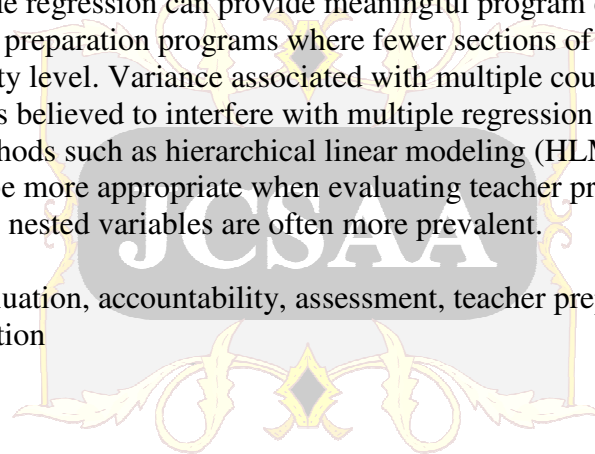# Multiple regression as a practical tool for teacher preparation program evaluation

Cynthia Williams
Texas Christian University

ABSTRACT

In response to No Child Left Behind mandates, budget cuts and various accountability demands aimed at improving programs, colleges and schools of education are in need of practical, quantitative evaluation methods which can be utilized internally to meaningfully examine teacher preparation programs and related coursework. The utility of multiple regression as a tool for linking coursework to teacher certification outcomes was examined in two separate case studies: one examined data from a smaller, private university and the other examined data from a larger public university. Grade inflation, missing or confounding variables, bivariate correlations, beta weights, statistical assumptions, and power were statistical considerations. Results indicated multiple regression can provide meaningful program evaluation information when examining teacher preparation programs where fewer sections of courses are offered, such as at the private university level. Variance associated with multiple course sections being nested in individual courses was believed to interfere with multiple regression results for public university analyses. Methods such as hierarchical linear modeling (HLM) and growth mixture modeling (GMM) may be more appropriate when evaluating teacher preparation programs at larger universities where nested variables are often more prevalent.

Keywords: program evaluation, accountability, assessment, teacher preparation, multiple regression, higher education

INTRODUCTION

No Child Left Behind (NCLB) mandates and increased scrutiny from higher education administrators have triggered one of the longest periods of educational reform in the United States (Paige, 2002; Paige, 2004; Spellings, 2006; Donaldson, 2006; Levine, 2006). Further, many colleges and schools of education are under internal pressure to enlarge enrollments in light of budget cuts, and to simultaneously improve curriculum to meet accountability demands placed on them by their universities and university systems, as well as by professional teacher accreditation bodies such as the National Council for the Accreditation of Teacher Education (NCATE) (Berry, 2006; Weaver, 2004; Trombley, 2003).

As a result, many colleges and schools of education are currently looking for practical ways to evaluate teacher preparation programs and coursework in order to identify areas of strength as well as areas which need strengthening. Thus, there is a growing need for a useful quantitative model which can link teacher preparation coursework to outcomes on teacher certification assessments. This study examines the utility and generalizability of multiple regression as a tool for evaluating university teacher preparation programs at the course and student level in a large public and a small private university in north-central Texas using state certification exam outcomes as the primary measure of student success. Further, this study examines the utility of the model in answering additional program-specific questions that may arise by participating institutions throughout the course of the study, such as: "Are additional, specific courses needed?," "Which courses best prepare teachers?," and/or "Which courses should be restructured?"

THEORETICAL FRAMEWORK

Program Evaluation Framework

Literature reveals an increased preference to use a variety of approaches when evaluating data, using both qualitative and quantitative methods. Approaches address both formative and summative aspects of specific programs (Astin, 1993; Fitzpatrick et al., 2003; Lincoln & Guba, 1989; Lynch et al., 1996). When addressing teacher preparation programs, it is essential to understand the process, such as teaching, and the impact, such as an outcome on a standardized assessment. However, with attempts to standardize and test students at multiple time points, the "big picture" regarding a teacher's impact on what a student actually learns is often blurred. Astin (1993) stated that in order to understand the relationships between processes and outcomes, researchers must also include input variables, which could be attributes pre-service teachers bring with them to a teacher preparation program, such as high school variables.

For this study, the theoretical framework stemmed from the input-environment-output theory regarding education (Astin, 1993). Although there are a number of models that have been formulated and used in the field of education to this end, Figure 1 (Appendix) conceptualizes the model generated by Astin and is often called the input-environment-output (I-E-O) model (1993). This was preferred is because it allows one to segregate, or account for, differences among input variables in order to reveal a more objective estimate regarding environmental impacts on educational outcomes. With this, more meaningful choices and decisions may be made regarding teacher preparation program implementation and evaluation. Astin's model is

both practical as well as simplistic in the sense that all program evaluation choices require comparative judgments. A decision to change something suggests a new element or environment will result in a better outcome. A decision to do nothing implies the status quo is believed to be as good as, or better than, other available alternatives. Either requires conceptualizing and comparing the alternatives. The aforementioned current issues described above illustrate it is time for the current educational climate to change.

In the state of Texas, pre-service educators are required to take and pass a battery of two or more assessments known as the Texas Examination of Educator Standards (TExES) in order to become certified. Further, in order for universities offering teacher preparation programs to remain in favorable standing with the state, at least 70% of their educational students as a whole as well as in subgroups (i.e., ethnic group), taking TExES exams required for initial certification must score a scaled score of 240 out of 300 possible points on each of their exams (Accountability System for Educator Preparation, 2005). The percent of students required to pass these exams will rise in coming years. Universities enrolling highly diverse populations consciously place themselves "at risk." For example, assume a program enrolls 100 students, with 10 of those being international students, and assume 96 out of 100, or 96% overall, pass their TExES exams, resulting in an overall 96% pass rate. However, if the 4 who did not pass were all from the same subgroup of international students that would generate a 60% pass rate for that subgroup, and the school or college of education would be in jeopardy of losing its ability to train pre-service teachers. Universities are aware there is greater reward in admitting students who are most capable of passing teacher certification exams and denying those who are not. This awareness, and possibly even the practice, is in direct contradiction to the ideals which fuel the current educational reform movements. According to NCLB, K-12 schools and districts are charged with believing all children can learn and teaching all children to learn. Thus, the question is: Should universities be charged to practice what they preach? With this premise, can all pre-service teachers who meet admission requirements be successful, and it is the responsibility of the institution to provide evidence to this end? Thus, the need for an enhanced method of evaluating teacher preparation programs is extremely relevant within the context of current educational reform efforts.

PURPOSE OF THE STUDY

Because of shortcomings with many current evaluation processes, the purpose of this study was to test the value of multiple regression as a quantitative method for connecting pre-service teacher characteristics to subsequent TExES outcomes. This model was believed to be capable of providing data-driven conclusions regarding the relationships between individual components of teacher preparation programs and initial certification. Further, multiple regression was selected as the model of choice due to the fact that many education faculty should already be familiar with this method, and by utilizing this approach a college or school of education could avoid the need to consult with and pay for the services of an external evaluation company.

There are key implications for such a model. First, one may be able to predict an individual's success in a teacher preparation program prior to admission. Second, one may be able to determine the effectiveness of each course within the context of an overall university program, something not accomplished in the past. Third, the model may make it possible to predict student outcomes on Early Childhood (EC-4) TExES Pedagogy and Professional Resposibilities (PPR) exams, which can further serve to flag university students, either pre-

admission or during teacher candidacy, who may be in need of additional preparation.

Although results derived from specific implementation of the model in this study cannot be generalized beyond each program at stake, the model itself can be individualized to evaluate any teacher preparation program in order to determine the impact of each individual component of a program, the efficacy of the overall program, and the likelihood of success in the field.

Thus, with the aim of improving program effectiveness, the purpose of this project was to examine the utility of a model, based on the I-E-O framework, which can be used to evaluate traditional early childhood teacher preparation programs at the individual course level and at the student level to predict subsequent success on TExES PPR exams. This study attempted to answer the following research question: Do grades earned in early childhood teacher preparation courses predict success on EC-4 TExES PPR certification exams based on institution type?

There were primarily two limitations associated with this study. First, although this model may be useful in evaluating a variety of traditional teacher preparation programs, the results from this study can only be generalized to the specific early childhood programs from which data was collected. Second, this model only examined relationships between outcomes on standardized tests associated with teacher certification in Texas. This research makes no claim to the construct validity of these standardized. However, these assessments were included because they are the only assessments currently utilized by teacher preparation programs associated within the state of Texas. Other factors, such as personality conflicts with professors, test anxiety, and other sources of measurement error, may have influenced results obtained.

