regression to determine the independent predictive validity of the two admission indices – Academic Achievement vs. Non-Academic Achievement – for clinical performance ratings when all potential confounders were taken into account. I also determined the proportion of variance in clinical performance ratings explained by each predictor.

Results

Descriptive Statistics

Means for academic, demographic and the two admissions predictor indices are shown in Table 1. The study population was more likely to be Caucasian and female. The distribution of grade point averages (BCPM) was skewed. Since this variable was a component of the academic achievement index, the index scores also clustered towards their upper limit (Table 1). In contrast, the MCAT scores were more normally dispersed. Scores on the "non-academic" index were not heavily skewed but were restricted in range; 75% of scores were greater than 9 (scale 0-16). The 17 students who were excluded from analysis based on missing interview data were not systematically different from the remaining cohort in terms of age, sex, race, BCPM, MCAT scores, or clinical ratings (data not shown).

Table 2 shows the means for the three clerkship performance summary measures: clinical skills, medical knowledge, and professionalism. The evaluation form for family medicine evaluation was slightly different from that in Internal Medicine and Pediatrics in two ways: it lacked ratings corresponding to the 'medical knowledge' dimension, and the professionalism ratings employed a binary system (equivalent in practice to a pass/fail standard). Therefore, only the clinical skills summary measure was analyzed for this clerkship. Overall, the summary measures did not disperse normally and were limited in range. This was particularly true for

Internal Medicine, in which 75% of students received an average clinical skills rating of 4 or higher on a scale of 1-5. Despite different rating scales, (Table 2) ratings in each clerkship spanned approximately 60% of their possible range. The summary measures were derived from integer value ratings; therefore the individual evaluations in each clerkship took on the following whole numbers: Internal Medicine 3-5, Family Medicine 4-8, Pediatrics 5-9. These limited ranges created significant ceiling effects for the linear regression analysis.

Demographics Variables

Table 3 presents the relationship between demographic variables and clerkship summary measures. Since most summary measures and predictor variables were not normally dispersed, these relationships were analyzed with parametric (ANOVA or t-tests) and non-parametric (Kruskal-Wallis or Wilcoxen rank sum) tests. Both methods of analysis gave similar results in terms of statistical significance (with one exception, discussed below). Means are reported here since they are more interpretable.

In Internal medicine and Pediatrics, sex was significantly related to clinical performance, with females performing better. This pattern appeared to be reversed in Family Medicine, where males received higher average clinical ratings than did females; however, this difference did not reach statistical significance. In contrast, race was significantly related to clinical performance in Family Medicine, with Caucasians receiving higher average ratings than did non-whites. Race was not significantly related to average ratings in either Internal Medicine or Pediatrics. Age was not significantly related to clinical performance in either Internal Medicine or Pediatrics, but did correlate significantly with performance in Family Medicine under the assumption of a normal distribution (Pearson's $r = -0.37$ $p <0.01$), but not when using non-parametric tests(Spearman's rho $= -0.07$, $p = 0.53$).

Table 4 shows the correlations between admission predictors and clinical performance summary measures. When treated as independent predictors, neither total MCAT scores nor BCPM correlated with average clinical ratings in any clerkship. In contrast, the composite academic achievement index correlated significantly with average clinical skills ratings in Family Medicine ($r_s = 0.23$, $p<0.05$). This suggests a possible interaction between test scores and grade point average; in other words, the combination of these statistics may be more useful than either alone. Pediatrics ratings demonstrated the opposite trend; in this clerkship, scores on the non-academic index correlated significantly with clinical skills ($r_s = 0.22$, $p<0.05$), and more weakly with medical knowledge and learning ($r_s = 0.19$, $p<0.10$). Neither predictor index correlated with Internal Medicine clinical ratings.

Regression Analysis

Both admission indices and three demographic variables were entered stepwise into multiple linear regression models to predict ratings on the clinical performance dimension. Table 5 summarizes the standardized beta coefficients and the total proportion of variance explained by the combination of the variables. The unique proportion of the total variance explained by each predictor can be determined by comparing adjacent models. For example, race, sex, and the two admissions indices explained 17% of the variance in clinical skills ratings for Internal Medicine. Dropping race from the model, sex, the academic achievement index, and the non-academic index explained 14% of the variance in clinical skills ratings, suggesting that race accounted for 3% of the total variance in clinical skills ratings on this clerkship.

Demographic Variables: Sex was the only predictor than reached statistical significance for Internal Medicine ratings (standardized $\beta = -0.40$, $p<0.05$), explaining 12% of the variance in clinical skills scores. Thus, when age, race, and both admissions achievement indices are taken

into account, men received, on average, clinical skills ratings that were 0.4 points lower than did women. Considering that 75% of students received scores within a 1-point range (either a 4 or 5 out of 5), this statistically significant difference translates to a practically meaningful difference in terms of the ratings. Sex was also important, albeit to a lesser extent, in determining clinical skills ratings in Pediatrics (standardized β = -0.23, p<0.05). Thus, men scored approximately one-fifth of a point lower on average than did women, controlling for other variables.

In contrast, clinical skills ratings in Family Medicine were influenced significantly by age (standardized β = -0.35, p<0.05) and race (standardized β = 0.17, p<0.10). This means that older students received lower ratings on average than did younger students; specifically, clinical skills ratings decreased by 0.35 points for each 1-year increase in age at matriculation. Regarding race, the beta coefficient means that Caucasians received scores nearly one-fifth of a point higher on average than did students of other races and ethnicities. Sex was not a statistically significant predictor of Family Medicine ratings.

Admissions Indices: In this first test, the admissions indices showed variable and inconsistent predictive power. The non-academic achievement index predicted performance on Pediatrics (standardized β = 0.18, p=0.09) but did not predict performance in the other two clerkships. In contrast, the academic achievement index predicted clinical skills ratings in Family Medicine (standardized β = 0.18, p=0.07) but not performance ratings in either Internal Medicine or Pediatrics. Neither admissions tool reached statistical significance at the 0.05 level.

Discussion

Demographic Variables

A common finding in the literature is that women tend to perform better than do men during the pre-clinical years and on clinical assessments. [40-43] This study's results support that trend. Sex was the only significant predictor of clinical performance across multiple clerkships. If women consistently performed better on average than men, one would expect this difference to be reflected the Family Medicine ratings as well. However, the lack of a correlation between sex and performance on this clerkship may reflect the low statistical power of this study rather the absence of a true difference.

Older age predicted lower ratings in Family Medicine; this finding is at odds with extant research. Studies have shown either no difference in performance between older and younger students, or that older students tend to outperform younger ones. [40,44] Intuitively, this makes sense – one might expect older students to be more mature and have more work experience in a professional setting than their younger counterparts.

Some evidence exists that ethnic minorities have a greater likelihood of academic difficulty than do non-minorities. [45,46] Therefore it is not altogether surprising that Caucasians received higher average ratings than did students of other ethnic backgrounds in Family Medicine, but our definition of race was crude, and the study was not designed to detect differences in performance based on racial or ethnic background. The absence of interaction between race and ratings on the other clerkships makes it unlikely that race plays a stable or significant role in determining clinical performance.

Admission Indices

In keeping with past studies, traditional measures of academic achievement had minimal predictive validity for clinical performance. Considered as independent predictors, BCPM and MCAT scores did not correlate with clinical performance ratings on any of the three clerkships. However, when combined, these measures were a significant predictor of performance in Family Medicine. This finding was not robust – i.e. not stable across clerkships – but nevertheless warrants attention. Previous work has shown that both grade point average *and* standardized test scores independently predict performance, and that the effects of these measures are additive. In contrast, this study implies an interaction between the two variables. Some evidence suggests that GPA may be a better indicator of work ethic than of native intelligence [47], In contrast, the MCAT is commonly thought of as more of an aptitude test, although it certainly requires discrete foundational knowledge. The two metrics likely reflect overlapping but distinct traits and abilities. An effect-modification type interaction between the GPA and MCAT scores is both conceivable and worth further study.

