DOCUMENT RESUME

ED 229 411 TM 830 309

AUTHOR Wolf, Fredric M.; And Others
TITLE Comparison of the Incremental Validity of the Old and New MCAT.
PUB DATE Apr 83
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Admission Criteria; Biology; Chemistry; *College Entrance Examinations; Comparative Analysis; *Competitive Selection; Correlation; Medical Education; *Medical Students; Predictive Measurement; *Predictive Validity; Regression (Statistics); *Test Validity
IDENTIFIERS *Cross Validation; *Medical College Admission Test; National Board of Medical Examiners; Test Revision

ABSTRACT

The predictive and incremental validity of both the Old and New Medical College Admission Test (MCAT) was examined and compared with a sample of over 300 medical students. Results of zero order and incremental validity coefficients, as well as prediction models resulting from all possible subsets regression analyses using Mallow's Cp criterion, were subjected to cross-validation analyses by randomly dividing two medical school classes into screening and calibration samples. Results supported the incremental validity of both the Old and New MCAT. Coefficients were generally larger for the New than for the Old MCAT. Prediction models of National Board of Medical Examiners Part I and Part II performance, comprised of the New Biology and Chemistry subtests and the Old Science and General Information subtests were cross-validated. Prediction models of clinical evaluation clerkship performance were equivocal. (Author)

Comparison of the Incremental Validity of the Old and New MCAT

Fredric M. Wolf, Judith G. Calhoun,
Bruce R. Maxim, Wayne K. Davis
University of Michigan
Medical School

Alan L. Hull
Case Western Reserve University
School of Medicine

Running Head: Incremental Validity

Presented at the meeting of the American Educational Research Association, Montreal, April 1983. Correspondence may be addressed to the authors at the Department of Postgraduate Medicine and Health Professions Education, University of Michigan, G1208 Towsley Center, Ann Arbor, MI 48109.

Comparison of the Incremental Validity

of the Old and New MCAT

Abstract

The predictive and incremental validity of both the Old and New Medical College Admission Test (MCAT) was examined and compared with a sample of over 300 medical students. Results of zero order and incremental validity coefficients, as well as prediction models resulting from all possible subsets regression analyses using Mallow's C_p criterion, were subjected to cross-validation analyses by randomly dividing two medical school classes into screening and calibration samples. Results supported the incremental validity of both the Old and New MCAT. Coefficients were generally larger for the New than for the Old MCAT. Prediction models of NBME Part I and II performance, comprised of the New Biology and Chemistry subtests and the Old Science and General Information subtests were cross-validated. Prediction models of clinical evaluation clerkship performance were equivocal.

A number of studies recently have compared the ability of the Old and the New MCAT to predict student achievement in medical school. Since the New MCAT was first used in 1978, most studies have focused on the cognitive outcomes of student performance in medical school during the basic science curriculum. As students admitted on the basis of their New MCAT Scores, along with other admission criteria, progress through the medical curriculum, the contribution of the new MCAT in predicting other outcome measures such as clinical performance is beginning to be studied (e.g. Carline, Cullen & Scott, 1982; Carline et al, 1983; Hull, Calhoun & Maxim, 1981). Findings presented in the Association of American Medical Colleges' New Medical College Admission Test Interpretive Manual (1977), as well as the results from local analyses, indicate that the test has strong predictive validity for cognitive outcomes during the first and second years in medical school. McGuire (1980) investigated the relationship of the New MCAT to the criterion of class standing at the end of the freshman year. Results of a correlation analysis revealed that all of the New MCAT subscales except for Skills Analysis: Reading correlated significantly with class standing ($p < .001$). Similar results were also found for undergraduate GPA and undergraduate science GPA. Among the group of predictors used to create a revised admissions prediction index, maximum predictability was achieved by undergraduate GPA and the New MCAT Science Problems score, which is calculated from problem-solving items involved in the biology, chemistry and physics components. As Jones and Thomae-Forgues (1981) point out, the predictive ability of the New MCAT in relation to performance in the basic medical sciences is not surprising in view of the heavy emphasis placed in development on the science and medicine-related content relevance of test items. Performance in medical school has long been considered to be related to science achievement; therefore, those who succeed in the first science-based academic quarter have little trouble completing their medical studies (Cullen et al, 1980). Furthermore, Dawson-Saunders and Doolen (1981) and Jones and Thomae-Forgues (1981) discussed the New MCAT's potential value as a predictor of

clinical performance. Due to the increased emphasis on interpretation and problem solving in the new format, they suggested that the new MCAT may result in measures which are more closely associated with the information gathering, evaluation, and utilization skills required during the clinical experience.

Hull et al (1981) found validity coefficients of the New MCAT subtests with NBME Part I scores (basic science) to be consistently larger than the validity coefficients of undergraduate grades with NBME I. Carline et al (1983) also found the New MCAT subtests to be superior to grades in validity coefficients for NBME Part II scores (clinical science). In reviewing the success of the New MCAT in predicting medical school performance, Jones and Thomas-Forgues (1982) noted several patterns in the America Medical College Application Service data set. Among these patterns were the MCAT's ability to predict NBME Part I performance better than undergraduate college grades and that "predictions of medical school course grade performance based on MCAT scores and undergraduate college grades are better than those based on either one alone" (p.6).

Friedman and his collegues (1980, 1981) have used the method of incremental validity (Sechrest, 1963) to illustrate the utility of both New and Old MCAT scores in improving the amount of variance accounted for in measures of medical student performance beyond that accounted for by other preadmission measures. They found that the explained variance in examination performance during the first two years was proportionately greater when the New MCAT was used in place of the Old MCAT in a stepwise regression analysis after all other admissions variables were included. Their analyses also revealed that the New MCAT's incremental predictive power was higher for nationally standardized outcome measures as compared to that for locally prepared achievement tests. When NBME Part I examination scores on the Microbiology and Anatomy subtests were used as criterion measures, the most valuable New MCAT predictors were the Biology, Science

Problems, and Skills Analysis: Quantitative assessment subtests. Erdmann (1980) characterized the results of the "first round" of MCAT studies as encouraging, and called for the next phase to focus on "the range of relationships between test scores and various criterion measures ..., the consistency of these findings over time, and the pattern of relationships with performance measures obtained at successive stages of the medical education process" (p. 464).

Among the preadmission measures included as predictor (independent) variables in the two incremental validity studies previously reported were undergraduate grade-point average, selectivity of undergraduate school, marital status, age, "quantitativeness" of undergraduate major, parental education and income, and hometown community size, as well as MCAT subtest scores (Friedman & Bakewell, 1980; Friedman & Porter, 1981). Dependent (criterion) variables included a composite first year medical school examination score and two NBME Part I subtest scores (Microbiology and Anatomy). While many of the non-MCAT preadmission measures were significant predictors of, and explained additional significant variance in, these criterion measures in multiple regression analyses, some of these measures (e.g., marital status, parental education and income, hometown community size) are of questionable utility for making admission decisions, given their sensitivity and potential legal implications. However, the results reported by Friedman and colleagues are encouraging because the inclusion of MCAT subtest scores explained additional significant variance in medical student performance even when non-traditional admission measures were included as predictors. It is likely, therefore, that their results were conservative because the additional non-traditional predictor variables reduced the amount of potential incremental variance available to be accounted for by the MCAT subscores.

