Journal of Teacher Education
1–14

DOI: 10.1177/0022487117702577
journals.sagepub.com/home/jte
SAGE

Predictive Validity and Impact of CAEP Standard 3.2: Results From One Master's-Level Teacher Preparation Program

Carla M. Evans[1]

Abstract

This study investigates the predictive validity and policy impact of Council for Accreditation of Educator Preparation minimum admission requirements in Standard 3.2 on teacher preparation programs (TPPs), their applicants, and the broader field of educator preparation. Undergraduate grade point average (GPA) and Graduate Record Examination (GRE) scores from 533 program graduates in one master's-level TPP were examined for their ability to predict graduate GPA and the effect minimum admissions criteria had on enrollment. Findings indicate that only undergraduate GPA is moderately related to a program graduate's success, controlling for student background characteristics. The study also finds that implementing GRE scores as a criterion in admissions decisions significantly reduces the number of admitted candidates so that the program may no longer be financially sustainable. These findings suggest many negative consequences may result from minimum admission requirements and more research is needed to evaluate the potential impact on other TPPs, teacher labor markets, and student learning outcomes.

Keywords

preservice teacher education, standards, teacher characteristics, educational policy, quantitative research

Introduction

Since the report, *A Nation at Risk* (The National Commission on Excellence in Education, 1983), many reformers have called for more rigorous teacher preparation program (TPP) admission standards to increase the quality of the nation's teachers (American Federation of Teachers, 2012; Holmes Group, 1986; Levine, 2006; National Commission on Teaching and America's Future, 1996). In this reform initiative, TPPs are positioned in a gatekeeping or screening role, charged with selecting only the "best and brightest" teacher candidates from among those applying. This position is based on research that argues teachers have the single-largest impact on student achievement of any school-based factor (Rivkin, Hanushek, & Kain, 2005) and that selecting the best applicants to become teachers is more effective than trying to dismiss them once they are already teaching (Barber & Mourshed, 2007). Improving teacher quality through policies that raise the entry bar into educator preparation is situated as a solution to many social and political problems, including meeting diverse student needs, addressing educational achievement disparities, and increasing international competitiveness (Cochran-Smith & Villegas, 2015). More selective and rigorous TPP admissions standards are one policy lever currently promoted for systemic P-12 education reform (Brabeck & Koch, 2013).

In July 2013, for example, the consolidated national accrediting agency for TPPs in the United States, Council for the Accreditation of Educator Preparation (CAEP), released five new accreditation standards (CAEP, 2013). Among the new standards, Standard 3.2 requires TPPs to increase the rigor of admission standards and to select only candidates who have demonstrated "high academic achievement and ability" (CAEP, 2013, p. 8). The academic achievement standard specifies that the provider set admissions requirements ensuring that the average undergraduate grade point average (UGPA) of an accepted cohort of teacher candidates seeking initial licensure meets or exceeds a minimum of 3.0 *and* (as originally written)

> the group average performance on nationally normed ability/achievement assessments such as ACT, SAT, or GRE is in the top 50 percent from 2016-2017, is in the top 40 percent of the distribution from 2018-2019; and is in the top 33 percent of the distribution by 2020. (CAEP, 2013, p. 8)

[1]University of New Hampshire, Durham, USA

Corresponding Author:
Carla M. Evans, Education Department, University of New Hampshire, 62 College Road, Morrill Hall 308, Durham, NH 03824, USA.
Email: carla.m.evans@gmail.com

Standard 3.2 was recently revised by the CAEP Board in June 2016 (effective July 1, 2016) as a result of a report by Teacher Preparation Analytics (Coble, Crowe, & Allen, 2016), as well as focus groups and interviews with concerned stakeholders in the education field (CAEP, 2016). Now all candidates admitted to a program must meet the academic achievement requirements either at admission or prior to graduation. Also, the group average performance for each cohort of teacher candidates must only be "within the top 50% for the reading, mathematics, and writing" as of 2016-2017, except the writing top 50% requirement is not effective until 2021 (CAEP, 2016). The CAEP Board also reclarified that American College Testing (ACT), Scholastic Aptitude Test (SAT), or Graduate Record Examination (GRE) were only examples of potential subject assessments and any "national, state, or valid and reliable alternative Educator Preparation Program-created tools" could be used as long as the admitted cohort group mean meets or exceeds the 50% threshold (CAEP, 2016, para. 4). Colleges and universities falling below the threshold in one standard will be put on probation and those falling below the standard in two or more standards will be denied CAEP accreditation (CAEP, 2013; Sawchuk, 2013).

Given the high-stakes implications of the newly revised TPP accreditation admission standard, the use and interpretation of teacher candidates' UGPA and very different subject area ability/achievement tests needs to be validated (Messick, 1985; Shepard, 1997). For example, what is the predictive validity of the minimum criteria in CAEP Standard 3.2 to teacher candidates' success in a TPP? Furthermore, what are the consequences of raising the entry bar into teaching on TPPs, their applicants, and the broader field of educator preparation? For example, how many applicants may be denied entry because they did not perform well on a standardized test? How will decreasing the supply of applicants into TPPs and graduates into the teacher workforce affect the financial viability of TPPs and teacher labor markets?

The purpose of this study is to investigate the predictive validity and policy impact of implementing CAEP Standard 3.2 on one master's-level TPP. *Predictive validity*, for this study, is defined as the extent to which scores on the GRE predict performance in a TPP. *Policy impact* is defined as the intended and unintended consequences on enrollment resulting from the new CAEP admission standard. Data on UGPA and GRE scores of 533 program graduates from the University of New Hampshire's (UNH) postbaccalaureate fifth-year master's TPP over four cohorts of teacher candidates are analyzed in two ways: for their ability to predict graduate GPA (GGPA) and to assess the effect the minimum criteria would have had on the selected pool of teacher candidates from 2010-2014 (i.e., how many program graduates would have not been accepted).