This study did not show that non-academic measures are superior to measures of academic achievement as predictors of clinical performance. By the same token, the predictive power of the non-academic index was not inferior to that of the achievement index. Moreover, the non-academic achievement index was a significant predictor of clinical performance in Pediatrics. The magnitude of this finding is small, but the effect is in the direction anticipated by our hypothesis. This finding suggests that personal interviews and applications provide useful information about a student's future clinical performance. The correlations found in this study are compressed by ceiling effects. Additionally, the sample size was very limited. Both of these statistical weaknesses may have obscured significant relationships between predictor indices and

performance ratings. Moreover, this was a pilot study using a single rater - a serious threat to validity of the findings. Therefore the results reported here must be interpreted with caution.

The marked range restriction in clinical ratings speaks to an underlying issue in the evaluation of clinical performance. The ratings clustered towards the upper end of their range and, in practical terms, served as a pass/fail standard rather than a fluid spectrum reflecting differential degrees of clinical competency. One explanation is that the vast majority of students truly performed above average, exhibiting only minor differences in competence. If this is the case, the ratings should be redesigned with greater sensitivity in their upper ranges. Alternatively and more likely, preceptors err on the side of grade inflation and simply do not give average or poor ratings unless there are glaring deficiencies in performance. Furthermore, clerkship evaluations are standardized, but this does not minimize subjectivity on behalf of the preceptors. One would expect students of exceptional clinical skills to perform well across clerkships. However, only 4 out of 92 students received ratings above the 75th percentile on all three rotations. This finding could reflect true differences in performance. For example, a student who performs well in Internal Medicine may not do as well in an outpatient setting and consequently receive lower ratings in Family Medicine. Residual confounding could also explain this finding: the analysis did not control for the effects of rotation number – e.g. the order of clerkships during the third year – a variable known to influence clerkship performance. These points notwithstanding, since the student was the common denominator across clerkships, this variance likely issues from inter-rater subjectivity. The evaluation forms do contain comments to clarify each rating; still, one preceptor's definition of "excellent" may be equivalent to another's definition of "average". In order to address this problem, the evaluation method should be more standardized. However, getting preceptors both within and across clerkships to agree on

common evaluation standards is a daunting and logistically challenging task. The time and space for workshops or formal training on evaluation is already scarce.

Perhaps the most intriguing finding is that non-overlapping variable sets predicted performance in Family Medicine versus performance in Internal Medicine and Pediatrics. To our knowledge, no studies report an interaction between clerkship specialty and demographic variables. Since race and age are not modifiable, one would expect their effect on performance (if one exists) to be stable across rotations, similar to the trend for sex discussed earlier. Therefore the observed difference in this study may reflect artifacts of UNC's third year curriculum. Superficially, the Internal Medicine and Pediatrics clerkships are more similar to each other than either is to the clerkship in Family Medicine. The former two are inpatient rotations that require teamwork with residents; the latter is exclusively outpatient based, and students work in a one-on-one format with a single preceptor. Reflecting these differences, there are multiple evaluations for students in Internal Medicine and Pediatrics but only a single evaluation in Family Medicine. Moreover, students at UNC rotate through numerous clerkship sites around North Carolina and are frequently in rural or underserved areas for their family medicine rotation. These structural and logistic issues do not readily explain why race, age, and the academic achievement index were significant predictors of Family Medicine ratings while they lacked predictive validity in the other clerkships. However, the analysis did not control for rotation number, preceptor, or number of evaluators. In this small sample, interactions could exist among these variables and sex, age, and race. This potential for residual confounding emphasizes the complexity and practical challenges of accurately evaluating clinical performance in medical school.

The underlying assumption in this study was that of linearity – a straightforward correlation between the admission indices and clinical performance ratings. While simplistic, this model mimics the rationale of the medical school admissions process, in which many screening criteria explicitly use a linear scale and imply a simple, dose-response relationship between predictor and future performance. For instance, it is generally assumed that an applicant who scores a 37 on the MCAT is a more desirable than is one who scores a 30, all other things being equal. However, our crude scale for rating academic achievement employed crude cutoffs. That this tool had predictive validity for family medicine ratings suggests that, above a certain threshold, the predictive validity of the test dissipates. In other words, moving from a score of 26 to 30 likely reflects a greater change than moving from a 36 to 40.

Moreover, professional performance of any type is influenced by personality, learning styles, life circumstances, and other contextual variables. Thus, while the linear model is easy to grasp, the relationship between preadmission variables and outcomes is considerably more complex. Most importantly, a linear model ignores qualitative differences among applicants. For example, in terms of medical exposure, both quantitative *and* qualitative differences between a student who has worked over 1,000 hours as an EMT and one who has shadowed physicians and worked at several health fairs should be considerable. In other words, it is not only the *amount* of exposure but also the *type* of exposure that matters. The same is true of other criteria commonly thought to be relevant to becoming a physician. Future research should focus on how to incorporate these complex factors into meaningful models.

References

1. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Advances in health sciences education : theory and practice.* 2001;6(2):159-175.
2. Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med.* Jan 2007;82(1):100-106.
3. Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Acad Med.* Oct 2005;80(10):910-917.
4. Albanese MA, Snow MH, Skochelak SE, Huggett KN, Farrell PM. Assessing personal qualities in medical school admissions. *Acad Med.* Mar 2003;78(3):313-321.
5. Siu E, Reiter HI. Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in health sciences education : theory and practice.* Dec 2009;14(5):759-775.
6. Callahan CA, Hojat M, Veloski J, Erdmann JB, Gonnella JS. The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: a longitudinal study of 36 classes of Jefferson Medical College. *Acad Med.* Jun 2010;85(6):980-987.
7. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ.* Apr 20 2002;324(7343):952-957.
8. Kreiter CD, Kreiter Y. A validity generalization perspective on the ability of undergraduate GPA and the medical college admission test to predict important outcomes. *Teach Learn Med.* Spring 2007;19(2):95-100.
9. Silver B, Hodgson CS. Evaluating GPAs and MCAT scores as predictors of NBME I and clerkship performances based on students' data from one undergraduate institution. *Acad Med.* May 1997;72(5):394-396.
10. White CB, Dey EL, Fanone JC. Analysis of Factors that predict clinical performance in medical school. *Advances in health sciences education : theory and practice.* Oct 2009;14(4):455-464.
11. Alexander GL, Davis WK, Yan AC, Fantone JC, 3rd. Following medical school graduates into practice: residency directors' assessments after the first year of residency. *Acad Med.* Oct 2000;75(10 Suppl):S15-17.
12. Dirschl DR, Campion ER, Gilliam K. Resident selection and predictors of performance: can we be evidence based? *Clinical orthopaedics and related research.* Aug 2006;449:44-49.
13. Boyse TD, Patterson SK, Cohan RH, et al. Does medical school performance predict radiology resident performance? *Academic radiology.* Apr 2002;9(4):437-445.
14. Harfmann KL, Zirwas MJ. Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *Journal of the American Academy of Dermatology.* Nov 2011;65(5):1010-1022 e1012.
15. Paolo AM, Bonaminio GA. Measuring outcomes of undergraduate medical education: residency directors' ratings of first-year residents. *Acad Med.* Jan 2003;78(1):90-95.

