

A Subgroup Analysis of Predictors to Standards – Based Certification Examination Scores In Differing Principal Preparation Programs in Texas

**Elaine L. Wilmore
Dallas Baptist University**

While the need for quality school leaders is well known, there is an increasing shortage across the nation (Bowles, King, & Crow, 2000; Erlandson, 2000; Fenwick & Pierce, 2001; Million, 1998; Potter, 2001; Richardson, 1999; Steinberg, 2000; “Study Warns,” 1998; Winter & Morgenthal, 2002). Concurrently, many states are implementing difficult certification/licensure examinations (Texas Administrative Code; Accountability System for Educator Preparation 19§241.01a, 1999). This is particularly true in Texas where the state standards – based Examination for the Certification of Educators in Texas (ExCET), now known as TExES, is required for persons seeking initial or additional certifications in any field (TEC, sub.D: chap. 21.048a, TAC, subchap. M., ASEP 19§230.413, 1999). It is paradoxical that while there is an increasing need for more educators, the tests are also scaled to become increasingly difficult. Universities are under pressure to produce graduates that pass in totality as well as by race and gender subgroups (TAC, ASEP 19§229.3 a, 1, A, 1999; TAC, ASEP 19§229.3 e, 2, B, 1999).

The Texas system of educator preservice program evaluation is the Accountability System for Educator Preparation (ASEP). This system consists of two components: state generated exams active by 1998 and performance evaluations active by 2002. The performance evaluation states that poor accomplishment by any subgroup of students, even if overall scores are acceptable, puts the institution or program under review (Laitsch, 1998). Similar to *No Child Left Behind* requirements, programs that do not annually meet strict demographic criteria are labeled “under review.” If they do not improve they can lose their authority to prepare educators (TAC, ASEP 19§229.3 a, 1, A, 1999; TAC, ASEP 19§229.3 e, 2, B, 1999). This has obvious serious concerns and impact on all preparation programs (Littleton, 2000). While accountability systems have been criticized as being developed with little regard to what constitutes “accountability,” other issues abound with high stakes standardized testing. With the further implementation of *No Child Left Behind* these State of Texas issues have tremendous national implications. The increase of high–stakes testing in higher education as well as PreK–12 academics has an important effect on instructional improvement in both areas.

This testing impact has clear implications on educational policy for school leaders, states, legislatures, and the development, implementation, and evaluation of program standards. But do these certification exams accurately measure what they purport to measure? Is there a relationship between success

on certification exams and improved teaching and learning? Last, are there predetermined factors that significantly influence a student's likelihood of success on the exam? If so, how can this knowledge be used to improve principal preparation and Pre K – 20 teaching and learning? What should this knowledge have on policy development, implementation, and evaluation?

This study addressed the factors of GRE (Graduate Record Exam), race, gender, and undergraduate GPA (grade point average) as predictors of principal certification examination success at The University of Texas at Arlington, a diverse urban university. At the time of the study the university had three programs that lead to a master's degree and principal certification. The Regular program consists of graduate students, not in a cohort, who complete targeted degree plans on individual timelines. The other two programs, Scholars of Practice and Educational Leadership UTA, consist of cohorts of graduate students who progress through concurrent coursework and internships. In the Scholars of Practice program students retain their positions as teachers, but are assisted by their district in supported internship experiences within an 18-month period. In the third program, Educational Leadership UTA, students are selected by their respective public, private, or charter districts to serve as paid administrative interns for one year. Coursework is presented during the preceding summer. Intense full-day Wednesday seminars are utilized weekly throughout the academic year. Attempts are made in all three programs to connect theory to practice in an authentic, pragmatic, and practical manner.

Objectives of the Study

1. Are there significant differences between state principal certification examination scores among students in the Regular, Scholars of Practice, or Educational Leadership UTA programs?
2. Are the variables of race, gender, GRE, or undergraduate GPA predictors of state principal certification examination scores?
3. What, if any, is the value of standards-based testing? Could standards be more appropriately assessed in a different manner that can better quantify a greater link to teaching and learning?

Limitations of the Study

- The lengths of the internships and degree of district support are not factored.
- Whereas scores are studied over a five-year period for the Regular and Educational Leadership UTA programs, the newer Scholars of Practice cohort had data for only one year.
- Only one minority student and two males were in the first Scholars of Practice cohort. This cohort's lack of diversity affects race and gender variables and results for those factors.

