

FINAL REPORT

Identification, Development, and Validation of Predictors for

Successful Lawyering

Marjorie M. Shultz and Sheldon Zedeck, Principal Investigators

September 2008

© 2008 Marjorie M. Shultz and Sheldon Zedeck

NOT TO BE USED FOR COMMERCIAL PURPOSES
NOT TO BE DISTRIBUTED, COPIED, OR QUOTED WITHOUT PERMISSION OF AUTHORS

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of donors without whose generosity we could not have carried out this research: first and foremost, the Law School Admission Council research grants program; Berkeley Law School and Dean Chris Edley; the Gerbode Foundation; Hanson, Bridgett; Martha Faye Africa, and Drucilla Ramey-Stender.

We also thank our loyal and long serving National Advisory Board members (David Chambers, Phoebe Haddon, Thelton Henderson, Garner Weng, Kelly Brown, Pilar Ossorio, Robert Nelson, Beth O'Neil, Paul Sackett, and Leaetta Hough). We were also guided by our Social Science Research Advisors including Paul Sackett, Leaetta Hough, James Outtz, Paul Ekman, and Joyce Hogan and Hogan Assessment Systems.

We very much appreciate the cooperation of Shauna Marshall, Gina Barnett and others at Hastings College of the Law that enabled us to include Hastings alumni in the final phase of the study. Many Berkeley faculty and staff including in particular, Herma Hill Kay, who as Dean initiated the committee that conceived this project, as well as Louise Epstein, Ed Tom, Mary Kelleher-Jones, and Patricia Donnelly and their staffs all contributed many hours to bring this project to fruition.

Our outstanding Psychology Graduate Student Researchers Eunice Chang and Jamie Clark were central throughout the years, especially in managing the computer and data aspects of the study. Joseph Plaster, our Administrative Assistant, played a large and critical role in administration of multiple aspects of the research, as did Philip Ryan. Law Research Assistants Sarah London, Chhunny Chhean, Nicole Thomas, and Bill Kidder played important roles.

We are also grateful to the literally thousands of Berkeley students and Berkeley and Hastings College of the Law alumni who volunteered time and energy to participate in this project as well as the 4000 practitioners who rated their professional performance.

TABLE OF CONTENTS

I. Introduction	11
II. Purpose of the Study	15
III. LSAT and UGPA as Predictors of Law School Effectiveness	15
IV. Non-Cognitive Predictors of Job Performance	18
A. Personality and Related Constructs (Big Five or FFM)	19
B. Self-Monitoring	21
C. Dispositional Optimism	21
D. Situational Judgment	22
E. Biographical Information Data	23
F. Emotion Recognition	23
V. Method	24
Overview of Research Chronology	24
A. Phase I	24
B. Phase II	27
1. "Off-the-Shelf" Tests	27
a. Hogan Personality Inventory	27
b. Hogan Development Survey	29
c. Motives, Values, Preferences Inventory	31
d. Optimism	32
e. Self-Monitoring	32
2. "Tailor-Made" Tests	33
a. Emotion Recognition Test	33
b. Situational Judgment Test	33
c. Biographical Information Data, or Biographical Inventory (BIO)	35
C. Phase III: Validation Research	36

1. Participants.....	36
2. Data Collection.....	37
3. Predictors.....	38
a. Academic performance data.....	38
b. HPI, HDS, and MVPI	38
c. OPT.....	38
d. SMS	38
e. ER.....	38
f. SJT and BIO	38
4. Performance Measures	40
VI. Results and Discussion.....	42
A. Alumni Sample	43
1. Description of Sample	43
2. Basic LSAC Predictors and Law School Measures	43
3. New Test Measures Identified/Developed for the Current Research	44
a. BIO and SJT.....	44
b. Hogan Personality Inventory (HPI)	45
c. Hogan Development Survey (HDS)	45
d. Motives, Values, Preferences Inventory (MVPI)	45
e. Other Potential Predictors – Self-Monitoring Scale (SMS), Optimism (OPT), and Emotion Recognition (ER)	45
4. Intercorrelations among New Predictors.....	46
5. Correlation among the LSAC Measures and the New Predictors	47
6. Ratings of Lawyer Effectiveness	47
7. Intercorrelations among the Performance Rating Sources	51
8. Prediction of Lawyering Effectiveness.....	52

a. LSAT, UGPA and Index Score as Predictors of Lawyering Performance	53
b. New Tests as Predictors of Lawyering Performance	55
c. Moderator Variables	59
d. Incremental Variance	60
9. Prediction of Lawyering Effectiveness in Law School	61
a. LSAC Components as Predictors of FYGPA in Law School	61
b. Tests Identified/Developed for this Project as Predictors of Law School Performance	63
c. Incremental Variance.....	64
d. Summary of Results: Predicting Law School Performance Measured by Grades	64
B. Student Sample.....	66
1. Description of Sample	66
2. Basic LSAC Predictors and Law School Measures	66
3. New Test Measures Identified/Developed for the Current Research	66
a. BIO and SJT.....	67
b. Hogan Personality Inventory (HPI)	67
c. Hogan Development Survey (HDS)	67
d. Motives, Values, Preferences Inventory (MVPI)	67
e. Other Potential Predictors – Self-Monitoring Scale (SMS), Optimism (OPT), and Emotion Recognition (ER).....	68
4. Intercorrelations among New Predictors.....	68
5. Correlation among the LSAC Measures and the New Predictors	69
6. Ratings of Student Effectiveness.....	69
7. Intercorrelations among the Performance Rating Sources	71

8. Prediction of Lawyering Effectiveness as Measured in Law School.....	72
a. LSAT, UGPA and Index Score as Predictors of Lawyering Performance	73
b. New Tests as Predictors of Lawyering Performance	74
c. Moderator Variables	77
d. Incremental Variance	77
9. Prediction of Effectiveness in Law School as Measured by Grades	77
a. LSAC Components as Predictors of FYGPA in Law School	77
b. Tests Identified/Developed for this Project as Predictors of Law School Performance	78
c. Incremental Variance.....	78
d. Summary of Results: Predicting Law School Performance Measured by Grades	79
VII. Conclusion	79
A. Basic Data, Further Research and Options for Use	79
B. Potential Uses and Benefits.....	81
C. Limitations.....	82
D. The Context Reviewed: Reasons to Add Professional Predictors.....	82
1. Recent Developments in Legal Education.....	83
a. Increased Applicant Pool.....	83
b. Inclusion of Women and Minorities.....	83
c. Controversy and Litigation	83
d. Rankings.....	84
e. The Place of Law Schools in Major Universities	84
2. These Changes Affect Admission Policies	85
3. Problems Associated with Over Emphasis on LSAT Scores	86

a. Misplaced precision.....	86
b. Selection bias.....	87
c. Professional and Academic Emphasis.....	87
VIII. References	91

LIST OF TABLES

ALUMNI SAMPLE

- Table 1: Descriptives for Alumni Sample
- Table 2: Breakdown of Background Info by School
- Table 3: Means for School-Based Performance Measures
- Table 4: Means for SJT and BIO
- Table 5: Means for Hogan Personality Inventory (HPI)
- Table 6: Means for Hogan Development Survey (HDS)
- Table 7: Means for Motives, Values, Preferences Inventory (MVPI)
- Table 8: Means for SMS, OPT, and ER
- Table 9: Intercorrelations of Predictors
- Table 10: Correlations of UGPA, LSAT, and Index
- Table 11: Correlations of Predictors with UGPA, LSAT, and Index
- Table 12: Means for Self Appraisals
- Table 13: Means for Peer Appraisals
- Table 14: Means for Supervisor Appraisals
- Table 15: Means for Other (Peer + Supervisor) Appraisals
- Table 16: Means for Average of All Appraisals
- Table 17: Intercorrelations of Raters
- Table 18: LSAT and GPA Correlations with Performance
- Table 19: Hogan Personality Inventory (HPI) Correlations with Performance
- Table 20: Hogan Development Survey (HDS) Correlations with Performance
- Table 21: Motives, Values, Preferences Inventory (MVPI) Correlations with
Performance
- Table 22: Other Test Correlations with Performance
- Table 23: Self Stepwise Regressions

Table 24: Peer Stepwise Regressions

Table 25: Supervisor Stepwise Regressions

Table 26: Other Stepwise Regressions

Table 27: Tally of Variables from Stepwise Regressions

Table 28: Correlation of UGPA, LSAT, and Index with 1LGPA

Table 29: Correlations of UGPA, LSAT, and Index with 1LGPA – Berkeley
Alumni

Table 30: Correlations of UGPA, LSAT, and Index with 1LGPA – Hastings
Alumni

Table 31: Correlations of All Tests with 1LGPA – All Alumni

Table 32: Correlations of All Tests with 1LGPA – Boalt Alumni

Table 33: Correlations of All Tests with 1LGPA – Hasting Alumni

STUDENT SAMPLE

Table 34: Descriptives for Alumni Sample

Table 35: Means for School-Based Performance Measures

Table 36: Means for SJT and BIO

Table 37: Means for Hogan Personality Inventory (HPI)

Table 38: Means for Hogan Development Survey (HDS)

Table 39: Means for Motives, Values, Preferences Inventory (MVPI)

Table 40: Means for SMS, OPT, and ER

Table 41: Intercorrelations of All Predictors, UGPA, LSAT, and Index

Table 42: Means for Self Appraisals

Table 43: Means for Peer Appraisals

Table 44: Means for Supervisor Appraisals

Table 45: Means for Other (Peer + Supervisor) Appraisals

Table 46: Means for Average of All Appraisals

Table 47: Intercorrelations of Raters

Table 48: LSAT and GPA Correlations with Performance

Table 49: Hogan Personality Inventory (HPI) Correlations with Performance

Table 50: Hogan Development Survey (HDS) Correlations with Performance

Table 51: Motives, Values, Preferences Inventory (MVPI) Correlations with
Performance

Table 52: Other Test Correlations with Performance

Table 53: Self Stepwise Regressions

Table 54: Peer Stepwise Regressions

Table 55: Supervisor Stepwise Regressions

Table 56: Other Stepwise Regressions

Table 57: Tally of Variables from Stepwise Regressions

Table 58: Correlations of All Predictors with 1LGPA

Table 59: Modeling Demographic Breakdown by Top Percentages

I. INTRODUCTION

The role of standardized testing in all levels of education has been increasingly controversial over recent years.¹ Debate about how America's institutions of higher and graduate education should make admission decisions has been especially acute. How should they define merit and qualification? What is "fair" allocation of scarce educational resources? How important is achieving racially and ethnically diverse classes, and how do we define "diversity?" Should deprivation of economic and educational opportunity be considered? The Supreme Court's decision in *Grutter* (*Grutter v. Bollinger*, 2003) provided some answers, offering a temporary Constitutional reprieve for limited consideration of race in admissions. Three states (California, Michigan, and Washington) have since adopted constitutional strictures on affirmative action and others will vote soon (Morain & Ricardi, 2008). Institutions of higher education continue to struggle to find ways to achieve equity and excellence.

In law schools, this challenge is particularly intense. Educational institutions naturally seek to admit those with the strongest academic skills. Several Ivy League schools collaborated in the late 1940's to create an entry test to aid in choosing among law school applicants (LaPiana, 2001). Design and administration of that test, the Law School Admission Test (LSAT), was soon turned over to the Law School Admission Council (LSAC) – a member organization of the American Bar Association of accredited law schools, currently numbering 195 U.S. law schools plus 16 in Canada. Today, the goal of the LSAT is to predict first year law school grades and, combined with the undergraduate grade point average (UGPA), it explains about 25% of the variance in those initial grades. Law school admission decisions are heavily influenced by scores

¹ The No Child Left Behind Act and litigation over teacher licensing tests are well known examples; *Association of Mexican-American Educators v. California*, 2000.

on the LSAT, and undergraduate grade point average (UGPA), combined into an Index Score. Law schools do employ other criteria, but on a largely ad hoc basis, with little other than intuitive confidence that these “other” factors are related to success in law school or professional practice. The lack of empirical evidence to support use of these added factors leads many to dismiss them as too subjective to be fair.

The advent of institutional rankings like those published by *US News & World Report* has further amplified an already strong emphasis on test scores. One factor in determining rank is each entering class’ median LSAT score. Because rankings translate into better applicants, easier fundraising, and improved faculty recruiting, schools are tempted to prioritize high test scores in admission decisions.

Other factors accelerate the trend. Applications to law schools have risen substantially. Greater stratification among schools, and equivalent “tiering” in the jobs and salaries their students can command, means that applicant pools are particularly large at highly ranked schools. (LSAC Volume Summary, 2008). As the ratio of applicants to admits rises, smaller differences in test scores and grades become decisive despite the decrease in meaningful differentiation. Large numbers of highly qualified applicants create pressure for streamlined and defensible decision-making. With admission to law school ever more prized, litigation always in the wings, and controversy among faculty, boards, alumni, and various publics about criteria, the apparent precision of numeric indicators of (apparently objective)² merit exert an understandable pull, especially if no alternative measures are available.

The LSAT has been the most effective method yet developed to predict first year law school grades, but it is narrow in method and in goal. By the LSAC’s own description, the test evaluates mainly reading, analytic and logic-based skills that are

² Although the LSAT is empirical and objective in design and scoring, how to define “merit” remains in dispute.

important for success in law school (Law School Admission Council, 1999). Linda Wightman (1997), former Vice President for Operations, Testing and Research at LSAC from 1988-1996, noted the LSAT's clearly defined, narrow focus, and acknowledged that using it to predict anything other than first year law grade point average (FYGPA) is less valid and less appropriate. The LSAC repeatedly advises schools that over-reliance on the test, particularly on fine distinctions in scores, is inappropriate both statistically and as a matter of policy.

Although the LSAT and UGPA have value as predictors of first year grades, they do not account for all the factors that influence grades or other performance in law school. Additionally, they make no effort to predict success in the profession. Nor do most schools' attempt to assess the needs of the profession and the citizenry when choosing among applicants.

Commentators have criticized legal education generally and the admission practices that reflect it, for over-emphasis on academic and cognitive competencies. The perceived failure of legal education to prepare students for professional practice has been a focus of many complaints (ABA "MacCrate Report," 1992; Edwards, 1992; Sullivan, Colby, Wegner, Bond, & Shulman, 2007). Clinical education, including both services and clinical policy research and intervention, ameliorates this problem, but typically lacks the resources, depth, continuity and integration to make up for inadequacies. The academic "tilt" of law school and its reflection in admissions criteria has consequences also for career paths. In higher-ranked schools, even students who have other career objectives feel they must accept jobs with corporate law firms when those are offered, in order to gain adequate training that only comparatively wealthy employers can afford to provide. Students who receive no such offer or who refuse to postpone any alternate career preferences they may have, face a difficult path to professional competency. Once a career path is begun, it is economically and

psychologically hard to abandon it. The chain of incentives does not encourage wider or more equal distribution of legal services.

Other critics object to admissions practices they see as reinforcing racial and class privilege (Society of American Law Teachers, 2003; Sturm & Guinier, 1996). Research consistently shows that affluent white students perform better on standardized tests, including the LSAT, than their less advantaged or minority peers (Schmidt & Hunter, 1981, 1998; Wightman, 1997). Heavy emphasis on LSAT scores in admission decisions substantially reduces the presence of African-American and Latino students in law school and the profession, as well as diminishing the prospects of inclusion for those from most non-elite families (Kidder, 2000; 2003).

Law schools select and educate students in academic programs, but as the LSAC recognized 20 plus years ago, law school admissions decisions also choose who will be the nation's lawyers (Raushenbush, 1986). In most states and for most lawyers, law school graduation is essential to licensure (National Conference of Bar Examiners, 2008). Law school also serves as advanced education for people entering a wide variety of other careers in which the problem-solving, communication, advocacy, and social/interpersonal skills are central to legal training matter. In law and related roles, law graduates reap significant rewards and wield significant influence in business, government, and education, as well as in the legal and justice systems; admissions decisions have deep and far-flung implications for almost every aspect of American society. To base admission to law school so heavily on LSAT scores is to choose academic skills (and only a subset of those) as the prime determinant of who gets into law and law-related careers that demand many competencies in addition to test taking, reading and reasoning skills. Moreover, it allocates the scarce resource of legal education, along with its ensuing influence and privilege, on the narrow basis of skills that are heavily linked to wealth and class.

II. PURPOSE OF THIS STUDY

The limits and downsides of current admission practices, as well as the logic of law schools' role as professional schools, urge that research move beyond attempts to predict grades in law school. Legal education needs tools that can reliably identify, assess and predict proto-competencies for professional effectiveness. This project sought to demonstrate methods to do that. Combined with LSAT and Index scores, such instruments would enable law schools to select better prospective lawyers based on both academic and professional competencies, thereby improving the profession's performance of its many tasks in society and the justice system.

Our task, then, was to produce a richer set of instruments and measurements for use by law school admission officers. We sought to: (1) empirically determine what *lawyers* view as factors important to effective lawyering; (2) broaden the criteria for measuring merit based on those factors; (3) use what *lawyers* see as more and less effective lawyer behavior to form measures of professional effectiveness; (4) develop test instruments to predict at the time of law school admission, the new array of lawyering competencies; (5) increase incrementally the amount of explainable variance in law school success; and (6) identify measures that on their own or in combination with LSAC measures (LSAT, UGPA, and Index) can predict lawyering effectiveness.

III. LSAT AND UGPA AS PREDICTORS OF LAW SCHOOL EFFECTIVENESS

As previously indicated, research on tests for admission to law school has focused on the LSAT and UGPA as predictors of first-year grades. Though individual schools also use various other factors, these two predictors are available for every school, are standardized, and are relatively non-subjective. The LSAC has conducted much of the research on the validity of these predictors by cumulating data across a number of law schools.