16.   Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med.* Mar 2004;79(3):244-249.

17.   Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *The New England journal of medicine.* Dec 22 2005;353(25):2673-2682.

18.   Teherani A, Hodgson CS, Banach M, Papadakis MA. Domains of unprofessional behavior during medical school associated with future disciplinary action by a state medical board. *Acad Med.* Oct 2005;80(10 Suppl):S17-20.

19.   Doherty EM, Nugent E. Personality factors and medical training: a review of the literature. *Medical education.* Feb 2011;45(2):132-140.

20.   Donnon T, Paolucci EO, Violato C. A predictive validity study of medical judgment vignettes to assess students' noncognitive attributes: A 3-year prospective longitudinal study. *Medical Teacher.* 2009;31:e148-155.

21.   Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Medical education.* Apr 2007;41(4):378-384.

22.   Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical education.* Jan 2007;41(1):57-64.

23.   Meredith KE, Dunlap MR, Baker HH. Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of medical education.* Oct 1982;57(10 Pt 1):743-751.

24.   Vu N, Dawson-Saunders B, Barrows H. Use of a Medical Reasoning Aptitude Test To Help Predict Performance in Medical School. *Journal of medical education.* 1987;62(4):325-335.

25.   DeVaul RA, Jervey F, Chappell JA, Caver P, Short B, O'Keefe S. Medical school performance of initially rejected students. *JAMA : the journal of the American Medical Association.* Jan 2 1987;257(1):47-51.

26.   Elam CL, Johnson MM. Prediction of medical students' academic performances: does the admission interview help? *Acad Med.* Oct 1992;67(10 Suppl):S28-30.

27.   Smith SR. Medical school and residency performances of students admitted with and without an admission interview. *Acad Med.* Aug 1991;66(8):474-476.

28.   Hall FR, Regan-Smith M, Tivnan T. Relationship of medical students' admission interview scores to their dean's letter ratings. *Acad Med.* Dec 1992;67(12):842-845.

29.   Murden R, Galloway GM, Reid JC, Colwill JM. Academic and personal characteristics as predictors of clinical success in medical school. *Annual Conference on Research in Medical Education. Conference on Research in Medical Education.* 1977;16:181-186.

30.   Basco WT, Jr., Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. *Acad Med.* Jul 2000;75(7):743-747.

31.   Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. Correlations of family medicine clerkship evaluations and Objective Structured Clinical Examination scores and residency directors' ratings. *Family medicine.* Feb 1999;31(2):90-94.

32. Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *BMJ.* Feb 22 2003;326(7386):429-432.

33. Urlings-Strop LC, Themmen AP, Stijnen T, Splinter TA. Selected medical students achieve better than lottery-admitted students during clerkships. *Medical education.* Oct 2011;45(10):1032-1040.

34. Benor D, Hobfoll SE. Prediction of clinical performance: the role of prior experience. *Journal of medical education.* Aug 1981;56(8):653-658.

35. Dawson-Saunders B, Doolen DR. An alternative method to predict performance: canonical redundancy analysis. *Journal of medical education.* Apr 1981;56(4):295-300.

36. Dickman RL, Sarnacki RE, Schimpfhauser FT, Katz LA. Medical students from natural science and nonscience undergraduate backgrounds. Similar academic performance and residency selection. *JAMA : the journal of the American Medical Association.* Jun 27 1980;243(24):2506-2509.

37. Hall ML, Stocks MT. Relationship between quantity of undergraduate science preparation and preclinical performance in medical school. *Acad Med.* Mar 1995;70(3):230-235.

38. Sade RM, Lancaster C. The relation of undergraduate major to medical school admission and medical school performance. *J S C Med Assoc.* Jul 1983;79(7):401-404.

39. Smith SR. Effect of undergraduate college major on performance in medical school. *Acad Med.* Sep 1998;73(9):1006-1008.

40. Ramsbottom-Lucier M, Johnson MM, Elam CL. Age and gender differences in students' preadmission qualifications and medical school performances. *Acad Med.* Mar 1995;70(3):236-239.

41. Mawhinney BS. The value of ordinary and advanced level British school-leaving examination results in predicting medical students' academic performance. *Medical education.* Mar 1976;10(2):87-89.

42. Oggins J, Inglehart M, Brown DR, Moore W. Gender differences in the prediction of medical students' clinical performance. *J Am Med Womens Assoc.* Nov-Dec 1988;43(6):171-175.

43. Weinberg E, Rooney JF. The academic performance of women students in medical school. *Journal of medical education.* Mar 1973;48(3):240-247.

44. Haist SA, Wilson JF, Elam CL, Blue AV, Fosson SE. The Effect of Gender and Age on Medical School Performance: An Important Interaction. *Advances in health sciences education : theory and practice.* 2000;5(3):197-205.

45. McManus IC, Richards P, Winder BC, Sproston KA. Final examination performance of medical students from ethnic minorities. *Medical education.* May 1996;30(3):195-200.

46. Campos-Outcalt D, Rutala PJ, Witzke DB, Fulginiti JV. Performances of underrepresented-minority students at the University of Arizona College of Medicine, 1987-1991. *Acad Med.* Jul 1994;69(7):577-582.

47. Duckworth AL, Seligman ME. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological science.* Dec 2005;16(12):939-944.

Table 1: Descriptive Characteristics of Study Population at Matriculation n=92

	Mean (sd) or N(%)	Min	Max
BCPM*	3.62 (0.32)	2.5	4.0
MCAT**	33(3)	24	42
Age	24(3)	20	38
Female	62(57%)	-	-
White	86(81%)	-	-
Total Academic Achievement	4.46(1.34)	0	6
Total Apps	10.5(2.28)	5	16
* BCPM is the cumulative grade point average on a 4-point scale for courses taken in Biology, Chemistry, Physics and Math			
**The MCAT comprises 3 subsection each scored on a 1-15 scale; therefore the possible range for overall scores is 3-45			

Table 2: Descriptive Statistics of Clinical Performance Summary Measures n = 92

Clerkship§	Clinical Skills	Medical Knowledge	Professionalism
Internal Medicine Scale: 1-5	4.33(0.36)*	4.34(0.32)	3.74(0.23)**
Pediatrics Scale: 1-9	7.36(0.98)	7.36(0.95)	8.0(0.85)
Family Medicine Scale 1-8	6.60(0.80)	--	--
* mean(sd)			
** The professionalism scale for internal medicine only ranged from 1-4.			
§ Internal Medicine and Pediatrics had multiple evaluations per student. The total number of observations from which these summary measures are derived is as follows: Internal Medicine: 881, Pediatrics: 204, Family Medicine: 92			

Table 3: Relationship of Demographic Variables and Clinical Performance Ratings

Admissions Predictor	Internal Medicine			Pediatrics			Family Medicine
	Clinical Skills	Medical Knowledge and Learning	Profession-alism	Clinical Skills	Medical Knowledge and Learning	Profession-alism	Clinical Skills
Age, (r_s)§	-0.15	-0.10	0.03	-0.10	-0.01	-0.02	-0.07
Sex							
Men	4.34**	4.23**	3.69**	7.06**	7.18*	7.88**	6.75
Women	4.51	4.47	3.84	7.63	7.53	8.34	6.50
Race							
Caucasian	4.43	4.39	3.8	7.29	7.33	8.00	6.75**
Other Race/Ethnicity	4.40	4.39	3.8	7.46	7.33	8.08	6.25
*p<0.10, **p<0.05, § (r_s) = Spearman's rho							

Table 4: Bivariate Correlations between Admissions Predictors and Clinical Performance Ratings

