Complex Admission Selection Procedures for a Graduate Preservice Teacher Education Program

By Micki M. Caskey, Kenneth D. Peterson, & Jacqueline B. Temple

A perennial problem of preservice teacher education programs is to select the highest quality candidates and to deny admission to those unsuited to work in schools. In addition, for programs having more applicants than available openings, the professional gate-keeping task presents an opportunity to support high teacher quality. For these admission judgments most current teacher training program admissions rely upon completion of required courses, minimal academic achievement (GPA), letters of recommendation, and interviews. However, because of the lack of definitive agreement in the literature about how to select the best teacher candidates, there is a need to continue the search for most effective admission selection procedures.

Micki M. Caskey, Kenneth D. Peterson, and Jacqueline B. Temple are professors in the Graduate School of Education at Portland State University, Portland, Oregon.

The purpose of this study was to validate and refine a complex competitive admissions procedure for a graduate preservice teacher education program. Specific tasks included estimating reliability of components (e.g., ratings of recommendations and personal statements), exploring the relationships among a

battery of applicant measures and assessments, and using a simulation performance activity in lieu of the often-criticized interview procedure. The two criterion standards for this study were (a) decisions to admit or deny, and (b) student success in program.

Background

Many authors have called for complex data to be used for teacher preparation program admission decisions. These include measures or other information about academic success, personal characteristics, academic aptitude, success in occupations or volunteer work, and interpersonal skills. There are significant questions about how best to document, assess, and judge based on backgrounds and performances of candidates. In addition to concerns about candidate quality, there are questions of logistics and efficiency: what techniques are affordable in terms of applicant and decision maker time and money? This background section will briefly review literature key to the design and interpretation of this study.

Schmidt, Ones, and Hunter (1992), reviewed general personnel selection techniques that are used in educational program admissions, occupational hiring, and promotion decisions. They listed as possible predictors of future success: aptitude and ability tests, biodata, personality and related predictors, interviews, and additional assessment measures such as work samples and simulations. The task of teacher education program designers is to incorporate these possible techniques into a coherent, valid, reliable, and practical admissions procedure.

Hunter and Schmidt (1989) reported a meta-analysis study, which examined intercorrelations among predictor variables from a large group of studies on job success prediction. They found General Cognitive Ability to be correlated with measures of Job Knowledge (0.80), Job Performance (0.75), and Supervisor Rating (0.47). Also Job Knowledge is correlated with Job Performance (0.80) and Supervisor Rating (0.56). Hunter and Schmidt (1989) further found that Job Performance is correlated with Supervisor Rating (0.52). Multiple correlations of all the variables were predictive for Job Performance (R=0.82) and Supervisor Rating (R=0.57). These authors also showed how the validity of the correlation of various measures deteriorates under less-than-ideal measurement conditions. For example, the correlation of General Cognitive Ability and Job Performance lowers in empirical trials from 0.75 to 0.20 because of the net impact of necessary application imperfections (too few judges, limited time for testing, unintended bias) and sampling errors.

Interviews have long been a staple of personnel selection (Eder & Harris, 1999). Their face validity satisfies many intuitive concerns of decision makers. However, there is abundant literature to suggest that interviews are not a consistently reliable and valid selection process. Webster (1964) found that interviewers in his study typically made their personnel decision some four minutes into the hour-long process, and used the remaining time to justify the first impression. Dailey (1982) reviewed six summary papers concerning interviews as valid selection procedures

and concluded, "not one confirms the soundness of interviewing as ordinarily practiced" (p. 13). Messmer (1998) stated, "the issue of interviews being accurate performance indicators has been studied extensively over the past 40 years, but with mixed results" (p. 110). Shechtman (1988) found in teacher education program admissions that the predictive validity of the individual interview for initial teaching success was zero. Thus, it is clear from the literature that teacher education programs should seek admission data separate from (or in addition to) interviews.