Background on CAEP Standard 3.2

The purpose of the new admissions criteria is to screen out candidates based on general predictors of academic success and cognitive ability. The implied theory of action is that improving the quality of teacher candidates will result in more effective teachers and improved student achievement outcomes. The decision to significantly alter TPP admission standards was based, in part, on a National Research Council (NRC; 2010) report that argued one of the three highest leverage aspects of teacher preparation likely to have the "strongest effects" on student outcomes is the quality of teacher candidates (p. 180). This argument is based on empirical research that examines the relationship of teacher academic ability to student achievement.

Wayne and Youngs (2003) provided a comprehensive review of the literature on the relationship between teacher characteristics (including academic ability and test scores) to student achievement and find that the studies do not lead to "clear conclusions" (p. 100). On one hand, multiple studies document a positive relationship between teachers' verbal ability and student achievement through tests intended to measure literacy levels or verbal ability (Andrew, Cobb, & Giampietro, 2005; Ehrenberg & Brewer, 1995; Ferguson & Ladd, 1996; Greenwald, Hedges, & Laine, 1996; Hanushek, 1971). And yet at the same time, tests of basic skills or teaching abilities do not report such consistent findings (Rice-King, 2003). Some argue that teachers' academic abilities, as measured by standardized tests, are most relevant for "at-risk" students (Zumwalt & Craig, 2009). Overall, the education production process is complex, but many would agree that teachers' academic ability is relevant to teacher effectiveness and student achievement (Wayne & Youngs, 2003), which is why CAEP Standard 3.2 increases the rigor of admissions standards as a policy lever to improve K-12 student achievement outcomes.

There is also evidence that the academic ability and achievement of teachers has improved over the last 25 years. For example, Lankford, Loeb, McEachin, Miller, and Wyckoff (2014) examined how teachers' SAT scores have steadily increased in New York State since 1999. This finding is supported by earlier research that documents how the SAT profile of teachers has increased over time and that academically weak students are often screened out of teaching through teacher tests required for state licensure and/or certification (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008; Gitomer & Anderson, 2009; Goldhaber & Walch, 2013). In other words, state and federal policy initiatives that have required prospective teachers to take and pass certification tests may have contributed to improvements in the academic profile of teachers entering the profession. CAEP Standard 3.2 aims to capitalize and build upon these teacher quality initiatives by using rigorous admission criteria. But what have been the admissions criteria that TPPs have utilized in the past to select teacher candidates and what does research tell us about the predictive validity of those criteria?

Prior Research on Predictive Validity of TPP Admissions Criteria

In this section, the empirical literature on TPP admissions criteria is reviewed with a particular emphasis on the relationship of the CAEP admissions requirements examined in

this study to "success" in a TPP. Success in a TPP is typically operationalized in predictive validity studies as either first-year GGPA (Ji, 1998), success in student teaching (Andrew et al., 1996; Kosnik, Brown, & Beck, 2005), faculty ratings at the end of programs taking into account multiple factors (Casey & Childs, 2011; Caskey, Peterson, & Temple, 2001), observations of practice teaching (Casey & Childs, 2011), or employment (Smith & Pratt, 1996).

Common TPP Admissions Criteria

If a primary purpose of TPPs is to prepare high-quality teachers, then one approach to selection is to link the admissions criteria to the characteristics assumed to affect teacher quality. These include teacher content knowledge, verbal ability, and personal characteristics such as care and motivation (American Educational Research Association (AERA) Panel on Research and Teacher Education, 2005; Rice-King, 2003; Wayne & Youngs, 2003). Common TPP admissions criteria serve as proxies to assess these broad teacher characteristics.

Although admissions criteria vary by program level (undergraduate vs. graduate) and program type (traditional vs. alternative), in general, TPPs require an application, minimum grade point average (GPA), essay or personal statement, and/or passing a minimum basic skills test (U.S. Department of Education, 2013). Less common, as reported by states, are admissions requirements such as interview or letters of recommendation (U.S. Department of Education [USED], 2011). Some programs utilize results or recommendations from prerequisite education courses and grades in a subject area (e.g., Andrew et al., 1996; Caskey et al., 2001) with strength of verbal and written communication skills an often utilized admission criterion (Coble et al., 2016).

Although minimum SAT, ACT, or GRE scores were part of the information solicited from TPPs by the USED (2011, pp. 70-71), they are not listed as commonly reported admissions requirements. According to a recent survey of CAEP constituents (Coble et al., 2016), most programs do not require a minimum SAT, ACT, or GRE score for admission and there is a wide range of acceptable scores if the tests are used.

UGPA. UPGA is the most widely used criterion for admission to graduate-level TPPs (Casey & Childs, 2007) and often the most heavily weighted, especially for postbaccalaureate TPPs (DeLuca, 2012). UGPA is generally believed to measure academic achievement and predict success in the academic portions of TPPs (Casey & Childs, 2007). The information is readily available on transcripts, making it a relatively easy criterion to access.

In terms of predictive validity, since past academic success is usually an indicator of future academic success, we would expect there to be a relationship between UGPA and success in a graduate-level TPP. However, there is inconclusive evidence about the predictive value of UGPA to performance in a TPP. For example, some studies do not find a significant correlation between UGPA and first-year GGPA for Education majors (Ji, 1998) or significant group differences in UGPA between teacher candidates with outstanding versus weak performance in a TPP (Andrew et al., 1996). This may be the result of the limited variability in GGPA. Other researchers, however, have found UGPA is a good indicator of success in a teaching practicum (Kosnik et al., 2005) and there is at least a weak relation (r = .1- .3) between UGPA and performance in a TPP (Casey & Childs, 2011, 2007; Caskey et al., 2001). This mirrors the findings of the relationship between UGPA and performance in graduate school, in general (Kuncel, Hezlett, & Ones, 2001).