Admissions Predictor	Internal Medicine			Pediatrics			Family Medicine
	Clinical Skills	Medical Knowledge and Learning	Profession-alism	Clinical Skills	Medical Knowledge and Learning	Profession-alism	Clinical Skills
BCPM	-0.12	-0.13	-0.11	-0.010	-0.04	-0.07	0.07
MCAT	0.06	-0.01	-0.03	0.02	0.05	0.05	0.17
Composite Academic Achievement	-0.08	-0.06	-0.09	0.01	0.02	0.02	0.23**
Composite Non-Academic Rating	0.06	0.05	0.05	0.22**	0.19*	0.17	0.11
*p<0.10							

Table 5: Multivariate Regression between Admissions Predictors and Ratings for the Clinical Skills Dimensions

Admissions Predictor	Internal Medicine				Pediatrics				Family Medicine			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Age	-0.16	--	--	--	-0.09	--	--	--	-0.35**	--	--	--
Race	0.23	0.24	--	--	-0.03	-0.04	--	--	0.20**	0.23**	--	--
Sex	-0.39**	-0.39**	0.34**	--	-0.18	-0.19**	-0.23*	--	-0.08	-0.10	-0.03	--
Composite Academic Rating	-0.10	-0.07	-0.07	-0.03	0.01	0.02	0.02	0.02	0.18*	0.23**	0.22**	0.22**
Composite Non-Academic Rating	0.12	0.11	0.10	0.10	0.18*	0.18*	0.18*	0.18*	0.11	0.10	0.10	0.09
Overall R^2	0.19	0.17	0.13	0.01	0.09	0.09	0.08	0.04	0.25	0.13	0.06	0.06
* $p<0.10$ ** $p<0.05$												

Appendix A: Systematic Review and Comprehensive Results

Methods

Search Strategy

A literature search was performed to identify the available evidence regarding predictors of clinical performance within medical school. The literature generally uses the terms 'cognitive' and 'non-cognitive' to refer to applicants' academic achievement and personal qualities or other traditionally less standardized, testable characteristics. These terms therefore appear in this review. A recent systematic review and a meta-analysis address academic, or 'cognitive' predictors of clinical performance including the MCAT and GPA. [2,3] Therefore, this review sought to determine what is known about 'non-cognitive' predictors of clinical performance. I searched PubMED (MEDLINE) with the following terms: selection OR admission OR non-cognitive AND (clinical performance OR clinical success) AND medical school; school admission criteria AND (clinical clerkship or non-cognitive). I also searched Web of Science using the terms: admissions AND medical school AND clinical performance AND non-cognitive. Searches were limited to English language articles. Given the paucity of literature on this subject, no date limits were set on the searches.

Study Selection

The literature searches returned 141 unique articles. I reviewed the titles and abstracts of these articles to identify research likely to address the study question. I included articles if they represented original research, if the independent variables measured aspects of the application other than standardized test scores and/or grade point averages, (e.g. interview ratings, personal statements, etc) and if the study outcome or

dependent variable was clinical performance. Given the scarcity of research on this subject, I included studies of U.S. allopathic medical students as well as studies conducted in medical schools outside the U.S. (e.g. the United Kingdom, Australia, Canada, and Israel).

The focus of this review and the associated original study was to identify non-academic qualities useful in the selection of medical students – i.e. prior to matriculation. Therefore, studies that measured predictor variables after the start of medical school (e.g. personality tests taken by 2nd or 3rd year students) were excluded. Since I aimed to identify non-academic variables that correlate with undergraduate medical school performance, I excluded studies that measured clinical performance only as an outcome in residency. Articles that used achievement-based metrics (e.g. GPA and MCAT) as independent variables were included as long as they also included a well-defined, non-academic measure among the predictor variable set. I also excluded studies conducted on non-medical school populations (e.g. dental school, nursing school, and other health profession students). Finally, two studies were excluded because their full text was not accessible through UNC databases. [48,49] A recent, well-conducted systematic review examined the relationship between personality traits and clinical performance. [19] My search did not identify any new studies beyond the publication date of this review; therefore, I excluded articles that examined personality and clinical performance.

Data Abstraction and Quality Assessment

Through title and abstract review, 16 articles met the inclusion criteria. The references of these articles were hand-searched and reviewed to yield an additional two relevant articles. I critically appraised a total of 18 articles by identifying the study

design, population studied, and by assessing the results and evaluating the internal and external validity on a scale of poor, fair and good. Some studies measured outcomes other than clinical performance, but only those results pertaining to clinical performance were included. The internal validity, external validity, and the strength and appropriateness of the statistical analysis were the primary criteria used to evaluate the overall quality of each study.

Table 1 summarizes the included studies in terms of their research design, study population, independent variables, outcome measures and results. It also presents the overall quality assessment for each article based on the internal and external validity.

Results

Interviews

Characteristics of Reviewed Studies: Eleven studies published from 1978-2011 evaluated the correlation between interview ratings and clinical performance. Six of these studies used a retrospective cohort design; five followed students prospectively (need citations). Study populations ranged from 26 to 660 students. Seven studies were conducted at U.S. allopathic medical schools[23-25,27-30] three studies were conducted at medical schools in Canada [20-22] and one study was conducted at an unidentified institution.[26] Overall quality ranged from poor to good. A variety of methodological issues, which are discussed below, made it difficult to evaluate the internal validity of several studies.

The independent variable of 'interview ratings' was considerably heterogeneous; although there were broad similarities in terms of rating processes or interview structure,

no two studies used the same interview format or rating system to evaluate applicants. Interviews formats ranged from completely open-ended [25-27] to semi-structured. [20,21] Some studies provided elaborate detail regarding the interview process and scoring methodology, [20,21,23,29] whereas others provided only an opaque description of what comprised an interview rating. [25,27,30] This lack of transparency made it hard to assess the potential for confounding.

Importantly, several studies specified that interviewers were blinded to applicants' academic record and test scores during the rating process; others stated that interview ratings included a composite assessment of personal qualities and academics (e.g MCAT's and uGPA) or did not specify. In such cases it was difficult to account for the influence of academic achievement on the assessment of non-academic qualities. Consequently, these studies bear a greater potential for bias.

The outcome of 'clinical performance' was measured in a number of ways: preceptor ratings or narrative clerkship evaluations, OSCE scores, overall grades in required clerkships, Dean's letter ratings, and residency rankings.

Predictive Validity: The majority of included studies found low to moderate correlations between interview ratings and clinical performance. The significance and magnitude of these associations varied by both interview format and the criterion measures used for clinical performance. I identified four distinct criterion measures for clinical performance: narrative interviewer comments, clerkship grades, dean's letter or residency ranking, and OSCE scores. The study findings are grouped and reviewed according to these categories.

Studies that defined clinical performance in terms of narrative preceptor evaluations or likert-scale clerkship ratings reported small to moderate correlations between these measures and interview ratings (r = 0.2 – 0.6).[20,21,23,24] Within this group of studies, there was an interaction between interview format and clinical performance; in general, structured interviews correlated more strongly with performance than did unstructured interviews. The best evidence to support this conclusion comes from two prospective studies that directly compared unstructured and structured interviews within their respective study populations. [20,21]Both studies found small, non-significant correlations between unstructured interviews and clinical ratings. In contrast, both found moderate, statistically significant correlations between structured interview scores and clinical evaluations (r = 0.39-0.62). [20,21]

Importantly, the associations between interview ratings and clinical performance are likely underestimated given that all of the included studies examined only accepted applicants. While the small number of studies reviewed doesn't permit definitive conclusions, these data suggest that structured interviews hold greater promise for predicting future clinical performance than do unstructured interviews. Further studies comparing unstructured vs. structured interviews are needed to confirm this trend. Multi-school, prospective studies with larger numbers would also strengthen the validity and applicability of these findings. None of the reviewed studies used the interview in the actual selection of applicants (as one of the purposes of each study was to validate an innovative interview method). Therefore, prospective studies employing structured interviews to make admission decisions could more accurately estimate the ability of interviews to predict clinical performance.