The purpose of the present study was to examine the ability of the New Medical College Admissions Test (MCAT) to predict medical students' performance on measures of both basic and clinical science, and to compare the New MCAT's performance with that of

the Old MCAT. The specific research questions included: (1) do the Old and New MCAT scores correlate positively and significantly with (a) basic science performance as measured by Part I of the National Board of Medical Examiners' examination (NBME Part I), (b) clinical science performance is measured by NBME Part II, (c) house officer ratings of students' clinical problem solving skills, and (d) house officer ratings of students' clinical interpersonal skills?; (2) do the Old and New MCAT scores explain significant incremental (predictive) variance in these four performance measures beyond that accounted for by other preadmission variables?; and (3) is the predictive power of the New MCAT superior to that of the Old MCAT?

There are several distinctions between the present study and previous studies. First, only more traditional preadmission measures that are routinely used in making admission decisions were included as predictor variables in addition to MCAT scores. Second, criterion measures included faculty ratings of student clinical performance and student NBME Part II (clinical science), as well as Part I (basic science), total scores. The clinical performance and NBME Part II measures represent outcome measures in New MCAT validity studies that have just begun to be examined (Carline et al, 1983; Hull, Calhoun & Maxim, 1981). Additionally, there are methodological differences between this study and those previously reported in the literature. These differences are discussed in the following section and focus primarily on the type of regression analyses performed, cross-validation procedures, and the exclusion of the New MCAT Science Problems subtest in multivariate analyses.

Methodology

Instrumentation and Sampling

Preadmission measures, clinical clerkship evaluation ratings, and NBME Part I and II total scores were obtained for persons who entered the four year curriculum at The University of Michigan Medical School in 1977 and 1978. These will be referred to as the

classes of 1981 and 1982, respectively, their year of graduation. The preadmission variables included undergraduate science and nonscience GPA, MCAT scores for each component of the Old MCAT for the class of 1981 and New MCAT for the class of 1982, and a mean interview rating assigned by the medical school faculty members who interview applicants. These represent the quantitative preadmission predictor measures available for students at the time of their application to medical school.

The clinical evaluation ratings and NBME scores represent the criterion medical school performance measures examined in the study. Clinical evaluation scores were obtained from faculty ratings of student performance on a clinical evaluation form (CEF) completed for each student during their required third year clerkship in Internal Medicine. Although most disciplines use the CEF, Internal Medicine was selected as representative to control for variations among ratings attributable to clerkship disciplines. Because several faculty members and house staff complete CEFs for each student, one CEF was randomly selected for each student from all CEFs completed by house officers for that student during the last four weeks of his/her twelve week clerkship. House officer CEFs were selected because previous studies have indicated their evaluations tend to have higher inter-rater agreement and correlate higher with NBME Part II scores than do faculty evaluations (Hull, 1982). The two subscores of the CEF, one representing problem solving skills and the other representing interpersonal skills, were used in the analyses for each student. An analysis of the reliability and validity of the CEF is presented elsewhere (Dielman, Hull & Davis, 1980).

Total sample sizes ranged between 155 and 185 subjects for analyses pertaining to an entire class. Subjects were randomly divided into two sub-samples, a screening sample and a calibration sample, in order to cross-validate the results obtained in the multiple correlation/regression analyses (Kerlinger & Pedhazur, 1973; Lord & Novick, 1968) described in the following section. All data were analyzed for each sub-sample independently and again for the total combined sample.

Capitalization on chance in the development of a regression/prediction model based on sample correlations is a well known problem (Lord & Novick, 1968). Because these sample correlations are based not only on true correlation among the variables, but also contain sampling error, the multiple correlation typically "shrinks" when these variables are used on a new sample. Both Lord and Novick (1968) and Kerlinger and Pedhazur (1973) recommend cross-validation procedures to address this problem. Cross-validation necessitates obtaining two samples. The first sample is referred to as the screening sample, and is used to develop the regression equation and multiple R^2. The predictor variables of the second sample, referred to as the calibration sample, are then applied to the regression equation obtained from the screening sample to obtain predicted scores for the criterion variable. The observed criterion scores (y) for the calibration sample are then correlated with the predicted criterion scores (y'). This Pearson $r_{yy'}$ is analogous to a multiple correlation between the observed and predicted scores. In the present study, this procedure was applied twice in order to allow each sub-sample to constitute the screening (and calibration) sample. This "double cross-validation procedure is strongly recommended as the most rigorous approach to the validation of results from regression analysis in a predictive framework" (Kerlinger and Pedhazur, 1973, p. 284). Results of the two regression equations, multiple R^2s and $r_{yy'}$s obtained from alternate samples were then compared. Analyses of the data were performed retrospectively and were not used in making admission decisions.

Correlational and Incremental Validity Analyses

Pearson zero order correlations were computed to test the research hypotheses of a significant positive relationship between each of the MCAT subscores and the four criterion performance measures. Incremental validity (Sechrest, 1963) was examined by using a step-wise, hierarchical multiple regression analysis design involving a two step procedure. In the first phase, all non-MCAT preadmission variables were simultaneously

included in the analysis. Only after all these non-MCAT variables were included were the MCAT subscores simultaneously stepped in the second phase of the analysis. Four seperate analyses were performed, one for each of the criterion measures. These analyses permitted an examination of the usefulness of the MCAT subtests in explaining additional variance in the criterion measures beyond that already explained by the non-MCAT admission variables.

Three separate indices of MCAT incremental validity were calculated. The first index indicates the absolute amount of variance (as measured by multiple R^2) explained for each of the four criterion measures by the MCAT subtest scores when they are stepped into the multiple regression analysis after all the non-MCAT preadmission measures have been included (Sechrest, 1963). This index was determined using formula 1.

$$\text{Index 1} = (R^2 \text{ for all variables}) - (R^2 \text{ for non-MCAT variables}) \quad (1)$$
$$= R^2 \text{ added by MCAT}$$

The second index provides a measure of the proportional increase in performance variance explained by stepping in MCAT scores last in the regression analysis and was calculated using formula 2 below.

$$\text{Index 2} = \frac{R^2 \text{ added by MCAT}}{R^2 \text{ for non-MCAT variables}} \quad (2)$$

The third index provides a measure of the proportional increase in performance variance that is unaccounted for by the non-MCAT measures and that is explained by adding the MCAT scores to the regression analysis. This index was calculated using formula 3 below. Friedman and Porter (1981) argued for the inclusion of both of these later two indexes in order to minimize artifactual differences in incremental validity

between the two MCAT versions due to varying amounts of non-MCAT explained variance remaining for differing medical school classes.

