DOCUMENT RESUME

ED 068 469 TM 001 803

AUTHOR	Quirk, Thomas J.; And Others
TITLE	A Critical Review of Research Related to the National Teacher Examinations.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-72-15
PUB DATE	Apr 72
NOTE	52p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Classroom Observation Techniques; Criterion Referenced Tests; Educational Research; Effective Teaching; *Evaluation Techniques; *Measurement Instruments; *National Competency Tests; Rating Scales; Research Methodology; Teacher Behavior; *Teacher Evaluation; *Test Interpretation; Tests
IDENTIFIERS	National Teacher Examinations; NTE

ABSTRACT

A collection of articles related to testing instruments of the National Teacher Examinations (NTE) are reviewed. The report has been organized into four sections. The first discusses briefly the background and purposes of the NTE and the hazards of combining data on these tests in research studies. The second section is concerned with articles related to the concurrent validity of the NTE and their relationship to pre-service teacher preparation. The third section has to do with articles related to the predictive validity of the NTE in terms of in-service teachers. The last section summarizes the research findings. The report concludes, "Perhaps more important than revising principal and pupil rating scales is to conduct systematic studies of the relationship between the NTE scores of teachers and average residual achievement gain scores of pupils in their classes." An 89 item list of references is included. (Author/LS)

RESEARCH BULLETIN

RB-72-15

A CRITICAL REVIEW OF RESEARCH RELATED TO THE NATIONAL TEACHER EXAMINATIONS

Thomas J. Quirk
Barbara J. Witten
and
Susan F. Weinberg

Teacher Behavior Research Group

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
April 1972

ACKNOWLEDGEMENTS

This report is a publication of the National Teacher Examinations (NTE) Program at Educational Testing Service. Several colleagues at Educational Testing Service have provided valuable assistance in the preparation of this review. We should like to thank Betty Humphry, Eleanor Weiss, and Richard Majetic for their patience and encouragement and Nat Hartshorne for asking interesting questions and helping us with our writing style.

T.J.Q.
B.J.W.
S.F.W.

TABLE OF CONTENTS

INTRODUCTION

This report has been organized into four main sections. The first discusses very briefly the background and purposes of the National Teacher Examinations (NTE) and the hazards of combining data on these tests in research studies. The second section is concerned with articles related to the concurrent validity of the NTE and their relationship to pre-service teacher preparation. The third section has to do with articles related to the predictive validity of the NTE in terms of in-service teachers. The last section summarizes the research findings.

The reader who wants to find more detailed information about any article mentioned in this report is urged to study the more extended discussion of these articles in a separate annotated bibliography (54).

The National Teacher Examinations have been in existence for more than 30 years. Thus, no review of the articles related to these major testing instruments could be exhaustive. Some unpublished articles have disappeared; still others are of such poor quality that they do not deserve to be resurrected. In this review, we have chosen not to discuss articles that do not contain any correlational or statistical data involving the NTE; these articles cluster into the following sets: those that either describe the tests or discuss historical changes in them (82, 79, 15, 3, 62); those that compare the performance of candidates on the Common Examinations who specialized in one subject area against the performance of candidates prepared in others (17, 39, 76, 77, 51); articles that either support or criticize the NTE (1, 14, 24, 52, 58, 59, 86, 2, 89, 34, 30, 56, 9, 13,

23, 40, 6, 7, 41, 57, 71, 73); articles that discuss the statistical or normative properties of the NTE (20, 32, 60, 61, 64, 65, 66, 69, 88, 4, 31); and articles that discuss the NTE within the larger framework of teacher evaluation (12, 38, 55, 87, 44, 63, 67, 68, 26, 19, 5, 85).

Educational Testing Service (ETS) would like to collect systematically all research studies that deal in some way with the National Teacher Examinations. Thus, we would welcome information from interested readers about any studies that may have escaped our scrutiny.

THE NATURE AND PURPOSE OF THE NATIONAL TEACHER EXAMINATIONS

The National Teacher Examinations have been used to assess the knowledge of prospective teachers since 1940 when the examinations were first administered by the American Council on Education. In 1950, full responsibility for preparing, administering, and scoring the examinations was transferred to Educational Testing Service in Princeton, New Jersey.

The NTE consist of the Common Examinations, which offer subtests in Professional Education and General Education, and the Teaching Area Examinations (TAE), which measure understanding of subject matter and methods in 24 areas.

The major purpose of the National Teacher Examinations is to provide an independent assessment of the academic preparation of college seniors completing a four-year program in teacher education. The NTE have been used principally to assist in selection of teachers by local school districts and in the assessment by teacher-training institutions of the academic

preparation of their teacher-training candidates. The National Teacher Examinations are national, standardized, secure tests that permit comparison of candidates within the same institution and across different institutions within the limitations of the content sampled by the tests. For a discussion of how the tests are planned and constructed, the interested reader can find more detail in the Prospectus for School and College Officials (50), the National Teacher Examinations: Interpretation of Scores (49), and the Bulletin of Information for Candidates 1970-1971 (47).

Educational Testing Service does not set any passing or failing standards for any of the National Teacher Examinations. Only local institutions can make this type of decision based on an assessment of their local needs and on their own validity studies.

The National Teacher Examinations are not designed to measure teacher aptitude, interests, attitudes, motivation, maturity, or other personal or social characteristics. Nor are they intended to be a measure of classroom teaching performance. What a teacher knows about his teaching area of specialization may or may not indicate what he will do in the classroom.

Educational Testing Service recommends that the NTE not be used in decisions about retention, hiring, or tenure of experienced teachers. According to the NTE Guidelines for Using the National Teacher Examinations (48), "When an adequate and reliable record of the teacher's performance is available there is no need to attempt to predict his teaching abilities." Any individual who has had extended teaching experience, either as a full-time teacher or as a fairly regular substitute teacher, has demonstrated his teaching ability to a degree that is not measured by the NTE. For a more detailed

discussion of the use and misuse of NTE scores, the reader should consult the Guidelines.

The Common Examinations

The Common Examinations of the NTE provide scores in Professional Education and General Education, and a weighted combination of these two areas. Neither the Professional Education score, the General Education score, nor any of their subtest scores (Psychological Foundations of Education, Societal Foundations of Education, Teaching Principles and Practices, Written English Expression, Social Studies, Literature and the Fine Arts, and Science and Mathematics) have ever been equated to each other from form to form; thus, these scores should be used with caution in research studies since the applicability to future studies of findings based on these scores will be restricted to an unknown degree.

The Weighted Common Examinations Total (WCET) score is on a scale based on the scores earned by college seniors who took the Common Examinations in 1940. Whenever new items are introduced into the Common Examinations, the new form of the test is equated statistically to previous forms of that test. Thus, the WCET scores are statistically comparable from administration to administration going back to 1940.

The NTE Common Examinations scores have not always been properly used in research studies. For example, in several studies (72, 83, 21, 10, 36, 11, 25) subtest scores on the Common Examinations (Professional Education, General Education, Written English Expression, Science and Mathematics, and so forth) earned in different years were combined into a single set of data and measured against some criterion or norm group; since these subtest scores have never

been equated to each other from form to form of the Common Examinations, such a procedure is always improper. Sutcliffe (80) created a unique type of NTE score which has not been used in other research studies and is, therefore, of very limited value.

The Teaching Area Examinations

Twenty-four Teaching Area Examinations were offered by the NTE program during 1971. The scaled score for the Teaching Area Examinations is based on substantially all candidates who indicated at the national administration of the NTE in February 1964 that the TAE they took was in the field for which they were best prepared to teach. Since February 1964, each new form of the TAE has been equated to earlier forms of the same TAE to allow for differences in the difficulty and length of subsequent test forms. Since the Teaching Area Examinations cover different subject fields, scores on one cannot be compared with scores on another. Only scores for candidates taking the same TAE can be compared and only if they have taken this TAE since February 1964. No scores on any Teaching Area Examinations taken prior to 1964 can be compared, since the TAE were not equated to earlier forms of the test prior to 1964.