Theoretical Framework

Although there is considerable longitudinal discussion about the validity of the GRE as a predictor of educational administration graduate school success (Lindle & Rinehart, 1998; Nagi, 1975; Wendel, 1991), there have been no studies of its use on state licensure/certification examination passing rates for school administrators. Several studies explored the issue regarding the state certification examination for teachers. Nance & Kinnison (1988) found the ACT and GPA were reliable predictors on the Texas ExCET for undergraduate education majors. Chambers, Munday, Sienty, and Justice (1999) revealed that along with the variables of GPA, age, gender, critical thinking abilities, reading ability, and Texas Academic Skills Program (TASP) scores, each could be used as predictors of successful performance on the professional development ExCET for teachers. However, when each variable was tested separately, only critical thinking abilities were predictive of successful performance. Justice and Hardy (2001) examined characteristics of minority teacher education students taking the professional development section of the ExCET and found that undergraduate GPA was a weak predictor of performance.

Nearly 30 years ago, Nagi (1975) began looking at the validity of the GRE and the Miller Analogies Test (MAT) as predictors of completion of the doctoral program in educational administration at the State University of New York at Albany. Comparisons between the MAT and GRE with educational administration graduate school performance have continued as Wendel (1991) correlated these with measurements obtained through authentic assessment by the Assessment Center Project of the National Association of Secondary School Principals. House's (1997, 1998) studies of GRE and gender found that while the GRE generally was predictive of graduate performance in a number of cases, it under-predicted the achievement of female students and over-predicted the achievement of males. The work of Lindle and Rinehart (1998) found GRE analytic scores should be given more weight in educational administration admissions decisions.

In an increasingly diverse society, high-stakes certification exams have the same effects that similar testing has on Pre K-12 students. It is necessary that all educators be given equal opportunity to succeed within the system. An analysis of certification exams in Texas reveals that the percentage of minorities in the teacher pool is declining (Littleton, 2002). These data support Gillis' (1989) prediction over a decade ago that only 5% of the U.S. teaching force would be people of color by 2000 while at least 33% of the students would be minorities. As early as 1986, Cooper (1986) reported the difficulty that Black education graduates experienced with teacher certification exams. The Southern Regional Education Board project was established to train faculty in Black colleges to lead faculty and curriculum development in order to improve student performance on standardized tests and to increase the pool of minority public school leaders. This situation remains unchanged today which accounts for why the ASEP system

(TEC, sub.D: chap. 21.048a, TAC, subchap. M., ASEP 19§230.413, 1999) is so heavily weighted towards subgroup performance.

With the many problems and possibilities associated with standards and standards-based reform, Falk (2002) contends that standards-based reforms can better support learning if they are used to direct teaching toward worthy goals. Working with standards and standards-based assessments has stimulated teachers to clarify instructional purposes and utilize a deepened range of instructional strategies that support students' varying learning modes. Conversely, others assert standards-based testing has reduced teaching to test preparation and the false pursuit of a "one-size-fits-all" approach to learning. Yet, because of the current lack of viable alternatives to large scale testing, policymakers are likely to continue to rely on the use of standards-based tests for accountability (Littleton, 2000).

Many educational institutions are developing alternative ways of assessing educators in preparation for their jobs. Hill (2000) encourages the development of electronic portfolios in order to reflect a wider range of complex knowledge, skills, and dispositions that quality professional educators possess that scores on licensure exams fail to reflect. Similarly, another method of assessing teacher performance and student learning is the new Teacher Work Sample Methodology (TWSM). This is a quality assurance system that can assess what students learn, how well they learn, what progress they make in individual learning, and how students who are not making progress can be helped. The TWSM is a tool which links teacher effectiveness and preparation to student learning (Fredman, 2002).

Although strategies such as these have been utilized for teacher effectiveness, there have been no proven links of principal preparation and effectiveness to student learning. There have also been no empirical studies of race, gender, undergraduate GPA, and GRE as predictors of principal certification exam success. This study was necessary to provide equitable recruitment, admissions, instructional, and support services to potential school leaders and to determine if a statistically significant correlation exists among any of these factors. Analyses of the results are necessary to address equitable school leader recruitment and enhancement. In addition, the results of this and future studies are important to state and national policy and legislation on the utilization of certification exams for school leaders.

Methods, Techniques, or Modes of Inquiry

From 1996-2001 over 337 students completed administrative certification requirements at the University of Texas at Arlington, an urban university with approximately 23,000 students. An ANOVA with Scheffe's multiple comparisons test was performed to compare the mean examination scores of students in each program. Then a computer generated multiple regression analysis using SAS® software was then utilized to determine if there was a significant relationship between the GRE, race, gender, and undergraduate GPA in predicting

certification examination scores. Due to the small number of ethnic subgroups (Black, Hispanic, Asian, Indian) having taken the ExCET during the five year time period, their scores were combined to make them more statistically robust. The subsequent minority number was 60.