The validity of the LSAT and UGPA as predictors of the FYGPA criterion has consistent statistical support (Anthony, Harris, & Pashley, 1999; Dalessandro, Stilwell & Reese, 2005; Evans, 1984; Linn & Hastings, 1983; Schrader, 1977; Wightman, 1993). A relatively recent analysis (2001-2003 data) shows that the combination of LSAT and UGPA correlates approximately .47 with FYGPA in law school (Dalessandro et al., 2005), which explains approximately 22% of the variance, leaving 78% unexplained. By itself, the LSAT correlates .35 with FYGPA, while UGPA alone correlates approximately .28. These results are based on data from LSAC's annual validity studies provided by 165 law schools over two periods in 2003 and 2004. Because the data derive from a restricted sample (i.e., those who are actually selected for law school), they underestimate the validity of the combination. Dalessandro et al. (2005) demonstrate that correlation coefficients are higher when the data are based on an unselected sample. They also show that results vary by law school (in a range from .00 to .60 for the year 2003), and that, in general, the LSAT is a better predictor of FYGPA than is UGPA. Finally, results cross-validate, indicating that the regression equation for the combination of predictors is a useful model for predicting FYGPA for law school applicants. The Dalessandro et al. (2005) study replicates earlier findings (c.f., Anthony et al., 1999; Linn & Hastings, 1983; Powers, 1982; Wightman, 1993). Stilwell, Dalessandro, and Reese (2003) reported similar analyses for the LSAT/UGPA combination for the 2001-2002 period. These results show a multiple correlation of .49 for LSAT and UGPA as predictors of FYGPA as well as patterns similar to those reported by Dalessandro et al. (2005) with regard to type of sample and variability among law schools.

Much earlier reviews have also supported the use of the LSAT as a predictor of FYGPA (Evans, 1984; Schrader, 1977), demonstrating the consistency over time. A comprehensive meta-analysis was reported in 1983 by Linn and Hastings (1983). Their

analysis of data from 154 law schools showed a multiple correlation of .46. An important finding from this research was that the variability in correlation coefficients among the different law schools could be explained by the statistical artifacts of sampling error (differences in study sample sizes) and range restriction, and that correcting for these artifacts yielded estimates of validity in the .5 to .6 range.

Norton, Suto and Reese (2006) examined the differential validity of the LSAT and UGPA combination for different ethnic groups (African-American, Asian American, Latino, and white law students) in 2002, 2003, and 2004 entering law school classes. Using data from 183 law schools, with FYGPA as the criterion, they showed that the LSAT is not differentially valid for the groups studied. Furthermore, the differential validity results found similar patterns to those reported for other cognitive ability tests used in employment settings (c.f., Schmidt & Hunter, 1981). That is, when the regression equation for a combined group (minority and non-minority) is used to make predictions of academic success, the equation tends to over-predict minority students' performance. These findings replicate those of earlier studies (Anthony & Liu, 2000; Stilwell & Pashley, 2003; Wightman & Muller, 1990). Norton et al. (2006) also conclude that although the combination of LSAT and UGPA results in the most accurate prediction of FYGPA, none of the regression equations would systematically exclude African-Americans, Latinos, or Asian Americans.

Wightman (1997) contrasted projected outcomes of admission policies incorporating affirmative action with use of the weighted combination as the main determinant of admission. Based on data from 1990-1991, she concluded that sole reliance on LSAT and UGPA would result in systematic exclusion of minorities from law school classes. She does not suggest abandonment of the weighted LSAT/UGPA combination, but recommends that other, additional predictors be sought. We agree.

In sum, research data consistently show that the combination of the LSAT and

UGPA, and each separately, are valid predictors of the FYGPA criterion. But, these predictors are limited by use of the FYGPA as the only criterion. The strength of the obtained correlation between LSAT and FYGPA reflects, in part, that both measure the same abilities; that is precisely its aim.

As demonstrated by the literature and discussion above, the LSAT and UGPA have value as predictors, but they do not attempt to account for all the factors that contribute to law school grades or to broader performance in law school, and even less for success in lawyering. Wightman (1997) notes the LSAT's clearly defined, narrow focus, and states that using it to predict criteria other than FYGPA is both less valid and less appropriate. Wightman (2000) also argues that new assessments are needed to focus on other constructs that are not represented by the LSAT, that these new assessments should focus on the diverse abilities and skills needed to perform in school.

One attempt to use criteria other than FYGPA is found in Diaz, Glass, Ankkoff, and Tanofsky-Kraff (2001). The key aspect for their research proposal, however, was the identification and use of predictors such as state-anxiety, reactions to tests, measures of anxiety, and other non-cognitive predictors. These researchers used performance in examination milestones (grade for a contracts course and an oral argument rating) as criteria, but found that LSAT was not predictive of either. Results were not promising in that these indicators were not predictive of the criteria used in the study.

IV. NON-COGNITIVE PREDICTORS OF JOB PERFORMANCE

Effective lawyering, like effectiveness in any professional career, draws upon many dimensions of human intelligence -- in the wider sense of that term. As traditionally used, the category "cognitive" mainly encompasses academic and test-taking capability, especially verbal and numeric knowledge and reasoning. Overwhelming evidence shows that cognitive ability in this sense is a predictor of job

performance (Schmidt, 2002; Sackett, Schmitt, Ellingson, & Kabin, 2001). However, other – traditionally “non-cognitive” predictors like personality, interpersonal skills, and practical judgment – have been identified and found to be valid predictors of performance.

As noted above, a major concern with standardized cognitive tests such as the LSAT is the mean difference in performance between ethnic groups, particularly African-Americans. Generally, African-Americans score about one standard deviation below whites on measures of general cognitive ability, though this standardized mean score difference is reduced in high complexity jobs (Hough, Oswald, & Ployhart, 2001). Latinos also tend to score lower than whites on these types of measures, while Asians tend to score slightly higher than whites (Hough et al., 2001). Employment personnel research attempts to minimize disadvantage to members of racial, gender, or ethnic groups by combining valid non-cognitive measures of performance with traditional cognitive ability tests in the selection process (Hunter & Hunter, 1984; Ones, Viswesvaran & Schmidt, 1993; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Generally, subgroup differences are smaller, or non-existent, on non-cognitive measures such as biodata and personality inventories. Moreover, some evidence suggests that validity can be increased in some jobs if appropriate additional predictors, such as measures of social skills or personality traits, are used in combination with cognitive ability measures (Guion, 1987; Hunter & Hunter, 1984; Schmitt et al., 1997).

A. Personality and Related Constructs (Big 5 or FFM)

Strong evidence suggests that certain dimensions of personality are useful in predicting job performance. Generally, personality can be described as those traits, states, and moods that are stable and enduring over time, and distinguish one person from another (Allport, 1937). A broader conceptualization can encompass a person's strengths, weaknesses, values, and motivations (Hogan, Hogan, & Warrenfeltz, 2007).

Personality is important to performance because the degree to which an individual's personality fits with the requirements of a job or the values of an organization will have a significant impact on both success and satisfaction (e.g., Chatman, 1981; Kristof, 1996).

Much of the research on personality has embraced the Five-Factor Model (FFM; Big 5) of personality, which categorizes personality into five broad factors: Extraversion, Agreeableness, Conscientiousness, Neuroticism (Emotional Stability), and Openness to Experience (Saucier & Goldberg, 1998; Wiggins & Trapnell, 1997). Early meta-analytic work (Barrick & Mount, 1991; Salgado, 1997; Tett, Jackson, & Rothstein, 1991) found that personality holds some utility for predicting job performance. Barrick and Mount (1991) reviewed 117 studies and found personality-performance correlations ranging from .03 to .13 among the five facets of the FFM, with Conscientiousness being the strongest and most consistent predictor of job performance across professions. More recently, Hurtz and Donovan (2000) re-examined the relationship between personality and job performance. Hurtz and Donovan found that the mean sample-size weighted correlations ranged from .04 to .14 across dimensions, again with Conscientiousness having the highest validity. Conscientiousness is a general predictor of job performance, and other Big 5 traits predict job performance in specific types of jobs. In other words, different jobs call for different personality profiles and strengths (Hogan, Hogan, & Roberts, 1996). Importantly, at the Big 5 level, there are few ethnic differences.

Reported correlations (between Big 5 factors and job performance) of .13 and .14 are relatively small, but these findings mainly reflect bivariate relationships with criteria. Multiple regressions for the Big 5 as a *set* show correlations ranging from .1 to .45 (predicting individual teamwork having a .37 validity correlation, for example). Furthermore, individual Big 5 personality traits have specific facet-level characteristics that may have different relationships with job performance, obscuring the relationship of the higher-order personality dimensions to job performance. For example, the facet of

Conscientiousness has more specific facets such as Order, Impulsivity, Cognitive Structure, Play, Endurance and Achievement. If some facets correlate negatively and others positively, aggregate overall correlation of Conscientiousness to job performance may appear deceptively small (Tett, Steele, & Beauregard, 2003).

In addition to stable personality characteristics that predict an individual's ability to get along with others and achieve occupational goals, other behavioral tendencies can "derail" a person's career success (Bentz, 1985; Hogan & Hogan, 1997). These counterproductive behaviors are relatively stable over time and can predict various performance risks (Hogan & Hogan, 1997). Also central to a person's identity are interests, motives, and goals which can be important in predicting a person's success and satisfaction in a job. Generally, people prefer to work with others who share similar values and within compatible organizational cultures (Chatman, 1991; Kristof, 1996).

B. Self-Monitoring

Aspects of another trait, Self-Monitoring, seemed potentially salient to effective lawyering because of lawyers' distinctive professional responsibilities for representing clients ("role morality"). Self-monitoring of expressive behavior and self-presentation differ from the Big 5 traits. Individuals who are high self-monitors are good at learning what is socially appropriate in new situations, have good self-control of their emotional expression (facial and verbal), and can effectively use this ability to create the impressions they seek to create (Snyder, 1974; Snyder & Gangestad, 1986). Some evidence suggests that high self-monitors have more career mobility and success (Kilduff & Day, 1994), as well as higher ratings of job performance (Caldwell & O'Reilly, 1982; Caligiuri & Day, 2000).

C. Dispositional Optimism

Dispositional Optimism refers to a generalized tendency to expect positive and favorable outcomes in the future; conversely, pessimism refers to a tendency to expect

that bad things will happen in the future (Carver & Scheier, 1981). Optimism has been recognized as a fundamental component of individual adaptability because of its relationship with stress resilience and coping (Hobfoll 2002; Scheier & Carver, 1992).

Optimists are more confident and persistent when confronting any challenge, while pessimists are more doubtful and hesitant (Carver & Scheier, 2002). Some research indicates that optimism predicts lower levels of stress and depression for students making their transitions to the first year of college (Aspinwall & Taylor, 1992; Brissette, Scheier, & Carver, 2002). In terms of job performance, evidence suggests that Dispositional Optimism has a unique impact on both self-reported job performance and organizational performance appraisals (Youssef & Luthans, 2007). Optimism may be a valuable resource for law students and lawyers who face great time demands, high job insecurity, and poor organizational climate (Heinz, Hull, & Harter, 1999; Goldhaber, 1999; Makikangas & Kinnunen, 2003; Scheier & Carver, 1985; Schiltz, 1999; Xanathopoulou, Bakker, Demerouti, & Schaufeli, 2007;).

D. Situational Judgment

Understanding how potential students and employees would react in critical situations is important to predicting performance in the complex, conflict-ridden, and pressured roles of lawyers. Situational Judgment Tests (SJTs) present descriptions of hypothetical job-related scenarios, asking them to pick how they would handle the situation from a list of possible responses. The hypothetical situations are often developed by asking professionals in the field what critical situations they encounter in their jobs (Weekly & Ployhart, 2005).

SJTs are often paired with traditional cognitive ability tests in applicant selection settings because they have significant criterion-related validity and possess incremental validity beyond cognitive ability and personality measures (Chan & Schmitt, 2002; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). For example, Chan and

Schmitt (2002) found that the SJT had a significant .30 correlation with overall job performance and had an incremental validity of .21 for overall performance. Weekly and Ployhart (2005) found that the SJT was correlated .21 with overall job performance, and had a significant incremental validity of .18, above and beyond a cognitive ability test and a FFM personality inventory. Another important reason for the popularity of SJTs is that there are fewer ethnic differences than traditional cognitive ability tests (Clevenger, Pereira, Wiechtmann, Schmitt, & Harvey-Schmidt, 2001)

SJT are also drawing interest to predict student performance (judged by mission statement and educational objectives) in undergraduate schools (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). Oswald et al. (2004) showed that the SJT has validity above and beyond cognitive ability and personality for predicting college performance. Here, too, scores on the SJT show no significant score differences between different ethnicities.

E. Biographical Information Data

Past performance is often the best predictor of future performance. Biographical Information Data measures (BIO) offer structured and systematic methods for collecting and scoring information on an individual's background and experience (Mumford, 1994). Items vary both in the nature of the constructs measured (e.g., past attitudes; experiences) and in the type of response scale (e.g., frequency of behavior, amount, degree of agreement). Research has shown that BIO scales can predict both college GPA and job performance, and reflect fewer ethnic differences than standardized tests such as the S.A.T. (Oswald et al., 2004).

F. Emotion Recognition

Emotional intelligence, a currently popular organizational topic, targets the ability to regulate one's own emotions and perceive/understand others' emotions (Goleman, 1995). Some studies suggest that emotional intelligence predicts the performance of

students (Lam & Kirby, 2002) as well as job performance (Law, Wong, & Song, 2004; Slaski & Cartwright, 2002). Emotional intelligence could be important to lawyers who must manage interactions with clients, juries, judges and colleagues as well as “read and interpret” whether the communications between lawyers and others are being understood.

Questionnaires have typically been used to measure emotional intelligence through presentation of situations followed by choice of a response from among multiple choices (c.f., the measures used by Salovey, Mayer, Goldman, Turvey & Palfai, 1995). Ekman uses a different, less verbal approach (c.f., Ekman, 2004). Based on extensive research, Ekman developed a visual test to assess individuals’ speed and accuracy in recognizing various emotions on slides of faces.

V. METHOD

Overview of Research Chronology

Phase I identified a list of factors that practicing lawyers (including lawyers doing law-related jobs) as well as law faculty, law students, judges and clients viewed as important to effective lawyering performance. Our research also developed specific examples of more and less effective lawyering behavior into scales that could be used to evaluate the performance of a particular lawyer.

Phase II involved a survey for relevance, quality and range of tests available to predict job performance. We selected five “off the shelf” tests. We also wrote or substantially adapted three additional “tailor-made” tests to suit our purposes.

Phase III conducted validation strategies of the battery of tests to assess whether and which tests would validly predict lawyer effectiveness.

A. Phase I

To predict who will be an effective lawyer, we first needed to determine what comprises effective lawyering (Shultz & Zedeck, 2003). We did this empirically,

conducting hundreds of individual and then group interviews with lawyers, law faculty, law students, judges and some clients, asking questions like “If you were looking for a lawyer for an important matter for yourself, what qualities would you most look for? What kind of lawyer do you want to teach or be?” In a rolling process we gradually selected, added to, subtracted from, defined and redefined identified factors, seeking rough consensus through successive discussions with lawyers in many fields, settings and career stages. We distilled a list of 26 Effectiveness Factors important in the eyes of these varied constituencies, to being an effective lawyer.

Next, again using rolling interviews and focus groups, we asked for specific examples of more and less effective behaviors (“What behavior would tell you that a particular lawyer had or lacked effectiveness?”) on each of the 26 Factors. When we had gathered hundreds of examples, we asked Berkeley alumni (by email) to rate the examples on a 1-5 scale, according to how effective they thought the stated behavior was as an illustration of a given Effectiveness Factor. After receiving more than 2000 responses, and based on the mean and standard deviations of the responses, we developed scales of less effective to more effective behavior for each of the 26 Effectiveness Factors.

The products of Phase I were: (1) a comprehensive list of 26 Effectiveness Factors that are important to effective lawyering; (2) a set of 715 behavioral examples of performance that illustrate poor to excellent performance on each of the 26 factors; and (3) 26 flexible Behaviorally Anchored Rating Scales (BARS; Smith & Kendall, 1963) based on the 715 examples, with which an evaluator could assess the effectiveness of any given practicing lawyer.

The research team grouped the 26 Effectiveness Factors into eight categories thought to represent meaningful clusters, but the clusters were not statistically derived. For the purposes of this project; we are interested in maintaining the 26 factors as

distinct areas of measurement and not in the underlying construct of performance.

(Note: The groupings and labels are purely for convenience to facilitate subsequent discussion; other researchers are likely to derive different groupings and labels. The key point for this report is that we present the actual statistical results for each of the 26 Effectiveness Factors, so that a researcher can recombine the individual factors into any meaningful subset for that researcher's purpose.)

List of 26 Effectiveness Factors with 8 Umbrella Categories

1 : Intellectual & Cognitive
• Analysis and Reasoning • Creativity/Innovation • Problem Solving • Practical Judgment
2: Research & Information Gathering
• Researching the Law • Fact Finding • Questioning and Interviewing
3: Communications
• Influencing and Advocating • Writing • Speaking • Listening
4: Planning and Organizing
• Strategic Planning • Organizing and Managing One's Own Work • Organizing and Managing Others (Staff/Colleagues)
5: Conflict Resolution
• Negotiation Skills • Able to See the World Through the Eyes of Others

6: Client & Business Relations - Entrepreneurship
• Networking and Business Development • Providing Advice & Counsel & Building Relationships with Clients
7: Working with Others
• Developing Relationships within the Legal Profession • Evaluation, Development, and Mentoring
8: Character
• Passion and Engagement • Diligence • Integrity/Honesty • Stress Management • Community Involvement and Service • Self-Development

B. Phase II

Having identified job performance effectiveness factors and measurement scales for effective lawyering in Phase I, we sought in Phase II to identify tests that would predict actual lawyering performance. After reviewing a wide range of available tests, and after convening test development experts to advise us, we chose five “off the shelf” tests and wrote or substantially adapted three “tailor-made” tests.

1. “Off the Shelf” Tests

a. Hogan Personality Inventory (HPI; Hogan & Hogan, 2007) is a measure of normal personality based on the Five-Factor Model and is designed specifically for use with working adults. The HPI is composed of 206 true-false self-report items. Seven primary personality scales are scored on the HPI on the basis of Hogan and Hogan's (1991) reinterpretation of the five-factor model: Adjustment, Ambition, Sociability, Interpersonal Sensitivity, Prudence, Inquisitive, and Learning Approach. The main

difference between the HPI and the five-factor model is that it divides Extraversion into Adjustment and Ambition and divides Openness into Intellectance and Learning Approach. Items with similar themes are organized into a total of 40 subscales, and each subscale is scored on one of the seven primary scales. For example, the Adjustment scale contains eight themes including anxiety, guilt, complaints, moodiness, and irritability. Because the items in these sub-themes cluster together, they are referred to as Homogenous Item Composites (HICs).