One reason for the inconclusive findings may be a result of how selection effects confound the research because only those students who graduate from the TPPs (not just who are admitted) are typically included in studies. This results in a restriction of range for UGPA, which attenuates the magnitude of the relationship between UGPA and any outcome measure used. In addition, the overall success of most teacher candidates in TPPs, which results in a restriction of range in GGPA, has also been noted as another potential reason for weak correlations (Casey & Childs, 2011).

GRE. The GRE is designed to predict success in an academic setting rather than a more distal measure of career success. For example, GRE tests measure only a portion of the individual characteristics that are important for graduate study, including reasoning skills, critical thinking, and the ability to communicate effectively in writing (Educational Testing Service [ETS], 2014). Although graduate success is likely associated with later career success, the GRE was not developed to predict career success, but general cognitive ability (Kuncel et al., 2001). The GRE is not a widely used criterion for admissions to TPPs because most programs prepare undergraduates, not graduate students (USED, 2013). It is unclear how TPPs that use GRE scores in admissions decisions weight applicants' scores in relation to other admission criteria.

There is a large body of research conducted on the predictive validity of GRE scores (especially the Verbal and Quantitative sections) to success in graduate school (almost all focus on doctoral, not master's-level programs), usually defined as either first-year GGPA or cumulative GGPA (Bleske-Rechek & Browne, 2014; Feeley, Williams, & Wise, 2007; Grossbach & Kuncel, 2011; Holt, Bleckman, & Zitzmann, 2006; House & Johnson, 2002; Katz, Chow, Motzer, & Woods, 2009; Klieger, Cline, Holzman, Minsky, & Lorenz, 2014). The Analytical Writing section has received less attention in the literature because it was only instituted in 2002; older studies evaluated the predictive validity of the Analytical Reasoning section instead.

In general, studies that estimate the predictive validity of GRE scores and/or UGPA for success in graduate school typically use correlations to estimate the linear relationship between the predictors and academic performance in graduate school. Meta-analyses that have compared the predictive validity of standardized admissions tests (e.g., GRE, Law

School Admission Test [LSAT], Management Aptitude Test [MAT], etc.) and UGPA to first-year GGPA and cumulative GGPA have four consistent findings: (a) standardized admissions tests are effective predictors of performance in graduate school; (b) both standardized admissions tests and undergraduate GPA each predict important outcomes beyond grades; (c) standardized admissions tests predict most measures of student success better than undergraduate GPA; and (d) the combination of tests and grades yields the most accurate predictions of success (Kuncel & Hezlett, 2007). However, there is some evidence to suggest that the effects of GRE vary across type of graduate program (i.e., master's and doctoral programs; Kuncel, Wee, Serafin, & Hezlett, 2010).

In contrast, there is only a small body of literature (three studies) pertaining to the predictive validity of GRE scores to success in a TPP because most programs do not require GRE scores. In the first study, among other predictor variables, Andrew and colleagues (1996) explored the older GRE General Test sections (Verbal, Quantitative, and Analytical Reasoning) for their predictive value relative to performance as measured by faculty ratings in student teaching in a sample of 374 student teachers from four cohorts at the University of New Hampshire (UNH). Using *t* tests, they found there was no significant predictive value in GRE scores to performance in student teaching, but some predictive value in UGPA ($t = 1.48$, $p = .14$, $df = 95$).

In another study at UNH, Andrew and colleagues (2005) used the GRE-V to explore the relationship between verbal ability and teaching ability (as measured by supervisor ratings of student teachers) for about 120 student teachers in one cohort, but the correlation was not statistically significant. They did, however, find a significant correlation between the GRE Analytical Reasoning section and teaching ability ($r = .30$, $p = .01$, $n = 75$), but that section is no longer in the GRE revised General Test.

The final study was more broad and included 170 different graduate Education majors at a private university, including students majoring in curriculum and teacher preparation, educational administration, educational psychology, and counseling psychology (Ji, 1998). Only first-year GGPAs were used because not all the students had completed their program of study. The study found that GRE scores were moderately related to first-year GGPA ($r = .34$, $p < .05$) and that on average, the GRE-Q was the most predictive of first-year GGPAs on its own (predicting 16% of the variance). Using hierarchical multiple regression, the study found GRE-Q and GRE-V were positively associated with first-year GGPA, but age was negatively associated with first-year GGPA. Together, GRE-Q, GRE-V, and age explained about 33% of the variance in first-year GGPA ($F_{3,66} = 10.79$, $p < .01$).

One factor that may have attenuated the magnitude of the relationship between GRE scores and success in a TPP in each of these studies is that, similar to UGPA, selection effects can confound the predictive value of GRE scores. For example, many argue that correcting for range restriction becomes important when investigating the relationship between GRE scores and GGPA because if students are explicitly (or directly) selected into a program based on GRE scores, the limited range of scores examined (i.e., only those accepted and enrolled, not all who applied or who could have applied) results in underestimates of GRE validity coefficients (House, 1983; Kuncel et al., 2001; Kuncel et al., 2010).

In contrast to using correlational analyses, some have argued that a more straightforward way to approach testing the predictive validity of GRE scores for success in graduate school is to do away with the regression model entirely and convey the information on the utility of admissions tests using expectancy tables. Proponents of this method argue that even though predictors may account for only a small percent of the variance in GPA, they may still be important from a practical perspective (Bridgeman, Burton, & Cline, 2009). In this alternative method, the "incremental validity" of GRE scores above and beyond other criterion measures such as UGPA is examined by aggregating data across disciplines using expectancy tables (Bridgeman et al., 2009). This approach simply "ranks students on the predictors and criteria" and displays the information graphically (Bridgeman et al., 2009, p. 113). For example, what percentage of program graduates had low GRE scores (defined as below the bottom quartile) and low UGPA, but high GGPA (defined as above the top quartile)? Research with this alternative approach, however, does not tend to provide information that is not contained in regression analysis.