LITERATURE REVIEW

Modeling Teacher Impact on K-12 Outcomes Beyond Certification

The following studies illustrate current research examining the impact individual teachers had on K-12 outcomes during a given academic year, the cumulative effect of different teachers over time on an individual student, and the impact of teachers on students of differing achievement levels. These studies modeled relationships starting from the time of teaching employment and/or initial certification and tested the utility of a variety of different types of analyses. At the time of this study, literature modeling relationships between pre-admission variables, teacher preparation program variables and teacher certification outcomes was extremely limited. However, because several of the issues associated with the following research were related to the current project, their findings were considered directly relevant to mention.

Wright, Horn, and Sanders (1997)

Wright et al. (1997) examined whether or not K-12 test outcomes were a function of selected student and teacher impact covariates. Data was organized as: (a) 30 school districts including 9,900-11,000 third graders, 9,300-10,500 fourth graders, and 6,500-8,900 fifth graders; and, (b) 24 school districts including 13,500-14,100 third graders, 12,300-13,500 fourth graders, and 8,600-10.100 fifth graders. Tennessee's standardized achievement tests were used to model gains in reading, math, language, social studies, and science. Thirty models were fit across content areas in each of the grade levels. Gains were modeled at the student and classroom levels. Results indicated teacher effects were statistically significant in all models and student achievement was statistically significant in 26 of the models. Wright et al. (1997) interpreted findings to indicate teachers are the most important factor impacting K-12 outcomes.

Concerns. McCaffrey et al. (2003) identified the following concerns: (a) there were a limited number of variables in the study, so it could not be said with solid certainty that teacher impact was the sole factor contributing to K-12 outcomes; (b) no discussion of the alignment between standardized achievement tests and curriculum offered at schools from which data was collected was included; (c) no discussion of participating teacher's perception of the importance of the standardized achievement testing was included. Further, this study dealt with intact groups as teachers were assigned to classrooms, which violated an assumption of some of the analyses used, potentially biasing estimates (McCaffrey et al., 2003; Williams, 2005; Henson, 1998).

Rivkin, Hanushek, and Kain (2000)

Rivkin et al. (2000) attempted to address an aforementioned weakness: covariates do not adequately account for residual effects of schools and/or students, resulting in confounded estimates. They estimated true teacher impact on K-12 outcomes as separate from all other sources of variability by utilizing criterion-referenced mathematics achievement data from results on Texas state achievement tests. Data from 500,000 students in 2,156 elementary schools was collected. Three cohorts were followed over a 3-year period as follows: two cohorts were followed in 4th, 5th, and 6th grades, one cohort was followed in 3rd, 4th, and 5th grades.

Gain scores were generated and reported as uncorrelated with student, neighborhood, peer, and school impact. Individual averages, A_i, for differences in gain scores of subsequent grade levels were generated in order to remove the impact of these factors on growth (McCaffrey et al., 2003; Rivkin et al., 2000). A_i was dependent on individual teacher impact for two grade levels as well as residuals from grade-within-school effects (Rivkin et al.). Grade-within-school effects were removed by squaring the differences between grade levels across cohorts to generate D values. D values were then dependent variables, and teacher turnover rate was an independent variable. Statistically significant relationships were reported between D and turnover rates in all models, regardless of whether or not covariates were included. Rivkin et al. reported differences in academic gains across cohorts varied based on teacher turnover rates.

Concerns. First, lower bound estimates of true variance utilized in this study removed variance between schools and districts to potentially bias estimates (McCaffrey et al., 2003). Second, D functions best when the sample size of teachers is large. Because schools can only house a smaller number of teachers within them, the D estimate is positively biased if teacher effectiveness and turnover rates are correlated (McCaffrey et al., 2003). Third, because scores were not tied to a single developmental scale across grades, changes in scores could not be assumed to be directly representative of changes in achievement (McCaffrey et al., 2003).

Rowan, Correnti, and Miller (2002)

Rowan et al. (2002) used data from a national dataset known as: Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity 1991-1994 to test and compare models estimating teacher impact on K-12 outcomes: a 3-level nested ANOVA model, an adjusted covariate model, a gain score model, and a cross-classified model as described by Raudenbush and Bryk (2002). The study also examined the magnitude and stability of teacher impact on K-12 outcomes, explored variables and determined which accounted for classroom-to-classroom differences, and discussed ways in which results from K-12 outcomes could be used

to improve teaching methodology. Rowan et al. (2002) reported accounting for 60-61% of reliable variance in reading and 52-72% of the math variance using the cross-classified model.

Concerns. Rowan et al. (2002) reported covariate models can be misinterpreted as they assess teacher impact on achievement status, not on achievement itself. Thus, when little variance exists among students' growth rates, unreliable estimates result. Because students are generally grouped in classes of similar demographics, the opportunity for diverse growth trajectories is limited (Henson, 1998; Rowan et al., 2002; Williams, 2005).

McCaffrey et al. (2003) raised additional concerns regarding the way in which Rowan et al. (2002) calculated reliable variance estimates for the cross-classified model, stating they may have selected an estimate that was positively biased in order to generate favorable results. McCaffrey et al. (2003) discussed the vagueness in which the study handled missing data, how previous year's achievement was unaccounted for, and the omission of potential variables that could have contributed to variance.

Webster, Mendro, Orsak, and Weerasinghe (1998)

In this study, Webster et al. (1998) implemented: (a) a two-stage, two-level student-school HLM model to estimate school impact on K-12 outcomes and, (b) a two-stage, two-level student-teacher HLM model to estimate teacher impact on K-12 outcomes. Ten years of data from Dallas Independent School District (DISD) were utilized to this end. Of specific relevance to this study, the authors discussed utilizing ordinary least squares regression (OLS) stating OLS models were significantly better than analyses based on unadjusted test scores or student gain scores. In fact, the authors stated estimates generated utilizing unadjusted test scores or student gain scores are neither informative nor fair (Webster et al., 1998).

Concerns. Although the authors reported OLS and HLM models were moderately ($r \geq$ .86) correlated in some analyses, they also reported they were poorly ($r \leq$ .58) correlated in others, stating valuable information can be lost in OLS analyses when student data was aggregated at the school level prior to analysis (Webster et al., 1998). Also, the authors did not describe, at any point, samples in terms of numbers of students, teachers or schools which were included in the analyses used to substantiate claims.

McCaffrey, Lockwood, Koretz, and Hamilton (2003)

In their book, *Evaluating Value-Added Models for Teacher Accountability*, studies listed above were reviewed and a number of relevant topics which may influence value added modeling estimates were discussed: (1) how to select and specify effects within a variety of basic, available value added models; (2) how to deal with missing information in longitudinal data, noting the fact no one can be certain any model includes all variables which can impact K-12 outcomes; (3) how no standardized test measures achievement perfectly, and measurement error can bias estimates; and (4) how errors in estimates can result from sample variance as well as inappropriate model selection or specification. McCaffrey et al. (2003) predominately dealt with quantifying links between generalizable teacher preparation variables, such as type of degree or teaching certification, to certification outcomes and ultimately to K-12 outcomes.

Modeling Teacher Impact on K-12 Outcomes
From High School and Through Teaching Employment

The pioneering and ongoing research of George Noell examines relationships between teacher preparation programs and K-12 outcomes by utilizing the statewide public educational K-12 database, LEADS, as managed by Louisiana's Department of Education's Division of Planning, Analysis, and Information Resources (Noell & Burns, 2006; Noell, 2005; Noell, 2004).

Noell Pilot Work (2004; 2005)

Noell piloted a data system following teachers and students in 10 school districts from 2002-2004, with 8 of those campuses being the same for both academic years. For the 2002-2003 academic year (n = 286,223), Noell compared three analyses examining K-12 outcomes in areas of math and English-language arts, accounting for teacher impact: (1) analysis of covariance (ANCOVA); (2) a weighted ANCOVA; and (3) hierarchical linear modeling (HLM). Models mirrored the layered mixed effects models described by Tekwe, Carter, M. Algina, Lucas, Roth, Arlet, Fisher, and Resnick (2004; Noell, 2006). Student level variables predictors were: free and reduced lunch status; ethnicity; gifted/special education status; Title 1 reading eligibility; English proficiency status; and student scores on previous year's state standardized English language arts, science, social studies and math exams. Campus level variables were campus averages on state standardized achievement tests from the previous year, the percent of females per campus as well as the percent gifted per campus. Teacher level variables were new teacher, emergency certified teacher, regularly certified teacher or other.