Interpretation of HPI results is job-specific, with no formula for a “good” personality. High and low scores on scales are not necessarily better, and scores that lead to success in one job may hinder performance in another.

<u>Hogan Personality Inventory</u>	
Adjustment	Reflects the degree to which a person is steady in the face of pressure, or conversely, moody and self-critical (FFM: Emotional Stability).
Ambition	Evaluates the degree to which a person seems leader-like, status-seeking, and achievement-oriented (FFM: Extraversion).
Sociability	Assesses the degree to which a person needs and/ or enjoys social interaction (FFM: Extraversion).
Interpersonal Sensitivity	Reflects social sensitivity, tact, and perceptiveness (FFM: Agreeableness).
Prudence	Concerns self-control and conscientiousness (FFM: Conscientiousness).
Intellectance	Reflects the degree to which a person seems imaginative, adventurous, and analytical (FFM: Openness).
Learning Approach	Reflects the degree to which a person enjoys academic activities and values education as an end in itself (FFM: Openness).

b. Hogan Development Survey (HDS; Hogan & Hogan, 1997) assesses 11 performance risks that can interfere with a person's ability to build relationships and collaborate with others in a work atmosphere. The primary purpose of the HDS is to identify behavioral tendencies that could interrupt or "derail" a person's career success. The HDS has 168 items that comprise 11 primary scales: Excitable, Skeptical, Cautious, Reserved, Leisurely, Bold, Mischievous, Colorful, Imaginative, Diligent, and Dutiful. Because the scores on the HDS scales measure "negative" personality characteristics, higher scores on an HDS scale increase the chances that counterproductive work behaviors will be problematic for that specific performance risk. Across all types of jobs then, particularly in leadership positions, HDS scores and job performance are negatively correlated.

<u>Hogan Development Survey</u>	
Excitable	Reflects being initially enthusiastic about people or projects and then becoming disappointed with them. Result: seems to lack persistence.
Skeptical	Reflects being socially insightful, but cynical, mistrustful, and overly sensitive to criticism. Result: seems to lack trust.
Cautious	Reflects being overly worried about making mistakes and criticism. Result: seems resistant to change and risk adverse.
Reserved	Reflects being tough, remote, detached, and hard to reach. Result: seems to be a poor communicator.
Leisurely	Reflects being independent, ignoring others' requests, and becoming irritable if they persist. Result: seems stubborn, procrastinating, and uncooperative.
Bold	Reflects being entitled and having inflated views of one's worth. Result: seems unable to admit mistakes or share credit.
Mischievous	Reflects being charming, but manipulative and ingratiating. Result: seems to have trouble maintaining relationships and learning from experience.
Colorful	Reflects being dramatic, engaging, and attention-seeking. Result: seems preoccupied with being noticed, and may lack sustained focus.
Imaginative	Reflects thinking and acting in interesting, unusual, and even eccentric ways. Result: seems creative but often lacking good judgment.
Diligent	Reflects being conscientious, perfectionistic, and hard to please. Result: tends to disempower staff and subordinates.
Dutiful	Reflects being eager to please and reluctant to act independently. Result: tends to be pleasant and agreeable, but reluctant to support subordinates and co-workers.

c. Motives, Values, Preferences Inventory (MVPI; Hogan & Hogan, 1996)

evaluates the fit between an individual and the organizational culture, and directly assesses a person's motives. The MVPI has 200 items that comprise 10 primary scales: Aesthetic, Affiliation, Altruistic, Commerce, Hedonism, Power, Recognition, Science, Security, and Tradition. High scores (65th percentile and above) indicate those values and drivers that are most important to people. The relative degree of person-job fit can then be determined by comparing a person's higher scores on the MVPI with the values of the organization and profession. Unlike the HDS, high and low scores are not direct indicators of bad behaviors or bad work. Rather, they show the relative importance to employees of various values.

<u>Motives Values Preferences Inventory</u>	
Aesthetic	Concerns valuing creative and artistic self-expression. Interests are in quality, product look and feel, and attractive surroundings.
Affiliation	Concerns valuing frequent and varied social interaction. Interests are in social networking and feeling a sense of belonging to a group or organization.
Altruistic	Concerns valuing actively helping others, and improving society. Interests are in helping others, providing good customer service, and building a better workplace.
Commerce	Concerns valuing business activities, money, and financial gain. Interests are in earning money, realizing profits, finding business opportunities, and making investments.
Hedonism	Concerns valuing fun, good company, and good times. Interests are in pleasure, excitement, and variety.
Power	Concerns valuing competition, achievement, and being perceived as influential. Interests are in challenge, competition, and a lifestyle organized around worldly success.
Recognition	Concerns valuing fame, visibility, and publicity. Interests are in being known, recognized, visible, and famous.

Science	Concerns valuing ideas, technology, and rational problem solving. Interests are in new ideas, technology, and analytical approach to solving problems, and understanding how they work.
Security	Concerns valuing certainty, predictability, and risk free environments. Interests are in structure, order, predictability, and planning for the future.
Tradition	Concerns valuing similarity between the organization's and the employee's perspectives on tradition, history, and old-fashioned virtues.

d. Optimism (OPT) as measured by the *Revised Life Orientation Test* (LOT-R; Scheier, Carver, & Bridges, 1994), assesses generalized outcome expectancies, with higher scores indicating a more optimistic overall outlook on life (Scheier & Carver, 1985). The LOT-R consists of six items, three of which assess optimism and three reverse-scored items that measure pessimism, plus four filler items on 5-point Likert scales.

e. Self-Monitoring Scale (SMS; Snyder, 1974) is a 25-item true-false scale used to assess an individual's tendency to monitor and adapt self expression. We reworded five (of the 25) items to increase their relevance to lawyers; and included them as additional questions. These five items are:

- 1) It has always been easier for me to argue for things that I believe.
- 2) It has always been easy for me to speak convincingly about topics, even when I don't know much about them.
- 3) It has never been easy for me to fit in with different people and different situations.
- 4) As a child, I found it easy to convince others I was telling the truth, even when I wasn't.
- 5) I think I could convince someone I was telling the truth, even when I wasn't.

2. “Tailor-Made” Tests

We developed three new tests based on prototypes used in employment selection. These tests sought to predict which law applicants have and/or could develop the 26 competencies of effective lawyering. Because the tests would eventually be administered as part of law admissions processes,³ they could not rest on legal knowledge or lawyering experience but only on the more general factors. The “tailor-made” tests we developed for our study are:

a. Emotion Recognition Test (ER) was modeled after the Facial Action Coding System (FACS) developed by Paul Ekman (2004). We used stimuli (stock color photos of neutral and emotional facial expressions) generated in the laboratories of emotion research (e.g., Ekman, 2004, and Dacher Keltner, personal communication, 2006) to develop a 20 minute, computer test presenting faces of different people expressing one of ten emotions: Anger, Compassion, Contempt, Disgust, Embarrassment, Fear, Happiness, Sadness, Shame, and Surprise. Our ER instrument included two practice items and 76 test items. In each item, participants saw 1) a neutral facial expression, followed by 2) a very brief (1/6 second) change in expression reflecting one particular emotion, and 3) a return to the initial expression. Participants had five seconds to choose which of the 10 emotions appeared during the changed facial expression. Participants were instructed to respond based on their first impression – even if they did not think they observed a change, or were unsure of the emotion expressed. Groups of 19 faces appeared, with a 30-second pause between each group. Faces included individuals of various ethnic backgrounds and genders.

b. Situational Judgment Test (SJT) required multiple steps to construct. First,

³ Because this exploratory research was not longitudinal, we administered material to law students and graduates, but the ultimate goal is to develop a test battery for use with law school applicants in the admissions process.

the researchers individually and then collaboratively wrote approximately 200 hypothetical situations to reflect each of the 26 Effectiveness Factors. Sometimes, items from existing SJT measures (e.g., from Camara, personal communication, January 9, 2006; Motowidlo, personal communication, January 9, 2006) stimulated ideas for situations we could customize for lawyer effectiveness factors; however, we wrote many items as originals. For each item scenario, we developed 4-5 answer options representing a range of viable responses.

Second, we refined each item multiple times to ensure clear phrasing, elimination of ethnic/racial/gender biases, and balance among the Effectiveness Factors. We then pilot-tested the items with practicing lawyers to get feedback. Next, we again reviewed, revised and rebalanced the inventory of items, working first individually and then collectively. We also chose 2-5 Effectiveness Factors that we hypothesized would correlate with each item. In this way, we generated 72 SJT items. A single item example appears below. We determined that this particular example reflected competency in three areas: Influencing and Advocating, Developing Relationships, and Integrity:

You learn that a co-worker, Angela, who you helped train for the job, copied some confidential and proprietary information from the company's files. What would you do?

- a. Tell Angela what I learned and that she should destroy the information before she gets caught.*
- b. Anonymously report Angela to management.*
- c. Report Angela to management and after disciplinary action has been taken, tell Angela that I'm the one that did so.*
- d. Threaten to report Angela unless she destroys the information.*
- e. Do nothing*

Because our research participants were busy professionals, the time they could expend on our test battery was limited. We therefore created eight SJT test forms with 18 items each. Nine items on each test form overlapped with another test form (e.g., form #1 had items 1-18, form #2 included items 10-27, etc.).

c. Biographical Information Data, or Biographical Inventory (BIO) also required multiple steps to create. First, the researchers independently and collaboratively wrote approximately 200 BIO items designed to cover each of the 26 Effectiveness Factors. Again, although some items from existing BIO measures (e.g., from Camara, personal communication, January 9, 2006; Motowidlo, personal communication January 9, 2006) stimulated items we designed for lawyering performance, we wrote many items as originals. We developed 4-5 answer options to represent a range of viable responses to the given item.

Second, we refined each item multiple times to ensure clear phrasing, elimination of race or gender bias, and balanced representation of Effectiveness Factors. Once more, we pilot-tested the items with practicing lawyers. After the pilot work, we had an inventory of 80 BIO items. In order to verify that all Effectiveness Factors were tapped in the BIO items, the three researchers independently reviewed each item-factor link again, then together decided on the 2-5 Effectiveness Factors we hypothesized would be linked to each item.

An example of one BIO item is shown below. We determined that this particular example reflected competency in both Creativity and Problem Solving:

How many times in the past year were you able to think of a way of doing something that most others would not have thought of?

- a. *Almost never.*
- b. *Seldom.*
- c. *Sometimes.*

d. *Often.*

e. *Very frequently.*

Again, in light of participant time constraints, we created eight BIO test forms with 20 items in each. To increase sub-sample sizes, each BIO test form had 10 items that overlapped with another test form.

C. Phase III: Validation Research

In Phase III, we collected data to assess whether performance on our chosen predictor tests correlated with actual lawyering effectiveness (as assessed by ratings provided by the participant him or herself, as well as by the participants' supervisors and peers), and with law school performance. The BARS performance rating scales developed in Phase I (Shultz & Zedeck, 2003) enabled us to assess participants' current workplace performance. We also examined relationships between the LSAC measures (LSAT, UGPA, and Index), appraised lawyering performance, and law school grades. Alumni and students comprised the participant samples.

1. Participants⁴

15,750 people were invited via email and regular mail to participate in the research: 657 Berkeley then-enrolled second and third year students and all alumni from Berkeley or Hastings who graduated between 1973 and 2006 for whom the schools had contact information. Approximately 10,000 Berkeley, and 5,000 alumni from the Hastings College of the Law received emails. Additionally, to reach those with outdated email addresses, or to honor individuals' requests to receive communication by regular mail only, 7000 Berkeley and 4000 Hastings alumni were sent invitations through the U.S. postal service.

⁴ For reasons of practicality (funding, access, and available staff), we worked with only two law schools: Berkeley Law (Boalt) as our initial exemplar, participated in all phases of the research, and Hastings College of the Law alumni joined in the final validation stage.

2. Data collection

Test data were collected via a two-hour test battery online. Participants logged in to the test system with individual user name and password (included with their original invitation). After login, an informed consent form outlined the study and asked whether the individual agreed to participate. In compliance with the regulations of the University of California at Berkeley Committee for the Protection of Human Subjects, the consent form outlined the study procedure, as well as the risks and benefits associated with the study. Participants were informed that participation was voluntary, and responses would be kept confidential. If they agreed to participate after reading the consent form, they were asked to complete a demographic questionnaire covering their gender, age, ethnic background, law school, and category of law work (e.g., solo practice, small/medium/large firm, business, government, etc.).

We decided that no participant should be asked to spend more than two hours on the test battery. Accordingly, the computer randomly and evenly directed each participant to one of 40 different combinations of tests from the battery we created. Every participant's combination included the HPI and two to three of the other tests. One participant might take the HPI, HDS, BIO, and OPT. Another might do the HPI, MVPI, and SJT. And yet another participant might take the HPI, BIO, and ER. As a result, of course, the HPI had the largest sample size.

We administered the HPI to all participants as an inducement to participate. As a benefit of taking the research tests, participants could opt to receive a confidential interpretive report of their occupationally relevant strengths and shortcomings based on their responses to the HPI. As a further incentive, participants and performance raters (see below) could register for continuing education credit (MCLE) of from 1-3 units, two for lawyering skills and one for elimination of bias credit. The bias unit required that a participant read additional materials written by the researchers, to explain how

definitions of “qualification” can intensify or minimize racial adverse impact in selection processes. Those participants were also required to take a short true/false quiz after reading the materials about bias.

3. Predictors

a. Academic performance data: Based on participant and law school permission, we obtained LSAT scores, undergraduate GPA, and law school GPA from either the Law School Admission Council (LSAC) or files at the two law schools.

b. HPI, HDS, and MVPI: Hogan Assessment Systems provided scores for each scale measured on each of these three Inventories.

c. OPT: The LOT-R test has six items, three to assess optimism and three reverse-scored items to measure pessimism, plus four filler items, and uses a 5-point Likert scale. The overall OPT score was calculated by reverse-scoring the three pessimism scores, and summing responses to all 6 questions.

d. SMS: Our adapted version of the SMS consisted of 30 true-false questions. We scored responses by giving a “1” to answers keyed in the high self monitoring direction, and a “0” to those keyed in the low monitoring direction. Scores on all 30 items were summed to create a total SMS score, with higher scores indicating a greater tendency to self-monitoring.

e. ER: Our ER consisted of 72 items. We assigned a “1” to item responses choosing the correct emotion; and a “0” to other choices and to non-responses. ER scores were the sum of responses to all 72 items.

f. SJT and BIO: The critical issue for these tailor-made instruments is how to score these tests. One strategy is to develop a rational key, whereby “subject matter experts” identify the response items that are “best” for each situation. An alternative option is to use an empirical key (a version of “cross-validation”) by relying on responses from participants to determine the algorithm and scoring keys. We used the latter

method for both the SJT and BIO.

Because we created 8 SJT test forms, participants responded to only a small subset of the 72 SJT items. For example, 500 participants may have responded to Item #1 but only 400 to Item #2. To score each SJT item, we randomly selected 2/3 of the participants with a score on that particular item. We then used our hypotheses about which items would link to which Effectiveness Factor; suppose a given item was designed to represent the Effectiveness Factors of “Analysis and Reasoning” and “Negotiating.” We took the individuals’ average effectiveness scores (based on self, peer, and supervisory ratings) for these two factors to create a gross “performance score” for that participant (see next section on “Performance Measures”). We repeated the process for each item and each participant. Then, for all of those within the 2/3’s sample who had answered Item #1, we determined the relationship between their “performance scores” and their choice of answer option. We conducted analysis of variance, with the “performance score” as the dependent variable, and the response options of “a,” “b,” “c,” etc., as the independent variables and determined which items showed significant differences in responses. Those items that had significant relationships were retained for further consideration. For these retained items, we used responses chosen by better performers, then generated an algorithm to score all responses to each item from 0 to 4 points. On some items, two answer options received the same score. Each participant, therefore, received a score ranging from 0 to 4 on each SJT or BIO item.

When a key is developed and assessed on the same sample, a concern about scoring bias is raised. In this study, however, using the same sample for key development and subsequent validity assessment likely had minimal impact. Not all items were answered by all participants, so the 2/3 of the sample who responded to one item were not likely to be the same 2/3 who responded to another item. The result of

this process was that we retained 36 of the 72 items for the SJT and 62 of the 80 items for the BIO instrument. Participants received scores on SJT and BIO only for those items they answered and their scores on those items were averaged to create a mean SJT score and a mean BIO score.

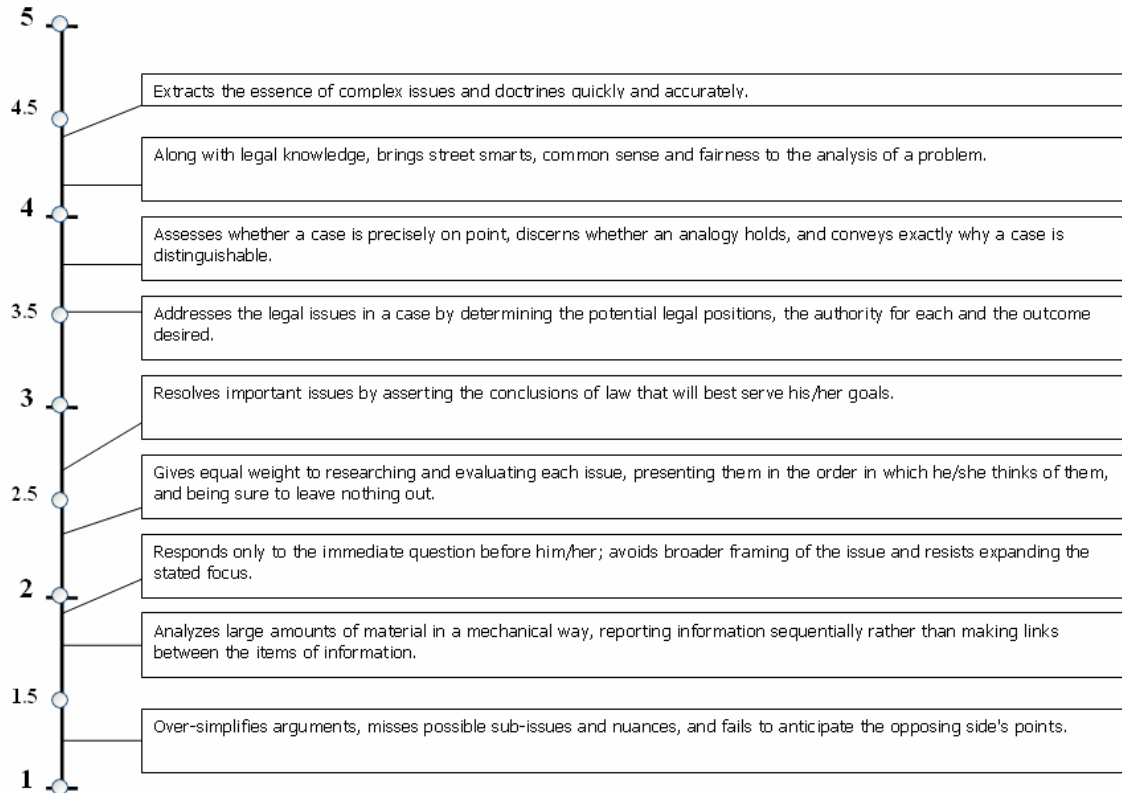
4. Performance Measures

To determine whether scores on the predictor tests related to on-the-job effectiveness, we needed ratings of each participant's work performance. We asked each participant to do a self-evaluation of his/her own lawyering effectiveness and to identify four other evaluators – two supervisors and two peers (with contact information) -- who could assess that participant's recent lawyering performance. These appraisals were collected online, after a participant completed his/her test items.