Current Practice in Teacher Education Program Admissions

Shank (1979) reported on data used in 187 teacher education program admissions before current educational reform efforts. He found that institutions used GPA (96.8 percent), language proficiency (77.5 percent), coursework (61.9 percent), interviews (53.8 percent), personal history (52.5 percent), personality and values measures (43.7 percent), reference letters (41.8 percent), and standardized tests (23.1 percent). Laman and Reeves (1983) found in 121 AACTE member institutions that a committee made admissions decisions in 53.7 percent of cases, while minimum criteria-based decisions were made in the remainder. These authors reported that program evidence included: GPA (95 percent), speech test (48.8 percnt), interview (41.3 percent), standardized test (41.3 percent), written language test (38 percent), physical examination (19 percent), and psychological examination (6.6 percent). By 1987, Lehmann and Phillips reported that 27 states mandated teacher candidate testing of academic achievement and aptitude.

Gunne and Peterson (1990) reported that an academic aptitude test ("General Cognitive Ability") was a useful contributor to teacher education program admissions. They found that the test scores correlated with program course instructors' candidate ratings (r=0.40), and the verbal section correlated with the education course GPA (r=0.35). Student teachers with high academic aptitude test scores were described as more satisfied with the program's expectations; more reasonable when responding to program demands; and more critical of program courses. Additionally, these students were less authoritarian in classroom discipline while student teaching; more realistic in their views about preventing teacher burnout; and possessed greater enthusiasm toward their future careers.

Smith and Pratt (1996) found that a structured personal statement (rated by trained analysts) combined with academic data was an important predictor of teacher education program success. Their procedure called for applicants to write a two-page statement of their reasons to become a teacher and accounts of life experiences considered relevant to teaching. Two judges assigned a score from 1 to 10 on each statement. "Teaching experience" and "travel experience" were influences on scoring, which was controlled for gender of applicant; statements were typed rather than hand written. Inter-rater reliability between pairs of judges was estimated to have a moderate correlation of r=0.57.

Andrew et al. (1996) reported a teacher education program admission system

that included undergraduate GPA, grades in education courses, Graduate Record Examination test scores, recommendations from professors and cooperating teachers in an early experience course, self-assessment paper, and a personal statement. They reported that approximately 25 percent of applicants are denied admission in their program. Their study found that the only factor predicting problems in student teaching was the cooperating teachers' judgment of weak performance in early field experiences.

While academic work certainly is a central task for school teachers, the dimension of personal qualities appears to be an equally important, and perhaps more difficult to measure, consideration for success in teaching. Andrew et al. (1996) listed positive personal qualities for teachers as: organization, creativity, reflection, commitment, hard work, initiative taking, positive attitude, enthusiasm, humor, and rich life experience. Negative personal qualities identified by these authors were: negativity in personal interactions, resistance to feedback, inability to accept criticism, lack of organization, low energy, lack of hard work, and limited life experience. Certainly personal characteristics have had a long history of study and assessment in teacher education (e.g., Barr, 1931; Boyce, 1915), but the central problem for current research is to find effective, reliable, and efficient ways of assessing personal characteristics for teacher education program admissions.

Small Group Selection Procedures

At least two teacher education programs have developed small group selection procedures, which add very important perspectives to understanding candidate quality. Roose, Mitchell, and Rudman (1985) reported a selection procedure that included "small group challenges" in which candidates were given a creative subject matter task to solve and present to a larger group of candidates. Examples of these tasks were to map something not visible, write a complex drama scene, or invent a device to protect an egg in a 30-foot drop. Then, the small group presented their solution and process to a larger group. During this presentation, program staff noted roles, styles, communication, and personal characteristics. These observations were combined with two personal interviews to acquaint two staff members with each candidate. The staff members then advocated for or against admission in a full staff session. The authors report success in screening for candidates likely to be well received by districts, and a lessening of staff time and energy in dealing with candidates unsuited for the complexities of human communication inherent in school teaching.