Noell concluded that although analyses generally yielded similar results, HLM analyses were regarded as more desirable for use, and suggested the strongest relationships existed between past and current achievement. Thus, students performed similarly year after year. Further, a negative relationship existed between years of achievement data and demographic variables; meaning as number of years of achievement data increased, the relative importance of demographic factors decreased. Also, students in K-12 classrooms with experienced teachers generally performed better on standardized achievement tests, but not always. One analysis revealed new teachers from a particular university were more successful at preparing K-12 students for math achievement tests than their experienced counterparts (Noell, 2004).

Noell (2006)

In subsequent analyses, K-12 performance on state achievement tests was assumed to be the result of the following: prior student achievement, student demographics, classroom context variables, teacher impact, and a school effect. Impacts on K-12 outcomes were examined at the teacher or classroom and school levels (2006). A third layer was included whereby teachers were grouped within schools, estimating the contribution a student's sole teacher had on only the learning assessed during that academic year. Statistically significant main effects were found for all of these and were retained. It is also interesting to note demographic variables collectively accounted for only 4% of variance found in corresponding achievement scores (Noell, 2006).

Noell (2006) then developed a model examining the classroom level on K-12 outcomes, utilizing many of the same variables: percentage of students who were male, minority, who received free and reduced lunch, were in special education, were gifted, or were Limited English Proficiency (ELP). Other variables included were class means on prior achievement values of standardized English Language Arts (ELA), math, science, and social studies tests. Similarly, the

classroom level model revealed performance on previous standardized tests was the best predictor of subsequent achievement outcomes (Noell, 2006).

Regarding overall teacher effects by years of teaching experience, general trends were: teacher impact increased dramatically over the first three years of teaching, and then leveled off (Noell, 2006). This suggests years of experience are not necessarily strongly correlated with K-12 outcomes. Further, Noell (2006) examined teacher preparation programs that had 10 new graduates out in the field, and analyzed traditional as well as university-based programs separately. Noell reported mean adjustments to K-12 outcomes expected based on a standard deviation of 50, as well as 95% confidence interval estimates.

Results indicated that of the 21 teacher preparation programs included (11 traditional programs; 10 alternative programs), none of them generating new teachers were reported to statistically significantly outperform the impact of experienced teachers in areas of ELA, math and science K-12 outcomes (Noell, 2006). Graduates of traditional programs outperformed alternative program graduates across all content areas. However, several preparation programs were reported to generate teachers who were comparable to experienced teachers. Wider confidence intervals were associated with outcomes in the content areas of math and science as compared to those of ELA and social studies. In addition, an empirical Bayes intercept residual, also used by Rowan et al. (2002), was estimated for each teacher, each year, to determine reliability at the individual teacher level. These estimates were considered lower bound due to the ongoing development of the model, with the expectation that subsequent multi-year averages would produce more reliable estimates in years to come (Noell, 2006).

Concerns. Noell (2004) noted several concerns about his research. First and foremost, because the kind of research he was conducting had not been done before, actual data as well as and a standard analytical approach toward examining said data was non-existent prior to his pilot study on the 2002-2003 data. Further, because NCLB mandates did not at that time, or even now, require standardized testing of children in grades PK-3, results could only be evaluated for grades 4-12 (Noell, 2004). Other concerns included the fact that in Louisiana students are given different types of standardized tests in different grade levels, making results from year to year not directly comparable. Although it was possible to standardize these results in a manner which makes them comparable, Noell believed the corrections required to do this would ultimately result in a weaker longitudinal model over time (Noell, 2004). Further limitations included the lack of options in terms of statistical packages available to analyze data and the issue of missing data over time as children and teachers move.

More work is needed to tie teacher preparation programs to K-12 outcomes. For example, Noell indicated although he desired to correlate teacher program admission variables (i.e., ACT scores) to K-12 outcomes, 69% of the teachers in his sample did not report ACT scores. Further, individual program courses were not tied to the model, making it impossible for higher education faculty and administrators to be able to pinpoint potential areas of program weakness which may be in need of restructuring. Without this kind of information, little improvement in the ways teacher preparation programs admit and educate pre-service teachers can be made.

METHODS

Power, Sample Size, Effect Size, and Statistical Significance

University administrators and early childhood program coordinators were consulted regarding the interpretation of power, sample size, effect size, and statistical significance. They requested the recommendations of Cohen (1988) be considered when drawing conclusions regarding statistical significance, effect size, power, and sample size. Because sample size could not be set apriori as entire populations are included in this study, power could not be selected apriori. Instead, results were interpreted in terms of statistical significance tests and effect sizes measures, with consideration of sample/population size and power included in concluding discussions. Statistical significance at the $p < .05$ and $p .< .01$ levels was sought. Effect size measures were interpreted as small, medium and large in the following manner: an R^2 of .01, .09, and .25 respectively (Cohen, 1988).

Replicability and Reliability

The Educational Testing Service (ETS), which writes the EC-4 TExES PPR, was contacted regarding reliability estimates. Reliability estimates at the university level did not exist for TExES examinations. However, because these measures are the primary standardized means of assessing teacher qualifications in the state of Texas, they were included in this study.

The extent to which study results would be replicable was also considered. The stability of programs over time was of primary interest to university administrators regarding the following issues: faculty turnover and ratio of full to part-time faculty, program requirement consistency across years, and stability of student type and admission requirements over time.

Normality

Non-normality was prevalent in many course distributions associated with both Private University and Public University in the form of grade inflation. As noted in the literature review, non-normality can bias estimates, especially when there is little variance associated with a distribution (McCaffrey et al., 2003; Young, 1990). Reasons for grade inflation as they individually related to each institution's teacher preparation programs were discussed with university administration. Normality in terms of skewness and kurtosis were considered prior to conducting each analysis associated in this study.

Omitted, Confounding and Missing Variables

Models which use observational data can skew estimates in two primary ways: (a) it is impossible to know that any given model includes all possible influences on dependent variables and that there are no confounding variables involved; and (b) incomplete, omitted or missing data can make it impossible to differentiate between effects (McCaffrey et al., 2003). Models which do not account for differences in student populations, such as differences in SES between two schools, can yield biased estimates, even when using complex multivariate models.

Further, because true teacher preparation program impact may be correlated with the types of students who are being taught, current models cannot separate existing contextual

effects from these effects (McCaffrey et al., 2003). Also, effects that result from school environments, school districts, and prior teachers are difficult to separate as well. If these effects are omitted from models they are, intentionally or unintentionally, subsumed by teacher preparation program effects. Thus, this may bias what researchers ascertain true effects to be (McCaffrey et al., 2003). University administrators were consulted regarding the interpretation of results and potential impact of omitted and confounding variables.

Bivariate Correlations vs. Structure Coefficients

Because this study was conducted on existing data intended to be utilized by universities, actual bivariate correlations between predictor and outcome variables were interpreted instead of structure coefficients. There were two reasons for this decision. First, Thompson's stated researchers need to interpret either beta weights with structure coefficients or beta weights with bivariate correlations with actual outcomes in order to address suppressor variables and multicollinearity issues (Thompson, 1992). Second, after talking with university administrators, it was decided the calculation and interpretation of bivariate correlations would potentially be easier to conceptualize as well as explain to faculty.

Suppressor Variables

One of the purposes of this study was to provide teacher preparation programs with a model which could be useful in revealing key courses which have an impact on program completion variables. As such, bivariate correlations with outcomes were considered apriori to many of the multiple regression analyses conducted. Although this approach may have made it more difficult to identify suppressor variables, the focus of this study was directed at narrowing the list of courses in order to predict those 2-3 courses most predictive of program completion outcomes. With this kind of information, program administrators could potentially track student progress in key courses more closely than other courses. Still, in order to be sensitive to the potential existence of suppressor variables, comparisons were consistently made between bivariate correlations with outcomes as well as beta weights associated with regression analyses during interpretation as Thompson (1992) suggested.

Early Childhood TExES PPR Outcomes

Inferences made as based on standardized student achievement measures are limited. First, no one assessment can give a true picture of all knowledge possessed or achievement accomplished, and the anxiety generated in students by the very high stakes tests introduces different amounts of measurement error for each test taker. At the time of this study in Texas, early childhood educators were required to take two exams in order to become certified: EC-4 TExES PPR and the EC-4 Generalist. The EC-4 TExES PPR examination was chosen for two reasons: first, scores from this exam were readily available from both Public University and Private University; and, second, the TExES PPR purports to focus more on pedagogy, or teaching ability, than the TExES Generalist exam, which focuses primarily on content.