In examining the literature about the NTE, several instances of the incorrect use of the NTE Teaching Area Examinations scores were discovered. Some authors compared a mixed group of TAE scores from different teaching-field specialties against a norm group (36) or against some criterion (81; 25). This practice is incorrect since the TAE scores taken in different subject-matter specialties are never comparable. Duncan (25) also correlated NTE Composite scores against grade-point averages, but since these composite

scores included the scores in Teaching Area Examinations in different subject-matter specialties, these correlations are also improper. Still others made the mistake of mixing scores on different nonequated forms earned by candidates in the same subject-field specialty (for example, different forms of the TAE in Education in the Elementary School) in order to compare these scores against some norm group (83) or against some criterion (21).

THE RELATIONSHIP BETWEEN NTE SCORES AND PRE-SERVICE TEACHER PREPARATION

The concurrent validity of the NTE has been studied in terms of the correlations between the test scores and success as an undergraduate, success as a graduate student, and the personal characteristics of the candidates. These studies are summarized in Table 1.

Seagoe (72) computed rank-order correlations between the NTE WCET scores and the qualifying examinations for candidacy for the doctorate that graduate students in the School of Education at UCLA took at the end of their general course work. The correlations between the WCET scores and the total score on the qualifying examinations was .78 during 1942-1945 (N = 11) and .26 during 1946-1947 (N = 19). Despite the small sample sizes, a cutoff score of 60 on the NTE was set for admission to graduate work.

Capps and DeCosta (10) studied the relationship between scores earned by 410 students on the NTE, Graduate Record Examinations (GRE), and undergraduate grade-point average (GPA) and grades in the four basic courses taken by all graduate students at South Carolina State College between 1948-1954. The best single predictor of graduate school success was the Advanced Education

TABLE 1

STUDIES OF THE CONCURRENT VALIDITY OF THE NTE

Date	Author	N	Location	Predictor	Criterion	Correlation
1949	Seagoe	11	UCLA	WCET	Qualifying Exams	.78
		19	UCLA	WCET	Qualifying Exams	.26
1955	Shea	110	Worcester State Teachers College	WCET	Coop. Gen. Cult. Test	.77
					ACE Psychol. Exam	.70
					Coop. Eng. Exam	.64
					Coop. Contemp. Affairs Test	.64
					Undergraduate GPA	.52
					Practice Teaching Grades	-.01
1957	Capps & DeCosta	410	So. Carolina State College	WCET	Grades in four basic grad. courses	.44
1958	McCamey	211	U. of Hawaii	NTE Prof. Info.	GPA in education courses	
					a) pre-school primary (N = 35)	.30
					b) elementary (N = 95)	.23
					c) secondary (N = 81)	.28
1962	Simpson	1636	State of Georgia	WCET	Age	.03
					Years of Teaching Experience	-.01
					Grades in Prof. Ed. courses	.36
					Grades in General Ed. courses	.33
					Grades in Teaching Field	.29
					No. degrees held	.29
1962	Pitcher	51-164	11 colleges & univ.	WCET	Undergraduate GPA	range .38-.74 wt. average .57
1963	Johnson	35 successful students	U. of Houston	WCET	GRE Adv. Test in Education	.51
					GRE-V	.77
					GRE-Q	.54
1964	Thacker	100	State of North Carolina	WCET	Undergraduate GPA	.48
					General Education GPA	.52
					Professional Education GPA	.45
					GPA for Prof. Ed. + Gen. Ed.	.54
					Major Field GPA	.35
1969	Elting	132	U. of Miami	WCET	GPA	not reported

Studies of the Concurrent Validity of the NTE, Table 1, continued

Date	Author	N	Location	Predictor	Criterion	Correlation
1967	Eissey	111	Florida State U.	WCET	Internship GPA	.01
					Undergraduate GPA	.23
					Professional Education GPA	.19
					Teaching Field GPA	.16
1967	Walberg	280	Illinois Teachers College	WCET	High School GPA	.10
					Undergraduate GPA	.36
					Practice Teaching Grade	-.04
1971	Duncan	62	East Tennessee State University	WCET	Psychology GPA	.62
					Education GPA	.58
					Major field GPA	.55
					Total GPA	.62

Test of the Graduate Record Examinations (r = .49), followed by the NTE Common Examinations Total Score (r = .44), and undergraduate GPA (r = .42). The multiple correlation using the GRE Advanced Test in Education, the GRE Aptitude Test, the NTE Common Examinations Total, and the undergraduate GPA was .59.

McCamey (45) correlated the 1957 NTE scores of the 1957 graduates of the University of Hawaii Teachers College (N = 211) with selected academic records. For the three curriculum levels, correlations between the Professional Information subtest* of the NTE Common Examinations with GPA in education courses was .30 for the pre-school-primary level (N = 35), .23 for the elementary level (N = 95), and .28 for the secondary level (N = 81). The correlation between this same NTE subtest and the total number of education units was .33 for pre-school-primary students, .12 for elementary education students, and .32 for secondary education students. A correlation of .63 was also reported between the NTE Professional Information subtest and the NTE Education in the Elementary School TAE. However, since the various forms of the Professional Information subtest of the Common Examinations have never been equated to each other, all of these correlations should be interpreted with caution even though all of these candidates took the same form of the NTE Common Examinations.

Simpson (75) compared NTE WCET scores with 21 personal characteristics of 1,636 candidates who took the NTE in Georgia in April of 1960. Correlations between WCET scores and some of these variables were .03 with age, -.01 with years of teaching experience, -.26 with total number of quarter hours in professional education during the B.A., .36 with average grade in professional

*This subtest is now called Professional Education.

education courses for the B.A., -.09 with number of quarter hours in general education during the B.A., .33 with average grade in general education during the B.A., -.14 with number of quarter hours in teaching field for the B.A., .29 with average grade in the teaching field for the B.A., and .29 with the number of degrees held. The comparisons between these same personal-characteristic variables and each of several different NTE Teaching Area Examinations were also given, and even though all of the candidates took the same form of the TAE in their respective subject-area specialties, these comparisons are of doubtful practical value since the TAE scores were not equated to each other from form to form until 1964. Similarly, subtests of the NTE Common Examinations (Professional Information, English Expression, and so forth) were compared with these personal characteristics, and even though these candidates took the same form of these subtests, it is impossible to draw firm conclusions about the results since these subtest scores have never been equated to each other.

Pitcher (53) correlated NTE WCET scores with GPA, excluding practice teaching grades, for college seniors enrolled in teacher preparatory curricula at a total of 11 colleges and universities during 1959-1961. The sample sizes ranged from 51 to 164, and the weighted average correlation between WCET and GPA was .57, with a range of .38-.74. Correlations and multiple correlations of the NTE subtests (Professional Information, English Expression, Science and Mathematics, and so forth) with GPA and correlations between the Professional Information subtest and the GPA based on professional education courses are also reported. Although each group took the same form of the NTE, these additional correlations are of limited generalizability.

Johnson (37) correlated NTE scores with GRE scores for 92 graduate students enrolled above the M.A. level in the College of Education at the University of Houston between 1945 and 1961 for whom complete data were available (out of the 279 cases). The WCET scores significantly distinguished (.01 level) between the medians of the 35 successful candidates who completed the doctoral program and of the 20 unsuccessful candidates who were not accepted to candidacy. The rank-order correlations between the WCET scores and the GRE for the successful students were .77 with GRE-V, .54 with GRE-Q, and .51 with the GRE Advanced Test in Education.

Elting (29) hypothesized a positive relationship between GPA, using the second 12 credits of undergraduate course work and the NTE for students in the Cuban Teacher Program at the University of Miami (N = 132) 85 percent of whom had Cuban degrees roughly equivalent to a U. S. bachelor's degree, since 66 percent of the students who scored above 500 on the NTE had grades of A or B, but the magnitude of the correlation is not reported.