$$\text{Index 3} = \frac{R^2 \text{ added by MCAT}}{1 - (R^2 \text{ for non-MCAT variables})} \quad (3)$$

Because the scores on the New MCAT Science Problems subtest are derived from a subset of the items that comprise three other New MCAT subtests, Biology, Chemistry, and Physics, this subtest is by definition linearly dependent upon these other subtests. Thus, while "scores on the six New MCAT areas of assessment are designed to be relatively independent and are purposefully reported seperately . . . . items from the Science Problems subtest contribute twice to New MCAT scores" (New MCAT Interpretive Manual, 1977). This issue has been addressed in several New MCAT validity studies (Hull, Calhoun & Maxim, 1981; Jones & Thomae-Forgues, 1981) by excluding the Science Problems subtest from multivariate analyses, while it has been included in most other studies (e.g. Carline, Cullen & Scott, 1982; Carline et al, 1983; Friedman & Bakewell, 1980; Friedman & Porter, 1981; Holley, 1981; McGuire, 1980; Molidor & Elstein, 1979; Molidor, Elstein & Scheifley, 1980). Psychometrically the problem is that the Science Problems subtest partakes of the same error component of the other subtests, violating the assumption of uncorrelated error variance, raising serious interpretative questions in multivariate analyses such as factor analysis (Gorsuch, 1974). When independent variables such as these are highly correlated in multiple regression analyses, "not only do the estimated regression coefficients tend to be quite imprecise, but the true regression coefficients tend to lose their meaning" (Neter & Wasserman, 1974). On the other hand, multicollinear variables have been included in the same analyses when strong rationale for their inclusion has been given. It is likely that the Science Problems subtest has been included in prediction equations used to make admission decisions at

many medical schools. A discussion of the incremental and predictive validity and usefulness of the New MCAT Science Problems subtest in predicting medical student basic and clinical science performance is presented elsewhere (Wolf, Calhoun, Maxim & Davis, 1983).

All Possible Subsets Regression Analyses

All possible subsets regression analyses (Frane, 1981) including the five New and four Old MCAT subtests are reported for each of the criterion measures. "The only way to be sure of obtaining the best n of N predictors would be to determine the multiple correlation for every such "set" by using an exhaustive procedure (Lord & Novick, 1968, p. 288). Until recently the economic cost of performing such analyses was prohibitive. However, "one major advance of the past decade in multiple regression has been the replacement of stepwise procedures with all possible subset searches for model selection, served by the C_p plot" (Wainer & Thissen, 1981, p. 313). Use of the Furnival-Wilson (1974) algorithm enables the identification of "subsets while computing only a small fraction of all possible regressions. Computer costs are comparable for stepwise regression for up to about 25 independent variables" (Frane, 1981, p. 264). For a discussion of some of the problems and issues related to stepwise procedures, see Cohen and Cohen (1975). Virtually all the studies encountered in the MCAT literature have used stepwise procedures in regression analyses, another distinction from the present study.

Mallow's C_p was the criterion used to identify the best subsets. The "best" subset is selected on the basis of an analysis of residuals that minimizes C_p based on the following formula (Daniel & Wood, 1971; Frane, 1981):

$$C_p = \frac{RSS}{s^2} - (N-2p') \qquad (4)$$

where

RSS= residual sum of squares for the subset of independent variables being tested

s^2= residual mean square based on the regression using all independent variables

p'= the number of variables in the subset, including the intercept, if any.

n= number of cases (sample size)

In addition, multiple R^2s and adjusted R^2s based on formula 5 were calculated.

$$\text{Adjusted } R^2 = \frac{R^2 - p(1-R^2)}{N - p'} \qquad (5)$$

where p = the number of independent variables when the intercept is set to zero.

These analyses enabled an examination of which potential preadmission measures the MCAT subtests and/or non-MCAT measures, were included in the "best" regression model for each criterion.

Results and Discussion: New MCAT

Results are presented seperately for the New and Old MCAT, respectively, before similarities and differences are summarized. Validity coefficients (i.e., correlations) among non-MCAT and New MCAT preadmission variables and the four criterion performance measures are summarized in Table 1 for both subsamples. Table 2 contains validity coefficients for all subjects (i.e., both subsamples combined). Among the preadmission variables, all New MCAT subtests were significantly correlated with each other in both subsamples and in the entire sample ($p < .05$, ranging between $r = .23$ and .74), except for the Reading - Biology correlation in subsample 2 ($r = .11$, n.s.). Undergraduate grades for science and non-science were significantly ($p < .01$) related in

both sample 1 (r = .61) and sample 2 (r = .69). Even though specific coefficients vary from sample 1 to sample 2, in general these results are consistent with prior research and support the validity of the interrelationships.

The findings regarding the preadmission interview rating were not replicated in the two subsamples. In sample 1, these ratings were significantly ($p < .05$) associated with both science and non-science GPAs, as well as with the MCAT Biology subtest. In sample 2, these associations were not significantly different than chance. However, these ratings were significantly associated ($p < .05$) with two other MCAT subtests, Physics and Chemistry. Thus the results of the relationships between the interview ratings and the other preadmission measures are equivocal. The same may be said for the associations between the grade point indexes and the New MCAT subtests. In sample 1, 11 of 12 validity coefficients were statistically significant, while only 3 of 12 were significant in sample 2. Thus the significant associations of science GPA with MCAT Physics and Chemistry performance and of non-science GPA with MCAT Chemistry performance were the only associations that were replicated in both samples.

Five of the 6 validity coefficients among the criterion measures were replicated in the two subsamples. The following associations were significant and positive in both samples: CEF-PS with both NBME I and II and with CEF-IP, and NBME I with NBME II. There was a consistent non-significant chance association between CEF-IP and NBME II in the two samples. Results for validity coefficients for CEF-IP and NBME I were inconsistent and thus ambiguous, as the coefficient was significant in sample 1 (r = .23, $p < .05$) and non-significant in sample 2 (r = .05). These results provide evidence of the concurrent validity of the CEF-PS, NBME I, and NBME II measures.

Validity coefficients between the two CEF criterion measures and the preadmission measures were generally disappointingly low in both subsamples, with only one coefficient attaining statistical significance in each sample (non-science GPA with CEF-IP in sample 1 and MCAT biology with CEF-PS in sample 2). Thus there appears to be little, if any,

zero order association between these clinical clerkship evaluations and preadmission measures.

All 12 validity coefficients between the 6 New MCAT subtests and NBME I and II performance were significant in sample 1, while 10 of 12 were significant in sample 2. These findings generally support the validity of the associations between the standardized admission measures (New MCAT) and the standardized medical school performance measures (NBME) that are found in most studies. However, consistent with prior research (Carline et al, 1983; Hull et al, 1981; Jones & Thomae-Forgues, 1982), these coefficients were consistently higher than those between undergraduate GPAs and NBME performance. The preadmission interview rating significantly related to NBME performance in sample 2, but not in sample 1.