Shechtman (1988, 1992, 1998) developed a two-hour group assessment procedure for predicting initial teaching success. This procedure was based upon the assessment center method, which often includes a battery of simulation experiences in which complex human performances can be practiced, judged, and assessed. The Shechtman procedure focused on three clusters of behaviors related to teaching performance: cognitive/intellectual, communication/language, and socioemotional/

feelings and interactions. Shechtman (1991) reported a predictive validity correlation of r=0.42 of group assessment rating with initial teaching success.

Additional Concerns about Admissions

Studies of teacher education program admissions also include considerations about decision making beyond predictive accuracy. For example, Benner, George, and Cagle (1987) recommended a more clinical or professional approach by using an admissions board. The recommended board would consist of faculty, teachers, students, and administrators. This board would look at personality data, basic skills, speech and hearing, academic records, and then hold interviews. The board would continue its involvement by acting as a mentoring team. In distinction to this clinical decision-making, Borman (1982) presented empirical evidence that a "mechanical composite," which adds up component values of evidence concerning personal behavioral assessment is a reliable predictor of ultimate performance.

Another concern about admissions is the need for extensive work on moral questions. Morality in this sense is not a certain kind of action but a view of conduct in connection with the effects it obtains. For example, Stengel and Tom (1995) raised issues about what should be included in admissions and retention of teacher candidates. One issue is the moral fitness of candidates: their honesty, caring, courage, fairness, and practical wisdom. Another issue is the number of students admitted to a program: can the institution provide the needed relationships, advocacy, and human interaction necessary to support growth of a competent professional. Other questions for researchers concern whose interests should be included at the point of admissions to the profession. In addition, the specific demands placed on applicants should be thought through; for example, the expense of extensive test programs may be an unfair barrier to less affluent applicants. The emotional stress of performing and revealing in a simulation situation should be understood, justified, and minimized. Decisions such as whether or not candidates should be called upon to judge each other (Shechtman, 1991) should be carefully considered and developed. Designers should consider what explanations should be given to those who experience the process, including feedback for those not accepted.

Finally, well-designed teacher preservice program admissions should take into account the resources that are available for decision-making. Study should be made about how to best manage time, materials, and dollar costs of the variety of people involved in the decisions. A variety of judges from other university departments, public school teachers and administrators, and county, state, and service-district level participation is desirable. Certainly cost-benefits analysis of money and time is an important part of research and development of educational programs (Thompson, 1980).

Implications of Literature to This Study

This study was designed to implement the best practices described in the

literature and to extend inquiry into practices described by researchers as needing further inquiry. For example, many authors have described the need for extensive data in determining applicant quality. This study included a wide variety of biodata, writing samples, standardized test scores, recommendations, and observation of candidates. However, our study did not include interviews or personality assessment tests because of the criticism of their use as found in the literature. Another example of the pertinence of research literature to this study is the pioneering work of small group selection procedures, a technique that seemed to deserve further work. Likewise, the ongoing question of rater reliability in scoring data calls for more consistent techniques and documentation of results. Finally, the constraints on resources for admissions decisions as described in the literature led us to examine our use of time, personnel, and dollar costs.

Methodology

The first task of analysis of this study was to estimate the relationships among selection variables (concurrent validity). This is important to understand how each worked and how much confidence to put into its future use. The methodology of this study featured content analysis of biodata, observation, reliability analysis, and multiple regression analysis. The second task of analysis was to estimate the effectiveness of each selection variable and their combinations in predicting success in the program (predictive validity). For this task the researchers employed multiple regression analysis of admission variables on program success. Both of these tasks were aimed to improve our admission procedures by increasing our knowledge of how our assessments and judgments worked with an actual population of applicants.

Population

The population for this study was 141 applicants to a four-quarter term (full year), graduate teacher preparation program at an urban university. All candidates had baccalaureate degrees in majors other than education; 70 (50 percent) were graduates of the institution in this study, 46 were graduates of other colleges and universities in the state, and 25 applicants were graduated out-of-state. Twelve ethnic minority students applied, which was 8.5 percent of the total. Fifty-four percent of applicants had undergraduate GPAs of 3.2 or higher (3.8-4.0, 6 percent; 3.5-3.79, 22 percent). Thirty-seven percent of applicants were in the 20 to 25-year age range, and 63 percent in the 25 to 55-year age range (30 percent, 30-55). Of the entire population, 82 were admitted and 59 were denied.

