Journal of Research on Leadership Education
Volume 3, Issue 1 (May, 2008)

PREDICTIVE VALIDITY OF THE GRE AND GPAS FOR A DOCTORAL PROGRAM FOCUSING ON EDUCATIONAL LEADERSHIP[1]

I. Phillip Young
University of California-Davis

Leadership in the public school setting is a topic of interest in the professional literature for a variety of stakeholders. One particular stakeholder group is those responsible for preparing educational leaders, i.e. faculty administering preparation programs. Research addressing the preparation of educational administrators has followed separate but related research streams.

One stream focuses on why practicing educators seek administrative positions. To illustrate, some studies address the decisions of teachers to become school administrators (Newton, Giesen, Freeman, Bishop, & Zeitoun, 2003; Newton, Keedy, & Winter, 2000; Winter, Keedy, & Newton, 2001), while other studies explore the decisions of practicing administrators to pursue advance level positions within the public school setting (Pounder & Merril, 2001; 2002; Winter, Keedy, & Newton, 2001; Winter & Morganthal, 2002). In general, these studies use tight experimental designs, manipulate experimentally certain contextual variables purported to influence the attraction of applicants, and advance current knowledge about why qualified job candidates choose either to remain in their current assignment or to seek leadership positions.

However, less attention is afforded to how individuals acquire the necessary qualifications/certifications to be considered as a viable candidate for a leadership position. One avenue for acquiring the necessary qualifications/certifications is a doctoral degree in

[1] This research was supported in part by the University of California Educational Research Center (UCERC). The opinions contained in this work, however, are those of the author and do not necessarily represent the endorsement of the UCERC.

educational leadership, and a doctoral degree in educational leadership provides not only an opportunity to fulfill necessary qualifications but may well provide a competitive advantage within the employment selection process. As such, research is needed about admission to and graduation from a doctoral program in educational leadership.

Admission to and graduation from a doctoral program in educational leadership involves a sequential process. This process begins when candidates fulfill formal application requirements. Consequently, proactive decisions about admission are a necessary but not a sufficient condition for graduation.

Admission decisions are based on a variety of predictors. Common predictors include "some type of standardized test, e.g., the Graduate Record Examination, the student's prior academic record, letters of recommendation, previous professional experience, and perhaps some type of writing sample. GRE scores and grade point average represent quantitative measures in the decision making process" (Malone, Nelson, & Nelson, 2001, p.3).

Although valid information is needed for all predictors used to admit students (Anastasi, 1988; Chronbach, 1990; Standards for Educational and Psychological Testing, 1999), the focus of this study is on only the quantitative measures as suggested by others (Creighton & Jones, 2001; Malone et al., 2001; Norton, 1994). Included among these measures are GRE quantitative scores, GRE verbal scores, undergraduate grade point averages (UGPA) and graduate grade point averages (GGPA). More specifically, a predictive validity analysis is performed to assess the utility of these variables for three separate groups: (1) those applying but rejected, (2) those accepted but failing to graduate, and (3) those accepted and graduating.

Background Information

Predictive validity studies addressing admission, matriculation, and graduation offer information to a variety of stakeholders. Some research suggests that only approximately 50% of those admitted ever complete all degree requirements (D' Andres, 2002; Dorn &

Papalewis, 1997; Marcus, 1997). "Failure to continue in a doctoral program is not only painful and expensive for a student, but is also discouraging for faculty involved, injurious to an institution's reputation, and results in a loss of high level resources" (Ivankova & Stick, 2007, p.94). By definition, valid selection decisions may well improve these percentages.

Indeed, no matter how systematized the admission process is for choosing among candidates from a procedure perspective, routine procedural processes are no assurance of substantive outcomes (validity). Substantive outcomes require valid empirical information. However, only scant empirical information exists for guiding the decision making of faculty within the context of doctoral programs focusing specifically on educational leadership.

What exists concerning admission and graduation from doctoral programs focusing on education (as opposed to educational administration) are certain caveats. Most of these bear on unique personal characteristics of individuals and on formal program requirements. Unique personal characteristics include age as well as employment status of doctoral level education students (e.g., Isaac, Pruit-Logan, & Upcraft, 1995), and formal doctoral program requirements involve navigational skills when choosing and satisfying hurdles imposed either by the examination processes (e.g., qualifying examinations) or by a dissertation chair (e.g., Baird, 1997; Lipschutz, 1993).

Less addressed within this general educational literature is empirical information focusing specifically on doctoral programs in educational leadership as related to the probable success of students based on their academic proclivity as defined by standardized test scores and by grade point averages. To provide empirical data that impart important information for faculty as well as for potential applicants is the focus of this study. Within this study, particular academic predictors purported to influence the probable success of doctoral students are assessed for a specific doctoral program focusing on educational leadership. As a result, important insights are provided to those seeking admission (applicants), to those granting admission (graduate faculty), and to graduate schools setting

admission standards for doctoral programs focusing on educational leadership by considering the utility of certain purported predictors.