INTRODUCTION OF THE MODEL

Participants

This study examined entire EC-4 populations of program completers who graduated with bachelor degrees and obtained initial certification during the 2005-2006 academic year, to include summer matriculation, at two institutions in north central Texas referred to as Private University and Public University. Only traditional routes toward initial certification at these institutions were examined. At Private University the sample size consisted of 46 EC-4 undergraduate students and at Public University the sample size consisted of 244 EC-4 undergraduate students. Total enrollment at Private University during the 2005-2006 academic year was roughly 8,500, with about 125 students graduating from all educational programs of study which led to bachelors degrees and initial teacher certification. Total enrollment at Public University during the 2005-2006 academic year was roughly 26,000, with about 450 students graduating from undergraduate educational programs leading to initial teacher certification.

Instrumentation, Variables, and Data Collection

Data was collected by personnel in each university, stripped by each institution of identifying information by existing university personnel, and given to the researcher for examination. Variables examined included: grades earned in key undergraduate educational courses and EC-4 TExES PPR outcomes.

Overview of the Model

The primary purpose of this study was to examine the relationship and impact an individual EC-4 teacher preparation program has with and on the teacher preparation program outcomes of EC-4 TExES PPR certifications examinations at the course and semester level. This model used a series multiple regression analyses to evaluate two teacher preparation programs housed within two separate colleges of education. Analyses were carried out independently for each institution, and then compared to look for similarities, differences, strengths, and weaknesses.

Multiple regression can be used to examine the relationship between several independent variables (IVs) and a single continuous dependent variable (DV) (Pedzahur, 1997). The general formula for least squares or ordinary least squares regression with one predictor is as follows, where a is a constant, b represents slope (regression coefficient or b coefficient), and X reflects a value for the IV, or a grade in a teacher preparation program course (Pedzahur, 1997):
$$Y^\wedge = a + bX$$
When more than one IV exists, as in the case for this study, the formula is (Pedzahur, 1997):
$$Y^\wedge = a + b_1X_1 + b_2X_2 + ... + b_pX_p.$$
$Y^\wedge$ represents the predicted DV value, or the continuous TExES score predicted for each pre-service teacher at the time of graduation. Analyses were run by teacher preparation program at each university examining EC-4 TExES PPR outcomes. Statistics included for interpretation purposes as associated with multiple regression analyses initially included bivariate correlations between predictor variables and outcome variables, the statistical significance F test, effect size

measures R^2 and adjusted R^2 (Henson & Smith, 2000; Vasquez, Gangstead & Henson, 2000). Examinations of beta weights, sample size and power were (Cohen, 1988; Thompson, 1992).

THREATS TO VALIDITY

Internal Validity

One primary possible threat to internal validity as identified by Campbell and Stanley (1963; 1969) was believed to potentially be associated with this study: selection. To avoid this potential threat to internal validity, this study included entire populations of program completers from Private University and Public University for the 2005-2006 academic years and made generalizations to those populations.

External validity

External validity refers to the extent to which a study can be generalized beyond a sample (Gall, Borg, & Gall, 1996). Specifically, population validity refers to the extent to which a study can be generalized beyond the sample from which data is taken. As stated previously, this model makes no claim to be a "one size fits all" and specifically states results can only be specifically generalized to the universities from which data is collected.

RESULTS

Private University

Descriptives were generated on all teacher preparation course level data to determine the breakdown of grades by course. Private University's EC-4 program inclusively provides theory, content and pedagogy courses. Table 1 (Appendix) shows coursework and grades earned at Private University in teacher preparation courses for 2005-2006 program completers ($N = 46$).

EC-4 coursework is blocked at Private University, and students take specific courses each semester. Blocks are known as: Junior1 (first semester of junior year), Junior2 (second semester of junior year), Senior1 (first semester of senior year) and Senior2 (second semester of senior year). EC-4 students specialize in either English as a Second Language (ESL) or Special Education (EDSP). Blocks were analyzed collectively for two reasons: (a) professors teaching in blocks often communicated with one another and coordinated or overlapped syllabi from different courses, and (b) blocked courses were considered complete sets of courses by university administration. A series of multiple regression analyses were conducted: first, by blocked semester to narrow down preliminarily predictive coursework variance by program; second, preliminarily predictive coursework variables were used to determine which courses, overall, were predictive of EC-4 TExES PPR outcomes. Correlations between courses recording letter grades (i.e., not pass/fail) and with $n > 10$ reported are found in Table 2 (Appendix). Courses reporting less than ten grades were excluded and normality was considered. Although OLS regression is quite robust to violations of the assumption of normality due to grade inflation (Field, 2000), variables reporting skewness greater than +3.0 or smaller than -3.0 were also excluded from further examination due the small population sizes associated with Private University.

Once courses with large amounts of grade inflation and small course numbers were removed from examination of the available Private University courses, multiple regression analyses were run by examining blocked courses. The sophomore, EC-4 Junior 1, EC-4 Junior 2 ESL, EC-4 Junior 2 EDSP, and EC-4 Senior 2 ESL semester blocks all reported statistically significant findings and large effect size measures, with individual blocks reportedly accounting for 0-62% of the variance associated with EC-4 TExES PPR outcomes. Table 3 (Appendix) reports multiple regression outcomes resulting from the examination these semester blocks. Following this, course variables weighing heaviest and possessing standardized $\beta > .3$ in statistically significant regression analyses as well as carrying statistically significant bivariate correlations with EC-4 TExES PPR outcomes ($p < .05$) were included in a final analysis for the EC-4 program. There were no outstanding discrepancies between bivariate correlations and standardized beta weights suggesting suppressor variables. Also, although the courses of EDEC 42223 and EDEC 30233 possessed $\beta > .3$ in statistically significant block analyses and were statistically significantly correlated with EC-4 TExES PPR outcomes, they were excluded from this final analysis because they were ESL specialization courses and not reflective of the greater EC-4 program as a whole.

A correlation matrix was generated to elaborate on multicollinearity between variables. A great deal of correlation was found to exist between several of the predictor variables. Because of both multicollinearity and large confidence intervals, it was decided that, for practicality purposes, Private University's courses known as EDEC 30103, EDEC 30014, EDEC 30213, and EDUC 30123 were similarly statistically significantly predictive of EC-4 TExES PPR outcomes. EDUC 20003 was the only course which did not possess a great deal of correlation with other EC-4 courses at Private University. See Table 4 in the Appendix.

Thus, the final analysis utilized EDEC 20003: "Critical Investigations: Teaching," EDEC 30103, "Introduction to Early Childhood Education," EDUC 30123, "Educational Psychology," EDEC 30014, "Science and Mathematical Thinking Through Play and Creativity: Science," and EDEC 30213, "Language and Literacy: Early Literacy" as predictors of TExES EC-4 PPR outcomes. Results indicated statistical significance, $F (5, 25) = 7.360$ $p < .01$, with predictors accounting for roughly 51% of the variance associated with EC-4 TExES PPR outcomes. EDUC 30103, "Introduction to Early Childhood Education," and EDUC 30123, "Educational Psychology," were the most influential courses. Confidence intervals were wide, indicating the weights could vary greatly.

Public University

Following Private University analyses, descriptives were generated and examined using all teacher preparation course level data to determine the breakdown of grades by course for Public University. The EC-4 program at Public University also provides theory, content and pedagogy courses to its students. Pedagogy course data were not comprehensive for Public University's EC-4 program. Table 5 (Appendix) shows data only from reported pedagogy courses and grades earned at Public University in teacher preparation courses for 2005-2006 program completers ($N = 271$) by program area. Similar to Private University, the highest single grade earned by students in these teacher preparation courses for each student at Public University is shown.

Coursework for the EC-4 program is only blocked during the senior year when students take Professional Development School (PDS) courses. The first semester of the senior year is

known as PDS 1, and the second semester is PDS 2. This arrangement affords students flexibility to determine when they enroll in courses prior to the PDS year. However, once in the PDS year, students typically enroll in PDS1 and PDS2 courses consecutively. Thus, selected courses were examined as sets as: (1) undergraduate educational courses taken prior to admission into a teacher preparation program, (2) undergraduate educational courses taken post-admission, but prior to senior year, (2) PDS1 educational courses, and (3) PDS2 educational courses.