Duncan (25) correlated NTE WCET scores generated between July 1968 and July 1970 with quality-point average for 62 students from East Tennessee State University who successfully completed the four basic psychology courses, at least the four basic education courses, and had graduated from ETSU. Six of these students had at least one year of teaching experience. The WCET correlated .62 with the Psychology QPA, .58 with the QPA in Education, .55 with the QPA in major field, and .62 with the total QPA. The author also correlated NTE scores in Professional Education and in Psychological Foundations of Education with these QPAs, but since these NTE scores occurred across nonequated subtests of different test forms, these correlations are improper.

Similarly, correlations between these QPAs and the NTE Teaching Area Examinations and the NTE Composite scores are meaningless because scores on TAEs in different teaching fields are not equated and therefore cannot be mixed in data analyses.

Additional studies by Shea (74), Thacker (81), Eissey (28), and Walberg (84), which also contained data relevant to the concurrent validity of the NTE, are discussed in subsequent sections of this report.

THE RELATIONSHIP BETWEEN NTE SCORES AND THE IN-SERVICE PREPARATION OF TEACHERS

The predictive validity of the NTE has been studied in relation to four types of criteria: 1) supervisor ratings, 2) pupil ratings, 3) pupil residual gain scores, and 4) classroom observation. These studies are summarized in Table 2.

Supervisor Ratings

In a study done by Flanagan (33) in 1941, 22 school systems were selected that had at least two teachers whose WCET scores differed by as much as 100 points (N = 49). All of the teachers were employed in regular teaching positions when they took the NTE in 1940. The school superintendents were asked to obtain ratings from each of two supervisors, and the correlations between the WCET scores and the supervisors' overall judgment of the teachers' general effectiveness and desirability (ten-point scale) was .51. According to the author, the correlations were "around .50" (values not given) between these NTE scores and the supervisors' ratings of the teachers' reasoning and problem-solving ability, judgment and perspective in making decisions and choices, breadth of cultural education as reflected in conversation and

TABLE 2

PREDICTIVE VALIDITY OF THE NTE AND SUPERVISORS' RATINGS,

PUPIL RATINGS, PUPIL RESIDUAL GAIN SCORES, AND CLASSROOM OBSERVATIONS

Date	Author	N	Location	Predictor	Criterion	Correlation
1941	Flanagan	49	22 School systems	WCET	Supervisors' Rating	
					1) overall judgment of teacher's gen. effectiveness and desirability	.51
					2) reasoning and problem-solving ability	"around .50"
					3) judgment and perspective in making decisions and choices	"around .50"
					4) breadth and cultural education	"around .50"
					5) understanding of current social problems	"around .50"
					6) ability to offer wise guidance	"around .50"
1946	Lins	29	U. of Wisconsin	NTE	Composite Rating (3-5 raters)	-.15
1951	Ryans	192	Elem. teachers In-service teachers	NTE Subtest on Gen. Prin. & Meth. of Teaching	Principal Rating	
					a) observation blank	.17
					b) general eval.	.23
		165	Sec. Sch. teachers In-service teachers	NTE Subtest on Gen. Prin. & Meth. of Teaching	Principal Rating	
					a) observation blank	.13
					b) general eval.	.15
1954	Delaney	93	Elizabeth, N.J.	WCET	Composite Rating	.17
1955	Shea	110	Worcester State Teachers College	WCET	M-Blank Rating by Superintendent, Principal, or Supervisor	.45
1964	Thacker	100	State of North Carolina	WCET	Principals' Rating	.18
					Supervisors' Rating	.17
1967	Eissey	111	Florida State U.	WCET	Principals' Rating	
					a) 1st year	.10
					b) 3rd year	.10
					Internship Ratings	
					a) Directing Teacher	.14
					b) University Supervisor	.09

Predictive Validity of the NTE and Supervisors' Ratings, Pupil Ratings, Pupil Residual Gain Scores, and Classroom Observations (Continued)

Date	Author	N	Location	Predictor	Criterion	Correlation
1967	Walberg	280	Illinois Teachers College	WCET	Principal Rating	.00
					Supervisor Rating	
					a) Personal Characteristics	-.03
					b) Classroom Performance	.02
1968	Lewis	45	Sul Ross State Coll.	NTE	College Coordinator Rating	.18
1969	Carson	241	Houston Sch. Dist.	WCET	12-week Principal Rating	
					a) personal efficiency	.08
					b) social efficiency	.13
					c) prof. attitude	.09
					d) cooperation	-.03
					e) skill in teaching	.16
					f) class management	.04
					g) general rating	.11
					h) composite rating	.10
1946	Lins	26	U. of Wisconsin	NTE	Pupil Ranking	-.30
1946	Lins	7	U. of Wisconsin	NTE	Pupil Average Residual Gain	.45
1970	Medley & Hill	53	Eastern U.S.A.	NTE Common Exams special subtests	Flanders: Lecturing Behavior	R_{xy}=.66

general teaching, understanding of current social problems, ability to offer wise guidance on the basis of sound individual and group analysis and knowledge of opportunities. The author also reports, not surprisingly, that the lowest correlations (values not given) were between the WCET and ratings of the teachers' health; physical appearance and poise; energy, enthusiasm, and drive in school work; quality of speech and voice; sense of humor; congeniality of adjustment to associates; neatness of work and classroom; integrity of character.

Lins (43) studied 58 female students who graduated from the University of Wisconsin in 1943, were certified to teach, and were teaching in Wisconsin high schools during 1943-1944. During their first year of teaching, a composite of the independent ratings of at least three evaluators out of a team of five members (two from the School of Education, one from the State Department of Public Instruction, a member of the Department of Educational Methods, and the superintendent or principal of the school) was collected for each teacher using the Wisconsin M-Blank (1940 edition) on a five-point scale. In addition, two staff members of the university and the school principal rated the teachers on a Guide Sheet of five-point scales (as a director of learning, as a friend and counselor of students, as a member of the school staff, as a member of the community, and as a person) and a total rating score. The more interesting zero-order correlations between the predictors and the supervisors' ratings are given in Table 3.

The NTE score is not specified, and so we are assuming that it is the WCET score. Thirty-nine of these teachers were teaching in Wisconsin during their second year after graduation, and 34 principals (12 of these teachers

had changed schools) rated the teachers during their second year of teaching. The correlations between the first and second year of teaching are given in Table 4.

Table 3

Correlations between Predictors and Criterion*

Predictor	N	Correlations with Composite M-Blank during 1st year of teaching (based on 3-5 raters).
H. S. Rank	55	.33
NTE	29	-.15
Undergraduate GPA	58	.31
Education GPA	58	.29
Major Field GPA	58	.33
Practice Teaching	58	.25
Guide Sheet	58	.80

*From Lins (43)

Table 4

Correlations between First- and Second-Year Ratings*

Rating during Teaching	Principal Rating-Second Year of Teaching	
	M-Blank	Guide Sheet
First Year of Teaching		
(a) Principal rating		
M-Blank	.20	--
Guide Sheet	--	.29
(b) Composite during 1st year of teaching		
M-Blank (3-5 raters)	.27	.37
Second Year of Teaching		
Principal rating on Guide Sheet	.77	

*From Lins (43)

Ryans (70) studied a target population consisting of the 1,296 in-service teachers (exclusive of those participating in statewide certification programs) who reported during the NTE administered in 1949 that they had had one or more years of teaching experience. An observation blank was sent to the school principal who was asked to rate each teacher with respect to opposing sets of characteristics on each of three dimensions: pupil behavior, teacher personal-social behavior in the classroom, and teacher behavior indicative of intellectual and educational background of the individual. An additional rating was made by the principal of the overall "general evaluation" of the teacher on 18 dimensions. Data on junior high school teachers were eliminated, thus limiting the study to 192 elementary and 165 secondary school teachers. The correlation between the ratings on the observation blank and the general evaluation were .83 for both the elementary and secondary school teachers. For the elementary school teachers, the NTE subtest on General Principles and Methods of Teaching correlated .17 with the observation blank and .23 with the general evaluation. For the secondary school teachers, the same NTE subtest correlated .13 with the observation blank and .15 with the general evaluation. The differences between the means of "high" (upper 27 percent) and "low" (lower 27 percent) groups on the NTE subtest were significant on the observation blank (.05 level) and the general evaluation (.01 level) at the elementary level, but the differences were not significant at the secondary level. Mean point-biserial correlations between the items of the NTE subtest and the two rating instruments are reported, and the author correctly comments that these low discrimination indices with the external criteria are not surprising when one considers the probable low reliability

of the ratings, the low reliability of individual test items, and the fact that the NTE subtest measures only a small part of the teacher's overall effectiveness. Since the NTE subtest scores of the Common Examinations have never been equated to each other, the results of this study should be interpreted with caution even though all the candidates took the same form of the test.