Participants, along with the supervisors and peers they named, were asked, via computer, to rate the performance of participants on the 26 Effectiveness Factors. BARS for each of the 26 Effectiveness Factors provided examples of different levels of performance on that Factor. Raters were asked to select the score (ranging from 1 to 5 in .5 increments) that best represented the participant's level of performance on that Factor. All raters were instructed to rate as many Factors as possible, and to skip Factors that were not relevant to the job or about which their knowledge was insufficient.

Raters were provided with detailed instructions about how to use the BARS. They were told that the particular examples on any scale might not *literally* apply to the participant-ratee's work or setting, but that the examples should be used by analogy to illustrate **levels** of performance from 1 (poor) through 5 (excellent). Below is an example:

Factor # 1: Analysis & Reasoning



As shown in the example above, the first factor scale was "Analysis and Reasoning." The lowest level example for this factor reads, "Analyzes large amounts of material in a mechanical way..." To rate the individual, appraisers were asked to read scales "from the bottom up", asking themselves, "Based on my observation and knowledge of this individual's performance, do I believe he or she would perform at the level of effectiveness reflected in this particular example?" When the raters came to a level of effectiveness that they believed the individual would **not** achieve, they were to

mark a value (in half-point increments) that represented the highest level that the rater believed the participant **would** achieve. Raters were asked to apply the standard based on the individual ratee's actual level of experience. For instance, the rater might rate someone with 10 years of experience a "2" on a given BARS scale, but evaluate that *same behavior* as a "3" for someone with less experience.

Note that the above BARS example is appropriate for the alumni sample, of practicing lawyers and law grads doing law-related work. For the student sample, the scales were adapted to be more consistent with student experiences, content, and context.

In sum, participants rated their own performance on a relevant subset of the 26 BARS, and also named two peers and two supervisors to rate the participant's performance on whichever of these same 26 BARS the rater could apply. We averaged the two peer ratings to create one Peer Appraisal score, and treated the two supervisor ratings similarly to create one Supervisor score. The Peer and Supervisor ratings on the 26 Effectiveness Factors were also averaged to create a unitary performance appraisal for each test participant which we labeled the combined "Other" rating. Additionally, all three rater perspectives were also averaged to create an "All" performance appraisal. Thus we had 5 different performance evaluation results: Self, Averaged Peer, Averaged Supervisor, Other (average of peers and supervisors), and "All" (average of all peers and supervisors and self).

VI. RESULTS AND DISCUSSION⁵

The overarching goal of the research project was to conduct an exploratory study on specific samples to determine whether new types of law school admission tests, and/or batteries of these tests, have the potential to predict actual lawyering performance. If the results yielded positive outcomes and trends, then we would urge

⁵ Cells shown in yellow in any table represent a statistically significant result.

additional research on a larger, more representative sample. This section of the Report presents and summarizes results from multiple rater perspectives (e.g., Self, Peer, and Supervisor evaluations) and for multiple subgroups (e.g., Berkeley vs. Hastings, males vs. females, etc.) to provide different views of the data. Similar patterns of results independent of group breakdowns would support conducting further research in a larger, broader sample.

A. ALUMNI SAMPLE

1. Description of Sample

Table 1 presents the background characteristics of the sample of 1148 participants. Overall, the sample was composed of mainly Berkeley (64.3%), female (56.8%), Caucasian (68.5%) practicing attorneys, with the largest number in large firm (16.6%) or government (13.7%) practice. All areas of expertise were represented with the most frequent specialization being litigation/advocacy (29.1%).

Table 2 presents the background characteristics by law school attended – Berkeley or Hastings. Divided by school, the pattern of findings was similar to the overall profile except that whereas Berkeley alumni worked in large firms or government positions, Hastings alumni worked in solo or small firm practice or government positions.

2. Basic LSAC Predictors and Law School Measures

Two statistical issues affect the data presented below:

(1) The sample included graduates from a 33 year period, which meant that the raw LSAT scores were derived from different versions of the test and different scoring templates (score ranges from 10-48, 120-180, and 200-800). Accordingly, we standardized scores (z-scores) within the templates to determine a common standardized score across the sample.

(2) Over the course of time covered in the study the two schools varied their weighting of the components of the Index Score. We weighted LSAT and UGPA

information 50-50 throughout.

Table 3 presents averages and standard deviations for the components used by LSAC -- LSAT, UGPA, and the Index Score (a school-specific composite of the previous two) -- by total sample, law school, gender, and race/ethnicity. Berkeley participants had higher UGPA and LSAT scores, males showed no pattern of practical differences compared to females, but Caucasians and Asians had a pattern of higher test scores compared to African-Americans and Hispanic participants. These results are consistent with findings reported in other research (Norton et al., 2006).

3. New Test Measures Identified/Developed for the Current Research

This section presents results for the different types of predictors identified or developed specifically for this research project. We selected those tests to broaden the types of abilities measured, with the goal of predicting success in law school and professional performance. Of particular interest was whether any patterns of differences among the subgroups emerged, which might suggest a particular measure might yield statistically and practically different results for some subgroups (e.g., males vs. females). We examined the latter, practical differences by determining effect size measures [e.g., d (difference between subgroup means/standard deviation) statistics, or r^2 that reflects amount of variance explained by subgroup membership].

a. BIO and SJT

Table 4 presents results for the BIO and SJT tests, which were tailor-made based on our Phase I research that derived Effectiveness Factors important for lawyering (Shultz & Zedeck, 2003). Average scores on the BIO test yielded similar findings for both Hastings and Berkeley alumni, for females and males, and for Caucasians and Asian/Pacific Islanders. African-Americans scored highest on the BIO, and Hispanics scored lowest, although the differences among the four groups are not statistically significant. Table 4 also shows no real differences among the subgroups for

the SJT, except that Hispanics scored higher than other ethnic groups ($r^2 = .01$). In general, the results show no practical differences for SJT and BIO based on gender and ethnicity, a finding consistent with the literature (Clevenger et al., 2001) for these types of tests.

b. Hogan Personality Inventory (HPI)

Table 5 shows the HPI results for the various subgroups. Across the seven scale scores, the only modest patterns to emerge were that females generally scored more positively on three dimensions (Adjustment, Sociability, and Intellectance) than males (r^2 's of .03 or less) and that Caucasians scored somewhat higher on Learning Approach than did Hispanics and African-Americans (r^2 's of .01 and .04, respectively).

c. Hogan Development Survey (HDS)

Table 6 presents the means and standard deviations for the 11 scale scores on the HDS. In general, no clear pattern of differences emerged except that African-Americans were significantly lower than Caucasians on the Cautious scale ($r^2 = .03$) and higher than Caucasians on the Mischievous, Imaginative, and Diligent scales (r^2 's of .02 or less). Males scored higher than females on Reserved ($r^2 = .04$) and Mischievous ($r^2 = .01$) while females scored higher on Diligence ($r^2 = .03$).

d. Motives, Values, Preferences Inventory (MVPI)

Table 7 presents the means and standard deviations for the 10 scales of the MVPI. Again, patterns of differences existed in only a few instances. Males scored somewhat higher than females on the scales of Commercial ($r^2 = .06$), Power ($r^2 = .03$), Recognition ($r^2 = .02$), and Scientific ($r^2 = .02$). African-Americans scored lower than Caucasians on Scientific ($r^2 = .02$).

e. Other Potential Predictors – Self-Monitoring Scale (SMS), Optimism (OPT), and Emotion Recognition (ER)

Table 8 presents the means and standard deviations for three tests chosen

because of their potential to measure specific components of performance: SMS, OPT, and ER. As with the other tests developed/identified for this project, in general, there were no clear or practical patterns of differences. Two noteworthy differences were that females scored better than males on the ER ($r^2 = .03$), and African-Americans scored higher than Caucasians on OPT ($r^2 = .01$).

CONCLUSION #1: In general, race and gender subgroup performance did not substantially differ on the new predictors. There were some exceptions that might be re-assessed in a larger study.

4. Intercorrelations among New Predictors

Given that there were no consistent patterns in mean differences based on race/ethnicity or gender among the tests, examination and presentation of intercorrelation results are provided for the total sample only.

Table 9 presents the intercorrelations among all of the new predictors studied in the research project. For HPI, the strength of the intercorrelations among the seven scales ranged from .00 to .44; for the HDS, the intercorrelations among the 11 scales ranged from .00 to .49; for MVPI, the strength of the intercorrelations among the 10 scales ranged from .03 to .55. For each test, the range of correlations includes negative correlations. In general, these correlations suggested that, overall, the three measures – HPI, HDS, and MVPI – were measuring relatively different characteristics.

Some tests yielded a single score. For BIO, the correlations between it and the other predictors ranged from .00 to .39; for SJT, the correlations between it and other predictors ranged from .01 to .21; for SMS the correlations ranged from .00 to .50; for OPT, the correlations ranged from .02 to .54; and for ER, the correlations ranged from .00 to .13. Each of the correlational patterns contained some negative correlations.

The above pattern of results suggests that the different, potentially new predictors were measuring abilities and characteristics that are relatively independent of

each other. This further suggests that one or more of the subscores or tests might correlate with aspects of performance that were *not* being tapped by the components used by LSAC (LSAT, UGPA, and Index Score). Support for this view is found, in part, by examining the relationships between these LSAC component measures and the new predictors (presented and discussed in the next section).

CONCLUSION #2: Results showed that new predictor tests were, for the most part, measuring characteristics that were independent of one another.

5. Correlation among the LSAC Measures and the New Predictors

Table 10 shows the correlations among the three components used by the LSAC— LSAT, UGPA, and Index Score. The intercorrelations among these three measures ranged from .20 (between LSAT and UGPA) to .78 (for the relations between the components and the Index) for the total sample. Table 11 shows that these same predictors had correlations that ranged from .00 to .37 with the new predictors studied in this research. However, approximately 74% of the correlations were below .10; also, a number of the correlations were negative.

The pattern of correlations among the three traditional and the new predictors suggests some degree of independence. The lack of overlap in the existing and new measures suggests that different traits and abilities were being measured, and that the tests predicted different aspects of performance. Ultimately, a test battery using different test measures could explain significant incremental variance above and beyond that which is explained by any single test.

CONCLUSION #3: The new predictor tests showed some degree of independence between the traits and abilities that they, as compared to LSAT, UGPA and Index Score, measure.

6. Ratings of Lawyer Effectiveness

Tables 12 - 16 present the means, standard deviations, and sample sizes for the

performance effectiveness measures as gathered from different sources – Self, Peer, and Supervisor –, displayed by total sample, school, gender, and race/ethnicity. Table 12 focuses on Self-Appraisals, Table 13 on Peer Appraisals, and Table 14 on Supervisor Appraisals. Given the three independent sources of evaluations, it is feasible to combine some of the sources into meaningful groups for further analysis. Consequently, an “Other” group was formed by averaging the ratings for a participant based on the Peer and Supervisor evaluations (see Table 15). Another combination averaged all the performance ratings from all sources for a participant, forming an evaluation score we named “All” (see Table 16). [Note that for the Peer and Supervisor ratings, the measure used in analysis could represent the appraisal of a single evaluator (if only one was named by the participant or if only one of two named responded) or the average of two evaluators (if both named members of a rater subgroup responded)].

Table 12, the research participants’ Self-Appraisals, shows relatively few practical differences and no consistent patterns such that, for example, one subgroup is always higher than another subgroup. Where there were significant differences, Berkeley participants tended to evaluate themselves more highly than Hastings participants on three Effectiveness Factors (Analysis and Reasoning; Writing; and Passion and Engagement) , males more highly on seven dimensions (Creativity/Innovation; Problem Solving; Influencing and Advocating; Speaking; Strategic Planning; Negotiation Skills; and Integrity), females more highly on five dimensions (Listening; Organizing and Managing One’s Own Work; Organizing and Managing Others; Advising Clients; and Evaluation, Development, and Mentoring) – all r^2 ’s .01 or less), and Hispanics more highly on 11 dimensions when compared to Caucasians (r^2 ’s .02 or less).

Table 13 presents the appraisal results provided by Peers named by participants. Peers rated Berkeley participants more highly on 13 of the 26 Effectiveness Factors.

Peers rated males more highly on six dimensions (Analysis and Reasoning; Problem Solving; Researching the Law; Fact Finding; Writing; Speaking; and Negotiation Skills) while peers rated females more highly on four dimensions (Evaluation, Development, and Mentoring; Passion and Engagement; Diligence; and Community Involvement and Service). Compared to Caucasians, African-Americans rated higher on Able to See the World Through the Eyes of Others and Developing Relationships; Hispanics rated lower on Diligence and Asian Americans rated lower on Creativity, Strategic Planning, and Passion and Engagement. Nearly all the r^2 's are .01 or less.

Table 14 presents the appraisal evaluation results from the perspective of Supervisors named by the participants. When rated by Supervisors, Berkeley participants as well as males received higher evaluations on 14 dimensions (Berkeley higher on Analysis and Reasoning; Creativity/Innovation; Problem Solving; Researching the Law; Fact Finding; Influencing and Advocating; Writing; Listening; Strategic Planning; Negotiation Skills; Passion and Engagement; Diligence; Integrity; and Community Service; Males higher on Analysis and Reasoning; Creativity/Innovation; Problem Solving; Researching the Law; Influencing and Advocating; Negotiation Skills; and Integrity). Supervisors rated Caucasians more highly on six dimensions than Asian Americans (Analysis and Reasoning; Creativity/Innovation; Problem Solving; Researching the Law; Influencing and Advocating; and Integrity). The r^2 's are generally at .01 or less.

Table 15 presents results of appraisals by the "Other" category (combined averaged ratings by Peers and Supervisors). The pattern is similar to others: Berkeley alumni and males were evaluated more highly, and race/ethnic differences appeared between Caucasians and Asian Americans.

Table 16 presents evaluation results from the "All" category (combined averaged ratings from all three sources -- Self, Peer, and Supervisor). The patterns are the same

as above.

The first row in each table (12 -16) shows the average rating for participants across the 26 dimensions. This can be viewed as a “Global Effectiveness” rating. On this measure, Hispanics rated themselves higher than Caucasians. Peers and Supervisors rated Berkeley participants higher than Hastings participants. Examining both “Other” and “All” ratings, Berkeley participants were rated higher than Hastings alumni and Caucasians were rated higher than Asian Americans.

Examination of these five tables (12 - 16) reveals the following: (1) within a rating subgroup, no evidence suggests that one particular participant subgroup was consistently higher or lower on any of the Effectiveness Factors; (2) across rating subgroups, Berkeley alumni and males tended to have higher performance ratings; males are generally evaluated as higher on Negotiations Skills and Passion and Engagement regardless of who provided the rating; and (3) differences among race/ethnicity subgroups show no consistent pattern across the 26 Effectiveness Factors. The tables also show that, in general, (1) the averages were rather high, at the 4-point mark on a 5-point scale, and (2) the Self Appraisals were lower (on 24 of the 26 Effectiveness Factors) than the Peer or Supervisor ratings.

The patterns presented in this section are somewhat equivocal in providing guidance about breakdowns of data for subsequent future analyses. However, because the study was exploratory, and the patterns do not overwhelmingly suggest that subgroups can be ignored or that analysis on the total sample would be adequate by itself, we will present our subsequent analyses and results based on various demographic (e.g. men vs. women), background and experience (Berkeley vs. Hastings) subgroups, and rater perspectives (e.g., Self vs. Other). If the same predictors prove useful across the entire set of analyses, the convergence in results would provide very strong support for future research.

CONCLUSION #4: Performance ratings from all sources showed some differences in participant-subgroups (school, race, gender, etc.) performance.

RECOMMENDATION #1: Larger follow-up research should retain subgroups analysis at least through the next stage of study.

7. Intercorrelations among the Performance Rating Sources

Table 17 summarizes the agreement among the various rater groups. The table shows that the correlations between Peers and Self across the 26 dimensions ranged from .08 to .38; for Self and Supervisor the correlations ranged from .09 to .41; for Peer and Supervisor, the correlations ranged from .12 to .34; for Self and Other, the correlations ranged from .12 to .43; for Self and All, the correlations ranged from .62 to .84; for Peer and Other, the correlations ranged from .80 to .91; for Peer and All, the correlations ranged from .72 to .83; for Supervisor and Other, the correlations ranged from .81 to .90; for Supervisor and All, the correlations ranged from .72 to .81; and, for All and Other, the correlations ranged from .88 to .92.

The results shown in Table 17 indicate that when each of the rater groups is compared to a different rater group (e.g., Peers to Supervisors), agreement among the different rater perspectives is low to modest. This pattern of results may imply that analysis of the validity of the tests should be studied separately for each rater group. An alternative would be to consider the reliability of the ratings and to determine whether combinations (e.g., the “Other” or “All”) present a more reliable estimate of performance than appraisal from any single rater source.