Correlations between individual courses recording letter grades (i.e., not pass/fail) and courses with 10 or more grades were reported, as well as correlations between courses and EC-4 TExES PPR scores, and can be found in Table 6 (Appendix). Note that although several undergraduate educational courses are statistically significantly correlated with one another, none of the undergraduate educational courses associated with Public University's EC-4 program were statistically significantly correlated with EC-4 TExES PPR outcomes except for the negative correlation associated with TExES EC-4 PPR outcomes and the undergraduate course of DFST 4233: Guidance of Children and Youth ($r = -.133$, $p < .05$, $N = 230$).

Again, normality in the form of skewness and kurtosis was considered apriori to conducting analyses. Skewness ranged between -5.361 and -.910 and kurtosis ranged between -.210 and 40.874 as associated with selected Public University's EC-4 undergraduate courses ($n = 213$-240). However, because multiple regression is quite robust to the violation of the normality assumption and because the EC-4 program population at Public University is six times larger than the EC-4 population a Private University, all undergraduate educational course variables were included in regression analyses regardless of normality statistics. Multiple regression analyses were conducted by program status under the rationale that if it were possible to predict EC-4 TExES PPR success or failure throughout the program, one could better target students in need of extra EC-4 TExES PPR preparation. However, none of the program status course sets were statistically significantly predictive of EC-4 TExES PPR outcomes: courses taken prior to program admission: $F (3, 219) = 2.111$, $p > .05$: courses taken post-admission but prior to PDS1: $F (3, 202) = .111$, $p > .05$, and PDS 1 courses, $F (3, 231) = 1.974$, $p > .05$. PDS 2 courses were not examined as they were pass/fail, and there was no variability among these outcomes as all students must pass in order to become a program completer in the first place. Given weak correlations between individual courses and TExES outcomes, results were not surprising.

Following this, an aggregated regression was run using all nine EC-4 courses as predictors of TExES EC-4 PPR outcomes regardless of program status. Because there were more than 200 EC-4 program completers at Public University, the large number of variables did not violate a statistical assumption (Field, 2000). Even still, statistical significance was not obtained despite the larger population. In addition, the correlation between cumulative GPA and TExES EC-4 PPR outcomes was calculated and found statistically insignificant ($N = 242$, $r = .077$, $p > .05$). A detailed discussion of these results as they relate to power calculations can be found in the concluding section. Additional results can be found in Table 7 (Appendix).

CONCLUSIONS

Replicability and Reliability

Because reliability estimates were not available at the university level for EC-4 TExES PPR outcomes, the stability of programs over time was of interest and university administrators were questioned regarding the following: faculty turnover and ratio of full- to part-time faculty,

program requirement consistency across years, and student type. For Private University, administration reported faculty turnover was low, courses/sections were typically taught by the same full-time professors from year to year with few adjuncts, student type was considered consistent, and programs of study remained relatively stable in terms of requirements across academic years. For Public University, university administration reported faculty turnover was considered moderate, courses were taught by both full-time and several adjunct professors each year, student type was considered consistent across academic years, and programs of study were reported to remain relatively stable in terms of requirements across the years. In addition, courses at Public University often had multiple sections associated with them, sometimes as many as 9 or more. With such a program delivery setup, it is impossible for any full-time faculty carrying a typical 3-3 or 3-2 teaching load to be responsible for teaching all sections of any one course during a semester. Utilizing multiple instructors to deliver multiple sections of the same course introduces a large amount of error associated with resulting course grades.

Grade Distributions

Grade inflation existed at both universities. Most students made *A*s in coursework. Although some of this may have been attributed to instructors being too lenient, the existence of such a phenomenon ultimately could not be known. What was known was that students must all meet the same requirements in order to be admitted into a teacher preparation program, which generated more homogeneous populations of students. As such, one would expect them to earn similar grades in educational courses. Universities reported frequencies for grades for all program completers, regardless of teacher preparation program type (EC-4, 4-8, 8-12, and EC-12), for the 2005-2006 academic year. Students at Private University earned 1049 *A*s (80%), 244 *B*s (19%), 18 *C*s (1%) and 1 *D* (<1%). Students at Public University earned 2361 *A*s (79%), 532 *B*s (18%), 105 *C*s (3%), and 7 *D*s (<1%).

Discussion with higher education administrators suggested grade inflation appeared to exist for a number of reasons. First, students in EC-4 programs were a more homogeneous group due to the fact students all met the same admission requirements. Second, students in the EC-4 programs all possessed the same career passions. Third, the content commonly taught in educational courses was considered subjective, often making it difficult for faculty to clearly distinguish between a completely "right" answer and a "wrong" answer. Fourth, several courses within educational programs reported the dichotomous outcomes of pass/fail as course grades. And, fifth, coursework associated with EC-4 programs often included cooperative learning activities and projects for which all students participating in a group received the same grade.

With regards to this study, grade inflation was found to be a slightly more concerning at Private University, especially in light of smaller sample sizes. This was believed to be the result of two factors: (1) the admission rate for Private University is extremely low, which further served to make program populations homogenous groups (i.e., in 2006, general college admission was only 1 in 8 applicants, and teacher preparation program admission was only 1 in 2 applicants beyond this); and (2) program populations were small, ranging from 20-46 program completers. As such, the variables of skewness and kurtosis were carefully considered apriori to running analyses for both universities, but particularly at Private University.

University Comparisons

The purpose of this study was to use multiple regression to examine the predictive utility of pre-admission and program variables regarding subsequent success on the EC-4 TExES PPR. Because of smaller program sizes Private University, the EC-4 program was examined by blocked semester. At Public University, the EC-4 program was examined in course sets (pre-PDS, PDS1, and PDS2).

A primary issue with this study was the wide variance associated with correlations between course grades and TExES PPR outcomes for both universities. At Private University, 13 out of 22 courses associated with the EC-4 program were statistically significantly correlated with EC-4 TExES PPR outcomes (n = 18-41, p <. 05). Five of these courses served to collectively account for more than 50% of the variance associated with EC-4 TExES PPR outcomes (see Table 4). Further, after additional analyses, two Private University courses were predictive of TExES PPR outcomes regardless of program area and accounted for roughly 20% of the variance associated with TExES PPR outcomes: EDUC 30143, "Child and Adolescent Development" and EDUC 30123, "Educational Psychology" (F = 4.746, p < .05).

In contrast, coursework at Public University did not prove to be a useful tool in predicting EC-4 TExES PPR outcomes. In fact, only one course was significantly correlated with EC-4 TExES PPR outcomes: DFST 4233, "Guidance of Children and Youth" possessed a surprising negative correlation with EC-4 TExES PPR outcomes (r = -.133, p < .05). In examining other programs, only one other course had a significant correlation with TExES PPR outcomes, but that correlation was based on only 7 valid cases and was viewed with great caution: EDSE 3830, "Teaching/Learning Process and Evaluation" as associated with the EC-12 program only (r = .884, p < .01). Neither of these courses generated noteworthy effect size measures.

To examine multiple regression analyses where statistical significance was not found at Public University, post-hoc power analyses were conducted. Power ($1 - \beta$) is said to be an analysis's ability to detect an effect, assuming an effect truly exists, and is dependent on effect size, sample size, and the significance criterion selected for an analysis (Bell, 2000). Power is also referred to as the probability a statistical test will generate statistically significant results (Cohen, 1998). In social science research, power of .80 is commonly, but not always, sought when developing the research design apriori. Because this study included entire populations of program completers, sample size could not controlled. Specifically, the EC-4 analyses which utilized 242 subjects and 9 predictor variables were examined. Calculations indicated that if an R^2 effect of .35 or .15 had existed, the power associated with detecting this would have been >.99. Even if an R^2 effect of .02 had existed, the power to detect this would have been .29. These results illustrate that, for Public University, regression would have been very likely to reveal either moderate to large effect sizes had they existed.

However, a primary factor contributing to the discrepancy between power and regression results was believed to be related to course delivery. At Private University courses are taught primarily by full-time faculty, and faculty turnover is low. In addition, generally there are only 1-3 sections of any one course being taught during a semester, and one faculty member is generally responsible for teaching all sections of that course. As such, sections at Private University were considered to be relatively consistent in terms of professors, course content, and delivery. In contrast, the teacher preparation programs at Public University are much larger in terms of student population and section offerings, faculty turnover was considered moderate by university administration, and the use of adjuncts in teacher preparation programs was frequent.