Delaney (21) studied the relationship between scores on the NTE, a standardized interview, and an evaluation of education and experience with teaching success for 93 teachers selected for employment in the elementary schools of Elizabeth, N. J. during 1940-1948. A 15-20 minute interview was conducted informally by 5 to 8 members who were either teachers, principals, or supervisors. Each member rated the candidates independently on nine personality traits (voice and speech, appearance, alertness, ability to present ideas, judgment, emotional stability, self-confidence, friendliness, and personal fitness for position); the values for each scale ranged from 15 to 75 points, and the ratings of the interviewers were averaged. Teaching success was determined by ratings made by principals and supervisors on five-point scales in four areas: working control, skill in teaching, cooperation, and preparation and growth. An overall rating was also assigned. A composite rating of teaching success was obtained by averaging the last rating by a principal, the average rating by all the principals with whom the teacher had taught, the last rating by the elementary education supervisor (who observed all of the teachers in the study), and the average rating by the three supervisors of elementary education who served during that time. An average of 56 ratings by at least 4 raters was available for the group with at least 4 ratings by principals or supervisors for each teacher. The

composite ratings of teaching success were correlated .17 with the NTE WCET scores, .43 with the interview scores, and .16 with the experience and training scores. When the NTE scores were combined with the interview scores and the scores on the evaluation of training and experience, they added less than .01 to the multiple correlation of .45 between these two variables and the composite ratings. For 81 teachers, the correlation between the average ratings on the original interview and the ratings on the same rating form by the principals in whose schools they taught from 1 to 5 years was .49 on the total score, with the correlations on the 9 personality traits ranging from .30 for friendliness to .48 for emotional stability.

Shea (74) studied the correlations between several predictors and success in teaching for 110 graduates of Worcester State Teachers College. Teaching success was measured by the rating on the M-Blank by either superintendents, principals, or supervisors at the end of the first year of teaching. Undergraduate GPA had the highest correlation with the criterion of teaching success ($r = .50$), followed by the WCET ($r = .45$). The NTE was moderately correlated with the Cooperative General Culture Test ($r = .77$), with the ACE Psychological Examination ($r = .70$), with the Cooperative English Examination ($r = .64$), with the Cooperative Contemporary Affairs Test ($r = .64$), and with undergraduate GPA ($r = .52$). The NTE were not closely associated with practice teaching grades ($r = -.01$), nor was the undergraduate GPA ($r = .31$). The correlation between practice teaching grades and the M-Blank was .38. A factor analysis was performed using the subtests of the NTE Common Examinations and the other

variables; the dates during which the NTE were taken are not given, but even if all candidates took the same form of the NTE, the results would have to be interpreted with caution because the subtest scores of the NTE Common Examinations have never been equated to each other from form to form.

Thacker (81) studied a random 10 percent sample of seniors who prepared for teaching, qualified for teaching certificates, graduated from colleges and universities in North Carolina in 1960, and who were teaching in North Carolina during 1960-1961. In 1960, all applicants for a teaching credential in North Carolina were required to take the NTE. Of 155 teachers in the study, 145 were found in separate schools while two teachers were found in each of five schools. Complete data were available for 126 teachers (100 white, 26 black). Because the size of the black sample was so small, the discussion of the findings will be limited to the 100 white teachers, of whom 58 percent taught in secondary schools and 42 percent taught in elementary schools. Scores on the spring 1960 administration of the NTE were correlated with seven measures of teacher preparation and effectiveness: 1) principals' ratings of teachers after one year of teaching experience (81 percent return rate); 2) supervisors' ratings of teachers during student teaching (76 percent return rate); 3) undergraduate GPA; 4) GPA for general education (language, literature, history); 5) GPA for professional education (courses that satisfied the professional education requirements for a Class A certificate); 6) GPA for professional education and general education courses combined; and 7) GPA for major field. Two years after the teachers had graduated from college, the college supervisors of the teaching and practicum phase of training were asked to rank their students on potential as a teacher,

by size of group, on the basis of the records that they had maintained while the teachers were undergraduates. Similarly, the principals were asked to rank their teachers on overall effectiveness as a teacher based on their records. The resulting ranks were converted to T-scores, and the correlations between the NTE WCET scores, principals' ratings, and the other criteria are given in Table 5.

Table 5

Correlations between Criteria and Predictors*

Criteria	NTE WCET	Principals' Ratings
Principals' Ratings	.18	
Supervisors' Ratings	.17	.03
Undergraduate GPA	.48	.08
General Education GPA	.52	.19
Professional Education GPA	.45	.05
GPA for combined Prof. & Gen. Ed.	.54	.17
Major Field GPA	.35	.01

*From Thacker (81)

Correlations are also reported between the NTE subtests of the Common Examinations and principals' ratings. Although all candidates took the same form of the NTE, these subtest scores have never been equated to each other and therefore these additional correlations are of doubtful generalizability to other test forms.

Eissey (28) studied 111 teachers who were certified to teach, had graduated from Florida State University during 1960-1961, had taken the

NTE upon graduation, and who were teaching in Florida during 1961-1962 and also during 1963-1964. Fifty-one percent of the teachers were teaching in grades 1-6 and 49 percent in grades 7-12. At the end of the first year of teaching, the principals' average rating was computed for a series of five-point scales on personal qualifications (emotional stability, health, and so forth for 10 items), teaching skills (plans, creative ability, and so forth for 11 items), relations with others (cordial, respected by, and so on for 6 items), professional ethics and performance (attitude, carries out policies, and so forth for 5 items), moral and social ethics and performance (moral standards, and so forth for 5 items), and a total score. At the end of the third year of teaching experience, the principals' average rating was computed for a series of two- or three-point scales on personal qualifications (health, appearance, and so on for 7 items), relations with others (respected by pupils, professional ethics and so on for 5 items), teaching skills (knowledge of subject, control of pupils etc. for 6 items), and a total score. A total score on eight items rated during internship by the directing teacher and a total score on these same items rated during internship by the university supervisor were also available. The correlations among these variables are given in Table 6.

Walberg (84) studied 280 students in their last year of elementary teacher training at Illinois Teachers College. During the last week of the student-teaching semester, the student teachers were rated by their principals and also by their field supervisors on a six-point scale for 10 personal characteristics of effective teaching: initiative, reliability,

Table 6

Correlations between Criteria and Predictors*

	Principal Rating End of 1st Year	Principal Rating End of 3rd Year	NTE WCET	Univ. Super. Rating
Directing-Teacher Rating	.10	.01	.14	.54
University Supervisor Rating	.18	.13	.09	
Internship GPA	.15	.14	.01	
Undergraduate GPA	.25	.15	.23	
Professional Education GPA	.22	.17	.19	
Teaching Field GPA	.16	.15	.16	
NTE WCET	.10	.10		
Principal Rating: End of Third Year	.21			

*From Eissey (28)

industry, open-mindedness, cooperation, personal appearance, emotional stability, social adaptability, leadership, and courtesy, and the ratings were summed for the items. The field supervisors also rated the student teachers on a three-point scale for classroom performance on 10 items: classroom management, discipline, motivation, curriculum, personal adjustment, planning, procedures, teaching, records, and responsibility, and these ratings were summed for an overall performance rating. The resulting correlations are given in Table 7.