Data

Ten selection measures constituted concurrent and predictor data for this study. Included were: undergraduate grade point average, letters of recommendation that address candidate's interpersonal skills and academic potential, a personal state-

ment of competence and intent, standardized test scores on a test of basic academic skills (reading, writing, mathematics), an overall admissions file rating, and performance on a group problem-solving simulation activity. Also included in this analysis, but not used as a selection variable, was the candidates ranking of the importance of multicultural education relative to other possible teacher preparation goals, such as working with parents or countering school violence (these data are used in the program to design instruction). Variables of gender, preference of grade level of teaching, and ethnicity were included in this analysis, but not used in admission decisions. Although scores on subject-matter knowledge tests were used in the admissions procedure, the non-uniform scale score reporting precluded analysis in this study. Finally, completion of prerequisite academic courses was a consideration in admissions but not included in this analysis.

Personal statement. Candidates submitted a two page (minimum) statement of their intent, interest, motivation, and background to become teachers. Two raters independently scored the quality of the statements on a 7-point scale based first on the content and secondly on the craft of their writing; the sum of the two judgments was used as the statement score.

Recommendations. Three letters of recommendation were required. The letters reported suitability to be a teacher and, in some cases, aptitude for graduate programs. Priority was placed on recommendations from school settings, with the writer having observed the candidate working with students. Some recommendations reported work rather than educational settings. Two raters independently scored the quality of the statements on a 7-point scale; the sum of the two judgments was used as the recommendation score.

Simulation activity. Small groups of applicants (4-6) completed a problem-solving simulation activity as part of the admissions decision. Each group was given 20 minutes to discuss and create a collaborative response to a written prompt, such as planning a 2-hour-per-day, 6-week course for 20 students. Observers for the small group simulation included teacher education faculty, school of education administrators, public school principals, faculty from other campus departments, and a researcher from an educational service district. Candidates were observed and rated for group and process skills, including communication and idea development. Candidates were rated on a 5-point scale for global quality using a discussion and consensus report of two or three observers in each group. Scores for analysis were reported in tenths (e.g., 4.3).

Standardized tests. Candidates were required to complete standardized, nationally normed tests of basic educational skills and subject matter knowledge. Basic skill scores were reading, writing, mathematics, and their test total. Subject-matter knowledge tests were of general knowledge for elementary level candidates, and specialty area tests for middle- and high-school level applicants.

Overall file quality rating. Two reviewers independently rated the overall quality of the applicant's admission file. This judgment included consideration of GPA, recommendations, course work, standardized test scores of basic academic skills and subject matter knowledge, essay, evidence of important life experiences related to education, and unique information supplied by each candidate. Two raters independently scored the quality of the files on a 7-point scale; the sum of the two judgments was used as the overall file score.

Data Analysis

There are several ways to analyze the admissions data to better understand the judgments made by the faculty. One method is to consider the ten data sources as independent and separate indicators of quality. For this approach we used descriptive and correlational analysis with individual selection criteria measures. A second method of analysis is to consider the possibility that the data sources were not unrelated, but in fact expressed in different ways some fewer number of underlying traits or factors inherent in the measures or recognized by the raters in their analytical judgments. Using this point of view, the data were subjected to factor analysis. Third, the separate data sources could be related to the empirical decision to admit or deny, a judgment requiring the expertise of the admission faculty. A discriminant analysis was performed to estimate the reliability of the judgments made by the faculty and to better understand the role each data source may have played in this complex decision making process about candidate quality. Finally, the reliability of independent judgement ratings of individual performance components, such as essay and references, was a question for study.