Although information was unavailable regarding which professors taught the various sections of courses associated with the 2005-2006 program completer population, course offerings for the 2007 academic year confirmed there were many sections of courses taught by various professors each semester. For example, in the Fall 2007 semester, there were 5 sections capped at 40 students each and 5 different instructors teaching EDEE 4340, "Social Studies in Grades EC-8" and 9 sections capped at 30 students each with different instructors teaching EDEE 4330, "Science in Grades EC-8." Even if syllabi were strictly standardized, error would still be associated with course grades as attributed to the differing teaching philosophies, backgrounds, and instructor specializations. As such, this error was believed to have contributed greatly to the lack of significant correlations between course grades and TExES PPR outcomes.

A second possibility for these findings may stem from course classification. Upon further examination of Public University coursework, some of the course descriptions and course titles reflected more content emphasis than pedagogy emphasis. For example, the courses EDEE 4340, Social Studies in Grades EC-8, and EDEE 4350, Mathematics in Grades EC-8, were labeled as pedagogy courses, not content courses. These results bring to question whether or not course content associated with these analyses were more predictive of EC-4 TExES Generalist scores than EC-4 TExES PPR scores. Course syllabi for all required courses and EC-4 TExES Generalist scores were not available at the time of this study to examine and help answer this question.   However, Public University's Student Advising Office confirmed that several of these courses are actually methods courses, but because they are taught in conjunction with an internship and student teaching experience they are viewed and labeled as pedagogy courses.

However, in looking at courses at both universities which were significantly correlated with EC-4 TExES PPR outcomes ($n > 10$), courses courses did have commonalities: they were (1) developmental theory driven, and (2) taken prior to admission into the College of Education (Private University) or PDS blocks (Public University). As such, college administration may want to consider examining grades earned in theory based courses more carefully prior to student admission to teacher preparation programs or field-based component portions of programs.

PROGRAM IMPLICATIONS AND RECOMMENDATIONS

Implications for Private University EC-4 Program

The model proved to be valuable in revealing pre-admission variables and course variables which faculty may want to consider more closely as they monitor student progress for Private University. Further, the structure and delivery of courses which did not prove predictive of program completion outcomes may warrant review and/or revision of course syllabi. However, because it was not possible to link program completer data to subsequent K-12 outcomes, caution is warranted when interpreting weak correlations between program variables and EC-4 TExES PPR outcomes. The EC-4 TExES PPR examination is just one instrument which only measures specific domains and competencies as they relate to pedagogy and professional responsibilities. The possibility exists that courses which correlated weakly with the EC-4 TExES PPR actually correlate more strongly with other desirable K-12 outcomes.

For example, all sections of EDEC 30033, "Play and Creativity" have been taught by the same tenured professor for several years. The focus of this course is to encourage students to think about issues creatively, critically, reflectively, and thoughtfully. A variety of hands on activities and assignments are incorporated into this course where students are actively engaged

in complex problem solving and higher order thinking tasks. Further, grades are determined by student option: students desiring an *A* must pass/fail a seven assignments, students desiring a *B* must pass/fail six assignments, and students desiring a *C* self select and pass/fail even fewer assignments. Thus, the correlation between EDEC 30033 and EC-4 TExES PPR outcomes was not statistically significant ($r = .185$, $p > .05$) because grade outcomes were limited by course structure: grades were largely determined by student initiative and assignment outcomes were dichotomous. However, because no data existed on K-12 outcomes, it is not possible to know whether or not skills learned in this course informed the teaching strategies of program completers in their own classrooms. As such, weak correlations between courses and EC-4 TExES PPR outcomes are not necessarily an indication of course ineffectiveness.

Implications for Public University EC-4 Program

The use of multiple regression proved to be problematic for Public University because courses were delivered in multiple sections and a large number of adjuncts as well as full-time faculty were teaching course sections. The use of more complex statistical approaches is needed to explore future research involving Public University data. The very nature of HLM techniques may have been more appropriate in modeling section level effects of a course. However, HLM techniques were not appropriate for this particular study for the following reasons: (a) program completer data existed only for those graduating during the 2005-2006 academic year; (b) course sections were capped at 30-40 students; (c) participants could take several courses out of sequence at Public University, which likely resulted in smaller course section sizes available for analysis as associated with this single program completion academic year.

FURTHER RECOMMENDATIONS

Because not all of variables contributing to a student's performance were ultimately known, a suggestion for future research is to dialogue with university administration, faculty, and students further to explore additional variables which may interact with teacher preparation impact on an individual program completer. For example, Private University is a highly selective institution. As noted previously, Private University program completers performed better on EC-4 TExES PPR exams than Public University program completers. Further, students are more likely to receive individualized assistance from full-time faculty and administration at Private University than those at Public University. Does this increased involvement with students contribute to better success on TExES PPR outcomes? Does the selectivity associated with Private University admissions contribute to better success on TExES PPR outcomes?

Another issue in need of attention is that of grade inflation. The less variation there is among grades, the less predictability a course possesses. A suggestion for beginning this program improvement dialogue regarding steps to move away from grade inflation would be to connect a variety of College of Education administrators and faculty with administrators and faculty from other colleges where grade inflation is not particularly evident. Young (1990) reported courses which focus on natural science and engineering generally possess less grade inflation than courses which focus on the humanities. Connecting administration and faculty between these departments at each institution may serve as a useful mechanism to addressing the issue of grade inflation as related to teacher preparation programs. In addition, it may be appropriate to consider the utility of employing +/- grades as a means to provide programs with

additional variance regarding about student progress and course utility.

Although results from this study cannot be generalized beyond the programs from which they were obtained, it is possible to make the following generalizations regardless of program. First, the nature of course delivery impacts the ability an individual course can have in predicting EC-4 TExES PPR outcomes. Multiple regression does not appear to be appropriate when using course grades to predict EC-4 TExES PPR outcomes at institutions where multiple sections of courses are offered and delivered by multiple instructors, especially in cases where adjuncts are utilized and who are generally unaware of overall program goals. Further, this model may not work when teacher preparation programs are not tracked. This does not necessarily indicate students are not learning when course delivery involves multiple sections or where programs are not tracked. However, the error introduced from multiple instructors teaching multiple courses or lack of tracking makes it impossible to partition variance appropriately.

A second generalization is as follows: even in cases were few course sections were offered and delivered by the same instructor, it is worth noting that just because a course was not statistically significantly correlated with EC-4 TExES PPR outcomes does not necessarily mean the course is not valuable. It only means, for certain, that the course content or structure was not correlated to content covered on the EC-4 TExES PPR examination.

These generalizations lead into a main recommendation for future research. In order to determine the true impact any teacher preparation course has on an individual, a method for linking teacher preparation program variables and the individuals who enroll in them to the K-12 standardized achievement exams ultimately associated with the students taught by these same teacher preparation program completers is needed.   As mentioned in the literature review above, HLM may show promise to this end. In addition, growth curve and developmental trend analyses regarding human behaviors can be utilized to study intra-individual differences in growth trajectories, as growth mixture models can examine student variables nested in classrooms, schools, and districts (Meredith & Tisak, 1990; Rogosa, Brandt, & Zimowski, 1982; Muthén & Curran, 1997; Bryk & Raudenbush, 1992; Willett & Sayer, 1996). Further, repeated measures ANOVA where data collected at specific points in time can be analyzed via univariate or multivariate techniques may prove useful as well (Williams, Nimon & Allen, 2007).

Thus, in response to No Child Left Behind (NCLB) mandates and increased scrutiny from higher education administrators, there is real need for a practical model to evaluate teacher preparation programs at the course level. This study illustrated that multiple regression can be a useful tool to examining smaller teacher preparation programs where students are tracked across semesters, and course delivery occurs with tight section level alignment. Such a model can help university administrators and faculty to evaluate course alignment with a variety of outcomes as needed, such as teacher certification examination outcomes, cumulative GPA, and ultimately someday in the future, subsequently with K-12 outcomes as well.

REFERENCES

Accountability system for educator preparation. (2005, September 7). Retrieved December 27, 2006 from
http://www.sbec.state.tx.us/SBECOnline/edprep/asep.asp?width=1024&height=768.

Astin, A. W. (1993). *What matters in college? Four critical years revisited.* San Francisco, CA: Jossey-Bass.

Bell, G. V. (2002). *Statistical rules of thumb.* New York: John Wiley & Sons.

Berry, B. (2006). New teacher quality study doesn't tell whole story. National Council for the Accreditation of Teacher Education. Retrieved December 9, 2006 from
http://www.ncate.org/.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1998). *An introduction to HLM: Computer program user's guide* (2nd ed.). Chicago, IL: University of Chicago.