Predictor Variables

Most doctoral programs in educational leadership require applicants to submit indicators of their academic proclivity for doctoral work as part of the initial application process. These indicators include grade point averages and standardized test scores. Purportedly, these academic measures capture proclivity from two separate but related perspectives: (1) past academic performance (GPAs) and (2) future academic potential (GRE).

Past academic performance. Defining predictors for past academic performance are transcripts denoting performance at the undergraduate level and at the graduate level. For example, Creighton and Jones (2001) reported that 194 institutions require a minimum undergraduate grade point average of 3.00, 124 institutions require a minimum undergraduate grade point average of 2.75, and 132 institutions require a minimum undergraduate grade point average of 2.50 for admission. These findings suggest a great deal of variability across institutions. In some instances, these pre-establish standards may be made from a policy as opposed to an empirical perspective to reflect institution values.

Future academic potential. With respect to future academic potential of applicants for admission to a doctoral program in educational leadership, the most common measure is scores from a standardized test such as the GRE (Malone, et al., 2001). According to Norton (1994), great deference is afforded to standardized test results in the admission process. However, information about doctoral program expectations on standardized test is vague, especially as compared to undergraduate grade point averages (see Creighton and Jones, 2001).

To illustrate, the web pages of the top programs in educational leadership listed by the US News and World Reports (2004) failed to indicate a minimum score on the GRE for

applicant consideration. Furthermore, follow-up e-mails to these programs confirmed this omission with a single exception. The University of Wisconsin-Madison indicated on its web page that a combined score lower than 1100 on the GRE requires a written explanation by the applicant describing reasons for their level of performance that can be considered within the admission process (University of Wisconsin, n.d.). However, this additional procedural requirement is undocumented either by an institutional report or published in the professional literature.

In fact, the practice of using a combined GRE score for admission is questionable in the absence of a validity study. Guidelines, by publishers of the GRE address this specific procedure when describing inappropriate uses of the GRE. Noted specifically it is inappropriate to use "any measure involving a summation of verbal, quantitative, analytical, analytical score, or any subtest of these scores without first conducting and documenting a validity study for each measure" (p. 7, http://www.gre.org/scoreuse.html).

Beyond concerns about potential predictors of academic performance are issues related to the criterion variable. Validity assessments focus both on predictor and criterion measures. Noted in this body of literature are certain shortcomings relative to the criterion variable.

Criterion Variable

Most research addressing the predictive validity of academic measures for doctoral programs focuses on a single criterion. The most common criterion variable in this research is first year grades in a doctoral program. "Rarely are validity studies carried out using anything more than first year grades" (Smith & Garrison, 2005, p. 630).

Procedurally, the most common statistical approach is a multiple regression where first year grades are regressed on academic predictors. Shortcomings with this criterion variable and this statistical approach are noted. With respect to the criterion variable, "this

research does not include those students who were not admitted" (Smith & Garrison, 2005, p. 633).

For the statistical approach, "the existing research most often relies on the statistical technology concerning 'variance' to communicate the predictive power of test, leaving open the very real possibility of misusing the results in policy and legal settings" (p. 633). To address both of these shortcomings, this study examines a different criterion and uses an alternate statistical technique. First, the criterion variable in this study consists of three mutually exclusive categories: (1) those applying but rejected, (2) those accepted but failing to graduate, and (3) those accepted and graduating.

Second, by including this different classification scheme for a criterion variable, an alternate statistical technique is required. The statistical technique is a discriminant analysis. Rather than relying solely on variance accounted for via regression analysis, the discriminant analysis produces classification statistics for each category of the classification variable.

Validity Assessments for Academic Predictors and Classification Levels

Validity, as a psychometric property and as assessed in this study, is well documented from a procedural perspective in the professional literature according to established standards (Standards for Educational and Psychological Testing, 1999). Of importance to note in this literature is that validity is always associated with a population (a particular educational leadership program) and not with a specific set of predictors (GRE & GPAs) independent of a population, *per se*. That is, what is valid in one circumstance may or may not be valid in another circumstance because validity is population specific (Chronbach, 1990).

This restriction does not diminish, however, the importance of this type of study that focuses on a specific population and a particular set of predictors in the body of research involving admission to a doctoral program in educational leadership because there is a void

in this body of literature. To partially fill this void, Anastasi (1988) indicates that validity is the results from an accumulation of evidence rather than a specific study. "In principle, just as no one study alone can prove an experimental hypothesis true, no one study by itself can definitively establish validity" (Bryant, 2002, p. 102) of academic predictors for doctoral programs in educational leadership.

As such, two research questions are set forth, and these questions are submitted to empirical tests via null hypotheses. First, the viability of academic predictors (GRE & GPAs) is explored for doctoral applicants rejected, admitted but not graduating, and those graduating, and second, of concern is if findings from these assessments are sample specific. To address these research questions the following null hypotheses are subjected to empirical tests.

