DOCUMENT RESUME

ED 353 291 TM 019 046

AUTHOR Riggs, Iris M.; And Others
TITLE An Assessment of Selection Criteria Validity for a Teacher Education Program.
PUB DATE Apr 92
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; College Entrance Examinations; *College Students; Correlation; Elementary Secondary Education; Grade Point Average; Higher Education; *Methods Courses; *Predictive Validity; Regression (Statistics); Standardized Tests; Statistical Significance; Student Evaluation; Student Teachers; *Teacher Education; *Test Validity
IDENTIFIERS California; California Basic Educational Skills Test; NTE Core Battery

ABSTRACT

The predictive validities of criteria for entrance into a teacher education program were studied, based on the hypothesis that criterion specificity is positively related to the predictive validity of the selection measure. Subjects (N=512) were drawn from approximately 1,200 students in the teacher education program of a small California state university. Measures from student files included the following predictor variables: (1) undergraduate grade point average (GPA); (2) scores on the California Basic Educational Skills Test (CBEST); (3) scores on the National Teachers Examinations; and (4) prerequisite education course grades. Three composite criterion scores of performance were based on student teaching performance as measured by university supervisors and resident teachers. Univariate correlations and multiple regression enabled the simultaneous consideration of all predictors and their relative importance. Consistent with the study's hypothesis, the methods courses were superior in their ability to account for variance in the criterion measures. Overall, GPA was also significantly correlated with all 3 criteria, but the standardized tests fell short of significance in 10 of 12 trials. Implications for selection of teacher candidates are discussed. Three tables present study data, and there is a 47-item list of references. (SLD)

An Assessment of Selection Criteria Validity for a Teacher Education Program

Iris M. Riggs

Matt L. Riggs

Ruth A. Sandlin

California State University San Bernardino

A Paper Presented at the
American Educational Research Association's
Annual Meeting
April 1992, San Fransisco

Running Head: Selection Criteria

INTRODUCTION

The validity of admission criteria utilized for entrance into a teacher education program is vitally important. Present national and state policies have called for higher and higher admission standards in response to reform reports calling for higher quality teachers. These standards often take the form of higher grade point averages or some form of standardized tests. The U.S. Department of Education reports that more than 20 states now mandate some form of testing for candidates to be admitted into teacher education programs (Office of Education Research and Improvement, 1987, p. 14).

While a variety of criteria may be used for admission to teacher education programs, little is known about the true objective value of these measures in reference to the selection of higher quality candidates. In many systems, selection practices appear based on convenience of predictors available and/or "common sense" decisions about what measures "should" predict good teaching performance. Little research has been conducted to support the predictive validity of newly implemented selection measures.

Validity does not refer to the selection criteria score or test itself, but to the quality of the inferences or decisions based upon the selection test score (APA, 1977; APA, 1985). The selection process is not valid if it does not result in the selection of those students who will make the best teachers. Unfortunately, supported by a misguided court system, content

validation procedures have often been substituted for methods which actually assess the ability of the selection scores to enable predictions relevant to future performance (Rosenfeld, Thornton, & Skurnik, 1986). Content validation is a process more suitable for making inferences about the quality of the process of initial test construction. Content validation is NOT suitable for making inferences about the meaning of the selection score or its ability to enhance the decision process (Tenopyr, 1977; SIOP, 1987). If the purpose is to infer how well a candidate will perform on the job, we must determine the mathematical relationship between test scores and some numerical index of job success (Lawshe, 1985; SIOP, 1987). This requires criterion validation. Anything less is described as unethical by division 14 of the APA (SIOP, 1987).

Nevertheless, most programs continue to use selection criteria that have never been subjected to the appropriate validation assessment. The use of invalidated admission standards may result in serious consequences, not only for the potential teacher education students, but also for the districts in need of a teaching force. If the utilized criteria have no relationship to actual teaching performance, potentially excellent teachers are prohibited from joining the profession.