Lewis (42) studied 45 student teachers at Sul Ross State College who took the NTE during the same semester that they did their student teaching. The correlation between the NTE scores and a rating by the student teacher's college coordinator of the teacher's success in student teaching was .18.

Table 7

Correlations between Predictors and Criteria*

Criteria	NTE WCET	Principal Rating of Personal Characteristics	Field Supervisor Ratings: Personal Characteristics	Field Supervisor Ratings: Classroom Performance
High School GPA	.10	.06	.06	.06
Seventh-term cumulative college GPA	.36	.07	.08	.08
Practice Teaching Grade	-.04	.17	.22	.18
Principal Rating: Personal characteristics	.00	--	.21	.29
Supervisor Rating: Personal characteristics	-.03	--	--	.20
Supervisor Rating: Classroom performance	.02	--	--	--

*From Walberg (84)

Unfortunately, the type of NTE score (Common Examinations, Teaching Area Examinations) is not specified, and so it is impossible to know if the scores were used properly. Further, the author does not describe the rating instrument; thus, it is impossible to make any judgments about its appropriateness or usefulness. Because of the lack of description of either the predictor or criterion, this study is of doubtful value.

Carson (11) studied a group of probationary teachers in Houston who had taken the NTE between 1957-1968. No teachers were included who were returning from a leave of absence or had previously worked for the school district as probationary teachers. The school principal rated the teachers at the end of the first 12 weeks of teaching and at the end of the two-year probationary

period on five-point scales for: 1) personal efficiency (health, voice, and so on for 7 items), 2) social efficiency (spelling, handwriting, and so on for 5 items), 3) professional attitude (reads professionally, attends professional meetings, and so forth for 3 items), 4) cooperation (with other teachers, and so forth for 3 items), 5) skill in teaching (lesson planning, conducting recitation, and so on for 8 items), and 6) classroom management (skill in discipline, neatness of room, and so on for 5 items); there was also a single overall rating called the "general rating," and scores on the 31 items were summed to form a "composite rating." The correlations between the NTE and the twelve-week principals' ratings are given in Table 8.

Table 8

Intercorrelations among Predictors and Criteria*

12-week Principal Rating	WCET (N=241)	Correlations between 12-week and 2-year Principals' Ratings (N=179)							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Personal Efficiency	.08	.53							
(2) Social Efficiency	.13		.57						
(3) Professional Attitude	.09			.43					
(4) Cooperation	-.03				.37				
(5) Skill in Teaching	.16					.42			
(6) Class Management	.04						.49		
(7) General Rating	.11							.42	.46
(8) Composite Rating	.10							.50	.54

*From Carson (11)

Correlations between the WCET scores and the two-year principals' rating were not reported. The author did report correlations, multiple correlations, and cross-validation results for the subtests of the NTE Common Examinations with principals' ratings, but the results must be discounted since these subtest scores have never been equated to each other.

Pupil Ratings

We were able to find only two studies that related NTE scores to pupil ratings. The Flanagan Study (33) discussed earlier also concerned ratings from at least five pupils who had taken a course from the teacher during the previous year. The pupils were not told that their reports would be used in appraising their teachers and they did not know which of their teachers were included in the study. No correlation coefficients were reported, but some interesting results appeared in the answers to the following questions:

Table 9

Relationship between Pupils' Attitudes toward Teachers and NTE Scores of Teachers*

Questions	NTE WCET Score	
	Below 600	Above 700
Which teachers seem to have a broad knowledge of other subjects besides the one you had with them?	30%	49%
Which of your teachers had the most pleasing personality?	24%	39%
	English Expression NTE Score	
	Below Average	Superior
Which teachers were most clear in presenting their ideas?	35%	51%

*From Flanagan (33)

Unfortunately, since the author does not state how the pupils were selected, we cannot know whether or not their responses were truly representative.

The Lins study (43) discussed earlier obtained pupil evaluations from a sampling of five to six pupils (the method of selection of the sample was not specified) who anonymously ranked their teachers from best to poorest among themselves. These ranks were then averaged for each teacher. The correlation between the composite pupil ranking and the composite ranking by the three to five evaluators was .28. Additional correlations between the pupil rankings and the other variables are given in Table 10.

Table 10

Correlation between Predictors and Criterion*

Predictors	N	Correlation with Pupil Ranking Composite
High School Rank	48	.06
NTE	26	-.30
Undergraduate GPA	50	.03
Major Field GPA	50	.05
Education GPA	50	.13
Practice Teaching	50	.06

*From Lins (43)

Pupil Residual Gain Scores

We were able to find only one study that related pupil residual gain scores on achievement tests to the NTE. The Lins study (43) correlated average residual pupil gain scores during the second semester on various standardized

achievement tests (biology, social studies, English, general science, civics) in the 27 classes taught by 17 of the teachers. Lins used pretest, I.Q., and mental age scores to produce a predicted gain score for each class of pupils. For those teachers who had two or more classes, a mean residual gain of the combined classes was used as the score for each teacher. Pupil average residual gain scores were correlated .06 with the pupil ranking composite and .19 with the composite rating of the three to five evaluators. Correlations with the other variables are given in Table 11.

Table 11

Correlations between Predictors and Criterion*

Predictors	N	Correlation with Pupil Average Residual Gain
High School Rank	16	.69
NTE	7	.45
Undergraduate GPA	17	.53
Major Field GPA	17	.55
Education GPA	17	.52
Practice Teaching	17	.21

*From Lins (43)

The correlation of .45 between NTE scores and average pupil residual gain scores is encouraging, but the extremely small sample size does not allow us to place much confidence in the results.

Classroom Observation

Only one study relating classroom observation procedures to NTE scores could be found. Medley and Hill (46) studied the relationship between teaching

style and subtest scores on the NTE Common Examinations. A group of 53 intern teachers in junior high and secondary schools (teachers of mathematics, English, science, and social studies) in a large metropolitan area in the Eastern United States were visited in their classrooms 4 times for about 30 minutes each by a pair of trained observers; one observer coded the teacher's behavior using Flanders' Interaction Analyses while the other used OScAR 4V. The observations were analyzed by a principal components analysis from which 15 scoring keys were built, 8 for OScAR and 7 for Flanders. Data for these 53 teachers and for an additional 38 teachers in the same program were analyzed and 11 of the 19 content areas of the NTE measured significantly different content. Multiple correlations were computed for each of the 15 classroom observation dimensions and these 11 NTE subtests. Only 2 of the 15 equations yielded significant correlations, and only 9 of the 165 beta weights were significant. A multiple correlation of .66 was obtained between Lecturing Behavior as measured by the Flanders technique and scores on the NTE. The beta weights in this equation indicated that teachers who score high on the science items lecture more, while teachers who score high on the teaching principles and practices items lecture less. Whether the results of this study would be replicated if different test items and teachers were used is a good research question.

A Note about Criterion Measures

No single criterion measure is sufficient unto itself in an occupation as complex and demanding as teaching. Scores derived from good paper-and-pencil tests of knowledge of teaching can tell us a great deal about whether a prospective teacher knows the important concepts and principles in a subject

area he would like to teach. However, while subject-matter knowledge may be a necessary prerequisite for successful teaching, it may or may not be a sufficient condition for successful teaching.

Some teachers may excel at one-to-one instruction with a pupil, some at managing large groups of pupils, some at producing a high degree of pupil interaction within small groups, some at stimulating independent behavior in pupils, and some at motivating the pupils in the lower levels of academic achievement. Any attempt to summarize such varied types of teacher behavior on a single rating scale would be foolhardy. Diverse criteria should be predicted separately, especially if different job responsibilities can be found for candidates with different profiles in teaching ability. Instead of trying to select the single best criterion, it is more realistic to think in terms of different types of multiple criteria within a criterion domain or multidimensional criterion space. Locating any individual teacher within this space requires describing the teacher's skills in terms of grade level, subject areas, types of pupils, and types of teaching situations.