In summary, the research hypotheses of a significant positive relationship between each of the MCAT subscores and the four criterion performance measures were rejected in relation to the two clinical evaluation measures, but accepted for NBME Parts I and II performance.

Insert Tables 1 and 2 about here

Incremental Validity Results

For sample 1, sample 2, and both samples combined, multiple R^2s indicated that all preadmission measures accounted for 8 percent, 16 percent, and 7 percent, respectively, of the variance in clinical problem solving evaluations (CEF-PS). New MCAT subtests accounted for the majority of this explained variance, 5 percent, 11 percent, and 4 percent in the three samples. These later percentages constitute the incremental validity of the New MCAT (index 1). Results summarized in Table 3 indicate , for example, that MCAT subtests explained 2.2 times more variability in CEF-PS ratings than did

non-MCAT measures (index 2), or 12 percent of the remaining variance not explained by the non-MCAT measures (index 3). Results for the CEF-IP criterion measure were similar to CEF-PS results, with the absolute amounts of variance accounted for being generally small. Cohen (1977) provides indexes of effect size for multiple R^2s for small (2 percent shared variance), medium (13 percent), and large (26 percent) effects.

Effect sizes improve considerably when NBME Parts I and II are the criteria. Multiple R^2s indicate that all preadmission measures explain between 25 percent and 47 percent of the variance in NBME I and II performance. Based on Cohen's criteria, these may be considered large effects. Again, MCAT subtests accounted for the majority of this explained variance, with the amount of variability additionally explained by the MCAT ranging between 13 percent and 37 percent (index 1). These effects thus may be considered to be medium to large in magnitude. The most dramatic effect occured in sample 1 where MCAT explained 7.4 times (740 percent) more variability than non-MCAT measures in NBME Part II performance (index 2). This amounted to explaining 39 percent of the remaining variance in NBME II performance once non-MCAT variance was removed (index 3). In summary, these findings clearly support the incremental validity of the New MCAT subtests in contributing to explained variance in the criterion measures.

Insert Table 3 about here

All Possible Subsets Regression Results

These analyses were performed to examine which preadmission measures were included in the best regression models for predicting each criterion. Based on the selection criteria of minimizing the C_p statistic for residuals, the following standardized regression models for CEF-PS were obtained for subsample 1 (equation 6), subsample 2 (equation 7), and the combined sample (equation 8):

PS.1 = .14 GPA.O + 6.32 (6)

PS.2 = .20 Rating + .19 GPA.S + .28 Biology - .27 Chemistry - 1.09 (7)

PS = .17 Rating + .17 Biology - .12 Chemistry - 1.34 (8)

The following models for CEF-IP were obtained for subsample 1 (equation 9), subsample 2 (equation 10), and the total sample (equation 11):

IP.1 = .31 GPA.O + 4.04 (9)

IP.2 = .23 Rating + .16 Biology - .21 Chemistry - 0.87 (10)

IP = .12 Rating + .17 GPA.O + .17 Biology - 0.16 Chemistry - 0.20 (11)

In comparing equations 6 and 7 between the two subsamples for CEF-PS and equations 9 and 10 for CEF-IP, it is evident that these models are inconsistent and not cross-validated. This is perhaps not surprising given the small amount of variance accounted for in the CEF measures by all the preadmission measures, individually or in combination.

The regression models developed for NBME performance faired somewhat better. The models for predicting NBME Part I peformance for each of the two subsamples and the combined sample are presented in equations 12-14 below.

NBME I.1 = .49 Biology + .26 Chemistry + 1.18 (12)

NBME I.2 = .25 Biology + .24 Chemistry + .17 Rating - 5.86 (13)

NBME I = .36 Biology + .26 Chemistry + .10 Reading + .12 Rating - 0.08 (14)

In examining equations 12 and 13, it is clear that both the New MCAT Biology and Chemistry subtests are good predictors and should be included in the model for NBME I. The result for the preadmission interview rating is ambiguous, as it was not validated in

subsample 1. The model represented in equation 14 for all subjects was selected on the basis of having the smallest C_p value (5.13), and should provide a more stable regression model than either of the subsample models (Kerlinger & Pedhazur, 1973; Moiser, 1951). However, the model comprised of just Biology and Chemistry resulted in a C_p value of 7.03. Combined with Frane's (1981) recommendation that only independent variables whose coefficients are significantly different from zero be retained, it may be unlikely that adding either the Reading (the beta coefficient of .10 was not significant, $p < .12$) or Rating (= .12, $p < .06$) subtests would result in predictions substantially different from excluding them from the model. This issue clearly necessitates further examination, as the beta coefficient for Rating approached statistical significance.

Models resulting for NBME Part II also contained similarities and differences, as evidenced by equations 15-17.

NBME II.1 = .41 Biology + .24 Reading + .20 Quantitative - 0.35 (15)

NBME II.2 = .30 Biology + .26 Chemistry + 2.38 (16)

NBME II = .38 Biology + .20 Reading + .16 Quantitative + .12 Rating - 5.92 (17)

Clearly the New MCAT Biology subtest is a component of the model for NBME II. Results for the MCAT Reading, Quantitative, and Chemistry subtests, and for interview ratings are not validated and remain equivocal.

Table 4 summarizes the C_p, multiple R^2, adjusted R^2, and r_{yy}, values for the best subset regression models reported above. The r_{yy}, coefficient of .65 for NBME I for sample 1 was obtained by correlating sample 1 (calibration sample) subjects' observed scores with their predicted scores based on the model derived with sample 2 (screening sample). In general, squaring the r_{yy}, coefficients from each sample and comparing them with the multiple R^2 or adjusted R^2 coefficients from the same sample indicates striking similarity and consistency for both NBME measures. The difference between multiple R^2s

for the two samples, as well as the difference between $r_{yy'}$ coefficients, provides an estimate of the amount of shrinkage of the multiple correlation. In general, shrinkage decreases as sample sizes increase (Kerlinger & Pedhazur, 1973). Even though the ratio of subjects to the number of independent variables was approximately 9 or 10:1 for the two subsamples, these samples may still be considered relatively small for the types of analyses performed. As data become available for the graduating class of 1983, it would be useful to replicate these analyses with the entire classes of 1982 and 1983 representing the two samples, in contrast to dividing the class of 1982 into two subsamples as reported here.