The present study builds upon and extends prior research on the predictive validity of UGPA and GRE scores in several ways. First, none of the predictive validity studies in Education to date examine the GRE Analytical Writing section because it was only instituted in 2002. Second, two of the three studies in Education examined both UGPA and GRE scores, but did so only using correlations and *t* tests. Instead, this study utilizes additional methods to analyze the predictive validity of the two criteria specified in CAEP Standard 3.2 for predicting success in a TPP. These methods include multiple regression and incremental validity analyses. And last, this study goes beyond investigating predictive validity and also examines the policy impact of instituting higher admissions standards on enrollment for a comparatively large sample size over four cohort of program graduates. These analyses have implications for how CAEP Standard 3.2 may affect TPPs, their applicants, and the broader field of educator preparation. As such, the following research questions guide the analyses in this study:

Research Question 1: How well do GRE scores and undergraduate GPA predict elementary and secondary teacher candidate performance in a master's-level TPP, controlling for student demographic characteristics?

Research Question 2: To what extent would a more rigorous admissions policy based on UGPA or GRE scores as a criterion impact the enrollment of a master's-level TPP and its applicants?

Data Sources and Methods

Context of the Study

The UNH is a public research university in the northeast with Carnegie classifications of high research activity and community engagement. As of the fall of 2015, UNH had a total undergraduate and graduate enrollment of about 14,600 students, approximately 86% of whom are undergraduates, 54% are female, and 48% are residents of the state. The overwhelming majority of students at UNH self-reported themselves as White (79.7%), with the second highest racial/ethnic group reported being Hispanic or Latino (3.3%)—a profile similar to the state. The entering freshmen class in the fall of 2015 had secondary school average GPAs of 3.39 and SAT score averages of 540 for Verbal and 554 for Math, which is above the national average (College Board, 2015).

UNH's TPP is a postbaccalaureate, fifth-year master's program with national accreditation. The program graduates, on average, around 130 students each academic year. Admitted teacher candidates are expected to complete a minimum of 32 credits: 20 credits of coursework in addition to a 12-credit yearlong student teaching experience. An admission rubric is used in recommending decisions to the graduate school about admission to the TPP on a rolling basis. The rubric has seven sections that are rated by two faculty members on a 4-point scale, 1 being *low* and 4 being *high*. The seven sections include UGPA,[1] grades in major, grades in education coursework (if applicable), GRE scores,[2] personal statement, three letters of recommendation, and both university professors' and cooperating teachers' recommendations from an early experience course taken prior to admission. There has been no research to date on the level of agreement between the two faculty raters or which of the seven admissions criteria on the rubric predict program success. The program is currently considering collecting the data necessary from student files to conduct such research.

The admissions rubric is designed to encompass various proxies for the knowledge, skills, and dispositions predicted to inform teacher candidates' success in a TPP (Andrew et al., 2005; Andrew et al., 1996; Casey & Childs, 2011, 2007). As such, an applicant could score below the minimum 3.0 UGPA and/or below the minimum expected GRE scores and still be admitted because other noncognitive factors and dispositions are considered. From speaking with faculty in the program, applicants tend to be considered holistically with no minimum admissions rubric score to qualify for admission to the program.

Study Sample

For this study, all the teacher candidates who completed UNH's TPP in elementary and secondary education between 2010 and 2014 were identified. Records on program applicants between 2010 and 2014 who were (a) denied entry (about 10% of total applicants), (b) admitted but chose not to attend (about 31% of total applicants), and (c) enrolled but did not finish the program (<1% of total applicants) were not available for inclusion in this study. Students were excluded from the study if they had international baccalaureates (i.e., no UGPAs) or they did not have GRE scores (n = 11). The final analytic sample included 533 program graduates who met the criteria for inclusion in this study.

Data Sources

UNH collects data on each student admitted and who completes the program for programmatic information and improvement, as well as accreditation purposes. The data for these analyses included graduates in four cohorts: 2010-2011, 2011-2012, 2012-2013, and 2013-2014. Variables in the data file included UGPA, GGPA, GRE scale scores (Verbal, Quantitative, and Analytical Writing), whether the graduate was in the elementary or secondary education program, as well as various student background and demographic characteristics (gender, race/ethnicity, and age). There were also data in the file on ratings from cooperating teachers and university supervisors from an early experience course taken prior to admission, as well as job placement information; however, because of extensive missing data and the unverified nature of the job placement information, these variables were excluded from analysis.

There was no data on student performance during the yearlong student teaching experience that could be used as an outcome measure alongside or instead of GGPA. For example, prior to 2014, the only way student performance in the clinical component of the program was collected was through "final triad" meetings. In these meetings, the teacher candidate would sit down with the university faculty supervisor and cooperating teacher and describe at least three ways in which they demonstrated competence with respect to the seven major TPP goals. A conversation would ensue in which the faculty supervisor and cooperating teacher would share their perspective on the student's performance relative to the TPP goals. The faculty supervisor would then summarize the meeting in a document by identifying at least three pieces of evidence that demonstrated how the student met each program goal. These "final triad" meetings were replaced in 2014 with a preservice teacher performance assessment, which is currently in the second phase of piloting.

Measures

This section describes the measures used in this study, including the admissions criteria from CAEP Standard 3.2, the covariates used in the analysis, and the outcome variable.

Predictor variables. Both UGPA and GRE scale scores of program completers were used as predictor variables in this study and centered on their sample means to aid interpretation. The

GRE General Test administered prior to August 1, 2011, has a Verbal and Quantitative score scale from 200 to 800 in 10-point increments, whereas the GRE General *Revised* Test administered on or after August 1, 2011, has a Verbal and Quantitative score scale from 130 to 170 in one-point increments. The Analytical Writing section in both versions is scored from 0 to 6 in half-point increments. For most of the analyses in this study, GRE sections were analyzed using scale scores separately by test version to determine whether a particular section or version of the GRE had more predictive value than another. To examine the policy impact, GRE scale scores were converted to percentile scores using ETS's two guides to the use of GRE scores (ETS, 2010, 2014). This conversion was done to create a variable that fit the new CAEP accreditation standard, which focuses on the cohort average GRE percentile score as compared with the national distribution.