The use of invalid standardized tests as selection hurdles has special implications for ethnic minorities. When potential teachers from diverse populations are excluded from entrance into teacher education programs on the basis of standardized test

scores, school districts are denied the opportunity to increase the diversity of their teaching force--an unnecessary and intolerable outcome. Cooper and Williams (1986) identified this disturbing trend in states that required competency testing for certification. The proportion of Black teachers within the teaching force decreased the longer the mandated testing had been in place. No similar trend was reported in states without mandated testing (Cooper & Williams, 1986).

The California State University system incorporated stricter standards in the form of higher GPA's, successful completion of the California Basic Educational Skills Test (CBEST), and passage of the National Teachers Exam (NTE) or an approved waiver program. Little research, however, has been done to examine the relationship between these scores and measures of subsequent student success in the teacher education program.

The few, rather limited studies available in the general literature that have attempted to investigate this question of predictability have provided mixed results. In two studies conducted by Olstad (1983), GPA was identified as a predictor of relative success in student teaching performance, while California Achievement Test scores were not supported as selection criteria.

Neither the NTE Test of Professional Knowledge nor student GPA predicted student teacher performance as measured by cooperating teachers in a study conducted by Dobry, Murphy, and Schmidt (1985).

Mayer's (1989) study of admission requirements reported no significant relationships between students exceptionally admitted to the teacher education program (due to low GPA) and time spent by university supervisors. University supervisors' evaluations, students' self-report of preparation time, completion of program, and teaching credential acquisition were also not significantly related to the predictor. However, classroom teachers' assessments of exceptional admits were found to be significantly lower than those of regularly admitted students. Results regarding specific predictors were unavailable due to the multiple categories used to define exceptional admits.

A preliminary study by Riggs and Riggs (1991) supported the use of previous academic achievement, both in the form of overall GPA and grades earned in early teaching methods courses. Standardized test scores received little support.

The question of admission standards is critical to reform efforts. Given the mixed results of the limited studies addressing this issue, the current project attempted to provide further evidence toward the resolution of this important question. Improvements in the current study over previous investigations include a much larger sample and a more reliable performance indicator based on multiple sources of evaluation.

Hypothesis

The hypothesis for this study is based both on results of previous research and the general principle that criterion specificity will be positively related to the predictive validity

of the selection measure. In the continuum of criterion specificity, grades from educational methods courses are based on developed abilities most exclusively related to the performance of tasks required by a teacher. The overall GPA is less specific, but nevertheless reflects achievement in a course of study designed for future educators. Broadly-oriented achievement tests designed to assess minimum competency levels of general developed abilities are the least specific. Consequently, it is predicted that methods course grades will be the best predictors of success in student teaching. GPA is expected to be less predictive, but still significantly correlated with the performance measure. The standardized achievement tests (CBEST subscales and NTE) are not expected to account for significant variance in the student teaching assessments.

METHOD

Subjects

Subjects (n = 512) were drawn from a population of approximately 1200 students admitted to the teacher education program of a small California state university.

Measures

Measures obtained from student files included the following predictor variables:

1) undergraduate GPA,
2) CBEST scores,
3) NTE scores,
4) prerequisite education course grades.

The prerequisite education courses were in educational

psychology, reading methodology, and math methodology. All three courses were taken prior to admission to student teaching.

Three composite criterion scores were based upon student teaching performance as measured by university supervisors and resident teachers. University assessment forms were completed by supervisors and resident teachers at the midpoint and at the end of each of two ten week student teaching blocks. University supervisors based their evaluations upon a minimum of five visitations, while resident teachers based theirs upon daily observations. A mailing requested additional evaluation of student teaching performance by supervisors and resident teachers.

Both parties assessed student teacher competencies as they are presented within the university's Student Teacher Evaluation Form. These competencies were described in depth within a Student Teaching Handbook that all supervisors, cooperating teachers, and student teachers utilized.

The five categories assessed within the Student Teacher Evaluation Form were: planning, instruction, evaluation, classroom organization, and classroom behavior (discipline). All categories were assessed for both individualized/small group instruction and large group instruction. Each category was rated with both of the following five-point scales:

Opportunities Taken

1 Nearly all opportunities to demonstrate progress taken.
2 Most opportunities to demonstrate progress taken.
3 Some opportunities to demonstrate progress taken.