Insert Table 4 about here

Results and Discussion: Old MCAT

Validity coefficients among the non-MCAT and Old MCAT preadmission variables and the four criterion performance measures are summarized in Table 5 for both subsamples. Table 6 contains validity coefficients for both subsamples combined. Among the preadmission variables, all Old MCAT subtests were significantly correlated ($p < .05$) with each other in both subsamples and in the combined sample ($r = .18$ to $.59$), with the exception of the correlation between General Information and Quantitative for subsample 2 ($r = .10$, n.s.). The only preadmission measure significantly associated with the preadmission interview ratings was the Old MCAT Quantitative subtest in both sample 1 ($r = .27$, $p < .01$) and sample 2 ($r = .23$; $p < .05$). Undergraduate science grade point averages were significantly associated with the Quantitative and Science subtests in both samples ($r = .46$ to $.58$; $p < .01$), and with the Verbal subtest in sample 1 only. Non-science grades (GPA-O) correlated significantly with the Old MCAT Science subtest in sample 1 only. GPA-O was consistently not related to the Verbal, Quantitative, and Science subtests in both samples.

Four of the six validity coefficients among the criterion measures were replicated in the two subsamples: there were significant positive associations between CEF-PS and CEF-IP, NBME I and NBME II, and CEF-PS and NBME II. CEF-IP and NBME II were consistently not related except by chance. The associations between NBME I and the two CEF measures were significant in sample 2, but not in sample 1, making these relationships equivocal.

Validity coefficients between the two CEF measures and the preadmission measures were again disappointingly low, as only the correlation between CEF-IP with the preadmission interview rating was significant (r = -.27; $p < .05$). However, this relationship was in the opposite direction than predicted and washed out when the two subsamples were combined in Table 6. Consistent with results previously presented in the summary of the New MCAT, there is little, if any, zero order association between the clinical clerkship evaluations and preadmission measures.

In contrast to the above results for CEF, 10 of 14 correlations in sample 1 and 9 of 14 correlations in sample 2 between NBME and preadmission measures were significant ($p < .05$). Cross-validated positive and significant associations included: NBME I with GPA-S, Old MCAT Verbal, General Information, and Science; NBME II with Old MCAT Verbal and General Information. Equivocal, ambiguous results were found for NBME I with Rating (r = .26, $p < .01$ in sample 2; r = .10 in sample 1), GPA-O (r = .22, $p < .05$ in sample 1; r = .11 sample 2), and Old MCAT Quantitative (r = .26, $p < .05$ in sample 1; r = .20, $p < .06$ in sample 2); for NBME II with Rating (r = .25, $p < .05$ in sample 2; r = .11 in sample 1), GPA-S (r = .28, $p < .05$ in sample 2; r = .22, $p < .06$ in sample 1), GPA-O (r = .24, $p < .05$ in sample 1; r = .10 in sample 2), and Old MCAT Science (r = .45, $p < .01$ in sample 1; r = .15 in sample 2). Consistent chance relationships were found between Old MCAT Quantative and NBME II in both subsamples.

Insert Tables 5 and 6 about here

Incremental Validity Results

Results summarized in Table 7 indicate that little variance in CEF performance was explained by the preadmission measures. Multiple R^2s including all predictors accounted for 4-6 percent of the varance in CEF-PS, and 4-15 percent in CEF-IP. In contrast to results in Table 3 for the New MCAT summary, Old MCAT subtests generally accounted for the same or less variance in CEF performance than did the non-MCAT measures.

Again, effect sizes improved considerably when NBME performance served as the criteria, with Old MCAT subtests contributing substantially to the amount of variance accounted for in both NBME Parts I and II beyond that accounted for by non-MCAT measures. R^2 added by the Old MCAT ranged between .10 and .25, which connote medium to large effects based on Cohen's (1977) criteria.

Insert Table 7 about here

All Possible Subsets Regression Results

Based on the criteria of minimizing Mallows C_p residual statistic, the following regression models resulted for CEF-PS for subsample 1 (equation 18), subsample 2 (equation 19), and the combined sample (equation 20):

$$PS.1 = .13\ GPA.S + 6.55 \quad (18)$$

$$PS.2 = .15\ GPA.S + 8.37 \quad (19)$$

$$PS = .13\ GPA.S + 7.20 \quad (20)$$

Undergraduate Science grade point average consistently emerges in the best model for predicitng CEF-PS in both subsamples and the combined sample. Equations 21-23 summarize the best models for CEF-IP. These results are ambiguous based on different models and the inconsistency in the sign of the preadmission interview rating in equations 21 and 22.

IP.1 = .21 GPA.O - .30 Rating + 20.52 (21)

IP.2 = .15 Rating + 1.69 (22)

IP = .15 GPA.O + 6.05 (23)

Just as in the New MCAT analyses, the regression models developed for NBME performance faired fairly well when they are compared for each subsample. Models for predicting NBME Part I performance are presented in equations 24-26.

NBME I.1 = .14 GPA.S + .56 Science - 2.71 (24)

NBME I.2 = .36 GPA.S + .30 Science + .20 General + .19 Rating - 13.75 (25)

NBME I = .23 GPA.S + .43 Science + .12 General + .10 Rating - 8.24 (26)

In examining equations 24 and 25, it is clear that both science GPA and the Old MCAT Science subtest are good predictors and should be included in the model for NBME I. Results for the Old MCAT General Information subtest and the preadmission interview Rating are equivocal. Models obtained for NBME II for subsample 1 (equation 27), subsample 2 (equation 28), and the combined sample (equation 29) are presented below:

NBME II.1 = .15 GPA.O + .35 Science + .21 General - 2.85 (27)

NBME II.2 = .28 GPA.S + .30 General + .16 Rating - 8.52 (28)

NBME II = .20 GPA.S+.29 General+.19 Science-.17 Quantitative+.17 Rating-8.02 (29)

In comparing equations 27 and 28 it can be seen that Old MCAT General Information should be included in the prediction model for NBME II. Results for GPA-O, GPA-S, MCAT Science and Quantitative, and the interview Rating are equivocal. Table 8 summarizes Mallow's C_p Multiple R^2, Adjusted Multiple R^2, and cross-validated composite correlations ($r_{yy'}$) for the best subset regression analyses using the Old MCAT. This table is analogous to Table 4 for the New MCAT and is interpreted similarly.

Insert Table 8 about here

Comparisons Between Old and New MCAT

Comparing the incremental validity results for the New MCAT summarized in Table 3 with results for the Old MCAT in Table 7, it is evident that the New MCAT explains a larger proportion of incremental variability in each of the four criterion measures than does the Old MCAT. This finding is consistent for all three indexes of incremental validity. This is also consistent with incremental validity results reported by Friedman and Porter (1981).

Both the Old and New MCATS did account for additional unique predictive variance in all four outcome measures when they were included in incremental validity analyses after the non-MCAT measures were included. Consistent with the zero order correlational analyses, these increments were significant in explaining additional NBME Part I and II variance, but non-significant for CEF-PS and CEF-IP variance.