(1) It is hypothesized that a specific linear combination of academic predictors (GPAs & GRE scores) fails to distinguish among those rejected, those admitted but failing to graduate, and those graduating from a doctoral program in educational leadership (development sample).

(2) It is hypothesized that the stability of this linear combination is sample specific and will be unstable when applied to another random sample drawn from the same population (hold out sample).

Method

Population. The population for this study is all applicants (n=203) seeking admission to a particular doctoral program in educational leadership between 1991 and 2000 and taking the GRE as a means for satisfying admission requirements.[1] A single program was used to defined the population of this study and to assess performance on academic predictors relative to acceptable practice in the measurement literature (Bryant, 2002; Chronbach, 1990; Standards for Educational and Psychological Testing, 1999). This

particular educational leadership program is located in a Pacific coast state and is co-directed by both the California State University and the University of California System as a joint effort.

Like many doctoral programs in educational leadership, a cohort model is followed. Each year, approximately 15 students are accepted. As such, standards are set on quantitative predictors each year by applicant pools even though applicants exhibit little variability on these predictors across time.

Most importantly, this doctoral program serves a large geographical region encompassing approximately 120 public school districts. Some of the applicants for this doctoral program come from an urban school district listed as one of the largest in the United States, while other applicants come from small rural school districts enrolling less that 500 students. Demographic and academic characteristics of the participants in this study are quite varied, and specific characteristics of applicants as well as their measures of performance on purported academic predictors are found in Table 1.

Table 1: *Descriptive Statistics for Demographic and Performance Indicators*

Variable	N	Mean	Standard deviation	Minimum	Maximum
Sex[1]	203	1.52	0.49	1.00	2.00
Age	203	34.60	9.00	21.00	63.00
UGPA	203	3.07	0.43	2.00	4.00
GGPA	203	3.66	0.28	2.70	4.00
GRE Verbal %	203	44.49	29.04	1.00	98.00
GRE Quantitative %	203	31.47	24.24	1.00	99.00

Note.[1] Males coded 1 and females coded 2.

The population addressed in this study encompasses the time spanning from 1991 through 2000, and this time period is used for two reasons. First, it includes the initial applicant pool for this particular program stemming from program inception (1991).

Second, it encompasses all those applicants seeking admission and afforded at least 5 years for completing a 3 year program during this time period.

For this particular program, two years of structured course work in a cohort group is required, and the third year focuses on dissertation research. By using five years rather than three years for program completion, a two grace period is provided. The use of a grace period is typical in this type of research, and a two year grace period is used by several studies (see, Ivankova & Stick, 2007; Smith & Garrison, 2005).

Procedure

Data for this study were obtained from archival sources maintained by the doctoral program under consideration. As part of the application process, all applicants applying to this particular program are required to submit evidence of past academic performance on certain purported predictors and probable future academic performance as measured by standardized test scores. Only those completing the application process are included in this study.

Academic predictors. Purported academic predictors required for program consideration and included within the scope of this study are measures of past academic performance (Creighton & Jones, 2001) and indicators of future academic potential (Norton, 1994). Past academic performance is assessed through an analysis of official transcripts as obtained from files of applicants, and past academic performance measures are calculated both for undergraduate and for graduate grade point averages. Undergraduate and graduate grade point averages are measured on a 4.0 scale with high grade point averages being indicative of greater academic performance than low grade point averages.

Future academic potential of applicants is measured by GRE scores. Both the verbal and the quantitative sections of this test are used. Because applicants took the GRE in different years with different norm groups, standardized test scores, as reported by

percentile measures, are used. Furthermore, these test scores must be timely as suggested by the publisher.[2]

Criterion Variable. The classification scheme in this study is group status varied three ways among mutually exclusive categories of applicants. Included within group one is those applicants completing the application process but denied admission to the doctoral program. Group two contains those applicants admitted to the doctoral program but failing to graduate within the timeframe, and group three includes those applicants that completed the doctoral program within the prescribed time sequence covered in this study.

Statistical Analyses

A discriminant analysis is used to analyze these data, and this technique is appropriate in two situations concerning group assignments and discriminating variables: (1) predictive and (2) descriptive. In a predictive discriminant analysis group membership is unknown and is predicted, while in descriptive discriminant analysis group membership is fixed *a priori* on the basis of classification variables under consideration (Silva & Stam, 2001). Within this study, group membership is fixed (rejected, admitted but not graduating, & graduating), and the discriminating variables are quantitative measures of academic success (Creighton & Jones, 2001; Malone, et al., 2001; Norton, 1994).

More specifically, the discriminating variables are undergraduate grade point average, graduate grade point average, performance on the quantitative section of the GRE, and performance on the verbal section of the GRE. Group statistics broken down by levels of the classification variable rejected (n=129), admitted (n=28), or graduated (n=46) and performance on GPAs and on the purported GRE predictors are found in Table 2. As can be calculated from information contained in Table 2 about the classification variable, approximately 36% of those applying are accepted, and of those admitted (n=74) 62% actually graduated.