The act of teaching is so complex that it is quite reasonable to expect measures that predict one particular outcome to be unsuccessful in predicting others. Separate criterion scores thus should function as partial criteria rather than as The Criterion. Further, the definition of the criterion itself can change over time. For example, changing a school program from one emphasizing large-group instruction to one stressing either individualized instruction or open spaces could easily change the ability of the predictor to estimate scores on the revised criterion.

Ratings of teacher performance have long been attacked and questioned in terms of their usefulness and accuracy. As Cronbach (16) has put it: "When a test fails to predict a rating, it is hard to say whether this is the fault of the test or of the rating." Ratings of teachers by school principals or by field supervisors of the teacher-training programs can easily reflect the degree to which the rater likes the teacher rather than the quality of the teacher's work. In some cases the rater may simply not know the facts about the teacher. Teachers' lunchrooms are filled with stories of teachers who claim that they were rated by someone who visited their classes a total of only 15 to 20 minutes during the entire school year. Such small sampling of the classroom behavior of the teachers can hardly be considered adequate.

Different raters attach different meanings to the traits on which they rate teachers. To one rater, "leadership" might mean relying on authority, dominance, and black-and-white decision-making. To another, it might mean encouraging pupils, effecting cooperative decisions between teachers and pupils, and ruling democratically. Moreover, the rating scale itself may be ambiguous. Rating "cooperativeness," "adaptability," or "sensitivity" on a scale from 0 to 20, or from poor to excellent, is hopeless unless clear descriptions of actual teacher behavior are given for each point on the scale.

The best way to obtain useful information from raters is to train them carefully on the definitions of the items, show them examples of actual teacher behavior for each item, and check the reliability of their ratings of actual classroom situations. It might also prove useful to use raters who do not know the teachers personally. A school principal, for example,

has a large personal investment in his own teachers and a number of reasons for wanting to see them succeed (not the least of which is a need to convince the school district superintendent that he is doing an excellent job of developing his teachers into outstanding members of the profession).

SUMMARY AND CONCLUSIONS

How can we summarize this diverse and confusing collection of articles related to the National Teacher Examinations? How well do the NTE correlate with undergraduate GPA? with ratings of an amorphous concept called "teacher effectiveness"? In the articles we reviewed, we found 16 correlations between WCET scores and undergraduate GPA; these correlations ranged from .23 to .74, with a median value of .55. Thus, we can say with some confidence that for the studies we reviewed the WCET scores are moderately correlated with success as an undergraduate as measured by course grades. Moreover, the WCET scores provide the added advantage of being comparable from form to form of the National Teacher Examinations and of providing a common measure of the training and learning experiences of teachers who are trained in different parts of the country and in programs with a considerable range of sophistication and quality. Course grades or grade-point-averages are typically selected as the criterion in research studies because these scores are easily obtainable and quantifiable, even though grades have been severely criticized for being contaminated with such personal factors as personality, attractiveness, general verbal ability, and handwriting skill, and further vary within departments and between instructors (35).

The course grades assigned for a student-teaching experience at a teacher-training institution have a certain attractiveness as a criterion measure. In practice, however, if almost all students enrolled in a teacher-training program receive a grade of A or B in student teaching, the spread of grades will be so small that prediction of individual differences could be quite unreliable. Further, because a predictor is highly correlated with an end-of-training outcome, such as grade in practice teaching, does not insure that the predictor will be highly correlated with important on-the-job criteria during a full-time teaching experience. The on-the-job criteria may be more demanding on the teacher than the typical demands placed on a student teacher. To the extent that this principle is correct, it is possible that the training that comes on the job may be sufficient to make the course grade in practice teaching a less useful predictor of full-time teaching performance. In this respect, long-term follow-up studies of graduates of teacher-training programs become an essential aspect of research studies designed to check on the effectiveness of the teacher-training program. We located only two correlations between WCET scores and grades in practice teaching and both were practically zero (-.01 and -.04); not surprisingly, knowledge of subject matter, as measured by the Common Examinations, does not appear to be highly related to whatever is summarized by the grade in practice teaching. A grade in a single course would not be expected to be very reliable in any case, especially when the criteria for differentiating between different levels of performance during student teaching remain pretty much undefined and subject to a wide variation from one college supervisor to another.

The correlations between WCET scores and ratings by college supervisors or principals during the student-teaching period are not very encouraging. We discovered six such correlations; they ranged from -.03 to .18 with a median value of .05. Obviously, the WCET scores do not predict these ratings very accurately.

The WCET scores do not correlate highly with ratings given by principals or supervisors during the first year of teaching either. We found seven such correlations with a range of -.15 to .45 and a median value of .11. The one correlation we found between WCET scores and principals' ratings at the end of the third year of teaching was only .10.

The ratings by college supervisors and the undergraduate GPA do not do much better in predicting on-the-job ratings of teachers. The two correlations that we found between college staff ratings and first-year principals' ratings were only .10 and .18; at the end of the third year of teaching, these same ratings by the college staff correlated only .01 and .13 with principals' ratings. The three correlations that we found between GPA and ratings during student teaching by field supervisors or principals were all either .07 or .08. The three correlations between GPA and principals' or supervisors' ratings during the first year of teaching ranged from .08 to .31 with a median value of .25. The two correlations between GPA and grade in practice teaching were only .14 and .31.

Why are NTE scores, college supervisors' ratings, and undergraduate GPA such poor predictors of a teacher's on-the-job ratings? The answer to this question is three-fold. First, in terms of the NTE, any score on a standardized test of knowledge in professional or general education

is bound to measure only a sample of the important qualities necessary to be a successful teacher, many of which have less to do with knowledge of subject matter than with management and planning strategies within the classroom.

Second, on-the-job ratings are notoriously unreliable, and their reputation is well-deserved. A closer look at two of the studies we reviewed will illustrate this point. In the study by Eissey (28), the correlation between the principals' ratings at the end of the first year of teaching correlated only .21 with their ratings at the end of the third year of teaching. Similarly, in the Lins study (43), the ratings by principals or supervisors during the first year of teaching correlated with an average value of only .28 with the principals' ratings during the second year of teaching. These low correlations provide us with two hypotheses, both of which are logically persuasive: Most likely the ratings by principals or supervisors are highly unreliable because of the lack of training of these raters as systematic observers in reliability studies and the ambiguity and generality of the items on which they rate teachers (both in meaning and in perception of the necessary behaviors). Moreover, it is quite likely that the teacher's behavior is changing over time, sometimes dramatically during the early years of teaching, because of the vast difference between the responsibilities of a student teacher and those of a full-time teacher and in some cases because of a lack of preparation for the problems that the teachers encounter during their early years of teaching. Unfortunately, until some systematic training of observers helps clear up the observer reliability question, we cannot test these two hypotheses; they are hopelessly confounded in the studies we have reviewed.

The rating scales we reviewed have one outstanding defect: The "composite rating" to which these scales so often refer is really a conglomeration of very disparate rating items that are summed to form a total score of unintelligible meaning. For example, how would you describe a composite rating score on poise, personality, classroom control, moral character, community relations, and conduct during recitation? Can one seriously expect the National Teacher Examinations to correlate with a teacher's neatness (11), moral standards (28), voice and speech (21), health (28), personal appearance (84)? Until the rating scales more closely relate to what the standardized tests themselves attempt to measure, such blind correlating of personal characteristics with NTE scores is not likely to produce fruitful research results. Wood (87, pp. 278-279) made this same plea over 30 years ago:

> To abandon examinations of intelligence, general culture, and professional information because they do not also measure personality, moral character, interest in children, and other important factors that determine teaching ability, would be as illogical as to abandon the use of the clinical thermometer and stethoscope because they do not measure a thousand other important diagnostic factors. We should avoid the naive error of judging the validity of such tests in terms of their correlation with available criteria of teaching success, just as the physician refuses to judge the validity of his thermometer in terms of the correlation of its readings with total health or life-expectancy estimates. The validity of the examinations should be judged by the accuracy with which they measure, not the total complex of teaching ability, but those parts which they are designed to measure....