In general, the incremental validity indexes reported in this study tended to be larger than those reported by Friedman and Bakewell (1980) and Friedman and Porter (1981). There were more non-MCAT measures included in their analyses, which not surprisingly accounted for additional variance in their criterion measures. This therefore,

left less remaining variance in their criteria for the MCAT to potentially account for than in the present study. Thus it is not surprising that the incremental validity indexes were larger in the present study.

Conclusions

Several inferences may be made from these results. MCAT subscores do account for predictive and incremental validity in medical school performance measures. This is particularly true for standardized measures of both basic and clinical science performance measured by the NBME examinations. This is less true for house staff clinical evaluation ratings. Several factors may mitigate against this later relationship, including the difference in format between the measures; MCAT uses a multiple choice format, while clinical evaluations are necessarily based on supervisor judgments indicated on behaviorally anchored rating scales. Restriction of range in performance, homogeneity of the sample, and low reliability of the measures could partially account for these weak relationships. Clearly, homogeneity of the sample and restricted range of performance on these measures are tenable explanations as a result of high admission standards and the generally high level of student performance. As Carline and his colleagues (1982, p. 208) have pointed out, "intervention of three years of study between entrance into medical school and completion of basic clinical training must act to decrease correlations between measurements. An additional limit on correlations is the inherent restriction of range in preselection variables; only higher-scoring students are admitted to medical training". Thus "the lack of large correlation coefficients does not offer significant evidence to negate the utility of the MCAT as an aid in the selection of academically able students" (Carline et al, 1983, p. 25). Low reliability is not as tenable an explanation because of the acceptable reliabilities of both the MCAT and CEF (New MCAT Interpretative Manual, 1977; Dielman et al, 1980). Additionally, the fact that the clinical ratings did correlate significantly with the NBME examinations in this study does provide support for their validity.

In summary, the findings of the present study support the following conclusions:

1. Both the Old and New MCAT explain additional incremental variance in medical school performance measures beyond that already explained by non-MCAT measures. This is particularly true for standardized criteria (i.e. NBME).
2. The New MCAT appears to have larger incremental validity coefficients than the Old MCAT.
3. The effect sizes appear to be medium to large when the criteria are National Boards. The effect sizes are very small when the criteria are clinical evaluation ratings.
4. Cross-validation analyses support the inclusion of the New MCAT Biology subtest in prediction models of NBME Part I and II performance. These analyses also support the inclusion of the New MCAT Chemistry subtest in the prediction model for NBME Part I only. Results for the other New MCAT subtests were equivocal.
5. Old MCAT analyses parallel New MCAT findings for NBME I in that the Old Science subtest, replaced by the New Biology and Chemistry subtests, cross-validated. The inclusion of the Old MCAT General Information subtest was cross-validated in the prediction model for NBME Part II.
6. Findings for prediction models for clinical evaluation performance ratings were equivocal and not cross-validated for either New or Old MCAT subtests.

Several methodological recommendations for future studies are suggested. These include (1) the substitution of all possible subsets regression analyses for stepwise procedures, (2) cross-validation of correlational/regression analyses, and (3) the exclusion of the New MCAT Science Problems subtest in multivariate analyses. Issues related to this third recommendation are addressed elsewhere (Wolf et al, 1983).

REFERENCES

Bulletin of Information and Description of National Board Examinations. Philadelphia: National Board of Medical Examiners, 1982.

Carline, J.D., Cullen, T.J. and Scott, C.S. Prediction of Clerkship Performance Using the New MCAT Examination: An Attempted Application of Canonical Redundancy Analysis. Proceedings of the 21st Annual conference on Research in Medical Education, Washington, D.C., 205-210, 1982.

Carline, J.D., Cullen, T.J., Scott, C.S., Shannon, N.F. and Schaad, D. Predicting Performance During Clinical Years from the New Medical College Admission Test. Journal of Medical Education, 58:18-25, 1983.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences (Rev. ed.). N.Y.: Academic Press, 1977.

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, N.J.: Erlbaum, 1975.

Cullen, T.J., Dohner, C.W., Peckham, P.D., Samson, W.E., Schwarz, M.R. Predicting First-Quarter Test Scores from the New Medical College Admission Test. Journal of Medical Education, 55:393-398, 1980.

Daniel, C. and Wood, F.S. Fitting Equations to Data. N.Y.: Wiley, 1971.

Dawson-Saunders, B., and Doolen, D.R. An Alternative Method to Predict Performance: Canonical Redundacy Analysis. Journal of Medical Education, 56:295-299, 1981.

Dielman, T.E., Hull, A.L., and Davis, W.K. Psychometric Properties of Clinical Performance Ratings. Evaluation and the Health Professions, 3:103-117, 1980.

Erdmann, J.B. Validating the MCAT. Journal of Medical Education, 55:463-464, 1980. (editorial)

Frane, J. All Possible Subsets Regression. In W.J. Dixon and M.B. Brown (eds.). BMDP: Biomedical Computer Programs. Los Angeles: Univeristy of California Press, 264-277, 1981.

Friedman, C.P., and Bakewell, Jr., W.E. Incremental Validity of the New MCAT. Journal of Medical Education, 55:399-404, 1980.

Friedman, C.P. and Porter, C.Q. Incremental Validity: The Old and New MCATs Compared. Proceedings of the 20th Annual Conference on Research in Medical Education, Washington, D.C., 251-256, 1981.

Furnival, G.M. and Wilson, R.W. Regression by Leaps and Bounds. Technometrics, 16:499-511, 1974.

Gorsuch, R.L. Factor Analysis. Philadelphia: W.B. Saunders, 1974.

Holley, C.D. Background Variables as Predictors of Acadmic Performance. Paper presented at the Educational Conference of the American Association of Colleges of Osteopathic Medicine, Kansas City, June 1981.

Hull, A.L. House Officer and Attending Staff as Evaluators of Medical Student Performance, Evaluation and the Health Professions, 5:87-94, 1982.

Hull, A.L., Calhoun, J.G. and Maxim, B.R. Predicting Medical School Performance Using the Old and New MCAT. Proceedings of the 20th Annual Conference on Research in Medical Education, Washington, D.C., 135-141, 1981.

Jones, R.F. and Thomae-Forgues, M. A Factor Comparison of Old and New MCAT Scales. Journal of Medical Education, 56:161-166, 1981.

Jones, R.F. and Thomae-Forgues, M. From MCAT to M.D.: Predicting Success in Medical School. In R. L. Linn (Chair), Interpreting Data through Externally Smudged Glasses: The Continuing Saga of Graduate and Professional School Validation Research. Symposium presented at the meeting of the American Educational Research Association, New York, March 1982.

Lord, F.M. and Novick, M.R. Statistical Theories of Mental Test Scores. New York: Addison-Wesley, 1968.

Kerlinger, F.N. and Pedhazur, E.J. Multiple Regression in Behavioral Research. N.Y.: Holt, Reinhart & Winston, 1973.

McGuire, F.L. The New MCAT and Medical Student Performance. Journal of Medical Education, 55:405-408, 1980.