Estimates of reliability of judgment ratings of References, Essay, and overall File were made by computing first order correlation coefficients (r) between ratings for each of these variables. Population Differences based on level of teaching interest (elementary or secondary) were examined with a series of t-tests; each comparison first checked for equality of subpopulation variance. Factor analysis was begun with a Bartlett Test of Sphericity to estimate suitability of the process for finding an underlying factor structure in the data. The factor analysis was a principal components analysis, with a varimax rotation and a lower eigenvalue acceptance limit of 1.0. The discriminant analysis began with test of equal group variance (Box's M) to assure assumptions of population characteristics. The discriminant analysis continued with a stepwise variable selection with the rule of minimizing Wilks' Lambda at each step. The variables used in this analysis were: Writing test, Reading test, Undergraduate GPA, Essay, References, and Simulation; these were selected from the correlation matrix as related variables. A canonical discriminant function was calculated, along with associated eigenvalue, coefficients with variables, total Wilks' Lambda, and classification results of percentages of grouped cases "correctly" classified (using this empirical discriminant function).

For the predictive validity question of this study, selected predictor variables

were included in a multiple regression analysis on the rating categories of candidates. This candidate scoring was made by the two group leaders who reviewed all program data (including instructor reports, field supervisor reports, and grades) and observed candidates in student teaching classrooms. Product-moment correlations were used to describe the individual indicators' relationship with the categories; a multiple correlation (R) was used to indicate explained variance for the entire set.

Findings

The findings of this study pertained to the two questions: first, how did the selection variables work concurrently in relation to the decisions to admit or deny, and second what was the predictive validity of the most salient selection variables and their total effect? This section first will briefly address the reliability of three admission variables that required the combined judgment of two raters, then present the empirical function of the variables in the admission decisions, next examine the predictive results of variables, and finally discuss the cost data.

Reliability Analysis of Multiple Judgments

Three of the admission assessments required the combined judgment of two raters on a 7-point scale. Correlations between raters were calculated on the variables of References (r=0.75), Essay (r=0.78), and File-overall (r=0.76).

We were interested in judgment procedures that could raise the reliability of this procedure. One method is to divert individual applicant judgments that vary to a third judge for a more reliable rating. The paired judgments which differed by 3 points on the 7 point scale can be referred to a third reviewer. If in these decisions, 8.5 percent of all paired reviews are eliminated from the reliability analysis, the remaining judgments (with agreements within 2 points) show a bivariate correlation of References 0.87, Essay 0.87, and File-overall 0.88. If the rule is to send all judgment pairs of 2 or 3 points (16.9 percent), the corresponding correlations rise to 0.92, 0.92, and 0.94, respectively. These data have been presented to the Program and Policy Committee for possible adoption.

Concurrent Validity of Variables in the Admission Decisions

Correlation Matrix of Individual Measures. The correlation matrix of ten selection variables and the admission decision is presented as Table 1. Inter-variable correlations ranged from 0.88 (Essay score and overall File score) to -0.22 (Rank of diversity inclusion in classroom and Reading score).

Factor Analysis of Measures. A factor analysis of the ten admission measures was performed to explore the possibility of underlying common factors (overlapping measures) in the predictor variables. This analysis provides information about how the measures work. It also possibly could lead to a reduced number of measures to be taken into account.

Table 1
Correlations Among Selection Variables

Data	Read-ing	Writ-ing	Math	Test Total	Essay	GPA	File	Simu-lation	Rank MC	Refer-ence
Writing	.23	—								
Math	.48	.19	—							
Test Total	.82	.62	.78	—						
Essay	.24	.17	.13	.20	—					
GPA	.25	.09	.24	.28	.42	—				
File	.18	.16	.06	.18	.88	.44	—			
Simulation	.14	.21	.11	.21	.02	.02	-.03	—		
Rank MC	-.22	-.06	-.11	-.19	.18	-.09	.06	-.10	—	
Reference	.23	.12	-.06	.12	.77	.42	.81	.01	-.02	—
Accepted	.20	.20	.08	.18	.68	.43	.62	.32	.13	.59

The Bartlett Test of Sphericity showed a value of 112.03, with a *p* of <.01 that the null hypothesis of no actual underlying factor structure, given this sample, was true. This means that the measures had an overlap of information and a possibility of reduction in number.