The argument that knowledge of methods of teaching is more important than knowledge of the content to be taught is specious. All of us would readily agree that every teacher should possess at least minimum competence

in knowledge of the subject areas that he is teaching, but how do you determine such minimum competence? The argument that the National Teacher Examinations measure only "mere knowledge" was dispensed with more than 30 years ago by Kandel (38, p. 755):

> They object to the tests of "mere knowledge" or of the kind of knowledge which "a scholar might be expected to know." The position is not novel; it is simply an echo of the American tradition of teacher preparation that a teacher need know nothing provided he knows how to teach.... Whence will he derive his content without proceeding in vacuo?.... Behind classroom procedures there must be a fund of something on which the teacher and pupils must draw; that fund all teachers must have; how they draw on that fund may vary with the current fashion, but "the what" cannot be discarded in favor of "the how."

How do you validate a test that purports to measure knowledge of concepts and principles necessary to be a "well-educated" person or a "well-educated" subject-matter specialist? One way is by arguing for the content validity of the test. The soundness of such arguments has been recognized in the Equal Employment Opportunity Program's guidelines (22). Since the argument for the content validity of any test is always a logical one, there is no such thing as a coefficient of content validity. A test is content-valid if a group of experts can agree that the test measures the objectives it is supposed to measure.

Given the low correlations between the National Teacher Examinations and ratings of on-the-job performance by principals and supervisors, it is difficult to justify the use of fixed cutoff WCET scores in considering salary raises of teachers as advocated by Eckelberry (27), for contract assignment by school districts as described by Carson (11), for provisional

teaching certificates as described by Starcher (78) and Boozer (8), and for differential rating on teaching certificates, as described by Crow (18), even if we allow for the unreliability of these ratings. Since we were unable to locate a single study that used scores earned on any of the Teaching Area Examinations after 1964, when these scores were first equated to each other, the use of fixed cutoff scores across TAE (Crow and Starcher) is especially arbitrary, blind, and inappropriate. Even if this practice had been in use only since 1964, it would still have screened out a different percentage of those candidates who took different TAEs--for example, a cutting score of 600 would screen out 30 percent of the college seniors who took the Mathematics TAE but only 15 percent of the college seniors who took the Biology and General Science TAE.

Perhaps more important than revising principal and pupil rating scales is to conduct systematic studies of the relationship between the NTE scores of teachers and average residual achievement gain scores of pupils in their classes. It is a reasonable hypothesis that the more a teacher knows about what he is teaching, the more his pupils will learn about it. Since pupil learning is one of the most important intended outcomes of the public schools, such studies would seem to be imperative.

REFERENCES

1. Anderson, W. A. The National Teacher Examinations - A criticism. Childhood Education, 1941, 18, pp. 179-181.

2. Anonymous. Is the National Teacher Examination service needed? News Bulletin for Classroom Teachers. National Education Association, Department of Classroom Teachers. (Montana Education, 1942, 18, p. 8.)

3. Anonymous. National Teacher Examinations. Illinois Education, 1945, 34, pp. 75-76.

4. Benson, A. L. The National Teacher Examinations in 1954. Journal of Teacher Education, 1954, 5, pp. 244-248.

5. Benson, A. L. The role of examinations in the preparation of teachers. The Journal of Teacher Education, 1959, 10, pp. 491-496.

6. Berg, H. D. Social studies: National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 1226-1227.

7. Blommers, P. Mathematics: National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 873-874.

8. Boozer, H. R. External examinations as predictors of competence. Journal of Teacher Education, 1965, 16, pp. 210-214.

9. Brownwell, W. A. National Teacher Examinations. The Fifth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1960, 5, pp. 632-635.

10. Capps, M. P. and DeCosta, F. A. Contributions of the Graduate Record Examinations and the National Teacher Examinations to the prediction of graduate school success. Journal of Educational Research, 1957, 50, pp. 383-389.

11. Carson, E. M. An analysis of National Teacher Examinations scores as predictors of teacher success in assignment. Unpublished doctoral dissertation, University of Houston, 1969.

12. Collins, E. R. Teacher selection by examination. Harvard Educational Review, 1940, 10, pp. 3-6.

13. Cook, W. W. National Teacher Examinations. The Fifth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1960, 5, pp. 635-638.

14. Crissy, W. J. A reply. Journal of Higher Education, 1941, 12, pp. 484-487.

15. Crissy, W. J. E. The National Teacher Examinations. Phi Delta Kappan, 1942, 24, pp. 353-356.

16. Cronbach, L. J. Essentials of psychological testing, third edition. New York: Harper and Row, 1970, p. 127.

17. Croon, C. W. Performance of the physical science candidates in the National Teacher Examinations. American Journal of Physics, 1941, 9, pp. 45-49.

18. Crow, E. R. Teacher examinations and the South Carolina certification program. Educational Record, 1947, 28, pp. 453-462.

19. Dailey, J. T. Development and applications of tests of educational achievement outside the schools. Review of Educational Research, 1953, 23, pp. 102-109.

20. Davis, F. B. The measurement of professional information among candidates for teaching positions. Education Administration and Supervision, 1941, 27, pp. 99-106.

21. Delaney, E. C. Teacher selection and evaluation - with special attention to the validity of the personal interview and the National Teacher Examinations as used in one selected community (Elizabeth, New Jersey). Unpublished doctoral dissertation, Columbia University, 1954.

22. Department of Labor: Office of Federal Contract Compliance Equal Employment Opportunity Program. Federal Register, 1971, 36, (77), pp. 7532-7535.

23. Derthick, L. G. National Teacher Examinations. The Fifth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1960, 5, pp. 638-639.

24. Douglass, H. R. National Teacher Examinations. Nation's Schools, 1941, 27, pp. 24-25.

25. Duncan, J. A statistical analysis of student grades in teacher-education curricula and certain scores on the National Teacher Examinations. Unpublished master's thesis, East Tennessee State University, 1971.

26. Durflinger, G. W. A study of recent findings on the prediction of teaching success. Educational Administration and Supervision, 1948, 34, pp. 321-337.

27. Eckelberry, R. H. Higher salaries for superior teachers. Educational Research Bulletin, 1949, 28, pp. 77-78ff.

28. Eissey, E. M. Selected variables in teacher preparation and subsequent evaluation of teaching performance. Unpublished doctoral dissertation, Florida State University, 1967.

29. Elting, R. A. and Butterfield, M. Academic performance of Cuban-teacher students at the University of Miami. College and University, 1969, 44, pp. 263-267.

30. Emens, J. R. National Teacher Examinations. Nation's Schools, 1947, 39, p. 47.

31. Fenstermacher, G. M. and Swineford, F. The National Teacher Examinations and the appraisal of teacher preparation. Journal of Teacher Education, 1958, 9, pp. 429-434.

32. Flanagan, J. C. An analysis of the results from the first annual edition of the National Teacher Examinations. Journal of Experimental Education, 1941a, 9, pp. 237-250.

33. Flanagan, J. C. A preliminary study of the validity of the 1940 edition of the National Teacher Examinations. School and Society, 1941b, 54, pp. 59-64.

34. Hill, H. H. The role of examinations in teacher selection. New Directions for Measurement and Guidance. Series I. Reports of Committees and Conferences. Washington, D. C.: American Council on Education Studies, 1944, 8, pp. 82-86.

35. Hills, J. R. Use of measurement in selection and placement. Educational Measurement, second edition, edited by R. L. Thorndike. Washington, D. C.: American Council on Education, 1971, pp. 680-732.