Barrett (2008) undertook an analysis of the project’s ratings within rater groups. He concludes that averaging the two Peer ratings for each performance dimension was reasonable; likewise, it was reasonable to average the two Supervisor ratings for each performance dimension. Additional analyses indicated that sufficient similarity existed between averaged Supervisor and averaged Peer ratings to average the two averages

to yield an “Other” rating viewpoint.

The results imply that Self ratings were somewhat different than the Other evaluations, again suggesting that perhaps subsequent validity analyses should be conducted separately for these two rating perspectives. An alternative would be to focus on an “All” rater group which could provide the most information on the participant’s performance.

CONCLUSION #5: Agreement between different rater groups was low to modest. Peer and Supervisor ratings were similar to one another although Self Appraisals differed.

RECOMMENDATION #2: Subsequent research should conduct validity analysis separately for appraisals by Self and Other (Peers and Supervisors).

8. Prediction of Lawyering Effectiveness

The key goal for this research project was to determine whether particular test types could predict on-the-job lawyering effectiveness (simple correlations) as well as how well a battery of tests (multiple correlations) could predict lawyering effectiveness. This section focuses on these questions.

This study generated a considerable amount of data on predictors and Effectiveness Factors. A legitimate presentation question is whether all possible analyses should be reported in this document.⁶ One option would be to present validity results for each predictor, for each participant subgroup, and for each performance rating subgroup. This would produce an enormous report overloaded with tables and output. After examining all of the analyses and results, however, it is our view that the conclusions regardless of, for example, the participant subgroup studied or the rater group used to evaluate performance, would lead us to the same recommendation, which is that no different recommendations would be made when all results are considered.

⁶ All statistical output and tables are available from the researchers.

This interpretation might suggest another option, that of presenting subsequent results from, for example, from only one rater perspective. Our research purpose, however, was to study whether one or more of our new tests and methods could be useful for predicting lawyering performance. The current research was exploratory. Therefore, any consistently positive result within or between participant or rater subgroups (e.g., Berkeley vs. Hastings) makes the picture clearer and provides stronger support for the need for more research.

Given the above assessment, we have opted to present results from different frameworks. We do this in the hope that we will persuade the reader, as we ourselves are persuaded, that the convergences in results from different rater and/or participant subgroups, enhance support for the overall conclusion that further validation research should definitely be conducted.

Given the analyses above, this section of the report on the validity of the predictors for lawyering performance will sometimes present results for the total sample, and in other instances, only for Berkeley or only for Hastings. The demonstrations will also differ according to rater groups – presenting each of the rater subgroups for some analyses, only the “All” combination for some, and only the “Other” rating combination for still different instances.

a. LSAT, UGPA and Index Score as Predictors of Lawyering Performance

Table 18 shows the zero-order correlations between the LSAT, UGPA, and Index score and each of the 26 Effectiveness Factors, as well as the average Global Effectiveness score (average of the 26 dimensions). The table shows data for each of the five rater subgroup perspectives – Self, Peer, Supervisor, Other, and All. Table 18 shows that the LSAT scores, taking into account more than one performance rating group, correlated with six Effectiveness Factors -- Analysis and Reasoning; Researching the Law; Writing; Networking; Integrity; and Community Service). For the first three of

these six performance factors, the correlations ranged from .08 to .16. This was to be expected, given that the LSAT specifically seeks to measure Analysis and Reasoning and has likely relationships or overlap with Writing and Researching. For the performance factors of Networking and Community Service, the correlations are negative and range from -.10 to -.14. This suggests that high scorers on the LSAT did not do well on two lawyer Effectiveness Factors, Networking and Community Service. Networking and Community Service both require interaction with others. It may be that those who scored highly on the LSAT were not viewed by the raters as devoting attention to Networking and Community Service or lacked the necessary skills. For Integrity, the correlation with the LSAT score was negative when performance was rated by Self appraisal ($r = -.09$), but positive when evaluated by other rater subgroups (ranging from .06 to .12). This discrepancy between rater subgroups may reflect distinctive characteristics of Integrity. Integrity, or especially its lack, concerns matters usually kept private, secret from others or the public. People may more likely try to manipulate others' perceptions of Integrity. A Self rating reflected inner knowledge of one's own secrets, but others likely appraised Integrity on the basis of outward manifestations. Characteristics peculiar to Integrity may, then, explain the discrepancies between Self and Peer or Supervisor ratings.

UGPA results showed fewer correlations than LSAT scores. In general, UGPA correlated most with Writing (r 's ranged from .09 to .12), with Managing One's Own Work (r 's range from .09 to .10), and with Diligence (r 's ranged from .09 to .11). Differences in correlations of UGPA and LSAT may reflect that the LSAT is a one day test, but UGPA depends on persistence and the ability to manage and apply oneself over four or more years. Diligence, time spent, and management of work could more readily substitute for "smartness" in the UGPA measure than in obtaining higher LSAT scores.

In sum, the LSAT, UGPA and Index were predictive of only a few of the Effectiveness Factors, mainly ones that overlapped with the LSAT's measurement targets. For example, the LSAT aims to evaluate analysis and reasoning and it correlated with performance appraisals of participants' "Analysis and Reasoning." When LSAT score was mixed with the UGPA in the Index score, correlations emerged with Managing Own Work, as might be expected. The LSAT and UGPA were not intended to predict lawyering effectiveness, but given our hypothesis that broader lawyering skills should be added to academic criteria when selecting the best qualified law school applicants, the important finding for us was that, for the most part, they did not.

CONCLUSION #6: The LSAT, UGPA and Index Scores were not particularly useful for predicting lawyer performance on the large majority of the 26 Effectiveness Factors identified in our research.

RECOMMENDATION #3: Because traditional indicators (LSAT, UGPA, and the Index Score) did not predict performance as a lawyer, other predictors focusing on prediction of post-graduate performance should be explored.

b. New Tests as Predictors of Lawyering Performance

Table 19 shows the correlations between the HPI scales and the 26 Effectiveness Factors as well as with the average Global performance across all 26 dimensions. In summarizing the results of this and subsequent tables that present correlations between particular predictors and performance measures, we will focus on correlations that were significant within at least three rater subgroups. Examination of the correlations in Table 19 shows that three of the HPI scales – Adjustment (r's ranged from .10's to .30's), Ambition (r's ranged from .10's to high .30's), and Interpersonal Sensitivity (r's ranged from .10's to high .20's) – correlated with 20, 20, and 14 Effectiveness Factors, respectively. The strongest correlations for Adjustment were with Stress Management ($r = .37$), Developing Relationships ($r = .19$), Seeing the World

Through the Eyes of Others ($r = .18$), Negotiations ($r = .17$) and Listening ($r = .15$). Ambition correlated relatively strongly with almost all of the set of 26 Effectiveness Factors noted above (e.g., with Creativity; Problem Solving; Practical Judgment; Fact Finding; Questioning and Interviewing; Influence and Advocating; Speaking; Strategic Planning; Negotiation; Networking; Passion; etc.). The strongest correlations for Interpersonal Sensitivity were with Questioning and Interviewing ($r = .22$), Listening ($r = .20$), Seeing the World through the Eyes of Others ($r = .24$), Developing Relationships ($r = .33$), and Evaluation, Developing, and Mentoring ($r = .22$). The most highly correlated HPI scales (Adjustment, Ambition, and Interpersonal Sensitivity) do NOT show a pattern of significant correlations with four of our lawyer Effectiveness Factors -- Analysis and Reasoning, Researching the Law, Writing, and Diligence. The first three of these were, however, correlated with the HPI scale of Learning Approach (r 's about $.10$), while the fourth is tapped by the HPI Prudence scale. Thus, five HPI scales would have potential to contribute to the prediction of many of the Effectiveness Factors.

Table 20 shows the correlations between the HDS scales and the 26 Effectiveness Factors. Only one HDS scale showed a consistent pattern and some promise, "Excitable," which concerns being overly enthusiastic about people/projects and then becoming disappointed with them. This scale correlated with 19 of the 26 Effectiveness Factors (r 's ranged from $-.10$ to high $-.30$'s). However, it also correlated $-.72$ with Adjustment on the HPI, suggesting that Adjustment and Excitable were measuring similar, albeit reversed, characteristics. The other scale that showed some promise is "Reserved." This scale correlated with seven Effectiveness Factors (r 's ranged from $-.10$'s to mid $-.20$'s). "Reserved" reflects being remote, detached and lacking awareness of feelings of others. Its correlation with our lawyer Effectiveness Factors such as Managing Others, Negotiation, Networking, Building Relationships, and Community Service are ones that we would expect it to predict -- awareness of others'

feelings was critical for efficacy in these performance areas. “Reserved,” however, was correlated (-.56) with Interpersonal Sensitivity, again suggesting overlap in what is being measured by the two scales. An issue for the future, then, is whether to concentrate on the two HDS scales of “Reserved” and “Excitable” or rely mainly on the HPI scales, which covered more of the Effectiveness Factors than did the two HDS scales.

Table 21 shows the correlations between the MVPI scales and the 26 Effectiveness Factors. Overall, this pattern of correlations was not as impressive as was the HPI, or even the HDS. The “Altruistic” scale correlated with five Effectiveness Factors (Creativity; Able to See the World Through the Eyes of Others; Passion; Integrity; and Community Service). Most of the correlations were in the .10s, but the MVPI scales of Affiliation and Altruistic correlated best with Community Service (r’s ranging from .16 to .42). The other MVPI scales that correlated with a small number of Effectiveness Factors (4) were Affiliation and Hedonistic (r’s ranged from .15’s to mid. 20’s). As with the HDS, the issue for the future would be whether to continue further research on the MVPI given its limited number of correlations with the lawyer Effectiveness Factors and its weak showing compared to the HPI. Though individual scales, such as Altruistic, correlate with some Effectiveness Factors (five), the one noteworthy pattern is with Community Service, where the correlations ranged from .16 to .42, depending on which rater subgroup was used.

Table 22 shows the correlations of the new tests (BIO, SJT, SMS, OPT, and ER) with the Effectiveness Factors. BIO scores showed correlations (in the .2’s and .3’s) with all Effectiveness Factors except Integrity and Stress Management. SJT scores showed correlations with all Effectiveness Factors other than Managing Others and Evaluation, Development, and Mentoring. The correlations were generally in the .10’s and low .20s. The impressive aspect of these results was (1) the large number of Effectiveness Factors that were predicted by both BIO and SJT tests, and (2) the fact

that the correlations were generally higher, though moderately, than the ones found for the instances in which the LSAT did have some relationship with a small subset of Effectiveness Factors.

Table 22 also shows correlations of the SMS, OPT, and ER with the Effectiveness Factors. The one predictor of these three that showed the most potential in this study was OPT which correlated positively with 13 of the Effectiveness Factors (*r*'s ranged from .10's to .20). Most notable are the correlations with Stress Management, Speaking, Networking, and Questioning and Interviewing. Because OPT correlated in the high .4's with the HPI Adjustment and Ambition scales, use of OPT and HPI might be duplicative. The other two measures, SMS and ER, did not show results that would suggest continuing pursuit.⁷

Overall, some of the new predictors identified or developed for this specific research project display results that argue for additional research. They do so because (1) they correlated with areas that were not predicted by the LSAT or UGPA alone, or as combined in the Index score, (2) their correlations were generally higher than ones obtained for the LSAT, UGPA and Index predictors, and (3) based on the literature and current research, these predictors tended to yield few if any mean differences by race/ethnicity and gender (Clevenger et al., 2001).

CONCLUSION #7: New predictors developed for this project correlated at a higher level with factors not well predicted by the LSAT, UGPA, or Index Score and showed little race or gender subgroup difference in results.

RECOMMENDATION #4: Based on the pattern of findings across different participant subgroups and from different rater subgroups, we recommend that future research focus on new predictors, especially HPI, BIO, SJT, and OPT.

⁷ It is possible that an SMS type test re-written specifically for law performance would show better results and that an ER test allowing longer time intervals, fewer emotions to choose from and more consistent face photographs would have improved results on those two tests.

c. Moderator Variables

Given Recommendation #4 (focus on four new tests), we examined through moderated regression, the relationship between the separate predictors of HPI (7 scales), BIO, SJT, and OPT and each of the 26 Effectiveness Factors of lawyer performance in order to determine whether there was differential validity for any participant subgroup. We conducted step-wise moderated regression by entering the predictor in the first step, the potential moderator (e.g., gender) in the second step, and the interaction of the two in the third step. Significant incremental variance on the third step provides evidence of differential validity. This process required over 1,000 analyses (26 dimensions x 5 rater groups x 10 predictor measures) for *each* potential moderator for the subset of HPI, BIO, SJT, and OPT. Results indicated few instances of significant incremental variance on the third step of the analysis. Where significant increments existed, the amount of variance was negligible (approximately 1% incremental variance).

With regard to the strength of these results, we note that reliance on graduates of only two schools limits the generalizability of the findings. In addition, small sample sizes for the ethnic groups limit the opportunity (low statistical power) to identify significant race/ethnic differences. Nevertheless, the results were consistent with the literature which has found few if any differences between genders and race/ethnic groups on the types of predictors studied in this project, however future research with larger more representative samples should examine moderated variable effects.

CONCLUSION #8: Consistent with the literature, results from analysis of this sample showed that the new predictors studied here showed no practical differences among race or gender subgroups.

RECOMMENDATION #5: Because predictors of professional effectiveness are important, further research on these new types of tests should be vigorously pursued.

d. Incremental Variance

In this research, we sought to determine whether a battery of tests could be formed that would explain variance in ratings of actual lawyer performance. Initially, we intended to examine whether the tests identified/developed for this project would yield incremental variance (hierarchical moderated regression) above what the LSAT, UGPA and Index score explain. However, given that the LSAT, UGPA and Index scores did not demonstrate many correlations with the lawyering Effectiveness Factors, we undertook step-wise regression analysis in which the order of entry into the analysis was determined by statistical relationships among the predictors and their correlations with the performance evaluations.

Table 23 presents the results of an analysis in which the LSAT, Index, HPI scales, BIO, SJT, and OPT were allowed to enter in a step-wise multiple regression to determine which combination, if any, of the predictors could explain Self-Appraisals of performance on each of the 26 Effectiveness Factors as well as on the Global average of all 26 dimensions. This table shows only those results that yielded significant incremental variance. An overview of this table indicates that a combination of two tests, and in some instances three tests, can produce multiple correlations with the Effectiveness Factors (and Global average for performance) in the range of the mid .20's to the high .30's. Tables 24 through 26 repeat the analytic strategy for Peer, Supervisor, and Other rater subgroups respectively. The multiple correlations when Peers provided the ratings (Table 24) ranged from about .15 to the mid .20's; for Supervisor ratings (Table 25) the multiple correlations ranged from the high .10's to the low .20's; while the multiple correlations for the Other ratings (Table 26) ranged from the mid .20's to the low .30's.

Table 27 summarizes the results of these step-wise multiple regression analyses. The main conclusion to be drawn from this summary is that SJT and BIO plus one or

more of the HPI scales (e.g., Ambition, or Interpersonal Sensitivity) yielded composites that predicted lawyering performance on a relatively large number of Effectiveness Factors. On the other hand, the LSAT and the Index did not demonstrate much value along with or in addition to the other potential tests in predicting lawyering performance. Taken as whole, the data suggest that SJT, BIO, HPI, and OPT have the best potential to predict lawyer performance effectiveness..

CONCLUSION #9: In multiple regression analysis, SJT, BIO and several HPI scales predicted many dimensions of Lawyering Effectiveness, whereas the LSAT and Index Score did not.

RECOMMENDATION #6: Further research should focus strongly on OPT, SJT, BIO and HPI predictors of professional performance.

9. Prediction of Lawyering Effectiveness in Law School

We obtained UGPA, LSAT, and Index scores on alumni participants as well as their FYGPA in law school. These data provided us with the opportunity to replicate, in part, the relationship between UGPA, LSAT, and Index with FYGPA. We could assess whether the current limited samples (Berkeley and Hastings alumni) generated findings consistent with the larger studies conducted under the auspices of the LSAC.

a. LSAC Components as Predictors of FYGPA in Law School

Traditionally, the LSAT, UGPA, and Index scores have been used to predict FYGPA in law schools (c.f., Dalessandro et al., 2005; Stilwell et al., 2003; Wightman, 1993). Results show that the correlations between LSAT, UGPA, and Index Score with FYGPA in law schools are .35, .28, and .25, respectively. The weighted composite of LSAT and UGPA (the Index score) yields a correlation of .49 with FYGPA.

In determining the validity of the LSAT, UGPA, and Index Score for predicting FYGPA for the current sample, we first examined the relationship between the law school components and FYGPA as moderated by gender, ethnicity, and school. These

analyses were similar to those performed to assess the relationship between the LSAT, UGPA, and Index Score and our Effectiveness Factors, particularly for whether there were gender or race/ethnicity differences. As above, these analyses were undertaken by conducting step-wise moderated regressions where the predictor (e.g., LSAT score) was entered in the first step, the potential moderator (e.g., gender) was entered on the second step, and the interaction of the two was entered on the third step. Evidence of differential validity is determined when there is significant incremental variance on the third step. Results of these analyses indicated no consistent practical differences. Accordingly, subsequent analyses and results are reported mainly for the total sample, with the exception that school differences are maintained to provide information for the participants in this project

Table 28 shows the correlations between the traditional admissions predictors and FYGPA for the total sample. These results showed reasonable consistency with other research findings: the LSAT correlated .42 with FYGPA, UGPA correlated .21 with FYGPA, and the Index correlated .42. The results for Berkeley grads (see Table 29) were: LSAT correlated .49 with FYGPA, UGPA correlated .24 with FYGPA, and the Index correlated .48 with FYGPA. For the Hastings sample (see Table 30), the results were: LSAT correlated .43 with FYGPA, UGPA correlated .31 with FYGPA, and the Index correlated .52 with FYGPA. The differences between the schools were not practically significant.

The composite of LSAT and UGPA yielded a multiple correlation of .44 for the total sample, .50 for Berkeley, and .53 for Hastings. In interpreting the data presented in this section, note that this project's results are based on only two schools (Hastings and Berkeley), whose percentages of ethnic minority participants is relatively small compared to white participants, while the LSAC's research findings are based on 165 schools. In addition, the Index Score for this project was formed on the basis of a 50-50 weighting

while the LSAC-reported research used indices as specified by the individual schools. Regardless of these caveats, it is reasonable to conclude that the current research replicates the finding that the LSAT, UGPA, and Index scores are useful predictors of FYGPA in law schools.