Molider, J.B. and Elstein, A.S. A Factor Analytic Study of the Old and New MCAT Examinations. Proceedings of the 18th Annual Conference on Research in Medical Education, Washington, D.C., 139-144, 1979.

Molidor, J.B., Elstein, A.S., Scheifley, V. The Old and New MCAT Examinations: What Do They Measure? Paper presented at the meeting of the American Educational Research Association, Boston, April 1980.

Moiser, C.I. Problems and Designs of Cross-Validation. Educational & Psychological Measurement, 11:5-11, 1951.

Neter, J. and Wasserman, W. Applied Linear Statistical Models. Homewood, Il: R.D. Irwin, 1974.

New Medical College Admission Test Interpretative Manual. Washington, D.C.: Association of American Medical Colleges, 1977.

Sechrest, L. Incremental Validity: A Recommendation. Educational and Psychological Measurement, 23:153-158, 1963.

Wainer, H. and Thissen, D. Graphical Data Analysis. In M.R. Rosenzweig & L. W. Porter (eds.). Annual Review of Psychology, 32:191-241, 1981.

Wolf, F.M., Calhoun, J.G., Maxim, B.R., and Davis, W.K. Predictive and Incremental Validity of the New MCAT Science Problems Subtest. Presented at the meeting of the National Council on Measurement in Education, Montreal, April 1983.

Table 1

Pearson Correlations Among Preadmission, New MCAT, Clinical Evaluation, and NBME I, II scores for Two Subsamples of Class of 1982

Sample 1

		Rating	GPA		BI	PH	CH	SP	RE	QA	CEF		NBME	
		1	2	3	4	5	6	7	8	9	PS 10	IP 11	I 12	II 13
1.	RATING	------	.22*	.22*	.22*	-.02	-.06	.11	-.08	-.03	.12	.07	.19	.04
2.	GPA-S	.16	------	.61**	.37**	.36**	.53**	.52**	.26**	.34**	.01	.10	.38**	.23*
3.	GPA-O	.07	.69**	------	.20*	.11	.23*	.22*	.22*	.10	.14	.31**	.18	.18
4.	MCAT-BI	.14	.07	-.09	------	.47**	.48**	.65**	.26*	.42**	.04	.12	.62**	.56**
5.	MCAT-PH	.21*	.24*	.02	.32**	------	.62**	.63**	.23*	.38**	.08	.11	.47**	.37**
6.	MCAT-CH	.25	.37**	.21*	.30**	.51**	------	.74**	.33**	.44**	-.01	.05	.49**	.36**
7.	MCAT-SP	.12	.09	-.07	.58**	.62**	.58**	------	.26*	.47**	.10	.14	.55**	.43**
8.	MCAT-RE	-.02	.13	.16	.11	.26**	.28**	.32**	------	.29**	.01	-.03	.30**	.40**
9.	MCAT-QA	-.14	.20	-.04	.23*	.35**	.38**	.45**	.26*	------	.12	.04	.32**	.43**
10.	CEF-PS	.19	.13	.04	.24*	-.00	-.07	.09	.03	.03	------	.72**	.27*	.32**
11.	CEF-IP	.18	.07	.05	.13	-.03	-.11	.05	-.04	-.04	.70**	------	.19	.23*
12.	NBME-I	.27**	.19	.01	.35**	.29**	.35**	.38**	.17	.24*	.46**	.16	------	.81**
13.	NBME-II	.24*	.12	.01	.37**	.21	.34**	.39**	.21*	.33*	.30**	.05	.78**	------

Sample 2

Note: RATING = preadmission interview rating; GPA = grade point average, S = science, O = other; MCAT = New version, BI = biology, PH = physics, CH = chemistry, SP = science problems, RE = reading, QA = quantitative; CEF = clinical evaluation form, PS = problem solving; IP = interpersonal skills.

*p < .05
< .01

Table 2

Pearson Correlations Among Preadmission, New MCAT, Clinical Evaluation, and NBME I, II scores for All Students of Class of 1982

		Rating	GPA							CEF		NBME
			S	O	BI	PH	CH	SP	RE	QA	PS	IP	
		1	2	3	4	5	6	7	8	9	10	11	12
1.	RATING												
2.	GPA-S	.19**											
3.	GPA-O	.14*	.64**										
4.	MCAT-BI	.19**	.25**	.06									
5.	MCAT-PH	.01	.30**	.06	.40**								
6.	MCAT-CH	-.01	.46**	.22**	.40**	.57**							
7.	MCAT-SP	.10	.35**	.08	.63**	.62**	.67**						
8.	MCAT-RE	-.06	.21**	.19**	.20**	.25**	.31**	.29**					
9.	MCAT-QA	-.03	.28**	.03	.34**	.36**	.42**	.46**	.28**				
10.	CEF-PS	.16*	.08	.09	.13	.04	-.04	.09	.02	.07			
11.	CEF-IP	.13	.07	.15*	.12	.03	-.05	.09	-.04	-.01	.71**		
12.	NBME-I	.18*	.29**	.09	.51**	.38**	.43**	.48**	.24**	.29**	.36**	.16*	
13.	NBME-II	.13	.19*	.08	.48**	.29**	.36**	.42**	.32**	.35**	.31**	.12	.80**

Note: RATING = preadmission interview rating; GPA = grade point average, S = Science, O = other; MCAT = New version, BI = biology, PH = physics, CH = chemistry, SP = science problems, RE = reading, QA = quantitative; CEF = clinical evaluation form, PS = problem solving; IP = interpersonal skills.

*p <.05
**p <.01

Table 3

Incremental Validity for New MCAT
for Class of 1982

Criterion Measure	Statistic	Sample 1	Sample 2	All Subjects
CEF - PS	Sample Size (n)	84	87	171
	R^2 non-MCAT	.03	.05	.03
	R^2 added by MCAT (1)	.05	.11	.04
	Total R^2	.08	.16	.07
	Incremental Validity (2)	1.67	2.20	1.33
	Incremental Validity (3)	.05	.12	.04
CEF-IP	Sample Size (n)	84	87	171
	R^2 non-MCAT	.10	.04	.04
	R^2 added by MCAT (1)	.04	.08	.04
	Total R^2	.14	.12	.08
	Incremental Validity (2)	.40	2.00	1.00
	Incremental Validity (3)	.04	.08	.02
NBME I	Sample Size (n)	92	93	185
	R^2 non-MCAT	.16	.12	.12
	R^2 added by MCAT (1)	.31	.13	.24
	Total R^2	.47	.25	.36
	Incremental Validity (2)	1.94	1.08	2.00
	Incremental Validity (3)	.37	.15	.27
NBME II	Sample Size (n)	81	85	166
	R^2 non-MCAT	.05	.07	.05
	R^2 added by MCAT (1)	.37	.19	.27
	Total R^2	.42	.26	.32
	Incremental Validity (2)	7.40	2.71	5.40
	Incremental Validity (3)	.39	.23	.28