The results of the principal components analysis are present in Table 2. Two separate factors with eigenvalues greater than 1.0 were identified. Factor one related to *background intent and competence* evidenced by references, essay, and grade point average. Factor two related to *verbal performance* evidenced in the simulation procedure and test of writing ability.

Discriminant Analysis. After reviewing the admission data, the selection committee decided to admit 82 of the 141 applicants (58.2 percent). Since the decision was made as a majority judgment, a discriminant analysis was performed

Table 2
Factor Analysis

Factor structure:

Factor	eigenvalue	% var	cum %
1	2.30	38.4	38.4
2	1.24	20.7	59.1

Rotated factor matrix:

	Factor 1	Factor 2
References	.887	.043
Essay	.882	.083
GPA	.686	.074
Reading test	.350	.534
Writing test	.103	.728
Simulation	-.116	.733

to compare the actual group membership of each applicant (admit or deny) with a predicted group membership based strictly on the most highly-correlated group of actual scores on criteria measures.

The test of equality of group covariance matrices showed non-significant differences (Box's M = 12.05, p = .08); thus the analysis could proceed without correction for group differences. The canonical discriminant function showed an eigenvalue of 1.23 and a Wilks' Lambda of 0.448 with an associated chi-square statistic of 53.44 (p=<.01), for the variables Essay, Undergraduate GPA, and Simulation. This means that a sufficiently strong analysis could be made for combining these variables. The canonical discriminant functions evaluated at group means (group centroids) were –1.79 (group 1) and 0.67 (group 2).

The discriminant analysis suggested that 82.09 percent of the actual admission decisions were correctly made according to this three variable solution. This represents an empirical backing to the validity of the selection process as carried out with this population and data set. The standardized canonical discriminant function coefficients and Classification results are presented in Table 3.

Predictive Validity of Selection Variables for End of Program Criterion Rating

At the end of the one year program, each of the 82 candidates was rated in quintiles by the two group faculty leaders. That is, each student teacher was placed in one of five equally-sized categories of "strength of performance in program." This description included quality of participation and achievement, as well as strength as a teacher. The program leaders knew the students best, having been an instructor for them in one or more classes, observed each in field placements, interacted with

Table 3
Discriminant Analysis

	Step entered	Discriminant function	Wilks' Lambda
Essay	1	.87	.583
Undergrad GPA	2	.51	.512
Simulation	3	.49	.448

Classification results:

	No. of Cases	Predicted Group Membership	
		1	2
Actual Group		Deny	Accept
Group 1	53	44	9
Deny		83.0%	17.0%
Group 2	81	15	66
Accept		18.5%	81.5%

Percent of "grouped" cases correctly classified: 82.1%

other educators on their behalf, and dealt with administrative matters throughout the year. These faculty leader judges participated in selection and were the academic advisors of the students.

The categorical rating (1-5) of each candidate was correlated with six of the admission measures (four were dropped because of redundancy-e.g., subtest scores of a total score measure). Next, a linear regression analysis was performed on the admission measures as predictors of the End of Program Rating. Table 4 presents these results. Using a multiple regression analysis, the six selection measures were found to have a multiple R=0.616 (R^2=0.380); thus, the six measures may be said to have predicted 38 percent of the variance in End of Program Ratings. This analysis demonstrated a small, yet acceptable predictive validity for these measures.

Time and personnel demands. The file review and rating took approximately 16 hours, using 4 to 5 reviewers at a time. Overall, 16 faculty participated in file review. All candidates were involved in simulations in groups of 4-6. Each simulation took 1 hour and 2-3 observers.

Conclusions and Recommendations

This analysis helped to better understand the decision-making processes used in admission to this teacher preparation program. In particular, it was disclosed that some selection measures overlapped with others, while others provided unique insights into applicant characteristics. This information permitted a simplification for future selections. Another understanding developed from this study was that some kinds of evidence were used more by faculty decision makers than were others. The predictive value of admission variables was found to be moderate, but positive and encouraging of further use with additional development. Finally, the relationships found in this study were in the same ranges as those reported in the literature by earlier researchers. This finding lends support to the contention that the admission procedures were reliable and valid.