CONCLUSION #10: Results essentially replicated the validity of the LSAT, UGPA, and Index Score for predicting FYGPA in our sample.

b. Tests Identified/Developed for this Project as Predictors of Law School Performance

Tables 31-33 show the correlations between each of the predictors identified or developed for this research and the FYGPA in law school, for the total sample, for Berkeley and for Hastings, respectively. Addressing the scale scores on the HPI, the results for the total sample (Table 31) showed that five (Adjustment, Ambition, Sociability, Interpersonal Sensitivity, and Learning Approach) of the seven scales correlated significantly with FYGPA, with the correlations ranging from .07 to .21. Of particular interest, however, was the finding that all but Learning Approach had negative correlations with FYGPA. The patterns of relationships were somewhat similar when Berkeley (see Table 32) and Hastings (see Table 33) samples were differentiated. For Berkeley, the pattern was the same as the total sample except that Adjustment was not correlated with FYGPA; for Hastings, Ambition and Sociability were not related to FYGPA.

For the 11 HDS scales, only Excitable correlated (.16) with FYGPA for the total sample, Imaginative correlated (-.15) for the Berkeley sample and Excitable correlated (.24) for the Hastings sample. For the 10 MVPI scales, only Hedonistic correlated (-.15) with FYGPA for the total sample, while Altruistic (-.20) and Hedonistic (-.19) correlated for Berkeley and Altruistic (.21) correlated for Hastings.

Of the remaining new scales, BIO correlated -.09, OPT correlated -.08, and ER

correlated .08 with FYGPA for the total sample. The SJT and SMS did not significantly correlate with FYGPA; the ER did correlate (.12) only for the Berkeley sample. None of these correlated for the Hastings sample.

CONCLUSION #11: Our new predictors showed few significant correlations with FYGPA, and of those that did exist, many (especially HPI scales, OPT and BIO) were negative.

c. Incremental Variance

In an attempt to determine whether any of the new predictors would yield incremental validity beyond that which is obtained by the LSAT alone or by the Index Score alone, hierarchical multiple regression was undertaken where tests were entered on a second step after first entering LSAT (separate analyses for Index Score). Significant increases in R^2 would suggest the potential for formation of a predictive battery of tests. Results indicated that five of the HPI scales (Adjustment, Ambition, Sociability, Interpersonal Sensitivity, and Learning Approach) and the OPT scales yielded significant increments above the LSAT, but each amounted to less than 1%.

For the Index, the results showed the same pattern, with the exception that Adjustment did *not* add incremental variance. For those results in which there were significant increments, the amounts were 1% or less.

CONCLUSION #12: New predictors added only slight (1%) incremental validity to LSAT and Index Score prediction of law school performance.

d. Summary of Results: Predicting Law School Performance Measured by Grades

In sum, the results for the LSAT, UGPA, and Index appeared to be good predictors of the traditional FYGPA measure of performance for the current samples. The new tests did not show consistently strong relationships on their own, or in conjunction with the LSAT, UGPA, and Index measures.

Questions can, however, be raised about the way performance is measured in the LSAC correlational research studies. The LSAT is a “paper-and-pencil” test that basically measures analytic and logical reasoning, along with reading. Law examinations dominate grades, especially in first year. Typically, exams require students to read fact patterns, identify and analyze legal issues, assemble evidence and arguments and sometimes to assess implications -- essentially the same abilities measured by the LSAT. The correlations between LSAT and FYGPA are, therefore, both unsurprising and somewhat circular, especially given that the LSAT was designed by asking teachers of first year law courses to identify which skills would yield high grades. The same narrow band of cognitive test-taking skills dominates in part because professors find it difficult for a number of reasons (e.g., limited to no opportunity to observe) to assess other types of abilities such as those identified in the lawyer Effectiveness Factors (e.g., Negotiations, Interviewing, Integrity, Problem Solving, Creativity, etc.) and those capabilities are not typically reflected in law school exams or grades (FYGPA). Given the criterion used, it would also be surprising if new non-cognitive measures could improve on the LSAT/UGPA Index correlation with first year grades.

CONCLUSION #13: For our sample, LSAT, UGPA and Index Score were good predictors of FYGPA. Our new predictors did not show strong correlation on their own nor did they add incremental validity to the LSAT, UGPA and Index Score.

RECOMMENDATION #7: Further research on new tests’ prediction of FYGPA is less important than research on predictors of professional effectiveness.

B. STUDENT SAMPLE

1. Description of Sample

Table 34 presents the background characteristics of the student sample, which included only Berkeley students; female (66%) and Caucasians (62.1%) who were in their second year of law school (2L; 59.1%). The most desired future employment setting for these students was a large firm (32.5%) with the intended types of practice being varied: civil rights (10.3%), corporate and business transactions (9.9%), intellectual property (8.9%), or criminal and criminal procedure (7.9%).

Note that the sample sizes of the minority student subgroups – African-American (7), Hispanic (19), Asian/Pacific Islander (32), and Native American (16) – were quite small compared to the number of Caucasians (126), making statistical comparisons of each ethnic group to Caucasians statistically problematic.⁸ Accordingly, subsequent race/ethnicity comparisons should be treated with extreme caution.

2. Basic LSAC Predictors and Law School Measures

Table 35 presents average and standard deviation results for the LSAT, UGPA, and Index Score by total sample, gender, and race/ethnicity. Table 35 shows no real difference in UGPA by gender or race; slightly higher LSAT scores for males and for Caucasians compared to African-Americans and Hispanics. The Index, which is a standardized score, and composed of the LSAT and UGPA scores, shows obviously similar patterns to the UGPA and LSAT. Differences in FYGPA are almost non-existent.⁹

3. New Test Measures Identified/Developed for the Current Research

This section presents the results for the different types of predictors identified or

⁸ The number of minority matriculants decreased substantially in the wake of California's 1996 passage of Proposition 209 banning affirmative action in state education and contracting. Recently, minority admit numbers at Berkeley have risen but not to pre-Proposition 209 levels.

⁹ Berkeley's law grading system contributes to obscuring LGPA differences. 60% of each class must receive a grade of Pass under the mandatory grading curve.

developed for this research project in order to broaden the types of abilities measured.

a. BIO and SJT

Table 36 presents the results for two of the tests that were tailor-made for this research (BIO and SJT). Consistent with the literature on SJT and BIO formats (c.f., Clevenger et al., 2001), results shows no real differences as a function of gender or race/ethnicity.

b. Hogan Personality Inventory (HPI)

Table 37 shows the HPI results for the various subgroups. Across the seven scale scores provided by the HPI, the only modest patterns to emerge were that females generally scored more positively on two dimensions (Interpersonal Sensitivity and Prudence) while males scored more highly on Intellectance (r^2 's = .07) and also, Caucasians scored somewhat higher on Adjustment than did African-Americans, though lower on Sociability (differences were not significant). Overall, there were no consistent patterns that suggested that one gender or race/ethnic group dominated on the set of characteristics measured by the HPI.

c. Hogan Development Survey (HDS)

Table 38 presents the means and standard deviations for the 11 scale scores on the HDS (because not all students took this test, the sample size for analysis was small relative to those who took the HPI; N = 63 total). In general, no clear pattern of differences between genders emerged; the sample sizes for race/ethnic comparisons were too small to provide meaningful interpretation.

d. Motives, Values, Preferences Inventory (MVPI)

Table 39 presents the means and standard deviations for the 10 scales of the MVPI (total sample of N = 60). Again, no patterns of consistent gender differences were seen in the results; Hispanics had lower scores on four of the scales (Affiliation, Altruistic, Power, and Tradition).

e. Other Potential Predictors – Self-Monitoring Scale (SMS), Optimism (OPT), and Emotion Recognition (ER)

Table 40 presents the means and standard deviations for three tests chosen because of their potential to measure specific components of performance: SMS, OPT, and ER. As with the other tests developed/identified for this project, no clear or practical patterns of differences appeared.

CONCLUSION #14: None of the new predictors showed consistent patterns of difference by race or gender.

4. Intercorrelations among New Predictors

Table 41 presents the intercorrelations among the LSAT, UGPA, Index and all of the new predictors studied in the research project. The correlation between UGPA and LSAT is effectively .00, due in large part to the minimal variance among the students in their UGPA. For HPI, the intercorrelations among the seven scales ranged from .00 to .55; for the HDS, the intercorrelations among the 11 scales ranged from .01 to .58; for MVPI, the intercorrelations among the 10 scales ranged from .00 to .52. For each test, the range of correlations included negative correlations.

Some tests yielded a single score. For BIO, the correlations between it and the other predictors ranged from .00 to .51; for SJT, the correlations between it and other predictors ranged from .00 to .48; for SMS the correlations ranged from .03 to .61; for OPT, the correlations ranged from .03 to .58; and for ER, the correlations ranged from .00 to .24. Each of the correlational patterns contained some negative correlations.

The results suggest that, as for the alumni sample, the new potential predictors measure abilities and characteristics that were relatively independent of each other. This further suggests that one or more of the subscores or tests might correlate with aspects of performance that are *not* being tapped by the LSAT, UGPA, and Index Scores. Support for this view is found, in part, by examining the relationships between

the LSAT, UGPA, Index measures and the new predictors (presented and discussed in the next section).

CONCLUSION #15: Intercorrelations among the new predictors showed that they measured traits that were relatively independent of one another.

5. Correlation among the LSAC Measures and the New Predictors

Table 41 also shows that the LSAT, UGPA, Index scores had correlations that ranged from .00 to .36 with the new predictors studied in the project although approximately 50% were below .10; also, a number of the correlations were negative.

The pattern of correlations found among the predictors suggests some degree of independence. Non-overlap in the measures demonstrates that different traits and abilities were being measured, creating the potential for the tests to predict different aspects of performance. This means that a varied test battery might be constructed that could explain significant incremental variance and measure a diverse set of abilities.

CONCLUSION #16: Traditional LSAT, UGPA, and Index Score on the one hand, and new predictors on the other, measured abilities independent of one another.

6. Ratings of Student Effectiveness

Tables 42 – 46 present the means, standard deviations, and sample sizes for the measures of performance effectiveness as gathered from different sources – Self Appraisals, Peer, and Supervisor Appraisals as well as Other and All composites --, displayed by total sample, gender, and race/ethnicity. Again, the total sample for the performance effectiveness measures was approximately 150; small samples of minority students suggest that race/ethnic group comparisons should be interpreted with caution. Table 42, the Self Appraisals, shows females to be higher on 15 Effectiveness Factors while males were higher on 11 Effectiveness Factors, but only three were significantly different -- males were higher on Analysis and Reasoning, and Problem Solving;

females higher on Community Service. However, the differences were not very large; the biggest difference was on the Community Service dimension where females were .38 points higher than males ($r^2 = .06$).

Table 43 presents the evaluation results provided by the Peers. These results show that females were rated more highly on 16 of the dimensions, but only one is a significant difference; the largest difference is on the Evaluation, Development, and Mentoring performance dimension (.22 of a point; $r^2 = .04$).

Table 44 presents the evaluation results provided by the Supervisors. Here we see a change in pattern. Males received higher evaluations on 15 dimensions, with four of the differences being significant; the largest difference was on the Evaluation, Development, and Mentoring performance dimension (males scored .52 of a point higher; $r^2 = .12$).

Table 45 presents the evaluation results generated by averaging the ratings provided by the Peers and Supervisors, the composite Other. These results show that males were rated more highly on 16 of the dimensions, but only two were significantly different. Males were rated higher on Analysis and Reasoning, while lower on Community Service ($r^2 = .02$ and $.03$, respectively).

Table 46 presents the evaluation results generated by averaging across all three rater subgroups to form the All rating. These results show females achieving higher ratings on 16 Effectiveness Factors, but only three were significantly different; the largest difference was on the Community Service dimension where females were .26 points higher than males; ($r^2 = .04$).

The first row in each table (42 - 46) shows the average Global rating across the 26 Effectiveness Factors by which rater subgroup is summarized; these results showed no statistically significant gender differences. This conclusion, however, illustrates the potential loss of important information when data are averaged across a number of

factors. Although the Global averages show no statistically significant gender difference, examining the five rater subgroups separately for each factor (tables 42 - 46), reveals consistent pattern of differences between males and females. Males scored consistently higher on 10 dimensions (Analysis and Reasoning; Creativity/Innovation; Problem Solving; Researching the Law; Fact Finding; Influencing and Advocating; Writing; Strategic Planning; Networking and Business Development; and Stress Management while females were consistently higher on 11 dimensions (Practical Judgment; Questioning and Interviewing; Listening; Organizing and Planning One's Own Work; Organizing and Managing Others' Work; Able to see the World Through the Eyes of Others; Providing Advice & Counsel & Building Client Relationships; Developing Relationships within the Legal Profession; Passion and Engagement; Integrity/Honesty; and Community Involvement and Service).

The tables also showed that, in general, (1) the averages were rather high, at the 4 point mark on a 5 point scale, and that (2) the Self Appraisals were lower on many of the Effectiveness Factors than the Peer or Supervisor ratings.

CONCLUSION #17: Appraisals by various rater subgroups showed patterns of gender difference with males or females higher on various Effectiveness Factors, but only a few were statistically significant. The Global average over 26 Effectiveness Factors, regardless of rater subgroup, showed no significant differences by gender.

RECOMMENDATION #8: In future research, preservation rather than aggregation of subcategories should be the analytical strategy.

7. Intercorrelations among the Performance Rating Sources

Table 47 summarizes the agreement among the various rater subgroups. The table shows that the correlations between Peers and Self across the 26 Effectiveness Factors ranged from .02 to .37; for Self and Supervisor, the correlations ranged from .00

to .40; for Peers and Supervisor, the correlations ranged from -.04 to .40; for Self and Other, the correlations ranged from .03 to .40; for Self and All the correlations ranged from .51 to .78; for Peers and Other, the correlations ranged from .81 to .96; for Peers and All, the correlations ranged from .71 to .90; for Supervisor and Other, the correlations ranged from .72 to .89; for Supervisor and All, the correlations ranged from .60 to .81; and, for All and Other, the correlations ranged from .86 to .94.

The results shown in Table 47 indicate low to moderate agreement when the rater subgroups are treated separately (Self, Supervisor, and Peer). These results provide further support for treating the rater groups separately for additional analyses. Given the reliability results reported by Barrett (2008), Other and All are reasonable combinations to study.

CONCLUSION #18: Analysis of separate rater subgroups showed low to moderate agreement among them.

RECOMMENDATION #9: Doing separate analyses of various rater groups should be continued in future research.

8. Prediction of Lawyering Effectiveness as Measured in Law School

Analysis of the student sample, like the alumni analysis, raises the same question: how well do the LSAT, UGPA, and Index Score predict performance in law school? But, more importantly, how well do they predict performance in law school other than through grades? For the student sample, we adapted and/or asked raters to analogize the 26 Effectiveness Factors developed for practicing attorneys to fit the context and content of experiences that law students encounter. Thus, appraisals of performance on the 26 Effectiveness Factors were not based on full-time job performance, but on work in clinics, part-time law jobs, research assistance work for faculty, student activities and groups, as well as on coursework.

a. LSAT, UGPA and Index Score as Predictors of Lawyering Performance

Table 48 shows the zero-order correlations between the LSAT, UGPA, and Index and each of the 26 performance measures, as well as the global average performance score; the table shows data for each of the five rater subgroups – Self, Peer, Supervisor, Other, and All. Correlations between several of the LSAC measures and performance appraisals from several rater sub-groups suggest that further research could be useful. Table 48 shows that, taking into account more than one performance rating group, the LSAT correlated with Analysis and Reasoning, Creativity, Problem Solving, Influence and Advocating, Writing, and Networking. The Effectiveness Factors that are most similar to what is measured by LSAT, such as Analysis and Reasoning, showed correlations with LSAT that ranged from .15 to .30.¹⁰ For the Networking Factor, the correlations were negative ranging from -.20 to -.30. This finding suggests that higher scorers on the LSAT did not do well in Networking. This result was similar to that found for the alumni sample, where we noted that Networking requires interaction with others. It may be that those who score highly on the LSAT are the type of students who do not devote attention to networking or lack the necessary skills to do so.

For UGPA, the results show somewhat fewer patterns of correlations than did the LSAT. In general, UGPA correlated with Practical Judgment (r's ranging from -.16 to -.20), Questioning and Interviewing (r's ranging from -.18 to -.19), Developing Relationships (r's ranging from -.17 to -.20), Integrity (r's ranging from -.17 to -.19), and Community Service (-.15 to -.26). All of these correlations were negative, which suggests that higher UGPA was related to poorer performance on the dimensions noted. Higher grades may reflect individuals who are more oriented to books and less toward real world practical skills. Questioning, Relationships, and Service all require

¹⁰ This result may reflect inclusion of faculty appraisers, who likely put emphasis on cognitive skills within factors like Creativity and Problem Solving – i.e. *intellectual* creativity or *intellectual* problem solving.

interpersonal skills and time away from studying.

The Index Score correlates with Analysis and Reasoning (r's ranging from .17 to .19), Writing (r's ranging from .16 to .22), Networking (r's ranging from -.20 to -.23), and Community Service (r's ranging from -.17 to -.28). Explanations for these patterns might be similar to those already mentioned

CONCLUSION #19: The LSAT, UGPA and Index Score predicted a few of the Effectiveness Factors, mainly ones where constructs overlapped between the LSAT, UGPA and Index and performance indices. However, the LSAT, UGPA, Index were not useful for prediction of the large majority of Effectiveness Factors.

b. New Tests as Predictors of Lawyering Performance

Tables 49 – 52 show correlations between the predictors identified/developed for this project and the 26 Effectiveness Factors as well as with the average Global rating. In summarizing the results of these tables (presenting correlations between particular predictors and performance on Effectiveness Factors), we focus on correlations that are significant with at least three rater sub-groups.

Examination of the correlations in Table 49 showed that three of the HPI scales – Ambition, Interpersonal Sensitivity, and Prudence – correlated with 6, 5, and 5 Effectiveness Factors, respectively. The strongest correlations for Ambition were with Networking and Passion (in the 30s). Interpersonal Sensitivity correlated relatively strongly with Developing Relationships and Community Service (.30s). For Prudence, the strongest correlations were with Managing One's Work and Diligence (.20s). The Sociability scale correlated with four Effectiveness Factors: Speaking, Managing Others, Networking, and Integrity. The first three of these factors showed correlations in the .20's while the correlations with Integrity were in the high .10's but negative. Among the HPI scales, Learning Approach was not a good predictor of any of the Effectiveness Factors. This was surprising because it was a good predictor for the alumni sample of

the Effectiveness Factors such as Analysis and Reasoning and Researching the Law.