UGPA serves as a proxy of subject-specific content knowledge since UNH's program culminates in a master's degree where each program graduate first earned a bachelor's degree in a specific content/subject area. The GRE, as mentioned earlier, serves as a proxy of general cognitive ability, and each GRE section tests a particular kind of ability assumed important in graduate school.

Covariates. Because a teacher candidate's academic performance can be influenced by which version of the GRE General Test they took, as well as various background and demographic characteristics, five covariates were included in the analysis. Four categorical variables were included: (a) older GRE General Test or the newer GRE *Revised* General Test (OLD = 1); (b) elementary or secondary education major (ELEM = 1); (c) female or male (FEMALE = 1); and (d) White or non-White (WHITE = 1). A categorical variable was used for race/ethnicity because 87.8% of the analytic sample indicated their race/ethnicity was White non-Hispanic—a reflection of the state's lack of racial/ethnic diversity. The race/ethnicities grouped as non-White included American Indian/Alaskan Native (.2%), Asian-Pacific Islander (.6%), Black non-Hispanic (.8%), Hispanic (1.1%), and not Hispanic or Latino (2.8%).

The final covariate was age, which was significantly positively skewed (s = 3.49, SE = .106) and leptokurtic (k = 12.89, SE = .211). After examining the bivariate scatterplot between GGPA versus age, the relationship appeared nonlinear. Age was transformed in multiple ways, but no improvement in R^2 was noted and therefore the untransformed variable was used in the analyses to simplify interpretation.

Outcome variable. In this study, the predictive validity of UGPA and GRE scores is represented by the estimated relationship between those scores and GGPA as it is in most predictive validity studies of UGPA and GRE scores (Kuncel et al., 2001; Kuncel et al., 2010) and because no other outcome measures were available.

Analytic Methods

Predictive validity analysis. To address the first research question about how well UGPA and GRE scores predict elementary and secondary teacher candidate performance in a master's-level TPP, the analysis was conducted in three phases:

Correlational analysis. Although other GRE predictive validity studies analyze correlations with and without corrections for restriction in range, there is no reason to believe that this sample does not contain the true range of GGPA or GRE scores for the population of all UNH's TPP graduates. For example, GRE scale scores in this sample included almost the entire score range for each GRE test version. As a result, exploratory analysis was conducted of bivariate relationships between GGPA and the predictor variables and covariates using only Pearson product–moment correlations.

Multiple regression analysis. In the second part of the analysis, a taxonomy of hierarchical multiple regression models were fit to determine the improvement, if any, in prediction for GGPA when different combinations of predictors and covariates are used. Models were fit beginning with the predictor variables more strongly associated with GGPA, and then adding covariates in as a group. Interactions were then tested between the covariates and predictor variables, as well as between the covariates. The final model for each GRE test version includes only the significant predictor variables, covariates, and/or interactions below the .05 alpha level. Influence statistics, residual analysis, and sensitivity checks were applied to the final models to check for violations of the assumptions of the linear regression model and atypical data points. No violations were noted or atypical data points removed.

Incremental validity analysis. Consistent with the standard approach (Bridgeman et al., 2009), students were divided into quartiles based on GRE scores and UGPA, and then the percentage of students in the top, middle two, and bottom quartile of GGPA were compared. For example, descriptive statistics were estimated on the percentage of elementary and secondary TPP graduates with low GRE scores (defined as below the bottom quartile of the sample), but high GGPA (defined as above the top quartile of the sample).

Chi-square tests were then used to test the relationship between GRE scores and GGPA. Students were categorized into groups: "high GRE" defined as the top quartile of the sample or "not high GRE" defined as any score below the top quartile of the sample. Students were also categorized into "high GGPA" defined as the top quartile of the sample or "not high GGPA" defined as any score below the top quartile of the sample. Cross-tabulating group results tested the null hypothesis that in the population of all the UNH's master's-level TPP graduates, there is no relationship between GRE scores and GGPA. For simplicity, because results of the incremental validity analyses were consistent with

the multiple regression analyses and did not add any new information, they will not be discussed further in this article. Regression results were also chosen to present because the effect of covariates and interactions on the relationship between UGPA and GRE scores with GGPA are not captured in the incremental validity analysis.

Policy impact analysis. To answer the second research question about the policy impact on a master's-level TPP and its applicants of implementing a more rigorous and selective admissions policy based on UGPA and GRE scores, a two-part analytic plan was used.

Frequency counts. Simple frequency counts were calculated to examine how many students would have been admitted and would not have been admitted by cohort to UNH's TPP if GRE scores were the determinant criterion under the newly revised CAEP Standard 3.2, which sets a top 50% threshold. UGPA was not analyzed as every cohort had an average UGPA above 3.0.

Two approaches to frequency counts were applied. First, a *strict individual cut score approach* was used so that no applicant with an average GRE percentile score below the top 50th percentile in a particular cohort would be admitted. Second, a *group score band approach* was used so that the program could select applicants falling below the minimum CAEP requirement, but the final cohort average would stay above the CAEP requirement because high performing students compensate for lower performing students. This provides flexibility for programs to accept students who do not meet the cutoff score, but whom they believe have the content knowledge, intellectual and verbal ability, and dispositions to become high-quality teachers. It may also meet the spirit of the CAEP requirement but have a less negative impact on TPPs. The group score bands used in this study were determined using GRE scores from the four cohorts in this sample. This analysis demonstrates the extent to which CAEP Standard 3.2 may affect the numbers of admitted students in the future, which has implications for the sustainability of TPPs and teacher labor markets.