This study and its findings helped the program to attend to reliability of judgments made on individual data sources. For the judgment reviews of Essay, References, and overall File, it is important to attend to the inter-rater reliability. In 83.1 percent of the two-judge decisions the rating was within 1 point on the 7-point scale. This level of agreement produced reliability estimates in the low .9 range. The rule of referring judgments differing by 3 or more is logistically feasible,

Table 4
Correlations of Six Selection Measures with End of Program Rating

	Essay	References	File	GPA	Simulation	Writing Test
End of Program Rating	.155	.402	.028	.108	.196	.304

and produces overall system judgments in the upper .8 range. More explicit directions can achieve other improvements in reliability for judges, e.g., quality of recommendations and overall files. This information has been given to the planning groups to implement in the next round of admissions.

Several measures were found to stand out in their correlation with admission decisions. This suggests that these dimensions were more influential than those measures with lower correlations with the decisions. First among these factors was the Essay statement, which corroborates the reports of Smith and Pratt (1996). With many well-qualified applicants (e.g., GPAs over 3.00 in non-education courses and high test scores), the remaining discrimination could be based on a variety of criteria. Essays were important because they revealed (a) motivation related to student needs, as opposed to self-interests, (b) congruence with the program and mission of the institution as a major urban university, (c) a vision of need or quality in schools, and (d) ability to express oneself in a compelling way. While Essays appeared to impress the decision makers, further study is needed to better understand the predictive validity of this source of information.

The simulation activity was an important part of admissions decision-making since it added a different kind of information to the admission process. In general, the simulation activity has strong face validity for admissions: this collaborative group task is much like the expectations of performance for program course work and field placements. It is recommended that this activity be continued and developed. In particular it will be important to estimate the reliability of judges. Although the number of judges in this study is the same as reported in the literature (Roose & Rudman, 1985; Shechtman, 1988), it is important to assure reliability of this important admissions criterion.

The discriminant function analysis worked well, predicting more than 80 percent of the admission decisions. However, its utility is more in understanding the underlying criteria used in selection rather than any mechanical use for selection (e.g., to make a mathematical selection decision rather than rely upon human judgment). It can be used to make validity estimates when criterion scores become more refined.

The additional analysis of subject-matter knowledge tests might have been helpful, but were excluded because the results were reported with widely varying scales. Test-taking ability, or General Cognitive Ability (Hunter & Schmidt, 1989), is an important consideration for selection. A measure such as subject-matter competence, additional to the basic skills analyzed in this study, might give a more stable predictor of future success.

While we were pleased with the information produced by the predictive validity analysis of this study, we are concerned with refining our measures to achieve higher levels. However, the kind of comparisons and analysis selected for analyzing predictive validity make a difference in these values. For example, the overall predictive validity of the admission variables might have been high if we

had merely predicted success or failure in the program. However, the completion rate of our program was 98.2 percent because of our highly-select candidates. Upon closer examination, the prediction of location in quintiles was not as accurate as expected. This may be a result of judgment inaccuracy in predictors or in criterion. In the future we look to refine both of these contributors to predictive validity (e.g., to have more judges and measures of program success).

The time, monetary, and logistical demands of teacher candidate admissions require deliberation and consideration. The need to compress simulations into a short period of time calls for advanced planning and commitments. The time and dollar costs should be balanced against the important needs to get a high-quality group of teacher candidates. The tasks and costs asked of candidates need to be considered. In this study, the costs of testing were borne by the applicants but were required by the state licensure agency. No additional candidate dollar costs were added by our performance expectations. More careful data should be gathered about cost-benefits in the procedures used for admissions.

There is a need to better develop the communication to applicants about the procedures, justifications, and individual results. While feedback was helpful, it was not always systematically provided. Efforts should be made to describe the process in steps and provide coherent, revealing, and thoughtful staged explanations to candidates. Descriptions of the procedures developed in this study could lend justification and support to this communication.