Data Sources or Evidence

Students were disaggregated per program for race, gender, GRE scores, and undergraduate GPA. Demographic data, GRE scores, and undergraduate GPA were obtained from Graduate College records. The State Board for Educator Certification (SBEC) supplied examination results for Educator Certification. Scores were then disaggregated based on all factors.

Objective 1 Methodology, Results, and Conclusions

Objective 1 questioned, “Are there significant differences between state principal certification examination scores among students in the Regular, Scholars of Practice, or Educational Leadership UTA programs?” To determine this, means for each group were computed and compared. Based on these results there was evidence to infer that at least two means differed, therefore there were significant differences among the means of the different groups. However, results thus far did not show which of the program means differed. To determine this, a Scheffe’s multiple comparisons test was utilized. This test showed differences in the mean ExCET scores between:

- The Scholars of Practice and Educational Leadership UTA cohorts
- The Scholars of Practice and Regular program

However, there were no differences in the mean ExCET scores between the Educational Leadership UTA cohort and the Regular program. Therefore, only the Scholars of Practice program showed significant differences with any of the other programs. As cited as a limitation of the study, this could be due to the lack of diversity in this program.

Objective 2 Methodology and Results

Objective 2 sought to determine, “Are the variables of race, gender, GRE, or undergraduate GPA predictors of state principal certification examination scores?” To determine this, means for each group were computed and compared.

Utilizing these results a regression model was generated with ExCET scores as the dependent variable. The independent variables were gender, race, GPA, and GRE scores. These results indicated that several conclusions can be drawn for the Regular program. Utilizing a procedure called an *F*-test in an ANOVA table (*p*-value) showed at least some of the parameters are non-zero. Therefore, there was a linear relationship between the independent and dependent variables in the model. Further, only the GRE variable was significant at the

0.05 level in the prediction indicating that ExCET scores will increase by 0.03 units for each additional GRE point.

These results indicated the following conclusions for the Scholars of Practice cohort. Because the p -value was $>.05$, there was no linear relationship between ExCET scores with gender, ethnicity, GPA and GRE. None of the variables were individually significant at the 0.05 level according to the regression analysis; consequently a prediction equation would not be useful to predict future ExCET examination scores. Therefore, neither race, gender, undergraduate GPA, nor GRE scores for the Scholars of Practice are predictors of ExCET examination scores.

Analysis indicated a regression model was appropriate and therefore was generated with ExCET scores as the dependent variable. The independent variables were gender, ethnic, GPA and GRE scores. These results indicated there is a linear relationship between ExCET scores with gender, ethnicity, GPA and GRE for Educational Leadership UTA students. Further, if everything else is held constant:

- o A female candidate will score 5.1 units higher on the ExCET than a male candidate.
- o A male candidate will score 5.1 units lower on the ExCET than a female candidate.
- o A student who is not White will score 3.6 units lower on the ExCET compared to a White student.
- o The ExCET score will increase by 0.02 units for each additional GRE point.

Although GPA was not statistically significant in the model, it was very close to being so. The adjusted coefficient of determination (R^2) of 45% shows that the model explains 45% of the variations in ExCET scores. All variables in the Educational Leadership UTA prediction equation were significant at the 0.05 level except GPA. Therefore, undergraduate GPA is the only variable found not to be significant. Race, gender, and GRE scores were shown to be significant in this cohort.

Conclusions for Objective 2. Data from all three programs were analyzed to see which variables, if any, were predictors of ExCET examination success. Prediction equations were shown to be appropriate and thus developed for both the Regular and Educational Leadership UTA programs. GRE was shown to be a predictor of ExCET scores in both these programs. However, there were no variables that were significant predictors of ExCET scores in the Scholars of Practice program. In addition, all other factors except GPA were also predictors of ExCET scores in the Educational Leadership UTA program. These results strongly indicate the highest predictor of overall ExCET success is students' GRE scores.

Objective 3 Methodology and Results

Objective 3 sought to determine, “What, if any, is the value of standards-based testing? Could standards be more appropriately assessed in a different manner that can better quantify a greater link to teaching and learning?” To determine this, an analysis of current literature and practice was interspersed with results from the first two objectives.

The literature currently shows no clear conclusions between the impact of teacher or principal preparation and effectiveness on PreK–12 student learning. Work is being done in various states including Texas to quantify this link, if any, particularly in teacher preparation (Littleton, 2000; Trinidad, Garza, & Gibbs, 2000). Hopefully progress can be made through such techniques as statewide teacher assessment models, tracking of students when they leave preparation programs into the schools, mentoring, portfolios, and other methodologies. However, at this time, there are no indicators that ExCET or TExES scores have any correlation at all on subsequent PreK – 12 teaching and learning.