Tests of group differences. Tests of group differences (*t* tests) were used to examine whether or not there were group differences in GGPA between those students who would have been admitted and would not have been admitted as a result of the academic achievement standard of the top 50% for reading, mathematics, and writing using either a strict cut score or score band approach. Data were analyzed by cohort and across cohorts, but because results were similar, only results of *t* tests across cohorts are reported in this article. Results of this analysis will demonstrate the extent to which the CAEP-proposed cut score of the top 50% relates to success in a TPP, which has implications for the meaningfulness of the CAEP cut score.

Findings

Descriptive Findings

The majority (58%) of the program graduates in this study were secondary education majors. Similar to the demographic characteristics of the state and the university, the majority of program graduates were White (87.5%). The average program graduate was 24 years old (SD = 4.58), although ages ranged from 20 to 50 years old. The means, standard deviations, and range of all the variables are presented in Table 1.

In general, GGPAs are higher than UGPAs, which is not surprising given that grade inflation is common in graduate school (Jewell, McPherson, & Tieslau, 2011) and students must maintain a minimum GPA of 3.0 to stay in UNH's program. Both cohort average UGPAs and GGPAs appear to be on the rise, however, with average UGPAs increasing the most from 3.4 for Cohort 1 to 3.55 for Cohort 4.

Program graduates from all four cohorts have a wide range of GRE scale scores on both GRE test versions with ranges including almost the entire scale. However, the mean GRE scale score does increase for each admitted cohort. In general, students tended to score significantly higher across all cohorts on the GRE Analytical Writing and Verbal sections as compared to the Quantitative section.

Predictive Validity Findings

The bivariate correlations between the variables of interest were examined prior to conducting the multivariate analysis. Bivariate correlations were generally low (Table 2), which, along with the R^2 values of the regression models, indicate that there is a lot of variability left to explain in GGPA. Overall, UGPA predicted the highest amount of variability in GGPA—about 14.5%. UGPA and GRE scale scores were weakly related to one another (r = .21-.34). GRE scale scores from each of the three sections did not predict much variability in GGPA, although the GRE-A section on both GRE test versions predicted about 3% of the variability in GGPA. The very weak relationships that exist between GGPA and GRE scores (r = −.01-.18) in this sample are similar to other studies at UNH on the predictive value of GRE scores to "success" in the TPP. For example, Andrew and colleagues (1996) found that GRE scores did not predict weak or outstanding performance in UNH's teaching internship. Andrew and colleagues (2005) later explored the relationship between verbal ability (as measured by GRE-V scores) and teaching ability (as measured by supervisor ratings) and found weak nonsignificant correlations for both the GRE-V (r = .23) and GRE-Q (r = .20).

Overall, regression analysis indicated two noteworthy findings (see Table 3). First, GRE Verbal and Quantitative sections are not statistically significant predictors of GGPA, but the GRE Analytical Writing section has a significant positive effect on GGPA in Model I. Once the interactions are added to the model, the Analytical Writing section is no

Table 1. Descriptive Statistics by Cohort (*N* = 533).

	Cohort 1 (2010-2011)	Cohort 2 (2011-2012)	Cohort 3 (2012-2013)	Cohort 4 (2013-2014)
n	153	134	133	113
Elem/Sec (%)	42/58	43/57	38/62	44/56
F/M (%)	80/20	77/23	84/16	74/26
White/NW/U (%)	89/3/8	90/6/4	91/6/3	80/8/12
M_{age}	24.4	23.95	24.27	23.61
(*SD*)	(4.9)	(4.23)	(4.99)	(3.99)
Range	21-50	21-48	21-50	20-46
M_{GGPA}	3.81	3.80	3.81	3.85
(*SD*)	(0.143)	(0.17)	(0.19)	(0.17)
Range	3.26-4.0	3.23-4.0	3.26-4.0	3.3-4.0
M_{UGPA}	3.40	3.46	3.45	3.55
(*SD*)	(0.32)	(0.32)	(0.30)	(0.27)
Range	2.63-4.0	2.61-3.98	2.25-4.0	2.86-3.97
GRE General Test scale scores (200-800)				
n	153	134	93	7
$M_{GRE\text{-}V}$	442.61	447.24	468.49	510.00
(*SD*)	(88.77)	(86.25)	(92.92)	(88.31)
Range	280-720	240-720	310-760	370-620
$M_{GRE\text{-}Q}$	517.58	532.54	545.70	534.29
(*SD*)	(110.52)	(120.70)	(116.17)	(143.28)
Range	260-800	260-800	220-800	360-760
$M_{GRE\text{-}A}$	4.06	4.04	4.17	4.07
(*SD*)	(0.69)	(0.62)	(0.64)	(0.35)
Range	2.5-5.5	2.5-6.0	2.5-6.0	3.5-4.5
GRE General Revised Test scale scores (130-170)				
n			40	106
$M_{GREr\text{-}V}$			149.23	150.28
(*SD*)			(6.5)	(6.03)
Range			138-163	137-169
$M_{GREr\text{-}Q}$			145.05	148.25
(*SD*)			(6.08)	(6.41)
Range			133-163	137-165
$M_{GREr\text{-}A}$			3.94	3.92
(*SD*)			(0.67)	(0.57)
Range			2.5-6.0	2.5-5.5

Note. *n* = number of students; Elem/Sec (%) = percentage of Elementary/Secondary education majors; F/M (%) = percentage of female/male students; White/NW/U (%) = percentage of White/non-White/ unspecified students; age = age in years at graduation; GPA = grade point average; GGPA = graduate grade point average; UGPA = undergraduate grade point average; GRE = Graduate Record Examination; GRE*r* = GRE General Revised Test; V = Verbal Reasoning; Q = Quantitative Reasoning; A = Analytical Writing.

longer significant. This finding is different from prior research for two reasons. First, in previous studies, the GRE was a significant predictor of GGPA in graduate school (Kuncel et al., 2001) and in one of the three studies in Education (Ji, 1998). Second, in those studies, GRE scores explained more of the variance in GGPA than UGPA, but in this study, that finding is reversed. It is unclear why the findings of this study are not consistent with earlier research. It may be the case that UNH's TPP requires applicants to take the GRE in order to apply to the program, but admissions decisions do not heavily weight applicants' scores. This would then create a situation where GRE scores are more of a formality rather than an admissions criterion. More research is needed to examine the predictive validity of the GRE in TPP admissions criteria in other programs with different student populations.