4 Few opportunities to demonstrate progress taken.
5 Opportunities to demonstrate progress seldom taken.

Attainment Demonstrated

1 Functions independently on most goals at level of fully qualified teacher.
2 Functions independently on several goals with occasional supervision needed.
3 Functions adequately on several goals under supervision.
4 Has difficulty attaining several goals even under supervision.
5 Has been unable to attain most of the goals.

An additional appraisal form mailed to both supervisors and resident teachers assessed the five performance components, but used a different response format.

The three final composite criterion measures were computed by collapsing multiple evaluations obtained from: 1) resident teachers, 2) university supervisors, and 3) a combination of both. Scores were converted to percentages of total possible points. Combination of the measures was justified by their high intercorrelations, especially within the first two groups (Dunnette, 1963). This combination is also supported by the fact that performance data based on multiple observations are more reliable/stable than those based on a single report.

Analysis

With the exception of Riggs and Riggs (1991), previous investigations categorized either their predictor and/or their criterion measures and tested criterion validity with either Chi-square analysis or analysis of variance. For example, student teaching performance rankings were trichotomized into "high, medium, and low" scores to enable analysis of variance on group

scores. This procedure reduces the power of such investigations because meaningful variance between members within each group is lost.

The present study employed more appropriate methods of analyses that enabled the retention of the full range of variation in all continuous measures. In addition to univariate correlations, multiple regression was used to enable the simultaneous consideration of all predictors and to assess the relative importance of each.

RESULTS

Descriptive statistics for all relevant variables are reported in Table 1. With the exception of method course grades (which were skewed due to a very disproportionate number of A's), all distributions were approximately normal. The skew and accompanying reduction of variance in the methods course measures reduces their potential to "covary" with the performance measures. This does not rule out their potential to perform as significant predictors, but it does limit the potential magnitude of the correlation-based statistics.

Pearson correlation coefficients between the eight potential predictor variables and the three composite criterion measures are reported in Table 2. As pairwise deletion of missing data was used, n's and exact p's are reported for each correlation. The predictor variables are ordered in reference to their rank in the magnitude of their average correlations with the criterion measures. A full correlation matrix which includes the

correlations within predictor and criterion groups is included in Appendix A.

Consistent with this study's hypothesis, the methods courses proved to be superior in their ability to account for variance in the criterion measures. The three courses correlated significantly with the three performance measures 12 of 12 trials. Correlations ranged from r = .10, p = .034 to r = .20, p = .000.

As predicted, overall GPA was also significantly correlated with all three criteria. The absolute magnitude of the correlations between GPA and teaching performance ranged between r = .14, p = .04 to r = .15, p = .03. The correlations were slightly less, on the average, than those produced by the methods courses, but the differences failed to reach statistical significance.

The standardized tests fell short of significance 10 of 12 trials. The CBEST reading and writing subscales produced statistically significant correlations of r = .12, p = .01 and r = .09, p = .035 with the performance evaluations submitted by the resident teachers. The maximum correlation resulting with any pairing between the NTE core battery and the teaching performance measures was r = -.029. All correlations with the NTE, though near zero, were negative in direction.

Stepwise multiple regression was also utilized to assess the potential predictive validity of selection test scores in optimal combinations. The NTE was omitted from these analyses, both

because it significantly reduced the viable sample size and due to its apparent impotency as a predictor as evidenced by the univariate correlations. The results of the multiple regressions are reported in Table 3. All of these analyses further support the superiority of the methods courses as predictors of student teacher success.

When predictors were regressed on the performance criterion produced by the resident teachers, the reading methods course (ED-345) entered the equation first. This variable accounted for 4% of the variance in performance. No other variables significantly increased R.

When performance as reported by the university supervisor was the criterion, the math methods course (ED-340) entered first, accounting for 3% of the variance. ED-345 entered second, significantly increasing the R to .22 and the R^2 to 5%.