Conclusions for Objective 3

Since there have been no definitive answers to whether standards-based certification testing produces principals who more positively impact student learning, the value of the tests remains in question. Similar to what is taking place on PreK-12 campuses, the heavy emphasis teachers place on preparing their students for the tests produces stress for all concerned. Anecdotal records indicate stress-related health problems for both faculty and students associated with preparing for and taking the exams.

The same situation now exists as professors and other programs or persons that prepare administrators spend course time preparing students to pass standards-based tests rather than to effectively implement the same standards for improved teaching and learning. Large scale assessment is strongly recommended by State Departments of Education to determine what value standardized certification testing has for serving the purposes for which they were developed. In today’s “test happy” social and political arenas, are standards-based certification exams valid and reliable for achieving their purposes or are we testing under the flag of “accountability” when a better model has yet to be determined?

Conclusions of the Study

This study showed there are significant differences between the Scholars of Practice and both the Regular program and Educational Leadership UTA cohort in mean examination scores. The Scholars of Practice cohort scored higher than both other groups. However, the lack of diversity and longitudinal data were noted as limitations of the study for this program. More importantly, there were no statistical differences between the Educational Leadership UTA and Regular students in their examination mean scores.

The GRE was a predictor of ExCET scores in both the Regular and Educational Leadership UTA programs, whereas there were no variables that were significant predictors of ExCET scores in the Scholars of Practice program. Further, all factors except GPA were predictors of ExCET scores in the more diversified Educational Leadership UTA program.

The *No Child Left Behind Act* is a clear indication that we have not seen the end of standards-based certification exams. This study demonstrates that the most significant predictor of how well students will do on the Texas ExCET exam is Graduate Record Exam scores. As additional states develop and implement other principal certification exams, the question must be asked, “If the most significant predictor of ExCET success in Texas is the GRE, why don’t we just look at a student’s GRE score and skip the test completely?” Yet even with that, no evidence exists to correlate test scores to an educator’s ability to lead, motivate people, or improve student learning performance.

Scientific Importance of the Study

This research is important because it studied students within one university who participated in three different preparation programs, but were measured by the same state certification examination. Results indicated there were no significant relationships between the variables of GRE, GPA, gender, and ethnicity in the Scholars of Practice cohort and that GRE was the only significant predictor in the Regular program. However, in the yearlong field-based Educational Leadership UTA program, all factors except undergraduate GPA were predictors of certification examination results. Further, gender was more significant in this program than the other two. Further research is indicated to determine causes of the discrepancy of predictors between programs, as well as to study the effects of the lengths of time spent and amount of district support provided in the internships if other variables are held constant. A limitation of the study was listed as the small number of minority and male students within the Scholars of Practice cohort that could result in skewed results for that program.

This research is scientifically significant as:

- Any type relationship between standards-based certification exams and educator effectiveness and preparation is sought.
- Strategies to enhance instructional strategies and student performance are developed, implemented, and assessed.
- Universities and other service providers strive to develop well-prepared and diversified school administrators for the increasing national shortage and the multi-faceted needs of a changing society.
- Further in-depth analysis towards specific program effectiveness is investigated, and
- The Scholars of Practice program grows, develops, and becomes more diverse.

Particularly with the implementation of *No Child Left Behind* high-stakes testing is having an increased emphasis on PreK-12 educational curriculum, instruction, assessment, and improvement. It is critical that policy and legislation developers understand the variables and nuances which impact who passes these tests. This study shows that overall GRE scores are a more likely indicator of success on the Texas principal certification exam than race, gender, or GPA. Further, large scale replicative studies are strongly recommended on a state and national arena to see if these results are consistent.

The subsequent question must be raised of why bother to test and stress students on this or other standards-based exams when outcomes are closely aligned with GRE scores that already exist? A case could be made to question if it is prudent to spend millions of dollars for test development and scoring when students who do well on the previously taken GRE will likely do well on the certification exam. The converse is also true. To discern the answers to these questions large scale state and national analysis and evaluation of scores and their predictors must be done. If results of those studies are aligned with this study, serious economic, psychological, ethical, policy, and regulatory decisions must be rethought. The time is now to assess and evaluate various bandwagon reform initiatives including certification testing.