A second noteworthy finding is that age moderates the relationship between UPGA and GGPA. This means that the effect of UGPA on GGPA varies as a function of a student's age. For example, Figure 1 illustrates the moderating effect of age on the relationship between UGPA and GGPA using Model VI for male, secondary education majors. Whereas the relationship between UGPA and GGPA is positive for both younger and older students in this sample, there is less of an effect of UGPA on GGPA for older students. This theoretically makes sense as

Table 2. Pearson Product–Moment Correlations Using Pairwise Deletion.

Variable	1	2	3	4	5	6	7	8	9	10	11
1. GGPA *n* = 533											
2. UGPA *n* = 533	.38**										
3. GRE-V *n* = 387	.12*	.26**									
4. GRE-Q *n* = 387	.13*	.22**	.39**								
5. GRE-A *n* = 385	.16**	.23**	.34**	.29**							
6. GRE*r*-V *n* = 146	–.01	.21*									
7. GRE*r*-Q *n* = 146	–.01	.34**				.56**					
8. GRE*r*-A *n* = 146	.18*	.27**				.35**	.28**				
9. ELEM *n* = 533	.16**	.02	–.22**	–.16**	–.24**	–.24**	–.29**	–.10			
10. AGE *n* = 533	.03	–.09*	.25**	–.03	–.04	.15	.03	.03	–.05		
11. WHITE *n* = 497	–.01	.05	–.01	.01	.03	–.05	–.02	–.01	.04	–.18**	
12.FEMALE *n* = 533	.26**	.20**	–.14**	–.11*	–.01	–.19*	–.31**	–.12	.30**	–.15**	.08

Note. GGPA = graduate grade point average; *n* = number of students; UGPA = undergraduate grade point average; GRE = Graduate Record Examination; V = Verbal Reasoning; Q = Quantitative Reasoning; A = Analytical Writing; GRE*r* = GRE General Revised Test; ELEM = elementary education major.
*$p < .05$. **$p < .01$.

older students are more likely to be nontraditional students whose past academic history may have less bearing on their academic performance in the program in comparison with students coming straight from an undergraduate program. Interestingly, before testing the two-way interactions between the predictors and control variables, age had a positive effect on GGPA (Models I, III, and V), which differed from the two prior studies on TPP admission criteria where the researchers controlled for age. In those studies, which did not test for interactions, age was negatively associated with first-year GGPA (Ji, 1998) and success in a student teaching practicum (Kosnik et al., 2005). More research is needed to (a) understand the magnitude and direction of the relationship between teacher candidate age and performance in a TPP and (b) explore whether or not age moderates the effect of UGPA on GGPA across settings and student populations as that may inform the predictive validity of admissions criteria such as UGPA based on the age of the applicant.

Policy Impact Findings

Based on the descriptive findings, using UGPA as an admissions criterion would not significantly affect enrollment at the TPP being studied because no cohort had an average below 3.0. In contrast, if this TPP had required a cohort average GRE score in the top 50% of the national distribution, 58% to 74% of each cohort would not have been admitted with a strict cut score approach and 15% to 34% of each cohort would not have been admitted using a score band approach (Figure 2). Under the score band approach, which is the most flexible and lenient, this equates to 163 out of the total 533 program graduates being denied entrance solely on their GRE scores not being in the top 50% of the national distribution—a 31% drop in enrollment. This number goes up to 360 out of 533 program graduates if a strict cut score approach had been used—a 68% drop in enrollment.

The mean GGPA of the 163 program graduates and the 360 program graduates who would have not been admitted was 3.78 and 3.79, respectively. Thus, students with lower GRE scores tended to perform at a high level in UNH's graduate program. That said, results of *t* tests suggest that there is a statistically significant difference in GGPA between students who perform above or below the 50th percentile on the national distribution using the strict cut score approach on the GRE ($t_{1,531} = 3.015$, $p < .01$, $d = .28$) and score band approach ($t_{1,531} = 2.952$, $p < .01$, $d = .27$; Table 4). This suggests that for UNH's fifth-year master's TPP, using the 50th percentile of the national distribution on the GRE as a cohort average admissions criterion may distinguish between top and weak academic performers, although .05 on a 4-point GPA scale (3.8 vs. 3.85) may not meaningfully distinguish top from weak performers.

Table 3. Hierarchical Multiple Regression Analysis Predicting Variance in Graduate GPA.

	GRE General Test (*n* = 385)		GRE Revised General Test (*n* = 146)		Combined GRE versions (*n* = 531)	
Model	I	II	III	IV	V	VI
Intercept	3.73**	3.75**	3.77**	3.74**	3.74**	3.74**
	(.02)	(.02)	(.03)	(.03)	(.03)	(.03)
UGPA	0.14**	0.17**	0.30**	0.19**	0.19**	0.20**
	(.03)	(.03)	(.05)	(.07)	(.02)	(.02)
GRE-A	0.03*	0.02	0.04	0.07		
	(.01)	(.01)	(.02)	(.03)		
FEMALE	0.08**	0.07**	0.05	0.08*	0.07**	0.07**
	(.02)	(.02)	(.04)	(.04)	(.02)	(.02)
ELEM	0.05**	0.04*	0.04	0.04	0.04**	0.03*
	(.02)	(.02)	(.03)	(.03)	(.01)	(.01)
AGE	0.003	–0.01*	0.01*	–0.01	0.003*	0.002
	(.002)	(.003)	(.004)	(.01)	(.001)	(.005)
AGE × UGPA		–0.02**		–0.08*		–0.02**
		(.01)		(.03)		(.004)
AGE × GRE-A		–0.004		0.02		
		(.002)		(.02)		
AGE × GEND		0.01**		0.03		0.01**
		(.003)		(.02)		(.003)
OLD					–0.06	–0.04
					(.04)	(.04)
R^2	19.9	25.1	28.8	32.7	19.4	24.5
F	13.34**	12.50**	7.98**	6.55**	25.27**	18.80**
df	7,377	10,374	7,138	10,135	5,525	9,521
ΔR^2		.05		.04		.05
F		8.66**		2.57		8.83**
df		3, 374		3,135		4,521