Campbell, D.T. (1969). Reforms as experiments. *American Psychologist, 24*, 409-429.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd Ed.). New York: Lawrence Erlbaum.

Donaldson, G. A. (2nd ed.). (2006). *Cultivating leadership in schools. Connecting people, purpose, and practice.* Williston, VT: Teachers College Press.

Field, A. (2000). *Discovering statistics using SPSS for Windows.* Thousand Oaks, CA: Sage Publications.

Fitzpatrick, J. L., Sanders, J.R., & Worthen, B.R. (3rd ed.) (2003). *Program evaluation: alternative approaches and practical guidelines.* Boston: Allyn & Bacon.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction.* White Plains, NY: Longman.

Garland, S. (2006, November 20). Uncertified teacher performing well, study finds. *New York Sun*. Retrieved December 9, 2006 from http://www.nysun.com/article/43827.

Henson, R. K. (1998, November). *ANCOVA with intact groups: Don't do it!* Paper presented at the annual meeting of Mid-South Educational Research Association, New Orleans, LA. (ERIC Document Reproduction No. ED 426 086)

Hussar, W. J., & Bailey, T. M. (2006). *Projections of education statistics to 2015*. (NCES Publication No. 2006-084). Washington, DC: U.S. Department of Education.

Kedem, B., & Fokianos, K. (2002). *Regression models for time series analysis.* Hoboken, NJ: Wiley-Interscience.

Levine, A. (2006). *Educating school teachers*. (The Education Schools Project). Princeton, JN: Woodrow Wilson National Fellowship Foundation.

Lincoln, Y. S., & Guba, E. G. (1989). Ethics: The failure of positivist science. *Review of Higher Education, 12*(3), 221-240.

Louisiana Department of Education (2006, October). *Types of teaching authorizations and certifications.* Retrieved January 17, 2007 from
http://www.doe.state.la.us/lde/uploads/7202.pdf.

Lynch et al. (1996). Quadrant analysis as a strategic planning technique in curriculum development and program marketing. *Journal of Marketing for Higher Education, 7*(2), 17-32.

McCaffrey, D. F., Lockwood, J.R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. Psychometrika, 55, 107–122.

National Council for the Accreditation of Teacher Education. (2006, November 14). *About us.* Retrieved December 27, 2006 from http://www.ncate.org/public/aboutNCATE.asp.

Noell, G. H. (2004). *Technical report of: Assessing teacher preparation program effectiveness: A pilot examination of value added approaches.* Retrieved January 17, 2007, from http://asa.regents.state.la.us/TE/technical_report.pdf.

Noell, G. H. (2005). *Technical report of: Assessing teacher preparation program effectiveness: A pilot examination of value added approaches.* Retrieved September 9, 2007, from http://asa.regents.state.la.us/TE/technical_report_200405.pdf.

Noell, G. H. & Burns, J. L. (in press). Value added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education.*

Noell, G. H. (2006). *Annual report of: Value added assessment of teacher preparation.* (Louisiana Board of Regents Publication No. CT-05/-06VAA-02). Baton Rouge, LA: Louisiana State University.

Paige, R. (2002). *Meeting the highly qualified teacher challenge: The secretary's annual report on teacher quality.* (U.S. Department of Education Publication No. ED-00-CO-0016). Washington, DC: U.S. Department of Education.

Paige, R. (2004). *No Child Left Behind: A toolkit for teachers.* (U.S. Department of Education Publication No. ED000632C). Washington, DC: U.S. Department of Education.

Pedzahur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction.* (3rd ed.). Toronto, Ontario: Thomson.

Peng, C. Y. J., & So, T. S. H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics 1*(1), 31-70.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2000). *Teachers, schools, and academic achievement*. (National Bureau of Economic Research Working Paper No. W6691). Cambridge, MA: National Bureau of Economic Research.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rowan, B. Correnti, R. & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the *Prospects* study of elementary schools. *Teachers College Board, 104*, 1525-1567.

Spellings, M. (2006). *The secretary's fifth annual report on teacher quality: A highly qualified teacher in every classroom*. (U.S. Department of Education Contract No. ED 04 CO-0059/0016). Washington, DC: U.S. Department of Education.

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

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., Arlet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11-36.

Thompson, B. (1992, April). *Interpreting regression results: beta weights and structure coefficients are both important.* Paper presented at the annual meeting of the American Educational Research Assocation, San Francisco. (ERIC Document reproduction Service No. ED 344 897).

Trombley, W. (2003). College affordability in jeopardy: The rising price of higher education. The National Center for Public Policy and Education.

Walsh, K. (2001). *Teacher certification reconsidered: Stumbling for quality.* (The Abell Foundation Publication). Baltimore, MD: The Abell Foundation.

Weaver, R. (2004, February 24). NEA President Reg Weaver responds to remarks by Secretary Rod Paige. (Statement, National Education Agency). Retrieved from http://www.nea.org/newsreleases/2004/nr040224.html.

Webster, W. J., Mendro, R. L., Orsak, T. H., Weerasinghe, D. (1998, April). An application of hierarchical linear modeling to the estimation of school and teacher effect. Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.

Willet, J. B., & Sayer, A. G. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation modeling: Issues and techniques (pp. 125–157). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Williams. C. S. (2005, February). Possible alternatives to ANCOVA when using intact groups. Paper presented at the annual meeting of Southwest Educational Research Association. New Orleans, LA.

Williams, C. S., Nimon, K. & Allen, J. (2007, January): Repeated measures: A primer for educators considering univariate and multivariate designs. Paper presented at the Hawaii International Conference on Education. Honolulu: Hawaii.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Young, J. W. (1990). Adjusting the cumulative GPA using item response theory. *Journal of Educational Measurement, 27*(2), 175-186.

APPENDIX

Table 1

Grade Distributions in Teacher Preparation Courses at Private University.

Program/Course	N	As	Bs	Cs	Ds	Fs	Pass
Courses Taken in All Programs							
EDUC 20003	103	83	19	1	*	*	*
EDUC 30113	93	80	13	*	*	*	*
EDUC 30143	91	59	28	4	*	*	*
EDUC 30123	117	79	36	2	*	*	*
EDSP 30603	124	99	24	1	*	*	*
EC-4 Certification							
EDEC 30001	41	*	*	*	*	*	41
EDEC 30014	41	31	9	1	*	*	*
EDEC 30023	41	38	3	*	*	*	*
EDEC 30033	41	24	17	*	*	*	*
EDEC 30103	36	23	13	*	*	*	*
EDEC 30143	18	16	2	*	*	*	*
EDEC 30203	38	38	*	*	*	*	*
EDEC 30213	38	30	8	*	*	*	*
EDEC 30223	38	34	4	*	*	*	*
EDEC 30233	23	19	4	*	*	*	*
EDEC 30234	38	36	1	1	*	*	*
EDEC 40013	3	2	1	*	*	*	*
EDEC 40023	3	3	*	*	*	*	*
EDEC 40033	2	2	*	*	*	*	*
EDEC 40133	3	3	*	*	*	*	*
EDEC 40154	3	3	*	*	*	*	*
EDEC 41103	32	26	4	2	*	*	*
EDEC 41113	41	39	1	1	*	*	*
EDEC 41123	36	34	2	*	*	*	*
EDEC 41143	27	26	1	*	*	*	*
EDEC 41153	41	37	4	*	*	*	*
EDEC 42213	41	27	14	*	*	*	*
EDEC 42223	25	21	4	*	*	*	*
EDEC 42233	3	2	1	*	*	*	*
EDEC 42236	33	*	*	*	*	*	33

* denotes no data reported

Table 2

Correlations Between Private University Coursework and TExES EC-4 PPR Outcomes for EC-4 Program Completers