In contrast, studies that defined clinical performance in terms of clerkship grades reported minimal to no correlation between these measures and interviews ratings. [25-27] Among these studies, 'clerkship grades' refers to a composite grade based on subjective clinical preceptor ratings and end-of-course examination scores. The extent to which this combined outcome reflects medical knowledge and/or clinical skills – communication, teamwork, and clinical reasoning, etc – is unclear. Given the strong evidence demonstrating a lack of correlation between interview ratings and shelf-exam scores [21,23], this combined outcome complicates the interpretation of the study findings.

All studies in this group[25-27] used unstructured interviews. One fair quality study compared the clinical performance of students who were accepted after interview to those students who were initially rejected following the interview (but were later accepted under unusual circumstances); the authors found no significant differences in clinical performance between the groups. [25] Another fair quality study compared the performance of students who were accepted using an interview as part of the admissions process to the performance of students admitted without the use of an interview. [27] The authors reported no differences in clinical performance between the two groups and proposed that medical schools should consider eliminating the interview as part of the admissions process. Notably, this institution still uses the interview in its selection efforts. [50] Both of these studies were limited by a failure to eliminate selection bias and a small sample size.

Altogether, this group of studies suggests that unstructured interviews have little practical value in predicting clinical performance. Higher quality evidence is needed to confirm this conclusion. Prospective studies using validated, reliable interview methods and more accurate measures of clinical performance would establish more definitive data.

Three studies used proxy measures for clinical performance. Two of these studies correlated interview ratings with used dean's letter. [28,29] To the extent that dean's letters correlate with specific clinical performance measures, they can be taken as valid and reliable surrogates for direct clinical performance ratings. Murden et al reported a correlation of 0.77 between the dean's letter rating and the clinical performance evaluations; this study found a small correlation ($r = 0.22$, $p< 0.01$) between interviewer ratings and dean's letter rankings. [29] The other reviewed study did not report a correlation between clinical performance ratings and dean's letter ratings. [28] The authors of this study reported a moderate correlation of ($r = 0.37$, $p=0.014$). These modest correlations are not surprising given that the outcome was not a pure measure of clinical performance but rather a composite metric that included contributions from preclinical grades. Both of these studies were limited by small sample sizes. Additionally, these correlations were found only for the extremes of the distribution and not for the entire study cohort.

In a fair quality study, Peskun et al reported that interview scores correlated significantly with family medicine rankings for graduating students. [22] The rationale for using residency rankings as proxy measures of clinical performance is that these rankings are heavily weighted towards perceived clinical skills. However, the authors reported the slope of the linear regression model; thus, it is unclear whether this statistically significant finding translates to a practically meaningful difference in clinical performance or ability.

As a group, these studies suggest that interviews can capture some enduring traits or abilities related to long-term success in medical school. The findings are consistent

with studies that used narrative preceptor comments as the clinical performance criterion. However, given their quality and design, these studies do not provide convincing evidence that interviews can predict clinical performance.

Several studies used OSCE scores as measures of clinical performance. [21,22,30] Two fair quality studies reported no significant correlations between interview ratings and clinical performance [22,30] Neither of these studies used structured interview formats, and one of the studies was underpowered (60% power to detect a 15% difference). [22,30] Interestingly, a good quality study by Reiter et al found no correlation between unstructured interview ratings and OSCE scores but a significant correlation between structured interview ratings and OSCE scores (standardized beta = 0.4, $p < 0.05$). [21] The authors did not discuss the extent to which the beta coefficient translates to a practical difference in overall performance or final grades therefore. Still, this finding echoes the trend described earlier among studies that used narrative preceptor comments as the clinical performance metric – the more structured the interview, the stronger its predictive validity for clinical performance.

The inconsistent correlations and variable quality of evidence do not permit firm conclusions about the overall predictive power of interviews for clinical performance. Nonetheless, several inferences stand to reason. The predictive validity of interviews varies significantly according to both the interview format and the criterion measure of clinical performance. If interviews are used to anticipate clinical performance, structured, rather than unstructured interviews, are preferable.

Interestingly, all included studies that found weak correlations between interview scores and clinical performance concluded that the personal interview was likely of little

use in selecting applicants. However, this conclusion presupposes the validity and reliability our clinical assessment tools – in reality the lack of correlation among these measures is likely bidirectional. In other words, the clinical assessment tools used in these studies may not be sensitive and specific measures of performance. Therefore, this group of studies points to the need to develop both more sophisticated interview tools as well as better methods of clinical assessment.

Research demonstrates that both OSCE scores and preceptor ratings positively correlate with residency director ratings. [31] It is therefore interesting that interviews correlate differentially with preceptor ratings and OSCE scores. OSCE's are more standardized and lend themselves to discrete (often correct vs. incorrect) responses whereas personal interviews tend to be more open-ended and permit and array of equally meritorious answers. In addition, OSCE's test a limited range of clinical skills whereas preceptors often asses a wider range of abilities. Therefore, one hypothesis to explain this observation is that the method of assessment (standardized vs. open-ended) influences the relationship between predictor and outcome. The finding that that structured interviews correlated with OSCE scores whereas unstructured interviews supports this hypothesis.

Components of Medical School Application: Biographical Sketch, Letters of References, and the Personal Statement

<u>Characteristics of Reviewed Studies:</u> A handful of studies examined the relationship between clinical performance and various components of the medical school application including the biographical sketch, letters of reference, and the personal statement. [21,22,32,33] Most of these studies used a 'non-cognitive assessment' as the predictor variable and clerkship grades as the measure of clinical performance. Several

studies defined 'non-cognitive assessment' as an overall rating derived from the biographical sketch – the portion of the application that details an applicant's extracurricular activities including work experience, medical exposure, research interests, volunteer activities. [21,33] The 'non-cognitive assessment' in other studies[22,32] included letters of reference and the personal statement in addition to the autobiographical statement.

Overall, these studies found negligible to small correlations between components of the medical school application and clinical performance. Three studies used a retrospective cohort design; one study followed a population of students prospectively. [33] Sample sizes ranged from 45-939. All of the studies were performed at medical schools outside of the U.S.. Given the small number of studies in this group, they will be discussed individually.

Predictive Validity: A poor quality study by Urlings-Stop et al compared the clinical performance of students admitted with and without the use of a non-cognitive assessment as a selection tool. Students selected with use of the non-cognitive assessment received higher mean clerkship grades than lottery admitted students (7.95 ± 0.03, 95% [CI] 7.90–8.00, versus 7.84 ± 0.02, 95% CI 7.81–7.87). [33] They also received grades of distinction 1.5 more often than did students admitted through the lottery system. This difference is likely underestimated due to ceiling effects – the clinical performance ratings were very skewed towards higher ratings. Still, it is difficult to interpret how meaningful the difference is based on the study description of the grading scale used to measure clinical performance. Thus, this statistical difference may not reflect a meaningful distinction in clinical skills or personal qualities.

This study was limited by significant, differential losses to follow-up (9.3% in the selected group versus 17.3% in the lottery group), as well as a failure to minimize the potential for selection bias.