Table 2: *Descriptive Statistics by Group Classification*

Classification	Variables	Mean	Standard deviation	N	Range
Rejected	GRE V %	38.54	27.82	129	92.00
	GRE Q %	27.98	23.78	129	97.00
	UGPA	3.06	.42	129	1.89
	GGPA	3.64	.28	129	1.30
Accepted	GRE V %	56.14	26.89	28	94.00
	GRE Q %	36.17	22.04	28	83.00
	UGPA	3.01	.51	28	1.92
	GGPA	3.66	.30	28	1.00
Graduated	GRE V %	54.00	29.54	46	93.00
	GRE Q %	38.39	25.31	46	91.00
	UGPA	3.16	.43	46	1.95
	GGPA	3.72	.27	46	1.00

With respect to the total number of available candidates (n=203) on this particular set of discriminating variables, the relationship between sample size and the number of discriminating variables (n=4) exceeds minimum requirements as set forth by noted authorities. More specifically, Tatsuoka (1970) indicated "another rule is that the total sample size should be at least two or (preferably) three times the number of variables used" (p.38). The total sample size available in this study exceeds this minimum requirement because it is over five times the number of variables used in this study.

Because the relationship between the available sample size (n=203) and the number of discriminating variable (n=4) exceeds minimum expectations as established by authorities (5:1 as opposed to 2:1 as a minimum recommendation, see Tatsuoka, 1970), a hold-out sample was constructed with this population (Stevens, 2002). A hold-out sample is used to provide a stability assessment of weights derived with the development sample when applied to the hold-out sample and addresses the second null hypothesis. Using a random selection process for the total population, 123 participants within this defined population (n=203) were selected at random for the development sample (3:1, number of participants/number of variables, 41 subjects per variable), and 80 participants were

delegated to the holdout sample subsequently used as a validation group (at least 2:1, sample size/number of variables, 20 subjects per variable, see Tatsuoka, 1970).

By using random sampling procedures to construct a development sample and a holdout sample, both null hypotheses as previously set forth in this manuscript can be assessed but not without additional considerations. The development sample of participants (n=123), containing unequal numbers with respect to the classification variable (rejected=78, accepted=17, and graduated=28), suggests the need for a test for equality of covariance matrices. Although a discriminant analysis does not require equal sample sizes across classification levels (Tatsuoka, 1970), according to Stevens (2002) when "group sizes are sharply unequal, it is important to check the homogeneity of covariance matrices assumption" (p. 293).

Results of Box's M (f=.61, p=.90) indicate the likelihood for homogeneity of covariance matrices and support the use of a discriminant analysis with this sample of participants (Stevens, 2002) given both the ratio of sample size relative to number of discriminating variables (Tatsuoka, 1970) and the homogeneity of covariance matrices (Stevens, 2002). For the first null hypothesis, two discriminant functions were produced by the discriminant analysis because the classification variable used in this study includes three levels (see Table 3). To test the statistical significance associated with these different discriminant functions, Wilk's chi-square tests were performed, and results of these chi-square tests indicate that both discriminant functions (X^2=29.2, df=8, p>.05; X^2=9.50, df=3, p>.05) are statistically significant. Of the relative percentage of variance accounted for among these two different discriminant functions, approximately 69% of this variance is associated with the first function (see Table 3) and approximately 31% of this variance is associated with the second function, and these functions yield independent canonical correlations of .39 and .28, respectively.

Table 3: *Discriminant Functions for Academic Predictors*

Functions	Eigenvalue	Variance accountable	Canonical correlation	Chi square	Degrees freedom
First	.18	69%	.39	29.20*	8
Second	.08	31%	.28	9.50*	3

*p ≥ .05

Given that four variables are used to define the two discriminant functions, attention is redirected to interpreting these findings. Attention is afforded both to structural coefficients and to canonical discriminant coefficients because each imparts different information. The former pertains to the relationship of individual academic predictors (GRE scores & GPAs) relative to each discriminant function, while the later has implications for the relationship between the discriminant functions and the classification variable (applying but rejected, accepted but not graduating, accepting and graduating).

An examination of the structure matrix for the first discriminant function reveals that both the verbal section of the GRE (r=.94) and graduate grade point average of candidates (r=.42) correlate significantly with this particular discriminant function (see Table 4). For the second discriminant function, an examination of the structure matrix coefficients indicates that both the undergraduate grade point average (r=.80) and the graduate grade point average (r=.47) account for a substantial amount of variance (see Table 4).

Table 4: *Standardized and Structured Matrix Coefficients for Discriminant Functions*

Variables	Structure Coefficients		Canonical Dis. Coefficients	
	Function 1	Function 2	Function 1	Function 2
GRE V %	.94	.07	.94	-.33
GRE Q %	.37	.16	-.01	.67
UGPA	.06	.80	-.15	.90
GGPA	.42	.47	.36	-.08

To assess for redundant information among these particular discriminating variables associated with each discriminant function, attention is afforded to standardized canonical

discriminant function coefficients. With respect to the first function, percentile scores on the verbal section of the GRE are the predominant contributor (.94) with considerably less unique variance associated with graduate grade point averages (.36) (see Table 4). More diversified is the second independent discriminant function accounting for only 31% of the variance and involving measures of undergraduate grade point average (.90) and the quantitative section of the GRE (.67) as indicated by standardized canonical discriminant function coefficients (see Table 4).