Overall, three to four HPI scales may offer benefit given that the correlations for these scales with lawyer effectiveness are slightly higher than those found for the LSAT, UGPA, and Index score predictors.

Table 50 shows the correlations between the HDS scales and the 26 Effectiveness Factors. The one scale that showed promise is “Reserved” which reflects being remote and detached, and a lack of concern or awareness about others feelings. This scale correlated with 10 of the 26 Effectiveness Factors, with a number of the correlations being in the .30s to .50s, and in the expected negative direction. The pattern suggested that those who lack concern for others are evaluated lower on Effectiveness Factors such as Managing Others, Developing Relationships, Networking, Community Service and Passion – all dimensions that the scale would be expected to predict.

Table 51 shows the correlations between the MVPI scales and the lawyer Effectiveness Factors. Overall, the pattern of correlations did not suggest much value for predicting student performance on lawyer Effectiveness Factors. Although an individual scale such as Altruistic correlated with Community Service ranging from .29 to .62, correlation patterns were insufficient to suggest the MVPI has value for predicting student performance as measured by the 26 Effectiveness Factors.

Table 52 presents, in part, the correlations between the non-cognitive, non-personality predictors (SJT and BIO) and the Effectiveness Factors. The BIO score correlated with Networking, Passion, and Community Service (range of .20s to .30s). SJT patterns provided insufficient basis to value it for predicting success in law school as measured by the Effectiveness Factors.

Table 52 also shows correlations between the SMS, OPT, and ER with the Effectiveness Factors. The SMS had few substantial correlations, but it did correlate

positively with Speaking (where one likely wants to manage impressions; r 's ranging from .29 to .41) as well as with Networking (where one also wants to manage impressions; r 's ranging from .21 to .32), but negatively with Integrity (managing impressions may negatively impact observers' views of student integrity).

The OPT had positive patterns with four of the Effectiveness Factors -- Networking; Evaluating, Developing, and Mentoring; Passion; and Stress Management (r 's basically in .20's, but reaching .48). OPT did correlate, however, with the HPI Adjustment ($r = .58$) and Interpersonal Sensitivity ($r = .49$) scales thereby suggesting it might be duplicative of the HPI. However, the OPT correlations with the four named Effectiveness Factors are higher than the HPI correlations with those same factors and would therefore be preferable to predict them.

The final measure, ER, did not show results that would suggest additional research.

Overall, some of the new predictors identified or developed for this specific research project displayed results that argue for additional research. They do so because (1) they correlated with areas that were not predicted by the LSAT, UGPA, and Index score, (2) their correlations were generally higher than the LSAT, UGPA, and Index score showed for the Effectiveness Factors with which they correlated, and (3) as noted earlier in the report, these predictors tend to not yield racial/ethnicity and gender mean differences.

CONCLUSION #20: Certain new predictors showed significant correlations with some Effectiveness Factors (3-4 of HPI scales; 1 of HDS scales; BIO, SMS, and OPT); others did not.

RECOMMENDATION #10: Further research on the tests that are predictive is especially warranted by the fact that they have higher correlations than LSAT, UGPA and Index Scores, and race/gender results are neutral.

c. Moderator Variables

Given the small sample sizes, moderated regression was not undertaken for the student sample.

d. Incremental Variance

Table 53 presents the results of an analysis in which the LSAT, Index, HPI scales, BIO, SJT, and OPT were allowed to enter in a step-wise multiple regression. The regression analysis sought to determine which combination (if any), of the predictors could explain Self ratings of performance on the 26 dimensions. Table 53 shows only the results that yielded significant incremental variance. An overview of the results for this table indicated that a combination of two tests, and in one instance of four tests, could produce multiple correlations with the 26 Effectiveness Factors and with the Global average performance, ranging from the mid .30's to the low .60's. Tables 54 through 56 repeated the analytic strategy for Peer, Supervisor, and Other, respectively. These tables showed similar patterns, though the multiple correlations approach the high .50's and not the .60's.

Table 57 summarizes the results of the multiple regression analyses. The main conclusion was that LSAT is a good predictor of performance for students in law school, particularly when Supervisors, Other, and All are the criteria, but that other predictors, such as some of the HPI scales and OPT might contribute to explained variance.

CONCLUSION #21: In multiple regression analysis, the LSAT is a relatively good predictor of Effectiveness Factors for students: some of the HPI scales and OPT might contribute to explanation of additional variance.

9. Prediction of Effectiveness in Law School As Measured by Grades

a. LSAC Components as Predictors of FYGPA in Law School

Table 58 shows the correlations between the LSAT, UGPA, and Index predictors and FYGPA for the current total sample. The results for the total sample showed

reasonable consistency with the research findings: LSAT correlated .36 with FYGPA, UGPA correlated .14 with FYGPA, and the Index correlated .40 with FYGPA.

As with our Alumni sample, the current research findings, which were based on uncorrected correlations and with a restricted sample on all measures, replicated the usefulness of the LSAT and Index as predictors of FYGPA in law schools.

b. Tests Identified/Developed for this Project as Predictors of Law School Performance

Table 58 also shows the correlations between each of the predictors identified or developed for this research project and the FYGPA in law school. Addressing the subtests, the results showed that two (Prudence and Learning Approach) of the seven subtests on the HPI correlate with FYGPA, with the correlations approximating .21. For the 11 HDS scales, only Mischievous correlated (-.31) with FYGPA. For the 10 MVPI scales, only Tradition correlated (.32) with FYGPA.

Of the remaining scales, SMS correlated .06, OPT correlated -.03, and ER correlated .02 with FYGPA, respectively. The correlations of FYGPA with BIO and SJT were -.05 and .11, respectively. None of these correlations was statistically significant.

c. Incremental Variance

Again, as with the Alumni sample, we undertook hierarchical multiple regression to determine whether any of the new predictors yielded incremental validity beyond that which was obtained by the LSAT alone or by the Index alone. Tests were entered on a second step after first entering LSAT (separate analyses for Index). Significant increases in R^2 would suggest the potential for the formation of a battery. Results indicated that only two scales, both HPI scales (Prudence and Intellectance), yielded incremental explanatory variance. Prudence yielded a 7% increase and Intellectance yielded a 2% increase. The Index results showed the same pattern.

d. Summary of Results: Predicting Law School Performance Measured by Grades

In summary, the results for the LSAC predictors were good predictors of FYGPA for the student sample studied. On the other hand, the new, potential tests did not show consistently strong relationships on their own, or in conjunction with the LSAT, UGPA, or Index measures.

CONCLUSION #22: As earlier research shows, the LSAT, UGPA, and Index Score are good predictors of FYGPA. Only a few scales or measures of the new predictor tests showed correlations and those were not substantial.

VII. CONCLUSION

A. Basic Data, Further Research and Options for Use

For this research study, data were collected:

- on multiple types of tests;
- from a large sample of practicing lawyers and law-related workers;
- with varied experience in terms of years, settings and practice areas;
- including a modest number of minority practitioners and students;
- and from peers, supervisors, and self asked to appraise these practitioners' current job performance on numerous dimensions. .

Our research results demonstrate that new tests have considerable potential to predict lawyer effectiveness which could, in turn, make important contributions to law school admissions decisions. Some of the new professional-performance-predicting tests produced very significant correlations indeed. For example, BIO scores showed correlations (in the .2's and .3's) with *all* Effectiveness Factors except Integrity and Stress Management. SJT scores showed correlations with *all* Effectiveness Factors

other than Managing Others and Evaluation, Development, and Mentoring. The correlations were generally in the .10's and low .20s. The impressive aspect of these results was (1) the large number of Effectiveness Factors that were predicted by the BIO and the SJT tests, and (2) that the correlations were generally higher, though moderately so, than those between the LSAT and the small subset of Effectiveness Factors that overlap with the LSAT and with which it had an expected relationship (e.g. Analysis and Reasoning, Researching the Law, Writing).

We believe the exploratory data reported here make a compelling case for undertaking large-scale, more definitive research. If the LSAC itself, or through contractors, more broadly researched and refined the new test battery and then offered it along with the LSAT, the Council could assure the new tests' validity and perhaps also recommend appropriate uses for the new scores. It could create a clearinghouse for different schools to share their experiences with combining these new predictors with the old.

The new tests used in conjunction with the LSAT and Index Score, could extend prediction beyond law school grades to project success in the practice of law. The new methods could predict professional performance using merit-based, theoretically justified selection factors that are also more racially neutral than current tests in their admissions processes. New measurements would also provide applicants, career placement officials and employers with more information about applicants' gifts and strengths.

If further research confirmed the validity of performance-predictive tests, then the new measures would open up an array of valuable options in admission practices. Tests have typically been used in top-down fashion, where the highest scores are selected first. Other alternatives could be explored perhaps with a pilot segment of an admissions class, or several: Member schools might, for example, use the LSAT and/or Index score to set an academic floor and then use the new scores and other file

materials to rank applicants who surpass that floor. Or, a school might use the LSAT to identify the top 20% (in terms of academic potential) and then combine the LSAT score with one or several of the new test scores into a new type of Index, using the combined information to admit applicants. Or, a school might wish to combine the Index Score and new test scores from the beginning in order to assure that its has selected its student body on the basis of relevant academic *and* performance-predictive factors, and has increased diversity compared to admission policies that predominately emphasize LSAT scores. Or, a school might establish minimum scores for each of multiple test instruments and require that an applicant achieve that minimum score on each to gain admission. In sum, the additional information supplied by the new scores could be used in a variety of ways, only some of which are suggested here.

B. Potential Uses and Benefits

Table 59 draws upon the alumni sample studied in this research project to illustrate the potential value of including new tests in admissions strategy. The table shows the number of admits – by gender and race – that would result from use of different instruments for admission decisions. To illustrate, if the LSAT were the only admissions test, and if it were used in a top-down fashion where the scores are presented in rank order, selection of the top 10% of the sample studied in this research project would yield 116 admits, 54.3% of whom would be female and 85.3% of whom would be white, .9% African-American, 4.3% Hispanic, and 6.9% Asian/Pacific Islander. In another example, if the Index Score were the only determinant of admissions, and were used in a top-down fashion to select 10% of this sample for admission, the chosen applicants would be 52.7% female and 87.5% white, 0% African-American, 4.5% Hispanic, and 6.3% Asian/Pacific Islander. By contrast, if the SJT were the only determinant of selection (N = 80 rather than 116 because not all in the sample took the SJT), 50% of the top 10% would be female and 68.8% white, with 7.5% being African-

American, 7.5% being Hispanic and 11.3% being Asian/Pacific Islander.

The remaining rows of Table 59 show results for different selection percentages and for different test combinations. For example, combining the tests of BIO and SJT, and using supervisor assessments of participants' lawyering effectiveness yields a group of 64 admits in the top 20% of scores, where 53.1% are female, 65.6% are white, 9.4% are African-American, 10.9% are Hispanic and 10.9% are Asian/Pacific Islander.

Inclusion of a battery of these new performance-predictive tests is in our view justified by the actual role and mandate of law schools as professional schools; it also has the salutary effect of significantly increasing diversity that is achieved by use of current tests alone.

C. Limitations

The reported research has several limitations:

(1) Results are based on two law schools only;

(2) Results reflect a restricted sample in that (a) all participants in the alumni sample were admitted to and graduated from law school; (b) all were law graduates practicing law or performing law-related jobs, which assumes they were reasonably successful – unsuccessful lawyers were not likely to participate; and (c) all were volunteer participants. Limitations such as these tend to underestimate correlations among the measures. However, the obtained correlations are sufficiently strong to strongly suggest the need for additional research.

D. The Context Reviewed: Reasons to Add Professional Predictors

Scholars and commentators on legal education have urged that the current criteria of merit for admission to law school, especially the LSAT, are too narrow in aim. Many would agree that assessing professional potential before admission would be a good idea, but no one has had a method to propose. Indeed, the prevailing view has been that the task is so difficult as to be flatly impossible. The research we report here

explored ways to assess and predict many dimensions of professional effectiveness and has yielded a rich harvest. We now briefly review developments in legal education that have *particular salience for admission policies*:

(1) Recent Developments in Legal Education

a. Increased Applicant Pool

In 1950, two years after the first use of the LSAT, 6,750 tests were administered, in 1955, 11,750. Shirley Abrahamson, Chief Justice of the Wisconsin Supreme Court, described admission in the fifties:

When I went to law school, it was said that there were two requirements for admission to most law schools: first, you had to have a college degree; and, second, you had to be breathing. And either requirement might be waived (Raushenbush, 1986).

In 2007-08, 142,331 LSAT tests were administered; 55,500 of 84,000 applicants were admitted to some ABA accredited law school. (LSAC Volume Summary Applicants: 1997-2007, 2008). Today it is harder to get into law school than to pass the bar; thus admissions decisions choose the nation's lawyers.

b. Inclusion of Women and Minorities

After many decades of official or *de facto* near-exclusion of women and minorities from legal education (Kidder, 2003), entry of these groups has enlarged the pool of applicants, making competition more intense especially for white males. Schools adopted various types of affirmative action to help offset past discrimination, cultural stereotypes, and lack of educational preparation that had hampered these groups' entry into law in larger numbers. Additionally, larger numbers of international students are coming to U.S. law schools.

c. Controversy and Litigation

A widening wealth gap, the erosion of the middle class, the fear of families that

their children may not keep up with parental hopes and expectations – all contribute to growing stress about gaining stable and remunerative work for the future. Many covet a legal education as a pipeline to high salaries, status and important jobs. Public scrutiny of admission policies is intense. Combat over affirmative action and other aspects of the “culture wars” creates a constant threat of litigation over the “fairness” of admission policies.

d. Rankings

When *US News and World Report* hit on the idea of ranking educational institutions to sell magazines, few would have predicted the stunning impact those rankings would have. Although disclosure of more information about schools is a good thing, the competitive concerns of everyone involved have turned “rankings fever” into an obsession. Higher rankings increase prestige, draw students, loosen alumni and donor wallets, give faculty ego points, and raise leverage within the university. Consequently, no matter where they place on the scale (except for a few iconoclasts like CUNY, New College, or Northeastern), schools want to move up the charts. Each, therefore, emulates those above them, from the bottom to the top of the scales.

e. The Place of Law Schools in Major Universities

Major research universities have increasingly decided that applied fields of study belong in non-University settings. Training for professional practice in such fields as the parish ministry, school teaching, architecture, and others have shifted to free-standing or less prestigious educational institutions. Business, medical and law schools, however, have been largely exempt from this trend. Increasingly, law faculties strive to be viewed as intellectual peers of academic colleagues. Faculty hiring, promotion and salary policies; student admission practices, and curricular policies reflect a significantly more theoretical and quantitative intellectual agenda than in the past. Other than in clinical programs, attention to training professionals for practice, examination of problems within

the profession and in society's provision of legal services garner less attention today than might once have been thought possible.

2. These Changes Affect Admission Policies

Each of the above changes has affected admissions policy and practice -- mostly in the direction of increasing the impact of the LSAT score on a candidate's chances. When a handful of Ivy League law schools created what became the LSAT, they sought a tool to screen for "legal aptitude" and a method to evaluate degrees from a widening and unfamiliar array of colleges (LaPiana, 2001). Never designed as the sole basis for admission decisions, the test was, instead, a way to judge an individual's ability to complete law study successfully. From that beginning, the LSAT score has become the most important criterion in gaining admission to law school, especially at the more prestigious schools. These schools graduate the lawyers who have the most opportunity for money, prestige, and influence in many dimensions of American and increasingly, global society. Lawyers protect and extend Americans' wealth; occupy judicial positions; constitute large percentages of legislators, governors, and presidents; advise and lead corporations and non-profits; and represent the government in civil and criminal justice systems. Today, although the LSAC continues formally to urge that the LSAT not be overused in selecting among law school applicants, the test's actual influence on admissions decisions is hard to overestimate. Examination of the factors described above helps to explain why.

The increased number of law applicants, including relatively new minority, women, and international contenders, makes selection more costly, time consuming and difficult. The desire for "efficiency," especially in expending faculty time, presses for quicker methods of comparison. Nothing is quicker or easier than comparison of standardized numeric indicators. Conflict about admissions criteria and appropriate definitions of "qualification" and "fairness" have become more frequent and heated.

Facing scrutiny and debate, and fearing litigation, faculties find an “objective” method of distinguishing among outstanding applicants appealing, especially at the top 30-50 schools. (Of course, the LSAT imports judgments about what constitutes “merit” in selecting law students and legal professionals; specifically, the LSAT prefers cognitive skills -- analysis, logic and reading -- over most anything else). Some evidence suggests that even when schools make policies that aspire to assure that LSAT scores should not drown out other indicators, test scores retain a greater weight than those policies intend (Kidder, 2000).¹¹

Rankings also create pressure to weigh LSAT scores more heavily. *US News and World Report* considers a number of dimensions in its rankings of law schools, but median LSAT scores of entering classes are the one that a school can most quickly and directly affect. With law professors caring a great deal about how university peers perceive them, and with law faculties more focused on and incited to emphasize academic agendas over professional agendas, emphasis on narrow cognitive predictors is very appealing (Rubin, 2008).