A fair quality study by Peskun et al reported a significant association between applicants' scores on a non-cognitive assessment and their internal medicine residency rankings (OR = 1.28, 95%CI (1.03 – 1.58)). [22] The authors also reported an association between the NCA ratings and OSCE scores (slope = 0.04, p<0.04). The non-cognitive assessment ratings were derived from applicants' biographical sketch, letters of reference, and a personal statement. Again, while these findings are statistically significant, it is unclear whether they amount to a practical difference in clinical skills.

A good quality study by Reiter et al. reported no relationship between ratings derived from the autobiographical sketch and clinical performance as measured by preceptor clinical ratings. [21] However, the inter-rater reliability of the application rating process was low (kappa = 0.45). Thus, the findings of this study may in part reflect the challenges of defining and evaluating 'non-cognitive' attributes rather than a true lack of association.

In a fair quality study, Ferguson et al performed a personality-based content analysis of applicants' personal statements and letters of reference. The authors reported a significant association between this data and clinical performance as measured by clerkship grades (r=0.23, p<0.05). [32] In contrast, the authors found no relationship between information in the letters of reference and clinical performance.

The paucity and variable quality of the evidence reviewed only allows tentative conclusions regarding the association of the 'non-cognitive' aspects of medical school

applications and clinical performance. Letters of references seem to have little predictive

value for clinical performance. No research has examined the personal statement as an

independent predictor of clinical performance rather than part of a composite variable.

These portions of the medical school application are required of applicants because they

presumably provide additional, useful information upon which admissions committees

can base selection decisions. However, evidence is weak or lacking that these factors

reliably and accurately forecast clinical performance. Key challenges to discerning the

true nature of the relationship between these components and performance are as follows:

Unlike the commonly accepted and standardized tools used to assess academic

achievement (e.g. multiple choice tests, GPA, graduation with honors etc), there are no

such metrics to assess the other portions of the medical school application. Each study

devised it's own evaluation method for the portions of the application aside from

academic achievement. Moreover, these studies were conducted at institutions with

dissimilar admissions processes and in a number of different countries. The pre-medical

educational environment across these settings could vary substantially. Thus, the

applicability of the findings is severely limited. Prospective, multi-institution studies

using validated and reliable assessment tools would strengthen the quality and certainty

of evidence that non-academic portions of the application significantly predict clinical

performance.

Educational Background and Life Experience:

Characteristics of Reviewed Studies: Several studies examined the relationship

between educational background and clinical performance. These studies were published

between 1980 and 1981; all used a retrospective design. Sample sizes ranged from 62 to 143. Two studies were conducted in the U.S [35,36] and one was conducted in Israel. [34] All three studies used undergraduate major or direct progression to medical school from collegiate institution (i.e. no interval time between college and medical school) as an independent variable. The studies used preceptor clinical ratings in required clerkships [34,35] or clerkship grades[36] as performance measures.

Predictive Validity: A poor quality study by Benor et al found that physics or biology majors received higher average clinical ratings than did non-science majors (87.0 vs. 85.0 vs. 79.0, p<0.05, respectively). [34] However, the authors do not address what these scores amount to in terms of pragmatic, clinical skills. In other words, do the differences in numeric grade differentiate superior clinical students from average students in a clinical context? This study was also limited by a failure to eliminate selection bias a small sample size (n=62).

A fair-quality study by Dawson-Saunders et al found no relationship between educational background and clinical performance. [35] The authors compared clinical ratings of two groups: students who either majored in a non-natural science discipline or did not matriculate directly to medical school, and students who majored in a natural science or went directly to medical school after completing college. They found no difference in the subjective clinical ratings received by the two groups.

Finally, a poor-quality study by Dickman et al also found no difference in clinical performance among natural science and non-science majors. [36] The authors did not describe the baseline characteristics of their populations and failed to minimize potential sources of selection bias.

Overall, the evidence suggests that one's educational background holds limited predictive value for future clinical performance. The mediocre quality and small quantity of research leaves considerable uncertainty regarding this conclusion. Studies with larger sample sizes and greater methodological rigor would provide more definitive answers regarding the relationship of educational background and clinical performance.

Common Limitations and Future Research

Ceiling effects are the most common limitation in the medical school admissions literature. Using populations of only admitted students restricts the range of predictor variables. By the same token, clinical performance measurements are frequently skewed, limiting the range of the criterion measure. Ceiling effects not only weaken the correlation among predictors and outcomes but also make it difficult to extrapolate the results of a particular study to a population of potential medical school students. Since the end-goal of medical admissions research is to improve the selection process, studies that actively employ new selection methods during the selection process are greatly needed. Prospective studies with long-term follow-up tracking the performance of rejected applicants who subsequently enroll at other medical schools, while logistically challenging, could reduce ceiling effects and thus more accurately evaluate the predictive validity of various admission tools. As the move towards competency-based education drives the development of and adherence to common methods of clinical assessment, it will theoretically become easier to compare the clinical performance of students across curricula.

Conclusion

This review manifests the challenges of identifying correlates of future clinical performance among medical school applicants. The literature is limited by the paucity of studies and an array of methodological issues including retrospective designs, small sample sizes, ceiling effects, and the failure to control for potential confounders. However, many of these deficiencies could be addressed by more or higher quality research. On the other hand, the heterogeneity of both the admissions factors and the outcomes assessed elucidates a more complex issue – how to define and measure qualities essential to the competent and ethical practice of medicine. While there is widespread agreement that we should measure more than discrete knowledge in applicants, medical students, and physicians, there is not standardized, scalable MCAT equivalent to measure personal qualities such as empathy or clinical skills such as teamwork and ethical decision making. The literature underscores the need for more refined admission tools as well as more sensitive and specific measures of clinical performance.

References

1. Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med.* Jan 2007;82(1):100-106.
2. Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Acad Med.* Oct 2005;80(10):910-917.
3. Hobfoll SE, Benor DE. Prediction of student clinical performance. *Medical education.* Jul 1981;15(4):231-236.
4. Milstein RM, Wilkinson L, Burrow GN, Kessen W. Admission decisions and performance during medical school. *Journal of medical education.* Feb 1981;56(2):77-82.
5. Doherty EM, Nugent E. Personality factors and medical training: a review of the literature. *Medical education.* Feb 2011;45(2):132-140.
6. Basco WT, Jr., Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. *Acad Med.* Jul 2000;75(7):743-747.
7. DeVaul RA, Jervey F, Chappell JA, Caver P, Short B, O'Keefe S. Medical school performance of initially rejected students. *JAMA : the journal of the American Medical Association.* Jan 2 1987;257(1):47-51.
8. Hall FR, Regan-Smith M, Tivnan T. Relationship of medical students' admission interview scores to their dean's letter ratings. *Acad Med.* Dec 1992;67(12):842-845.
9. Meredith KE, Dunlap MR, Baker HH. Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of medical education.* Oct 1982;57(10 Pt 1):743-751.
10. Murden R, Galloway GM, Reid JC, Colwill JM. Academic and personal characteristics as predictors of clinical success in medical school. *Annual Conference on Research in Medical Education. Conference on Research in Medical Education.* 1977;16:181-186.
11. Smith SR. Medical school and residency performances of students admitted with and without an admission interview. *Acad Med.* Aug 1991;66(8):474-476.
12. Vu N, Dawson-Saunders B, Barrows H. Use of a Medical Reasoning Aptitude Test To Help Predict Performance in Medical School. *Journal of medical education.* 1987;62(4):325-335.
13. Donnon T, Paolucci EO, Violato C. A predictive validity study of medical judgment vignettes to assess students' noncognitive attributes: A 3-year prospective longitudinal study. *Medical Teacher.* 2009;31:e148-155.
14. Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical education.* Jan 2007;41(1):57-64.

15. Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Medical education.* Apr 2007;41(4):378-384.

16. Elam CL, Johnson MM. Prediction of medical students' academic performances: does the admission interview help? *Acad Med.* Oct 1992;67(10 Suppl):S28-30.

17. http://brown.edu/academics/medical/admission/interview-process. Accessed 6/25/2012, 2012.

18. Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. Correlations of family medicine clerkship evaluations and Objective Structured Clinical Examination scores and residency directors' ratings. *Family medicine.* Feb 1999;31(2):90-94.

19. Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *BMJ.* Feb 22 2003;326(7386):429-432.

20. Urlings-Strop LC, Themmen AP, Stijnen T, Splinter TA. Selected medical students achieve better than lottery-admitted students during clerkships. *Medical education.* Oct 2011;45(10):1032-1040.

21. Dawson-Saunders B, Doolen DR. An alternative method to predict performance: canonical redundancy analysis. *Journal of medical education.* Apr 1981;56(4):295-300.

22. Dickman RL, Sarnacki RE, Schimpfhauser FT, Katz LA. Medical students from natural science and nonscience undergraduate backgrounds. Similar academic performance and residency selection. *JAMA : the journal of the American Medical Association.* Jun 27 1980;243(24):2506-2509.

23. Benor D, Hobfoll SE. Prediction of clinical performance: the role of prior experience. *Journal of medical education.* Aug 1981;56(8):653-658.

Table 1. Description and Critical Appraisal of Included Studies

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Basco, Gilbert, et al. (2000)	Retrosp-ective Cohort	Medical University of South Carolina Students matriculating in 1993-1995 (N=222)*	1)"Academic Profile", a score based on MCAT and GPA 2)"Selection Profile", the Average of 3 interview scores	Faculty evaluation of OSCE performance, Patient Satisfaction scores completed by OSCE standardized patients	Neither academic profile nor selection profile scores correlated significantly with OSCE faculty ratings or OSCE standardized patient satisfaction scores.**	Poor/Fair	Poor	Fair
Benor and Hobfoll (1981)	Retrosp-ective Cohort	Beersheva, Israel Community Training Program (N=62)	1)Age 2)Prior Miliary experience (officers vs. enlisted) 3)Premedical Concentration/ Major (physics/math vs. biology vs. humanities)	Subjective Clinical Ratings	1) Middle-age students received higher mean clinical ratings than very young or very old students. 2) Prior Military experience did not consistently correlate with clinical ratings 3) Science majors received higher mean clinical ratings compared with non-science majors.	Good	Poor	Fair

*This is the total sample size. However, each class was analyzed as a separate cohort. The respective class sizes for 1993, 1994, 1995 were 70, 81, 71, respectively.
.

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Dawson-Saunders and Doolen (1981)	Retrospective Cohort	Southern Illinois Univ School of Medicine Graduating Students in 1979-1980 (N=143)	Undergraduate Major	Grades in Required Clerkships	No significant correlation between Undergraduate Major and clinical grades	Fair	Fair	Fair
DeVaul, Jervey, et al. (1987)	Prospective Cohort	Univ Texas Medical School 127 Admitted Students and 49 Initially Rejected Students (N=176)	Interviewer and Committee Rating	Clerkship Grade (composite of NBME Shelf Score and Subjective Evaluations by residents and faculty)	No difference in clinical performance among those admitted and those initially rejected. Analysis was duplicated for top and bottom quartiles Important thing here is that all were prescreened according to academic achievement, extracurricular activities, pre-health advisor ratings, work and/or volunteer experiences that may relate to medicine	Fair/Good	Fair	Fair

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Dickman, Sarnacki, et al. (1980)	Retrosp-ective Cohort	SUNY Buffalo School of Medicine Members of Graduating Classes of 1977-1979 (N=96)	Undergraduate Major (natural science vs. non-science)	Clinical Grades (Dichotomized as either satisfactory or honors)	No difference in terms of clinical grades by undergraduate major (in this study clinical grades included subjective measures and shelf scores)	Fair	Poor/Fair	Poor
Donnon, Oddone-Paolucci, et al. (2009)	Prosp-ective Cohort	University of Calgary, Canada Class of 2007 (N=26)	Scores on a series of Medical Judgment Vignettes* Scores on a 6-member panel interview (rated 1-5)	Average rating on In-training evaluation reports (ITER) across 7 mandatory clerkships**	Average score from panel interview moderately correlated with the ITER category "fund of knowledge and understanding of disease mechanisms". (recall that interviewers had access to applicants grades) but did not correlate with 8 other ITER items Total score for the series of vignettes correlated significantly with clinical skills, problem solving, motivation, and professionalism as measured on the ITERs (range: r = 0.46 p<0.05 to 0.62 p<0.01)	Fair/Good	Poor/Fair	Fair/Good

* The medical judgement vignettes were designed to measure students' open-ended responses to medical scenarios that were classified into three broad categories: (1) major ethical dilemmas in medicine (end-of-life – euthanasia), (2) relationships with patients and their families (altruistic commitment and compassionate treatment), and (3) collaboration and clarification with staff and colleagues (dutifulness and understanding of medical relationships).

** For each required clerkship, an eight-item ITER was completed by the attending physician. The items were scored on a five-point scale from 'Unsatisfactory' to 'Outstanding' and were a reflection of the physicians' perceptions of clerks' knowledge of basic and clinical disease mechanisms (cognitive reasoning ability) to more practical skills related to history taking and physical examinations, communication skills with patients and families, and a sense of professionalism and responsibility.

48

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Dowell, Lumsden et al. (2011)	Retrospective Cohort	Glasgow and Dundee Medical Schools Scotland (N=335)	Personal Qualities Assessment (PQA) Scores*	OSCE scores	No relationship between OSCE scores and PQA. Extremes (PQA scores +/- 1.5 SD from mean scored 7.5% lower on average than non-extreme students (didn't define outcome very well).	Poor	Poor	Poor
Elam and Johnson (1992)	Retrospective Cohort	Unidentified Medical School Students who matriculated from 1984-1991 (N=649)	Admission Interview Global Rating	Third Year GPA	Admissions Interview Ratings were strongest predictor of third year GPAs ($R^2=0.05$, $p<0.0001$) and Fourth Year GPA ($R^2=0.04$, $p<0.01$). Compared with MCATs and undergraduate GPA	Fair/Good	Fair	Fair

*The Personal Qualities Assessment (PQA) is a battery of psychometric tests. It includes instruments designed to test a range of personal traits such as empathy, confidence, narcissism and aloofness. Other PQA instruments are designed to assess moral orientation, and other personal qualities such as honesty and integrity.

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Ferguson, James et al. (2003)	Prospective Cohort	Nottingham University Class that matriculated in 1995 (N=118)	Premedical Grades* Personal Statement & Letters of References Scores on Five Factor Personality Questionnaire**	Preclinical Grades Average Grade on 10 core clerkships	Preclinical Grades were strongest predictor of clinical performance (r=0.32, p<0.01) References did not correlate with clinical performance Personal Statement Information correlated with clinical performance (r=0.23, p<0.050) Conscientiousness was related to better preclinical performance (r=0.58, p<0.001), but not as strongly correlated with poorer clinical performance (r=0.26 p<0.05)	Fair/Good	Poor	Good
Hall, Regan-Smith et al. (1992)	Retrospective Cohort	Dartmouth College of Medicine (N=62)	Interview Scores (a composite of academic and non-academic criteria)	Dean's Letter Ratings	Significant Correlation Between Interview Score and Dean's letter Rating No significant relationship between GPA and interview ratings Significant Relationship between MCAT and Dean's Letter Ratings	Fair	Poor	Poor

*Due to differences in the UK grading system compared with the US, these premedical grades are similar, but not equivalent, to a premedical GPA.