Utility of these independent discriminant functions for classifying correctly the group membership of candidates (applying, admitted, or graduated) for this particular doctoral program focusing on educational leadership is assessed. Weights on the two discriminant functions derived with 123 applicants in the development sample are used to classify the group membership relative to this classification scheme (actual vs. predicted membership). Results of this analysis indicate that 68% of the individuals can be classified correctly using weights generated with the 123 applicants comprising the development sample (see Table 5) and reflects why the rejection of the first null hypothesis is suggested by statistical tests.

Table 5: *Classification Results for both Samples*

	Predicted Group Membership			
Classification	Rejected (1)	Accepted (2)	Graduated (3)	Total
		Development Sample[a]		
Rejected (1)	71.0	3.0	4.0	78
Accepted (2)	11.0	4.0	2.0	17
Graduated (3)	18.0	2.0	8.0	28
Rejected (1)	91.0	3.8	5.1	100
Accepted (2)	64.7	23.5	11.8	100
Graduated (3)	64.3	7.1	28.6	100
		Hold out Sample[b]		
Rejected (1)	41.0	3.0	7.0	51
Accepted (2)	10.0	0.0	1.0	11
Graduated (3)	14.0	2.0	2.0	18
Rejected (1)	80.4	5.9	13.7	100
Accepted (2)	90.9	0.0	9.1	100
Graduated (3)	77.8	11.1	1.1	100

Note[a.] 68% of selected original grouped cases correctly classified.

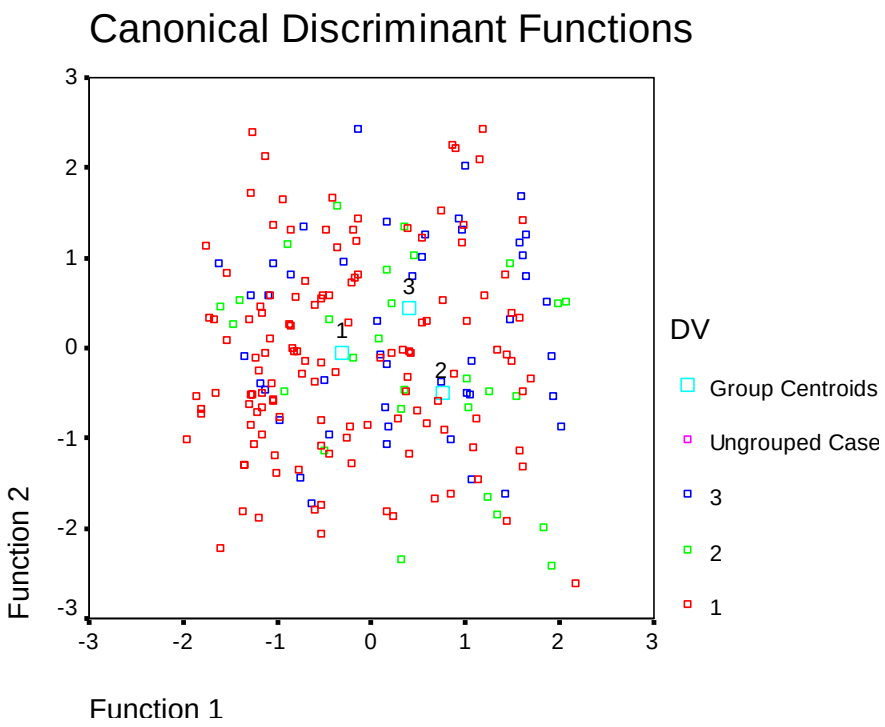
Note[b]. 54% of unselected original grouped cases correctly classified.

For the validation analysis designed to test the second hypothesis as set forth in this study, weights derived with the development sample (n=123) are applied to the hold out group of individuals assigned to the validation sample. This later group of applicants is

comprised of individuals (n=80) excluded from the initial analysis used to calculate the discriminant function weights (development sample) while maintaining minimum sample size requirements as suggested by Tatsuoka (1970) for a discriminant analysis. Results of the classification analysis performed with the hold out validation sample indicate that 54% of the applicants could be classified correctly using discriminant function weights calculated with the development sample (see Table 5) and suggest the rejection of the second null hypothesis.

Although groups used to define the classification variable are mutually exclusive, the manner in which these groups differ on these discriminant functions has important implications for practice in the field setting. These discriminant functions separate the groups on two levels. When group centroids are evaluated relative to the classification functions, these data indicate those rejected differ both from those accepted and those graduating and very little difference exists between the later two groups (see Figure 1).