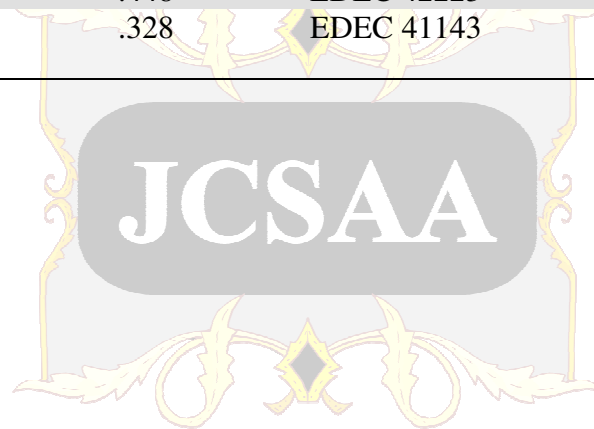| Course	r	Course	r
EDSP 30243	-.102	EDUC 20003	.323*
EDSP 30603	.407**	EDUC 30123	.690**
EDSP 40243	.349	EDUC 30143	.080
EDEC 30014	.567**	EDEC 30023	.504**
EDEC 30033	.185	EDEC 30103	.670**
EDEC 30143	.418	EDEC 30213	.548**
EDEC 30223	-.023	EDEC 30233	.694**
EDEC 30234	.404*	EDEC 41113	.296
EDEC 41123	.356*	EDEC 41153	.243
EDEC 42213	.448**	EDEC 42223	.654**
EDEC 41103	.328	EDEC 41143	.444*

Table 3

Private University EC-4 Program: Block Course Predictors of Scaled TExES PPR Outcomes

| | | | | | 95% Confidence Interval for β | |
| | | | | | Upper | Upper |
Course	b	β	t	Sig.	Bound	Bound
ESL/SPED Sophomore Prerequisites:						
EDEC 30103	18.433	.577	2.926	.014	4.567	32.300
EDEC 31043	18.667	.309	1.689	.119	-5.657	42.990
EDUC 20003	13.667	.226	1.237	.242	10.657	37.990
		$F(3, 10) = 8.708,\ p < .01,\ R^2 = .704,\ R^2_{adj} = .623$				
ESL/SPED Junior 1 Block:						
EDEC 30014	8.746	.333	2.403	.022	1.356	16.136
EDEC 30033	-.625	-.023	-.194	.847	-7.157	5.907
EDUC 30123	12.936	.452	2.692	.011	3.180	22.692
EDSP 30603	4.972	.126	.842	.405	-7.012	16.959
		$F(4, 34) = 10.735, p < .001, R^2 = .551,\ R^2_{adj} = .500$				
ESL Junior Block 2:						
EDEC 30223	-3.056	-.091	-.609	.550	13.559	7.447
EDEC 30233	16.949	.570	3.540	.002	6.928	26.971
EDEC 30213	10.449	.352	2.182	.042	.428	20.471
		$F(3, 18) = 9.054, p = .001, R^2 = .588,\ R^2_{adj} = .523$				
SPED Junior Block 2:						
EDEC 30223	-1.650	-.041	-.288	.775	-13.372	9.971
EDEC 30213	10.449	.352	2.182	.042	.428	20.471
		$F(2, 34) = 7.559, p = .002, R^2 = .302,\ R^2_{adj} = .262$				
ESL/SPED Senior Block 1:						
EDEC 41103	8.020	.345	1.451	.158	-3.286	19.326
EDEC 41153	-.987	-.025	-.106	.917	-20.098	18.124
		$F(2, 28) = 1.758, p = .190, R^2 = .108,\ R^2_{adj} = .047$				
ESL Senior Block 2:						
EDEC 42223	21.012	.654	4.151	<.001	10.539	31.484
		$F(1, 22) = 17.227, p < .001, R^2 = .428,\ R^2_{adj} = .403$				
SPED Senior Block 2:						
EDEC 42213	-2.333	-.078	-.143	.888	-.37.514	32.847
EDEC 41153	12.222	.418	.769	.456	-22.110	46.555
		$F(2, 12) = .915, p = .425, R^2 = .123,\ R^2_{adj} = -.011$				

Table 4

Private University: Results from Final Analyses.

Correlations Between EC-4 TExES PPR Outcomes and Final Analysis Predictors

Course	EC-4 PPR	EDEC 30103	EDEC 30014	EDEC 30213	EDUC 30123	EDUC 20003
EC-4 PPR	1.00					
EDEC 30103	.670**	1.00				
EDEC 30014	.567**	.376*	1.00			
EDEC 30213	.548**	.492**	.722**	1.00		
EDUC 30123	.690**	.609**	.513**	.478**	1.00	
EDUC 20003	.323*	.362*	.302	.315	.178	1.00

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Primary Overall Course Predictors of EC-4 TExES PPR Outcomes: Final Regression.

Course	b	β	t	Sig.	95% Confidence Interval for β Lower Bound	Upper Bound
EDEC 30103	8.816	.339	1.974	.059	.366	17.997
EDEC 30014	5.212	.207	.905	.374	6.620	17.045
EDEC 30213	.895	.029	.119	.906	14.586	16.376
EDUC 30123	7.861	.290	1.738	.094	1.437	17.159
EDUC 20003	4.853	.140	.935	.358	5.815	15.522

$F (5, 25) = 7.360, p < .001, R^2 = .586, R^2_{adj} = .506$

Table 5

Grade Distributions in Teacher Preparation Courses at Public University.

Program	Course	N	As	Bs	Cs	Ds	Fs	Pass
EC-4 Certification								
	DFST 3123	240	133	78	28	1	*	*
	DFEC 4243	244	221	21	1	1	*	*
	DFST 4233	232	212	18	2	*	*	*
	EDEE 3320	245	211	26	8	*	*	*
	EDEE 3380	213	144	64	5	*	*	*
	EDEE 4340	243	197	41	5	*	*	*
	EDEE 4350	242	208	29	4	1	*	*
	EDRE 4450	242	190	43	9	*	*	*
	EDRE 4860	242	226	13	3	*	*	*
	EDEE 4101	242	*	*	*	*	*	242
	EDEE 4102	242	*	*	*	*	*	242

Table 6

Correlations Between Public University Coursework and TExES PPR Outcomes for EC-4 Program Completers

	DFEC 4243	DFEC 4233	DFST 3123	EDRE 4450	EDRE 4860	EDEE 3320	EDEE 3380	EDEE 4340	EDEE 4350	TExES
DFEC 4243	1.00									
DFEC 4233	-.013	1.00								
DFST 3123	.081	.040	1.00							
EDRE 4450	.125	.081	.090	1.00						
EDRE 4860	-.004	-.032	.126	169**	1.00					
EDEE 3320	.008	.172**	.226**	.082	.107	1.00				
EDEE 3380	.023	.101	.335**	.085	.227**	.278**	1.00			
EDEE 4340	-.016	.201**	.170**	.372**	.378**	.142*	.118	1.00		
EDEE 4350	.012	.125	.142*	.174**	.217**	.143*	.200**	.219**	1.00	
TExES PPR	-.087	-.133*	.065	-.010	.118	.037	.043	.005	.091	1.00

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 7

Public University EC-4 Program: Aggregated Course Predictors of Scaled TExES PPR Outcomes:

Predictor	b	β	t	Sig.	95% Confidence Interval for β	
					Lower Bound	Upper Bound
DFEC 4243	-3.495	-.092	-1.259	.209	-8.971	1.980
EDEE 4350	5.648	.155	2.040	.043	.184	11.112
EDEE 3320	.382	.011	.142	.887	-4.920	5.683
EDEE 3380	-.151	-.005	-.065	.949	-4.777	4.474
EDEE 4340	-2.165	-.063	-.743	.458	-7.914	3.583
EDRE 4860	4.863	.097	1.209	.228	-3.075	12.801
DFST 3123	.902	.042	.528	.598	-2.471	4.275
EDRE 4450	.958	.027	.353	.724	-4.389	6.305
DFST 4233	-7.608	-.153	-2.021	.045	-15.037	-.180
EDEE 3320	-1.225	-.059	-.252	.804	-11.536	9.085

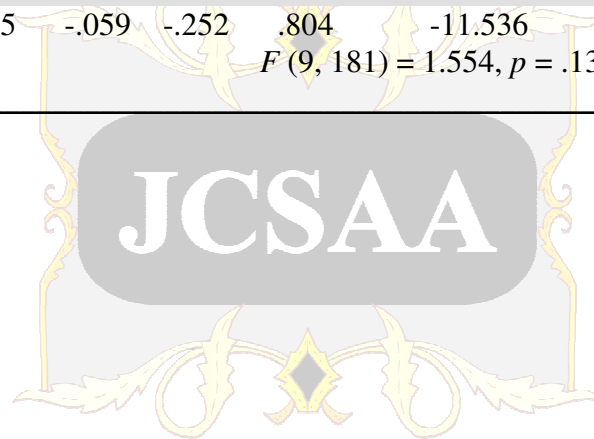$F(9, 181) = 1.554$, $p = .132$, $R^2 = .071$ $R^2_{adj} = .025$

Figure 1. Astin's input-environment-output (I-E-O) model (1993).