Using the total performance composite, the reading methods course once again entered first producing an R of .21. The educational psychology course entered second, swelling R to .24 and accounting for 6% of the variance in the performance measure.

DISCUSSION

Though the validity coefficients produced in this study were not of the magnitude that one might hope, they are of average size. Boehm's (1982) review of published criterion-related validity analyses reported that the average of validity coefficients from 176 studies was r = .219. The coefficients in this study were probably afflicted by maladies common to most

validation studies. The performance criteria were less than perfectly reliable and restricted in range (students scoring lowest in the predictor measures were not admitted to the teacher education program), and the most effective predictors were badly skewed. Nevertheless, the variance accounted for was not so small as to be of no practical importance. In certain selection circumstances, even a validity coefficient of .20 can significantly improve the "hit" rate in selection decisions (Saal & Knight, 1988).

It is reassuring to find the positive relationship between methods course grades and student teaching performance. Courses which teach pedagogy should relate to ultimate teaching performance. This finding suggests we might place more trust in our own ability as teacher educators to evaluate our students' ability to teach within the education course setting, thus producing data that can be used as a major selection criterion. Certainly, at the very least, prerequisite methods course grades should be given the same consideration as other selection variables.

Consideration of students' grades within methods type courses might prove especially helpful in the case of low overall GPA candidates. Present practice often "exceptionally" admits low GPA students to education programs or student teaching. When possible, final admission might be postponed for these students until they have had a chance to "prove" themselves within their methods classes. This study's results suggest that even a B

within such a class may be indicative of potential weakness in student teaching performance. Yet, the low GPA candidate who excels within methods coursework might be considered a worthwhile risk.

Selection committees might also consider collecting more selection data from the methods instructors in addition to the course grade. This is especially important given the skewed, restricted variance reflected in most class grade distributions. The Pre-teacher Assessment (Gerlach & Millward, 1989) was developed to collect and utilize observational field data within a methods course. The authors do not recommend its use for selection, but believe it facilitates undergraduate preservice teachers' development of teaching skills. Perhaps, though, this assessment or similar others could be used to help make selection decisions, especially when considering students with low GPA or low standardized test scores.

When selection practices are based on multiple objective criteria, the resulting selection decisions should be more valid and possibly more equitable. It has been demonstrated that under-represented student groups may be disadvantaged by many standardized tests (Cooper & Williams, 1986). While policy calls for a more diverse teaching force, current practice may lock out many potential teaching candidates. The use of performance-based evaluation in early methods courses in addition to other selection measures, would enable such candidates to better demonstrate their teaching potential.

Finally, the worth of selection criteria must be demonstrated to the public. National reports have raised many questions concerning the competence level of existing teachers and the teacher education programs which produced them. Validation studies must go beyond investigation of the relationship of predictor variables to student teacher performance. Preservice teachers' performance must be studied as they complete their first years of teaching within their own classrooms. Only then will teacher education programs be accountable to the public for the teachers they prepare.

REFERENCES

Boehm, V.R. (1982). Are we validating more but publishing less? (The impact of governmental regulation on published validation research--An exploratory investigation). Personnel Psychology, 33, 495-503.

Cooper, C. C., & Williams, B. J. (1986). The effects of teacher competency testing: Issues in certification and retention. Urban Education, 8, 127-133.

Dobry, A.M., Murphy, P.D. & Schmidt, D.M. (1985). Predicting teacher competence. Action in Teacher Education, 7, 69-74.

Dunnette, M. D. (1963). A note on the criterion. Journal of Applied Psychology, 47, 251-254.

Gerlach, G. J., & Millward, R. E. (1989). A new perspective for strengthening teaching skills: Pre-teacher assessment. Foreign Language Annals, 22, 145-148.

Haney, W. (1990). Preservice evaluation of teachers. In J. Millman & L. Darling-Hammond (Eds.), The New Handbook of Teacher Evaluation. Newbury Park, CA: Sage. p.46-61

Lawshe, C. H. (1985). Inferences from personnel tests and their validity. Journal of Applied Psychology, 70, 237-238.