Figure 1. *Group Centroids relative to Discriminant Functions for Group Classifications*



Discussion

A review of existing literature addressing the topic of this study suggests that it is largely an uncharted but an important area of investigation for doctoral programs focusing on educational leadership. Smith and Garrison (2005) noted that "rarely are validity studies carried out using anything more than first year grades" (p. 630) as a criterion variable and that "this research does not include those students who were not admitted" (p. 633). In light of these voids, this study uses a different criterion variable and fulfills a void in this body of literature.

Addressed in this study are those applying but failing to be admitted, those admitted but failing to graduate, and those admitted as well as graduating. To explore differences among levels of the criterion variable only quantitative predictors are addressed. According to Malone et al. (2001) "GRE scores and grade point averages represent quantitative measures in the decision making process" (p. 3) used to delimit initial applicant pools.

By affording considerations to these ways of variation involving admission, matriculation, and graduation, results suggest that the concerns of applicants and the practices of faculty are not entirely misplaced for the academic predictors considered in this study. However, performance on some of these academic predictors is found to be more important than performance on other academic predictors.

Performance on the verbal portion of the GRE is the most valid predictor for this particular doctoral program focusing on educational leadership. At first glance, this finding for the GRE verbal would seem to be a "self-fulfilling hypothesis" (Rosenthal & Rosnow, 1975). It was a self fulfilling hypothesis in the sense that graduate faculty members were privy to this information at the screening stage of the selection process and eliminated all low performing candidates on the basis of their verbal test performance on the GRE as well as on indicators of past academic performance (GPAs).

However, a review of the descriptive information depicted for the total pool of candidates under consideration (see Table 2) indicates that a number of those candidates admitted or graduated exhibited low scores on the GRE as well as on GPAs. Because some of those admitted and some of those graduated possessed low scores on the GRE and GPAs like those denied, these data refute, at least in part, the notion of a self-fulfilling hypothesis as a viable explanation for these findings.

Emerging from these findings is the importance of two variables for defining the first and most important discriminant function: GRE Verbal and GGPA. Definition of this function is subjective like factors derived from a factor analysis. However, common to both of these variables is a heavy reliance on verbal skills.

Across these results, a latent variable defined loosely as verbal as opposed to quantitative ability is suggested by an interpretation of variable weights derived from the discriminant analyses. The importance of verbal ability is reflected not only by standardized test measures (GRE Verbal) but reinforced by the graduate grade point average. Graduate grade point average of applicants encompasses, no doubt, emphasis on written

performances, assigned papers, and classroom projects when assigning grades that are considered ultimately within the graduate admission process.

Reinforcing the above speculation is the finding that standardized test scores (Verbal section of the GRE) and graduate grade point average exceeded always the importance of undergraduate grade point average given that the effects of a standardized test score and GGPA are considered *a priori* within the analyses. Undergraduate grade point average surfaces as a valid predictor only in a second discriminant function and then so only after graduate grade point average is considered.

For a multitude of reasons (youth, academic focus, undecided major, etc.), many undergraduate students may be less than serious scholars during their early college years (see Table 1). This is suggested, at least in part, by the lower mean undergraduate grade point average (M= 3.07, see Table 1) and the larger standard deviation for UGPA (SD= .43, see Table 1) when compared to graduate grade point averages (M= 3.7. SD= .28, see Table 2). Not to be overlooked within these data is the fact that undergraduate grade point average, unlike performance on the other academic predictors, is associated with distal as opposed to proximal academic performance on the part of applicants, and it would seem only logical that preference should be given to proximal as opposed to distal performance of potential doctoral students when making admission decisions by the graduate faculty.

Turning attention from results focusing on the rejection level of the classification variable to the admitted and graduated levels of the classification variable, these data are somewhat blurred. The lack of ability to distinguish clearly between those admitted and those graduated should not be surprising given that this study examined only a specific set of academic predictors and that many but not all candidates admitted exhibit the necessary academic abilities on these purported predictors (see Table 2). Given the prerequisite level of ability on the academic predictors as possessed by these two different groups (admitted and graduated), other factors obviously come into play that may differentiate between

those accepted and those graduating, and additional research needs to explore these other variables (e.g. letters of recommendations, experiential data, and/or goals of candidates).

Because these later two groups (admitted and graduated) are least distinguishable (see Figure 1), additional limited speculations are provided based largely on reflective practice of the author. It is quite possible that many of the academically able students lose interest in obtaining a doctoral degree due to life changes (e.g. marriage, mortgages, children), obtained ultimate career objectives during enrollment without the benefit of a doctoral degree (got the desired job, etc.), or failed to refine a workable dissertation project necessary for consummating the matriculation process for reasons other than academic proclivity (e.g., lack of mentoring, faculty expertise in an area of interest, etc.). To assess the validity of these reasons, the interactions among these reasons, as well as other potential causes, additional research is needed that goes well beyond the scope of this study.