References

Andrew, M., Lent, E., Moorehead, C., Moss, D., Singer, M., & Woolf, K. (1996). Predicting performance in a teaching internship. *Journal of Personnel Evaluation in Education, 10,* 271-278.

Barr, A.S. (1931). *An introduction to the scientific study of classroom supervision.* New York: Appleton Century.

Benner, S., George, T., & Cagle, L. (1987). Admission boards: The contribution of professional judgment to the admission process. *Journal of Teacher Education, 38*(2), 7-11.

Borman, W.C. (1982). Validity of behavioral assessment for predicting military recruiter performance. *Journal of Applied Psychology, 67*(1), 3-9.

Boyce, A.C. (1915). Methods of measuring teachers' efficiency. *The Fourteenth Yearbook of the National Society for the Study of Education* (Part II). Chicago, IL: University of Chicago Press.

Dailey, C.A. (1982). *Using the track record approach.* New York: American Management Associations.

Eder, R.W., & Harris, M.M. (1999). *The employment interview handbook.* Thousand Oaks, CA: Sage.

Gunne, C., & Peterson, K.D. (1990). Use of an academic aptitude test for teacher preservice admissions. *Teacher Education Quarterly, 17*(2), 85-92.

Hunter, J.E., & Schmidt, F.L. (1989). Meta-analysis: Facts and theories. In M. Smith & I.T. Robertson (eds.) *Advances in selection and assessment* (pp. 203-226). Chichester, UK:

John Wiley & Sons.

Kenny, D., & La Voie. L. (1984) The social relations model. In L. Berkovitz (Ed.) *Advances in Experimental Social Psychology, 18* (pp. 142-182). Orlando, FL: Academic Press.

Laman, A.E., & Reeves, D.E. (1983). Admission to teacher education programs: The status and trends. *Journal of Teacher Education, 34*(1), 2-4.

Lehmann, I.J., & Phillips, S.E. (1987). A survey of state teacher-competency examination programs. *Educational Measurement: Issues & Practices, 6*(1), 19-22.

Malvern, D.D. (1991). Assessing the personal qualities of the applicants to teacher training. *Studies in Educational Evaluation, 17*, 239-253.

McCutcheon, J.W., Schmidt, C.P., & Bolden, S.H. (1991). Relationships among selected personality variables, academic achievement, and student teaching behavior. *Journal of Research and Development in Education, 24*(3), 38-44.

Messmer, M. (1989). *Fast forward MBA in hiring.* New York: John Wiley & Sons.

Roose, D., Mitchell, S., & Rudman, M.K. (1985). Selecting the brightest and the best. *Phi Delta Kappan, 67*, 219-221.

Schmidt, F.L., Ones, D.S., & Hunter, J.E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627-670.

Shank, K.S. (1979). *Nationwide survey of practices in selection and retention of teacher education candidates.* ERIC Document Reproduction Service ED 167 539. Charleston, IL: Eastern Illinois University.

Shechtman, Z. (1988). Selecting candidates for teachers' training college: A group assessment procedure as an alternative to individual interviews. *European Journal of Teacher Education, 11*, 185-193.

Shechtman, Z. (1992). A group assessment procedure as a predictor of on-the-job performance as a teacher. *Journal of Applied Psychology, 77*(3), 383-387.

Shechtman, Z. (1998). Agreement between lay participants and professional assessors: Support of a group assessment procedure for selection purposes. *Journal of Personnel Evaluation in Education, 12*, 5-17.

Smith, H.A., & Pratt, D. (1996). The use of biodata in admissions to teacher education. *Journal of Teacher Education, 47*(1), 43-52.

Stengel, B.S., & Tom, A.R. (1995). Taking the moral nature of teaching seriously. *Educational Forum, 59*, 154-163.

Thompson, M.S. (1980). *Benefit-cost analysis for program evaluation*. Beverly Hills, CA: Sage.

Webster, E.C. (1964). *Decision making in the employment interview.* Montreal, Quebec, Canada: Eagle Publishers.