Table 4

Mallow's C_p, Multiple R^2, Adjusted Multiple R^2, and Cross-Validated Composite Correlations ($r_{yy'}$) for Best Subset Regression Analyses Using New MCAT Predictors

Criterion Measure	n	C_p	R^2	Adj R^2	$r_{yy'}$
CEF - PS					
Sample 1	84	-1.60	.02	.01	.07
Sample 2	87	1.80	.15	.11	.04
All Subjects	171	1.59	.05	.04	---
CEF - IP					
Sample 1	84	-2.13	.10	.08	.06
Sample 2	87	0.86	.09	.05	.05
All Subjects	171	1.98	.07	.05	---
NBME I					
Sample 1	92	2.79	.43	.42	.65
Sample 2	93	2.52	.22	.19	.42
All Subjects	185	5.13	.34	.33	---
NBME II					
Sample 1	81	0.11	.41	.39	.53
Sample 2	85	1.87	.20	.18	.41
All Subjects	166	2.75	.32	.30	---

Table 5

Pearson Correlations Among Preadmission, Old MCAT, Clinical Evaluation and NBME I, II Scores for Two Subsamples of Class of 1981

Sample 1

		Rating	GPA						CEF		NBME	
			S	O	VA	QA	GI	SCI	PS	IP	I	II
		1	2	3	4	5	6	7	8	9	10	11
1.	RATING	-----	.16	.13	.15	.27**	.02	.09	-.08	-.27*	.10	.11
2.	GPA-S	.18	-----	.59**	.23*	.58**	.15	.51**	.13	.14	.40**	.22
3.	GPA-O	.17	.42**	-----	.17	.15	.13	.25*	.13	.18	.22*	.24*
4.	MCAT-VA	.16	.15	.14	-----	.38**	.61**	.44**	-.09	-.05	.28**	.29**
5.	MCAT-QA	.23*	.46**	.09	.22*	-----	.26*	.43**	-.08	.01	.26*	.04
6.	MCAT-GI	.15	.08	-.01	.57**	.10	-----	.34**	.06	.12	.24*	.37**
7.	MCAT-SCI	.13	.54**	.15	.29**	.52**	.24*	-----	.07	.09	.62**	.45**
8.	CEF-PS	.09	.10	.06	-.01	.00	-.04	.13	-----	.66**	.17	.23*
9.	CEF-IP	.11	.01	.12	.02	.04	-.10	.08	.61**	-----	.00	.04
10.	NBME-I	.26*	.45**	.11	.28**	.20	.25*	.55**	.40**	.26*	-----	.80**
11.	NBME-II	.25*	.28*	.10	.26*	.04	.30**	.15	.30**	.10	.71**	-----

Sample 2

Note: RATING = preadmission interview rating; GPA - grade point average, S = science, O = other; MCAT = Old version, BI = biology, PH = physics, CH = chemistry, SP = science problems, RE = reading, QA = quantitative; CEF = clinical evaluation form, PS = problem solving; IP = interpersonal skills.

*p <.05
**p <.01

Table 6

Pearson Correlations Among Preadmission, Old MCAT, Clinical Evaluation, and NBME I, II scores for All Students of Class of 1981

		Rating	GPA					CEF		NBME I
			S	O	VA	QA	GI	SCI	PS	IP	
		1	2	3	4	5	6	7	8	9	10
1.	RATING										
2.	GPA-S	.17*									
3.	GPA-O	.15*	.53**								
4.	MCAT-VA	.15*	.20**	.16*							
5.	MCAT-QA	.25**	.52**	.12	.30**						
6.	MCAT-GI	.08	.12	.08	.59**	.18*					
7.	MCAT-SCI	.11	.52**	.21**	.36**	.48**	.29**				
8.	CEF-PS	.04	.13	.10	-.04	-.03	.02	.11			
9.	CEF-IP	-.05	.09	.15	-.01	.03	.02	.09	.65**		
10.	NBME-I	.18*	.45**	.18*	.30**	.29**	.27**	.58**	.27**	.12	
11.	NBME-II	.18*	.25**	.18*	.28**	.04	.34**	.31**	.26**	.07	.76**

Note: RATING = preadmission interview rating; GPA = grade point average, S = science, O = other; MCAT = old version, BI = biology, PH = physics, CH = chemistry, SP = science problems, RE = reading, QA = quantitative; CEF = clinical evaluation form, PS = problem solving, IP = interpersonal skills.

*p < .05
**p < .01

Table 7

Incremental Validity for Old MCAT
for Class of 1981

Criterion Measure	Statistic	Sample 1	Sample 2	All Subjects
CEF - PS	Sample Size (n)	76	79	155
	R^2 non-MCAT	.03	.04	.02
	R^2 added by MCAT (1)	.03	.01	.02
	Total R^2	.06	.05	.04
	Incremental Validity (2)	1.00	.25	1.00
	Incremental Validity (3)	.03	.01	.02
CEF-IP	Sample Size (n)	76	79	155
	R^2 non-MCAT	.13	.03	.03
	R^2 added by MCAT (1)	.02	.02	.01
	Total R^2	.15	.05	.04
	Incremental Validity (2)	.15	.67	.33
	Incremental Validity (3)	.02	.02	.01
NBME I	Sample Size (n)	93	88	181
	R^2 non-MCAT	.16	.33	.22
	R^2 added by MCAT (1)	.25	.15	.19
	Total R^2	.41	.48	.41
	Incremental Validity (2)	1.56	.45	.86
	Incremental Validity (3)	.30	.22	.24
NBME II	Sample Size (n)	78	81	159
	R^2 non-MCAT	.07	.14	.09
	R^2 added by MCAT (1)	.22	.10	.14
	Total R^2	.29	.24	.23
	Incremental Validity (2)	3.14	.71	1.56
	Incremental Validity (3)	.27	.12	.17

Table 8

Mallow's C_p, Multiple R^2, Adjusted Multiple R^2, and Cross-Validated Composite Correlations ($r_{yy'}$) for Best Subset Regression Analyses Using Old MCAT Predictors

Criterion Measure	n	C_p	R^2	Adj R^2	$r_{yy'}$
CEF - PS					
Sample 1	76	-0.46	.02	.00	.13
Sample 2	79	-2.53	.02	.01	.10
All Subjects	155	-0.86	.02	.01	---
CEF - IP					
Sample 1	76	0.84	.12	.10	-.27
Sample 2	79	-2.04	.02	.01	-.07
All Subjects	155	-0.86	.02	.02	---
NBME I					
Sample 1	93	-1.38	.40	.39	.57
Sample 2	88	5.53	.47	.44	.60
All Subjects	181	4.86	.41	.39	---
NBME II					
Sample 1	78	2.27	.27	.24	.40
Sample 2	81	2.18	.21	.18	.30
All Subjects	159	4.85	.23	.21	---