Implications and Limitations

Clearly, the specific set of academic measures assessed in this study are not the sole predictors used to delimit applicant pools for making screening decisions but are among the most common quantitative measures (Creighton & Jones, 2001; Malone, et al., 2001; Norton, 1994). Selection decisions should and do involve other sources of data about applicants, e.g., reference information, writing samples, and personal interviews as suggested by Malone, et al. (2001). In fact, some authorities (i.e., Heneman & Judge, 2006) as well as federal statues (i.e., EEOC , 1978) suggest that selection should be governed by multiple predictors but the utility of each predictor should be assessed relative to predictive validity (Anastasi, 1988; Chronbach, 1990; Standards for Educational and Psychological Testing, 1999), and this study focuses on some of the most common predictors used by faculty and graduate schools for doctoral programs in educational

leaderships, i.e. GPAs and GRE scores (Creighton & Jones, 2001; Malone, et al., 2001; Norton, 1994).

In keeping with these recommendations, only specific academic predictors are assessed in this study as part of the total selection process for doctoral candidates seeking admission to a particular doctoral program. The doctoral program under investigation is held constant like the specific predictors considered in this study but for a different reason. Most notably is that past research in this has suggested that validity studies are most informative when conducted within a program context as compared to across programs/institutions (Anastasi, 1988; Chronbach, 1990; Standards for Educational and Psychological Testing, 1999).

Beyond holding constant both the set of academic predictors investigated and the doctoral program under consideration, this study uses a research protocol that involves actual decisions from the field setting as well as assesses the stability of these decisions across random samples drawn from the same population. More specifically, archival data are used to capture known outcomes for the doctoral admission process, and equations are generated for capturing these decisions with a specific random sample of individuals implying the rejection of the first null hypothesis. Equations, as generated with this initial random sample, are applied to a subsequent sample, found to be stable (statistically significant), and suggest the rejection of the second null hypothesis.

Based on these results and by following this research protocol by other doctoral programs in educational leadership, important information is provided to potential applicants seeking admission to a doctoral program in educational leadership, to graduate faculty making admission decisions, and to graduate schools setting standards for a doctoral program in educational leadership. For potential applicants exploring educational opportunities afforded by a doctoral program in educational leadership, many are concerned about formal criteria involving grade point averages and GRE test scores prior to seeking admission or prior to taking the GRE. That is, they want to know *a priori* what it takes to be

considered as a viable candidate, and by following this research protocol such information can be provided to potential applicants as a means of attraction.

Almost without stating, graduate faculty must make proactive decisions about applicants seeking admission to a doctoral program in educational leadership. Published sources indicate that faculty members rely on specific academic predictors (Creighton & Jones, 2001; Malone, et al. 2001; Norton, 1994), at least in part, to make these decisions. To date, this reliance, in most all instances, is, no doubt, from a subjective as opposed to an empirical perspective. By using the predictive validity paradigm as set forth in this study, faculty for any particular doctoral program can be equipped with data from an empirical perspective to guide their decision making relative to the admission process in an informed manner.

Often overlooked within the admission process for doctoral students in educational leadership is the role of the graduate school (Ivankova and Stick, 2007). Graduate schools represent the university at large and are charged with protecting the intellectual/academic integrity of the institution within the admission process. Baring existing empirical information that is program specific as provided in this study, criteria for academic standards are set usually from a university as contrasted to a program perspective, and this study provides means and methods for tailoring requirements to specific programs.

Although validity assessments for individual programs are labor intensive, the importance of this undertaking cannot be over emphasized. This point is well sounded by publishers of the GRE. "Departments using the GRE scores for graduate admission, fellowship awards, and other approved purposes are encouraged to collect validity information by conducting their own studies" (ETS, 2004).

Finally, this study, like all studies, is subject to specific limitations, but many of these limitations are strengths as well as weakness because this investigation navigates relatively uncharted waters within the educational administration domain. More specifically, this study focuses on a particular educational leadership program but other researchers

(Nunnally, 1990) and other organizations (Educational Testing Service, 2004; Standards for Educational and Psychological Testing, 1999) indicate that validity assessments should be program specific, and this study examines a specific array of academic predictors that other studies suggest are used by most institutions of higher education (Creighton & Jones, 2001) and are afforded great deference within the admission process (Norton, 1994). Given the specifics of this study relative to program and academic predictors, any generalizations beyond these data should be made with caution pending additional research.

Notes

1. Scores on the GRE are deemed invalid that exceed five years by this publishing company (Educational Testing Service, 2004).
2. Some candidates took the Miller Analogy Test to satisfy admission requirements.

References

Anastasi, A. (1988). *Psychological testing and assessment* (2nd ed.). Boston: Allyn & Bacon.

Baird, L.L. (1997). Completing the dissertation: Theory, research, and practice. In M. Kramer, L.F. Goodchild, K.E. Green, E.L. Kratz, & R.C. Kluever (eds.), New Directions for Higher Education, San Francisco: Jossey-Bass.

Bryant, F.B. (2002). Assessing validity of measurement. In L Grimm & P. Yarnold (eds.). Reading and Understanding more Multivariate Statistics, American Psychological Association, Washington D.C.