References

- Bowles, B., King, D., and Crow, G. (2000 April). *Viable principal candidates: Superintendents' perspective*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Chambers, S., Munday, R., Sienty, S., & Justice, M. (1999). Predictors of success on the Texas state certification tests for secondary teaching. *College Student Journal*, 33(1), 10 – 16.
- Cooper, C. C. (1986). Strategies to assure certification and retention of black teachers. *Journal of Negro Education*, 55(2), 46 – 55.
- Erlandson, D. A. (2000, August). Texas faces principal shortage. *Texas Elementary Principals and Supervisors Association*, 1, 4.
- Falk, B. (2002) Standards-based reforms: Problems and possibilities. *Phi Delta Kappan*, 83(8), 612 – 622.
- Fenwick, L. T., & Pierce, M. C. (2001). The principal shortage: Crisis or opportunity? *Principal*, 80(4), 24 - 28.
- Fredman, T. (2002, February). *The TWSM: An essential component in the assessment of teacher performance and student learning*. A paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, New York, NY.
- Gillis, M.K. (1989, November). *Impact of admissions and certification exams on the Percentage of minorities in the new teacher pool: A case study*. A paper presented at the Annual Conference of the College

Reading Association, Philadelphia, PA.

- Hill, D. M. (2002, February). *Electronic portfolios: Teacher candidate development and assessment*. A paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, New York, NY.
- House, J. D., Gupta, S., & Xiao, B. (1997, November). *Differences in predicting of grade performance from Graduate Record Examination scores*. Paper presented at the Annual Meeting of the Illinois Association for Institutional Research.
- House, J. D. (1998, May). *Gender differences in prediction of graduate course performance from admissions test scores: An empirical example of statistical methods for investigating prediction bias*. Paper presented at the meeting of the Annual Forum of the Association for Institutional Research, Minneapolis, MN.
- Justice, M., & Hardy, J. (2001). Minority students and the examination for the certification of Educators in Texas. *Education, 121*(3), 592 – 596.
- Laitsch, D. (1998). *State evaluation of teacher preservice programs. Texas and Florida*. American Association of Colleges for Teacher Education, Washington, DC. (ERIC Document Reproduction Service No. ED417192)
- Lindle, J. C., & Rinehart, J. S. (1998, May). *Emerging issues with the predictive applications of the GRE in educational administration programs: One doctoral program's experience*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Littleton, M. (2000, April). *Accountability in teacher education: Systems and trends*. A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Million, J. (1998, April). Where have all the principals gone? *NAESP Communicator*, 21, 5.
- Nagi, J. L. (1975). Predictive validity of the Graduate Record Examination and the Miller Analogies Tests. *Educational and Psychological Measurement, 35*(2), 471-472.
- Nance, J. L., & Kinnison, L. R. (1988). An examination of ACT, PPST and ExCET performance of teacher education candidates. *Teacher Education and Practice, 5*(1), 25 – 30.
- Potter, L. (2001). Solving the principal shortage. *Principal, 80*(4), 34 - 37.
- Richardson, L. (1999, June 23). Principal: a tougher job, fewer takers. *The Los Angeles Times*, p.A1.
- SAS Institute Inc., (1989). *SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2*. Cary, NC: SAS Inc.
- Steinberg, J. (2000, September 3). Shortage of principals growing in U. S. Schools. *Fort Worth Star-Telegram*, p. 11A.

- Study warns of shortage of qualified candidates for principalship. (1998, May). *Copy Editor*, 55, 1.
- Texas Administrative Code Accountability System for Educator Preparation 19§241.01a. (Vernon 1999).
- Texas Education Code. Subtitle D: Chapter 21.048a. Texas Administrative Code. Subchapter M. Certification of Educators in General. ASEP Rule 19§230.413. General Requirements (Vernon 1999).
- Texas Administrative Code. Accountability System for Educator Preparation 19§229.3 a, 1, A (Vernon 1999).
- Texas Administrative Code. Accountability System for Educator Preparation 19§229.3 e, 2, B (Vernon 1999).
- Texas Administrative Code. Accountability System for Educator Preparation 19§229.3 e, 3 (Vernon 1999).
- Texas Administrative Code. Accountability System for Educator Preparation 19§229.3 f, 3 (Vernon 1999).
- Trinidad, S. M., Garza, R., & Gibbs, W. (2000, April). *Pre-kindergarten-16 educational accountability system: The Lone Star state's response and is anyone listening?* A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA).
- Wendel, F. C., et al. (1991). *Do graduate school admissions tests relate to assessment center dimensions?* (ERIC Document Reproduction Service No ED350695).
- Winter, P. A., & Morgenthal, J. R. (2002). Principal recruitment in a reform environment: Effects of school achievement and school level on applicant attraction to the job. *Educational Administration Quarterly*, 38(3), 319 - 340.