**The big 5 framework of personality traits refers to a psychological construct developed in the 1990's as a way to better study the relationship between personality and academic behaviors. For this study, the 5 personality dimensions assessed were: 1) intellect, 2)emotional stability 3) conscientiousness, 4)Surgency and 5)Agreeableness

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Meredith, Dunlap, et al. (1982)	Retrosp-ective Cohort	Arizona College of Medicine, Class of 1981 (N=85)	Average interview ratings on 5 subscales*	Content Analysis of Narrative Evaluations from 3rd Year Clerkships (Pediatrics &Internal Medicine only)	Admissions Interview evaluation explains twice the proportion of variance in clerkship narrative evaluations as do MCAT scores (0.104 vs 0.050)	Good	Fair	Good
Murden, Galloway et al. (1978)	Retrospe ctive Cohort	University of Missouri-Columbia School of Medicine Students who Matriculated from 1966-1970 (N=458)	Personal Characteristics evaluated through Interviews** Undergraduate GPA MCAT	Dean's Letter Ratings§	All Categories of Personal Characteristics correlated significantly with internship letter ratings (r = 0.18 – 0.22, p<0.0003)	Good	Fair	Good/Fair

*The subscales used were: (a) maturity, (b) individual achievement, (c) motivation/interest in medicine, (d) judged ability, and (e) interpersonal skills. Each variable was rated using a five-point scale, with 1 being unacceptable and 5 being outstanding. For every variable, each scale point was defined and accompanied by representative comments.

**Interviewers rated applicants on the following variables: (a) maturity, (b) nonacademic achievement, (c) motivation/interest in medicine, and (d) rapport. Each variable was rated using a five-point scale, with -2 being unacceptable and +2 being outstanding. Committee members did not have discrete definitions of these personal characteristics; they relied on informal operational definitions.

§The internship letter rankings correlated between 0.68-0.77 with clinical clerkship evaluations

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Peskun, Detsky, et al. (2007)	Prospect-ive Cohort	University of Toronto Students who entered in 1994-98, and applied to Internal Medicine and Family Medicne programs in their graduating year (1999-2003). (N=660)	1)Non-Cognitive Assessment (NCA): Overall score (1-6), based on extracurricular activities, personal essay, and letters of reference 2)Admissions Interview Score	1) Residency Ranking in Internal Medicine or Family Medicine 2) OSCE Scores*	1) NCA predicts Internal Medicine Ranking (OR = 1.28, 95% CI[1.03 - 1.58]) 2) Interview score predicts Family Medicine Ranking, (Slope = 3.80, p<0.019) 3) NCA was a significant predictor of OSCE scores (slope = 0.04, p< 0.04) and interviewers ratings were significantly correlated with OSCE scores (Slope = 0.02, p< 0.02).	Good	Fair	Fair
Reiter, Eva, et al. (2007)	Prospect-ive Cohort	McMaster University Ontario, Canada Applicants who underwent MMI in 2002 and were accepted (N=45)	1) Ratings on Multiple Mini Interview (MMI) 2) Ratings from Personal Interview 3) Non-Cognitive Assessment of Autobiographical Sketch**	1) OSCE Scores 2) Preceptor Clerkship Performance Ratings	1) MMI was statistically predictive of clerkship performance, measured with both the average ratings assigned by clerkship directors (standardized beta 0.7, P < 0.001) and encounter card ratings provided by clinical preceptors (standardized beta 0.5, P < 0.01) 2) MMI was statistically predictive of OSCE performance (standardized beta 0.4, P < 0.05);	Good	Fair	Fair

*A 10-station OSCE was administered at the end of the 2nd academic year and evaluated students across multiple cognitive and non-cognitive domains. For the purpose of this study an OSCE score variable was defined as the average of all non-cognitive domain scores. The OSCE score variable ranged between 1 and 5 and was correlated with both admissions variables and residency ranking.

** The ABS comprised a series of 15 short answer format questions, to be completed remotely by all applicants in a non-invigilated setting.

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Shen, Comrey el al. (1997)	Retrosp-ective Cohort	UCLA Students who matriculated in 1985 (N=97)	Comrey Personality Scale Score*	1) Clerkship Grades (shelf tests & ward evaluations) 2) Weighted GPA (average of preclinical & clinical grades) 3) Overall evaluation Score (dean's letter evaluation)	1) Overall CPS score failed to predict any performance measure 2) Two CPS subscales were strongest predictor of clinical ward evaluations. 3) Two CPS subscales were the only significant predictors of overall evaluation score. (R=0.44, p<0.05) 4) Two subscales demonstrated quadratic relationship with overall evaluation	Good	Fair	Good/Fair
Smith (1991)	Retrosp-ective Cohort	Brown University School of Medicine Students admitted 1980-1982 with interviews and those admitted 1982-1985 without interviews (N=67 and 113, respectively)	Interview vs. No Interview	1) Honors Grades and Deficient Grades (those requiring some type of remediation) in Required Clerkships 2) Honors Grades in Elective Clerkships	1) No difference between those admitted with interview compared with those admitted without an interview in terms of Honors Grades. 2) The group that did not have a personal interview received fewer deficient grades in required Clerkships. 3) No difference between the two cohorts in the number of Honors Grades in Elective Clerkships	Good	Fair	Fair

*The Comrey Personality Scale measures eight personality factors along the following bipolar dimensions: 1) Trust vs. Defensiveness 2) Orderliness vs. Lack of Compulsion 3) Social Conformity vs. Rebelliousness 4)Activity vs. Lack of Energy 5)Emotional Stability vs. Neuroticism 6)Extroversion vs. Introversion 7)Mental Toughness vs. Emotional Sensitivity 8)Empathy vs. Egocentrism

Citation (Year)	Study Design	Study Population (Sample Size)	Independent Variables	Outcomes Measured	Results	Internal Validity	External Validity	Overall Quality
Urlings-Strop, Themmen et al. (2011)	Prospective Cohort	Erasmus MC Medical School Rotterdam, the Netherlands 1) Those admitted through a selection process that evaluated non-academic qualities as well as academics* (N=389) 2) Students who were admitted through a weighted lottery based on pre-medical GPA (N=938)	Method of Admission: Selected Group or Lottery Group	Mean grade on first 5 required clerkships**	Those students selected using a non-cognitive assessment in addition to cognitive measures achieved higher mean grades than those selected through lottery using only pre-medical GPA. Mean Grades: 7.95 ± 0.03, 95% [CI] 7.90–8.00 versus 7.84 ± 0.02, 95% CI 7.81–7.87;	Poor/Fair	Poor	Poor/Fair
Vu, Dawson-Saunders, et al. (1987)	Prospective Cohort	Southern Illinois Univ School of Medcine Graduating Classes of 1984-1985 (N=68)	Scores on a series of Clinical Reasoning Vignettes	Content analysis of Narrative Clerkship Comments from required 3rd year Clerkships	MRAT increased predictive validity (beyond MCAT and GPA) for 3rd year clerkship grades by 10% (p<0.10)	Good	Fair	Fair

*Applicants were evaluated on the following non-cognitive criteria: 1) activities in health care; 2) activities in management and organization; 3) activities related to a talent (such as music, sport or science); 4) (extracurricula academic education, and 5) additional pre-university education.
**Each student's clerkship performance was evaluated on a scale of 5-10 using a combination of patient-related assessments, an oral examination, and an overall clinical performance assessment by a supervising preceptor.