Mayer, D. (1990). Differential admission requirements and students achievement in teacher education. Journal of Teacher Education.

Office of Educational Research and Improvement (1987). What's happening in teacher testing? Washington, DC: U.S. Government Printing Office.

Olstad, R.G. (1983). Preservice teaching performance: A search for predictor variables. Final report. (Report No. 83-3), Seattle, WA: Washington University, Teacher Education Research Center. (ERIC Document Reproduction Service No. ED 231 810).

Riggs, I.M., & Riggs, M.L. (1991). Predictors of student success in a teacher education program: What is valid, what is not. Action in Teacher Education, 12, 41-46.

Rosenfeld, M, Thornton, R. F., and Skurnik, L. S. (1986). Analysis of the professional functions of teachers. Relationships between job functions and the NTE core battery. (Report No. ETS-RR-86-8), Princeton, N.J.: ETS, Center for Occupational and Professional Assessment.

Saal, F.E. & Knight, P.A. (1988). Industrial/Organizational psychology: Science & practice. Pacific Grove, CA: Brooks/Cole Publishing.

Society for Industrial and Organizational Psychology, Inc. (1987). Principles for the validation and use of personnel selection procedures (3rd ed.). College Park, MD: Author.

Standards for educational and psychological testing. (1985). Washington, DC: American Psychological Association.

Standards for providers of psychological services. (1977). Washington, DC: American Psychological Association.

Tenopyr, M. L. (1977). Content-construct confusion. Personnel Psychology, 30, 47-54.

Table 1: Descriptive Statistics for Predictor and Criterion Variables

Variable	n	Mean	Std Dv
Predictors:			
GPA	486	3.11	.45
CBEST - Reading	492	54.52	10.54
CBEST - Math	492	54.80	11.25
CBEST - Writing	492	46.31	7.91
Education Psychology	487	4.08	1.20
Math Methods	448	4.43	.89
Reading Methods	485	4.14	1.11
NTE Core Battery	189	667.80	5.64
Criteria			
Resident Teacher Eval	376	.89	.11
Supervisor Eval	403	.89	.11
Combined Evals	420	.88	.10

Table 2: Pearson Correlation Coefficients between Selection Predictors and Performance Criteria

	Criteria		
	Resident Teacher	University Supervisor	Combined Evals
Predictors			
Reading Methods	.2000 (364) p = .000	.1623 (390) p = .001	.1967 (407) p = .000
Educational Psychology	.1348 (363) p = .005	.1523 (389) p = .001	.1842 (406) p = .000
Math Methods	.1001 (335) p = .034	.1774 (363) p = .000	.1525 (379) p = .001
GPA	.1457 (355) p = .003	.1368 (383) p = .004	.1433 (399) p = .002
CBEST - Reading	.1182 (366) p = .012	.0564 (392) p = .133	.0652 (409) p = .094
CBEST - Math	.0598 (366) p = .127	.0348 (392) p = .246	.0421 (409) p = .198
CBEST - Writing	.0945 (366) p = .035	.0055 (392) p = .457	.0289 (409) p = .280
NTE	-.0293 (146) p = .363	-.0224 (158) p = .390	-.0050 (165) p = .475

Coefficient/(Cases)/1-tailed Significance

Table 3: Regression Analyses

Analysis #1: Dependent Measure - Resident Teacher Evaluation

	Variable	R	R Sq	Beta	T	Sig T
Step 1	Reading Methods	.20	.04	.20	3.73	.0002

Analysis #2: Dependent Measure - Supervisor Evaluation

	Variable	R	R Sq	Beta	T	Sig T
Step 1	Math Methods	.18	.03	.18	3.43	.0007
Step 2	Math Methods	.22	.05	.15	2.74	.0064
	Reading Methods			.13	2.36	.0186

Analysis #3: Dependent Measure - Combined Evaluation

	Variable	R	R Sq	Beta	T	Sig T
Step 1	Reading Methods	.20	.04	.20	3.80	.0001
Step 2	Reading Methods	.24	.06	.16	3.05	.0025
	Educational Psych			.14	2.72	.0069