3. Problems Associated with Over Emphasis on LSAT Scores

At points in the last half of the twentieth century, tests like the LSAT have helped to reduce subjective biases and improve access (e.g. religion, race, gender, attendance at a lesser known colleges) to law school but, at present, strong questions must be raised about the test's impact on law school admissions.

a. Misplaced Precision

The Law School Admission Council regularly admonishes schools not to over-rely on the LSAT score, and to use other factors in addition to the test. But LSAT scores

¹¹ In a study of UC law school admission statistics, Kidder (2000) found that in 1998, holding undergraduate institution and major constant, for applicants who had GPAs of 3.75 or more, a 5 point difference in LSAT score cut the chance of admission from 89% to 44% at Berkeley Law School; for the same year at UCLA, the chance of admission dropped from 66% to 10%.

dominate today's admissions decisions. More applicants means more scores at most points along the score scale. As scores cluster, decisions that depend heavily on test outcomes will risk being less valid. Choices between individual scores tend to rest on smaller actual differences, sometimes even leading decision-makers to distinguish between scores that fall within the statistical error of measurement for the test. This creates a potential fallacy of misplaced precision (the illusion of decisive precision), especially when LSAT scores explain only about 25% of variance in first year grades.

b. Selection Bias

Over-reliance on the LSAT score may create what Christopher Jencks of Harvard described as "selection system bias" (Jencks, 1998). Although items on the LSAT have been carefully vetted several times for any bias embedded in the *content* of the questions, the way schools use the LSAT has the potential to create race bias. Jencks defines selection system bias as selection based disproportionately on some factor (call it A) instead of or out of proportion to another factor (call it B), where A and B are (for sake of easy illustration) equally important to the output sought, and A and B have differing profiles of racial group performance. Disproportionate influence on A or B may then unfairly skew selection outcomes. Law school admission offers a persuasive example. Let A stand for cognitive, test-taking abilities and B for an array of factors important to effective professional performance. Research shows that white students outperform African-American, Hispanic, and other under-represented minority groups in cognitive, school-skill type tests. Scores on such academic tests, including the LSAT, are heavily correlated with socio-economic status and educational opportunity. By contrast, research (including our own) suggests that racial groups perform similarly in their jobs.

Using Jencks' category to analyze admission practices, law schools that overemphasize the LSAT score (A) and do not try to predict professional performance

may be described as engaging in selection system bias. Emphasizing academic tests on which whites excel without including tests that might predict lawyering performance on which race and gender do not significantly affect performance suggests that white-favoring admissions criteria greatly outweigh race-neutral professional performance admissions criteria. If valid prediction of professional performance is possible, in order to be race neutral, schools should factor professional lawyering skills into admission decisions along with academic predictors.

c. Professional and Academic Emphasis

As law schools gravitate away from emphasis on professional competencies toward more theoretical and principally academic emphases characteristic of other departments in the university, other policy questions arise. Rubin recently named the phenomenon by which a substantial portion of student tuition pays for faculty research a “cross subsidy” from students’ interests in professional training to faculty preferences for research and suggests curriculum should align more closely with faculty research interests (Rubin, 2008). The interests in academic research and professional training are not mutually exclusive, nor need all schools balance the two in the same way. But, arguably, the pendulum may have swung too far toward research interests.

Of course, a heavy proportion of theoretical and empirical research might arguably be fitting for a small number of research university law schools, but more closely examined, even that premise can be questioned. First, even at elite schools, the vast majority of graduates enter (and indeed, monopolize) elite jobs in law practice. Law schools and law jobs alike are extremely stratified, and therefore, often matched; consequently, top schools draw many students desiring to practice law or take a law related job. For example, Yale Law School has the highest percentage of graduates entering academia of any law school, and its percentage is 13%; at other top academic “feeder” schools, the percentages are substantially lower. Given that each tranche of

schools emulates the ones above it to increase prestige, recruitment, and financial support, the heavily academic orientation is not at all restricted to elite schools.

In the context of admissions, the faculty academic and research focus translates into strong support for high reliance in admissions on academic criteria (LSAT and grades). The result is that law schools weigh a standardized test score more heavily than either a) other professional schools (like business and medicine) or b) graduate academic departments (like philosophy, psychology, or economics) whose primary role is to educate the next generation of academics in the field.

Second, as a consequence of more research focus and more dominant academic criteria for admission, that new graduates will not be prepared to practice law is a truism. Graduates know they must get a job with a well-resourced employer in order to learn how to actually practice law. Students who want careers in public interest organizations, with service or advocacy groups, in solo practice or with small law firms face an ironic headwind. The employers they seek cannot afford, as a large corporate firm can, to spend several years training recent graduates before they actually become market-useful. Recent grads' main utility for large firms is the prestige they confer by dint of their graduation from good schools; their productivity need not be as high as their salaries in the early years. This suggests that students' job choices may potentially be distorted by more than the magnet of high salaries offered by elite firms. If they do get jobs they desire with less well-financed employers, or hang out their own shingle, they face a difficult path to professional competency as a result of law schools' emphasis on research rather than professional competence. Their clients will also get less effective services. Even if this distorting factor is small in percentage terms, large corporate firms should not be subsidized by professional schools' choices. Clinical programs now exist at most law schools and go some ways to ameliorate the problem. But clinical programs that receive equal time, pay and status, either for clinic professors or enrolled students,

are rare.

The discussion of academic research versus professional education need not and should not be either/or. Both are essential to fulfillment of law school responsibilities and opportunities. What is salient here is that *both* academic and professional indicators should be included in admission decisions. Because effective academic indicators like the LSAT are now the dominant force in admissions, exploration of professional predictors should become the next major agenda for law school admissions.

VIII. REFERENCES

- A.B.A. Section on Legal Education and Admissions to the Bar. (1992). Legal Education and Professional Development: An Educational Continuum ["MacCrate Report"].
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Anthony, L. C., Harris, V. F., & Pashley, P. J. (1999). *Predictive validity of the LSAT: A national summary of the 1995-96 correlation studies* (Law School Admission Council LSAT Technical Research Report TR-97-01). Newtown, PA: Law School Admission Services, Inc.
- Anthony, L.C., & Liu, M. (2000). *Analysis of differential prediction of law school performance by racial/ethnic subgroups based on the 1996-1998 entering law school classes* (Law School Admission Council Technical Research Report TR-00-02). Newton, PA: Law School Admission Services, Inc.
- Aspinwall, L.G. ,& Taylor, S.E. (1992). Individual differences, coping, and psychological adjustment: A longitudinal study of college adjustment and performance. *Journal of Personality and Social Psychology*. 63, 989-1003.
- Association of Mexican-American Educators v. California, 231 F.3d 572, 581-83 (9th Cir. 2000) (en banc).
- Barrett, P. (2008). Some analyses of personality, academic, and performance rating data from alumni and students with the faculty of law, University of California, Berkeley. Tulsa, OK: Hogan Assessment Systems.
- Barrick, M.R., & Mount, M.K. (1991) The Big Five personality dimensions and job performance: A meta analysis. *Personnel Psychology*, 4, 1-26.
- Bentz, V.J. (1985, August). *A view from the top: A thirty year perspective of research devoted to discovery, description, and prediction of executive behavior*: Paper presented at the 93rd Annual Convention of the American Psychological Association, Los Angeles.

- Brissette, I., Scheier, M.F., & Carver, C.S. (2002). The role of optimism in social network development, coping, and psychological adjustment during a life transition. *Journal of Personality and Social Psychology*, 82(1), 102-111.
- Caldwell, D. F., & O'Reilly, C. A. (1982). Boundary spanning and individual performance: The impact of self-monitoring. *Journal of Applied Psychology*, 67, 124-127.
- Caligiuri, P. M., & Day, D. V. (2000) Effects of self-monitoring on technical, contextual, and assignment-specific performance. *Group & Organization Management*, 25, 154–174.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer-Verlag.
- Carver, C. S., & Scheier, M. F. (2002). Optimism. In C. R. L. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 231-243). New York: Oxford University Press.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15(3), 233-254.
- Chatman, J. (1991). Matching people and organizations: selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36, 459-484.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey-Schmidt, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410–417.
- Dalessandro, S.P., Stilwell, L.A., & Reese, L.M. (2005). LSAT performance with regional, gender, and racial/ethnic breakdowns: 1993-1994 through 1999-2000 testing years (Law School Admission Council Research Report TR-00-01). Newton, PA: Law School Admission Services, Inc.
- Diaz, R. J., Glass, C. R., Ankoff, D. B., & Tanofsky-Kraff, M. (2001). Cognition, anxiety, and prediction of performance in 1st-year law students. *Journal of Educational*

- Psychology*, 93, 420-49.
- Edwards, H.T. (1992). The Growing Disjunction Between Legal Education and the Legal Profession, *Michigan Law Review*, 91, 34-78.
- Evans, F. R. (1984). *Recent trends in law school validity studies* (Law School Admission Council Report 82-1). Reports of LSAC Sponsored Research: Volume IV, 1978-1983, pp. 347-361, Newtown, PA: Law School Admission Services, Inc.
- Ekman, P. (2004). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. NY: Macmillan.
- Goldhaber, M.D (1999, July 5). Is the Promised Land Heav'n or Hell? *National Law Journal*, pp. A17.
- Goleman, D. (1995) *Emotional Intelligence*. New York: Bantam Books
- Grutter v. Bollinger, 539 U.S. 306 (2003).
- Guion, R. (1987). Changing views for personnel selection research. *Personnel Psychology*, 40, 199-213.
- Heinz, J.P, Hull, K, E., & Harter, A (1999). Lawyers and Their Discontents: Findings from a Survey of the Chicago Bar, *Indiana Law Journal*, 74, 735- 758
- Hobfoll, S.E. (2002). Social and psychological resources and adaptation. *Review of General Psychology*, 6, 307-324.
- Hogan, J., & Hogan, R. (1991). *Levels of analysis in big five theory: The structure of self-description*. Paper presented at the Sixth Annual Conference of the Society for Industrial and Organizational Psychology. St. Louis, MO.
- Hogan, J., & Hogan, R. (1996). *Motives, Values, Preferences Inventory manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., & Hogan, J. (1997). *Hogan Development Survey manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory manual* (3rd ed.). Tulsa,

OK: Hogan Assessment Systems.

Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, 51, 469-477.

Hogan, R., Hogan J., & Warrenfeltz, R. (2007). The Hogan Guide: interpretation and use of Hogan inventories. Tulsa, OK: Hogan Assessment Systems.

Hough, L.M., Oswald, F.L., & Ployhart, R.E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194.

Hunter, J., & Hunter, R. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 1, 72-98.

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

Jencks, C. (1988). Racial Bias in Testing. In Jencks, C. and Phillips, M. (Eds.). *The Black-White Test Score Gap*, pp. 55-85. Brookings Institution Press.

Kidder, W.C. (2000). The Rise of the Testocracy: An Essay on the LSAT: Conventional Wisdom, and the Dismantling of Diversity, *Texas Journal of Women and the Law*, 9, 167 - 217.

Kidder, W.C. (2003). Law and Education: Affirmative Action Under Attack: The Struggle for Access from Sweatt to Grutter: A History of African American, Latino and American Indian Law School Admissions, 1950-2000. *Harvard Blackletter Journal*, 19, 1-35.

Kilduff, M., & Day, D. V. (1994). Do chameleons get ahead? The effects of self-monitoring on managerial careers. *Academy of Management Journal*, 37, 1047-1060.

- Kristof, A.L. (1996) Person-organization fit: An integrative review of its conceptualizations, measurement, and implications, *Personnel Psychology*, 49, 1-49.
- Lam, L. T., & Kirby, S. L. (2002). Is emotional intelligence an advantage? An exploration of the impact of emotional and general intelligence on individual performance. *The Journal of Social Psychology*, 142(1), 133–143.
- LaPiana, W.P. (2001). *A History of the Law School Admission Council and the LSAT*. Retrieved September 8, 2008 from <http://www.lsacnet.org/publications/history-lsac-lsat.pdf>.
- Law School Admission Council. (1999). LSAT/LSDAS registration & information book. Newtown, PA: Law School Admission Services, Inc.
- Law School Admission Council Volume Summary Applicants: 1997-2007. Retrieved September 3, 2008, from <http://members.lsac.org/Public/MainPage.aspx?ReturnUrl=%2fPrivate%2fMainPage2.aspx>.
- Law, K. S., Wong, C., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology*, 89, 483-496.
- Linn, R. L., & Hastings, C. N. (1983). *A meta analysis of the validity of predictors of performance in law school*. Reports of LSAC Sponsored Research: Volume IV, 1978-1983, pp. 507-544, Newtown, PA: Law School Admission Services, Inc.
- Makikangas, A., & Kinnunen, U. (2003). Psychosocial work stressors and well-being: self-esteem and optimism as moderators in a one-year longitudinal sample. *Personality and Individual Differences*, 35(3), 537-557.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A

- clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- Morain, D. & Riccardi, N. (2008, July 29). Campaign '08: Race for the White House: A Lack of (Ballot) Initiative: GOP's Strategy of Using State Measures to Draw Conservatives to the Polls has Faltered, *Los Angeles Times*.
- Mumford, M. (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction*. Palo Alto, CA, US: CPP Books.
- National Conference of Bar Examiners & A.B.A. Section on Legal Education & Admissions to the Bar (2008). Comprehensive Guide to Bar Admission Requirements. Chart III (Erica Moeser & Margaret Fuller Corneille, Eds.) Retrieved September 8, 2008, from <http://www.ncbex.org/comprehensive-guide-to-bar-admissions>.
- No Child Left Behind Act of 2001, 20 U.S.C. 6301 et seq. (2002).
- Norton, L.L., Suto, D.A., & Reese, L.M. (2006). Analysis of differential prediction of law school performance by racial/ethnic subgroups based on 2002-2004 entering law school classes. (Law School Admission Council Technical Research Report TR-06-01). Newton, PA: Law School Admission Services, Inc.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679–703.
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89, 187-207.
- Powers, D. E. (1982). Long-term predictive and construct validity of two traditional predictors of law school performance. *Journal of Educational Psychology*, 74, 568-576.

- Raushenbush, W. (Ed.). (1986). Law School Admissions 1984-2001: Selecting Lawyers for the Twenty-First Century. Law School Admission Services, Inc.
- Rubin, E. (Ed.), (2008). Should Law Schools Support Faculty Research? *Legal Studies Research Paper Series*, Paper No. 08-038. Retrieved September 8, 2008 from <http://ssrn.com/abstract=1161273>.
- Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001) High-stakes testing in employment, credentialing, and higher education: Prospects in a post affirmative-action world. *American Psychologist* 56, 302-378.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30–43.
- Salovey, P., Mayer, J.D., Goldman, S.L., Turvey, C., & Palfai, T. (1995). Emotional attention, clarity and repair: exploring emotional intelligence using the Trait Meta-Mood Scale. In J.W. Pennebaker (ed.), *Emotion, disclosure, and health*. American Psychological Association, Washington, D.C.
- Saucier G., & Goldberg, L.R. (1998). What is beyond the Big Five? *Journal of Personality*, 66, 495-524.
- Scheier, M. F., & Carver, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219-247
- Scheier, M. F., & Carver, C.S. (1992). Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research*, 16, 201-228.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 1063-1078.

- Schiltz, P.J. (1999). On Being a Happy, Healthy and Ethical Member of an Unhappy, Unhealthy and Unethical Profession, *Vanderbilt Law Review*, 52, 871- 951.
- Schmidt, F.L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15, 187-210.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719–730.
- Schrader, W. B. (1977). Summary of law school validity studies, 1948-1975 (Report No. LSAC-76-8). Law School Admission Council. *Reports of LSAC Sponsored Research: Volume III, 1975-1977*, pp. 519-550, Princeton, NJ: Law School Admission Council.
- Shultz, M., & Zedeck, S. (2003). Phase I Final Report: Identification and Development of Predictors for Successful Lawyering.
- Slaski, M., & Cartwright, S. (2002). Health, performance and emotional intelligence: An exploratory study of retail managers. *Stress and Health*, 18, 63-68
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for ratings scales. *Journal of Applied Psychology*, 47, 149-155.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526-537.
- Snyder, M. & Gangestad, S. (1986). On the nature of self-monitoring: matters of

- assessment, matters of validity. *Journal of Personality and Social Psychology*, 51, 125-139.
- Society of American Law Teachers, (2003). SALT on the LSAT, Retrieved September 3, 2008 from <http://www.saltlaw.org/StatementLSATBrochure.pdf> .
- Stilwell, L.A., Dalessandro, S.P., & Reese, L.M. (2003). *Predictive Validity of the LSAT: A National Summary of the 2001–2002 Correlation Studies*. (Law School Admission Council LSAT Technical Research Report TR-03-01). (TR-03-01). Newtown, PA: Law School Admission Services, Inc.
- Stilwell, L.A., & Pashley, P.J. (2003). Analysis of differential prediction of law school performance by racial/ethnic subgroups based on 1999-2001 entering law school classes. (Law School Admission Council LSAT Technical Research Report TR-03-03). Newtown, PA: Law School Admission Services, Inc.
- Sturm, S., & Guinier, L. (1996). The Future of Affirmative Action: Reclaiming the Innovative Ideal, *California Law Review*, 84, 953-1036.
- Sullivan, W.M., Colby, A., Wegner, J., Bond, L., Shulman, L. (2007). *Educating Lawyers: Preparation for the Profession of Law*. San Francisco: Jossey-Bass.
- Tett, R.P, Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology*, 44, 703-742.
- Tett, R.P., Steele, J.R., & Beauregard, R.S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior*, 24, 335-356.
- Weekley, J.A., & Ployhart, R.E. (2005). Situational Judgment: Antecedents and Relationships with Performance. *Human Performance*, 18, 81-104.
- Wiggins, J.S., & Trapnell, P.D. (1997). Personality structure: the return of the Big Five. In R. Hogan, J. Johnson, and S. Briggs (eds.), *Handbook of Personality Psychology*. San Diego: Academic.

- Wightman, L. F. (1993). *Predictive validity of the LSAT: A national summary of the 1990-1992 correlation studies* (Law School Admission Council Research Report 93-05). Newtown, PA: Law School Admission Services, Inc.
- Wightman, L. F. (1997). The threat to diversity in legal education: An empirical analysis of the consequences of abandoning race as a factor in law school admission decisions. *New York University Law Review*, 72 (1), 1-53.
- Wightman, L. F. (2000). The role of standardized admission tests in the debate about merit, academic standards, and affirmative action. *Psychology, Public Policy, and Law*, 6, 90-100.
- Wightman, L. F., & Muller, D. G. (1990). *An analysis of differential validity and differential prediction for Black, Mexican American, Hispanic, and White law school students* (Law School Admission Council Research Report 90-03). Newtown, PA: Law School Admission Services, Inc.
- Youssef, C.M., & Luthans, F. (2007). Emerging positive organizational behavior. *Journal of Management*, 33(3), 321-349.
- Xanthopoulou, D., Bakker, A.B., Demerouti, E., & Schaufeli, W.B. (2007). The role of personal resources in the job demands-resources model. *International Journal of Stress Management*, 14 (2), 121-141.