Creighton, T.B. & Jones, G.D. (August, 2001). Selection of self selection? How rigorous are our selection criteria for educational administration programs. Paper presented at the Conference of National Professors of Educational Administration. University of Houston: Houston , Texas.

Chronbach, L. J. (1990). *Essentials of psychological testing*. New York: Harper & Row.

Dorn, S.M. & Papalewis, R. (1997). Improving doctoral retention. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

D'Andres, L.M. (2002). Obstacles to completion of the doctoral degree in colleges of education: The professors' perspective. *Educational Research Quarterly*, Vol. 25.3, 42-58.

Educational Testing Service (2004). Guidelines for the use of GRE scores. Retrieved June 21, 2007 from http://www.gre.org/scoreuse.html.

Equal Employment Opportunity Commission. (1978). Guidelines on employee selection procedures. Federal Register, 35, 12333–12336.

Heneman, H. G. & Judge, T. A. (2006). *Staffing Organizations* (4th ed.). Boston, MA: Irwin/McGraw-Hill.

Isaac, P.D., Pruitt-Logan, A.S., & Upcraft, M.L. (1995). The landscape of graduate education. In M.Barr (Series Ed) & A.S. Pruitt-Logan, & P.D. Issac (Vol. Ed.), *New Directions for Student Services*: No. 72, Winter.

Ivankova, N. V. & Stick, S. L. (2007). Student persistence in a distributed doctoral program in educational leadership in higher education. *Research in Higher Education*, Vol. 48, 1, p93-135.

Lipschultz, S.S. (1993). Enhancing success in doctoral education: From policy to practice. In L.L. Baird (Ed.), *New Directions for Institutional Research*, San Francisco: Jossey-Bass.

Malone, B. G., Nelson, J. S., & Nelson, C. V. (2001). Completion and attrition raters of doctoral students in educational administration. Ed 457759.

Marcus, M.B. (1997). Half a doctor, *US News and World Reports*, April 17.

Newton, R.M., Giesen, J., Freeman, J., Bishop, H., & Zeitoun, P(. 2003). Assessing the reactions of men and women to attributes of the principalship. *Educational Administration Quarterly*, Vol 39, 4, 468-503.

Newton, R. M., Winter, P.A., & Keedy, J.L. (2000). Teacher attraction to school councils service in Kentucky: Implications for national reform. *Planning and Changing*, Vol. 32 No. 1, 84-103.

Norton, M.S. (1994). *Student recruitment and selection in educational administration programs. Arizona State University*. (ERIC Document Reproduction Service No. ED. 366 087).

Pounder, D.G. & Merrill (2001). Job desirability of the high school principalship: A job choice theory perspective. *Educational Administration Quarterly*, 37, 27-57.

Rosenthal R., & Rosnow, R. (1975). *Primer of methods for the behavior sciences*. John Wiley & Sons, Canada.

Silva, P.D. & Stam, A. (2001). Discriminant analysis. . In L Grimm & P. Yarnold (eds.). Reading and Understanding more Multivariate Statistics, American Psychological Association, Washington D.C.

Smith, D. G. & Garrison, G. (2005). The impending loss of talent: An exploratory study challenging assumptions about testing and merit. *Teachers College Record*, 17, 4, 629-653.

Standards for Educational and Psychological Testing (1999). *American Educational Research Association*, Washington DC.

Stevens, J.P. (2002. *Applied multivariate statistics for the social sciences (4th ed.)*. Lawrence Erlbaum, Mahwah:NJ.

Tatsuoka, M.M., (1970). Discriminant analysis. *Institute for personality and ability test*,,Champaign, IL..

University of Wisconsin, (n.d). Admission requirements and application procedures. http://www.education.wisc.edu/elpa/admissions/general_adm.html.

US News and World Report (2004). Graduate school rankings. (http://www.usnews.com/usnews/edu/grad/rankings/edu/brief/edusp01_brief.php.

Winter, P.A. (1996). Applicant evaluations of formal position advertisements: The influence of sex, job message, content, and information order. *Journal of Personnel Evaluation in Education*, 10, 105-116.

Winter, P.A. & Jaeger, M.G. (2004). Principal selection decisions made by teachers: The influence of principal candidate experience. *Journal of school Leadership*, 14(4), 411-433.

Winter, P.A., Keedy, J., & Newton, R.M. (2001) Teachers serving on school decision making councils: Predictors of teacher attraction to the job. *Journal of School Leadership*, 10(3), 249-263.

Winter P.A. & Morganthal, J.R. (2002). Principal recruitment in a reform environment: Effects of school achievement and school level on applicant attraction to the job.

I. Phillip Young is a Professor of Education at the University of California-Davis and Director of a Joint Doctoral Program in Educational Leadership between the University of California and the California State University-Fresno. His area of research is human resource functions related to selection and to compensation. He has published a leading textbook now in its 9th edition "The Human Resources in Educational Administration" and over 60 referred articles addressing selection and compensation within the public school setting.