36. Hunter, E. C. Results of the National Teacher Examinations at Tulane University, 1948-1966. Journal of Psychology, 1968, 68, pp. 193-201.

37. Johnson, B. G. The prediction of success in the doctoral program in the College of Education at the University of Houston on the basis of objective test scores and quality point averages. (Doctoral dissertation, University of Houston) Ann Arbor, Michigan: University Microfilms, 1963. No. 64-2817.

38. Kandel, I. L. The teacher's right to be ignorant apropos of the criticisms of the National Teacher Examination. School and Society, 1940, 51, pp. 753-756.

39. King, H. V. The performance of the Latin group in the 1940 National Teacher Examinations. Classical Journal, 1941, 36, pp. 357-361.

40. Koerner, J. D. The miseducation of American teachers. Boston: Houghton Mifflin Co., 1963, cf. especially pp. 254-256.

41. Larson, W. S. Music Education: National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 623-626.

42. Lewis, J. Tests of teachers. The Texas Outlook, 1968, 52, pp. 20-21.

43. Lins, L. The prediction of teaching efficiency. Journal of Experimental Education, 1946, 15, pp. 2-60.

44. Martin, L. O. Teacher colleges can select students with superior ability. School and Society, 1946, 63, pp. 318-319.

45. McCamey, J., Jr. The correlation between certain academic factors and scores on the 1957 National Teacher Examinations of the 1957 graduates of the University of Hawaii Teachers' College. Unpublished masters thesis, University of Hawaii, 1958.

46. Medley, D. M. and Hill, R. A. Cognitive factors in teaching style. Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, March 4, 1970.

47. National Teacher Examinations. Bulletin of information for candidates 1970-71. Princeton, New Jersey: Educational Testing Service, 1970.

48. National Teacher Examinations. Guidelines for using the National Teacher Examinations. Princeton, New Jersey: Educational Testing Service, 1971, p. 8.

49. National Teacher Examinations. The National Teacher Examinations: Interpretation of scores. Princeton, New Jersey: Educational Testing Service, 1969.

50. National Teacher Examinations. Prospectus for school and college officials. Princeton, New Jersey: Educational Testing Service, 1970.

51. Pearson, C. A. and Ryans, D. G. The performance of science teachers on the 1941 National Teacher Examinations. The Science Counselor, 1942, 8, pp. 2-4ff.

52. Pilley, J. E. The National Teacher Examination Service. The School Review, 1941, 49, pp. 177-186.

53. Pitcher, B. The relation of academic success in teacher preparatory curricula to scores on the NTE Common Examinations. ETS Statistical Report 62-63. Princeton, New Jersey: Educational Testing Service, 1962.

54. Quirk, T. J., Witten, B., and Weinberg, S. F. The National Teacher Examinations: An annotated bibliography 1940-1971. Research Memorandum 72-4. Princeton, New Jersey: Educational Testing Service, 1972.

55. Reed, C. R. The role of examinations in teacher selection. Educational Record, 1941, 22, pp. 44-53.

56. Rivlin, H. N. National Teacher Examinations. The Fourth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1955, 4, pp. 799-802.

57. Roberts, H. English language and literature: National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 558-560.

58. Rowland, A. L. The proposed teacher examination service. Harvard Educational Review, 1940, 10, pp. 283-288.

59. Rowland, A. L. Examinations or credentials in hiring classroom teachers. Nation's Schools, 1941, 27, pp. 53-54.

60. Ryans, D. G. The professional examination of teaching candidates: a report of the first annual administration of the National Teacher Examinations. School and Society, 1940, 52, pp. 273-284.

61. Ryans, D. G. The 1941 administration of the National Teacher Examinations. School and Society, 1941, 54, pp. 361-368.

62. Ryans, D. G. New features of the National Teacher Examinations program. Journal of Educational Research, 1945, 39, p. 155.

63. Ryans, D. G. Notes on teacher selection: sources of information about the candidate. Education Administration and Supervision, 1946, 32, pp. 333-342.

64. Ryans, D. G. Appraising teacher personnel. Journal of Experimental Education, 1947, 16, pp. 1-30.

65. Ryans, D. G. The 1948 National Teacher Examinations. Journal of Experimental Education, 1948, 17, pp. 1-25.

66. Ryans, D. G. An analysis of teacher examination scores of college seniors who expect to become teachers. American Psychologist, 1949a, 4, p. 288. (Abstract)

67. Ryans, D. G. Local selection. Review of Educational Research, 1949b, 19, pp. 210-218.

68. Ryans, D. G. The function of examinations in the selection of teachers. School Executive, 1949c, 68, pp. 39-41.

69. Ryans, D. G. The use of the National Teacher Examinations in colleges and universities. Journal of Educational Research, 1949d, 42, pp. 678-689.

70. Ryans, D. G. The results of internal consistency and external validation procedures applied in the analysis of test items measuring professional information. Educational and Psychological Measurement, 1951, 11, pp. 549-560.

71. Schultz, H. A. Art education: National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 616-618.

72. Seagoe, M. V. The prediction of success in a graduate school of education. School and Society, 1949, 69, pp. 89-93.

73. Seashore, H. National Teacher Examinations. The Sixth Mental Measurements Yearbook, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1965, 6, pp. 943-947.

74. Shea, J. A. The predictive value of various combinations of standardized tests and subtests for prognosis of teaching efficiency. Washington, D. C.: Catholic University of America Press, 1955.

75. Simpson, H. D. An analysis of the relationship between scores attained on The National Teacher Examinations and certain other factors. (Doctoral dissertation, University of Georgia) Ann Arbor, Michigan: University Microfilms, 1962. No. 63-7465.

76. Spaney, E. The performance of the mathematics candidates in the 1940 National Teacher Examinations. Mathematics Teacher, 1941, 34, pp. 8-12.

77. Spaulding, G. The achievement of the modern language candidates in the National Teacher Examinations. Modern Language Journal, 1941, 25, pp. 361-367.

78. Starcher, G. National Teacher Examinations: A certification instrument in West Virginia. Journal of Teacher Education, 1959, 10, pp. 102-106.

79. Stoddard, A. J. The selection of teachers from the national viewpoint. Educational Record, 1940, 21 (suppl. 13), pp. 144-151.

80. Sutcliffe, H. N. A study of some correlations existing between the four year indices of selected seniors and graduates-in-service of the Rhode Island College of Education and their National Teacher Examination scores: 1940-1950. Unpublished master's thesis, Rhode Island College of Education, 1952.

81. Thacker, J. A. A study of the relationship between principals' estimates of teaching efficiency and scores on National Teacher Examinations, academic averages, and supervisors' estimates of potential of selected teachers in North Carolina. (Doctoral dissertation, University of North Carolina, Chapel Hill) Ann Arbor, Michigan: University Microfilms, 1964. No. 68-3661.

82. Townsend, M. E. An experiment in the professional examination of teachers. School and Society, 1939, 50, pp. 537-541.

83. Tully, A. G. A study to evaluate the achievement of the senior students of the Rhode Island College of Education in the National Teacher Examination programs during the years 1940-1950. Unpublished master's thesis, Rhode Island College of Education, 1952.

84. Walberg, H. J. Scholastic aptitude, the National Teacher Examinations, and teaching success. Journal of Educational Research, 1967, 61, pp. 129-131.

85. Williams, R. F. Should teachers be tested? Virginia Journal of Education, 1968, 61, p. 10.

86. Winetrout, K. The National Teacher Examinations, 1941. Journal of Higher Education, 1941, 12, pp. 479-484.

87. Wood, B. D. Making use of the objective examination as a phase of teacher selection. Harvard Educational Review, 1940, 10, pp. 277-282.

88. Wood, B. D. Scores on National Committee Teacher Examinations, 1940 and 1941. School and Society, 1941, 54, pp. 625-627.

89. Wood, B. National Teacher Examinations...a reply to Dr. Anderson. Childhood Education, 1942, 18, pp. 227-230.