Note. Models I to IV included GRE Verbal and Quantitative scale scores and Models V to VI included all GRE test sections combined, but the parameter estimates were zero and therefore not included; all two-way interactions between the predictor variables and control variables were tested, but only significant interactions were reported; unstandardized parameter estimates are followed by standard errors in parentheses. GPA = grade point average; GRE = Graduate Record Examination; *n* = number of students; UGPA = undergraduate grade point average; A = GRE Analytical Writing; ELEM = elementary education major; OLD = older version of the GRE General Test administered prior to August 1, 2011.
*$p < .05$. **$p < .01$.

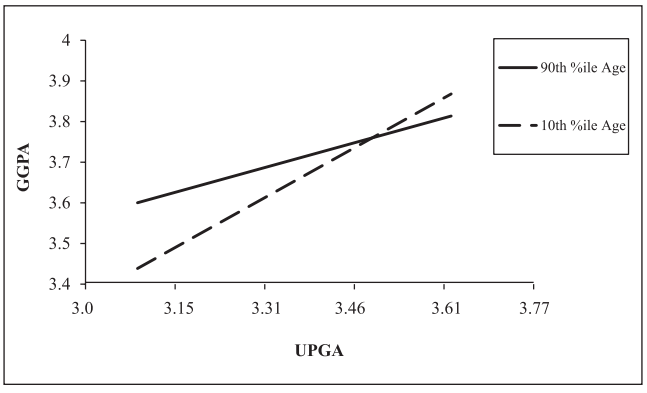

Figure 1. Line plot showing the moderating effect of age on the relationship between UGPA and GGPA for male, secondary education majors at the 90th percentile (28 years old) and 10th percentile (21 years old) of the sample for age using Model VI parameter estimates (*N* = 533).
Note. GGPA = graduate grade point average; UGPA = undergraduate grade point average.

Discussion

Results suggest that UGPA is the best predictor of strong performance in UNH's TPP for these four cohorts of graduates. In general, GRE scores did not explain a significant amount of the variability in GGPA except the positive effect of the GRE General Test Analytical Writing section on GGPA in one regression model. Because these findings differ from past research on the predictive validity of the GRE to graduate school success, and the results of this study do not generalize beyond this university, research using nationally representative data from TPPs is an important next step.

Results also suggest that the refined CAEP cut score of the group average performance on the GRE in the top 50% may be a useful indicator of future academic performance for this TPP because results of *t* tests indicated there were differences in GGPA between those students who scored below and above that cut score. However, as noted above, the difference of .05 on a 4-point GPA scale (3.8 vs. 3.85) may not meaningfully distinguish between top and weak performers. Furthermore,

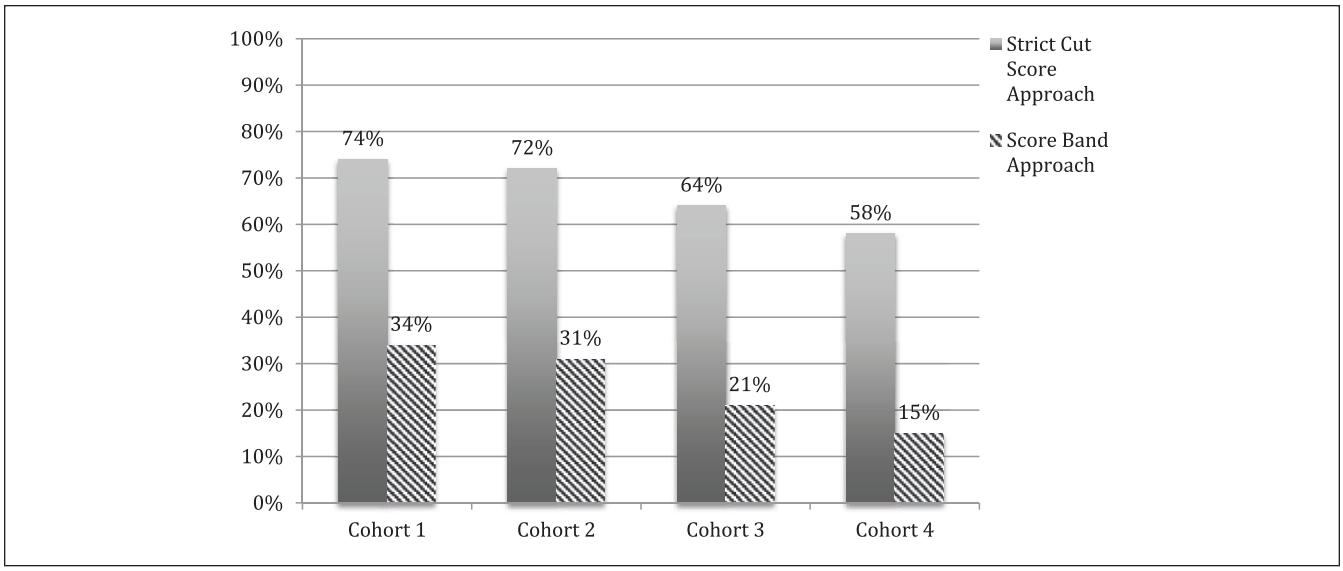

Figure 2. Percentage of students not admitted based solely on GRE scores using a strict cut score (top 50th percentile of the national distribution) or score band approach (≥34th percentile) by cohort (N = 533).
Note. GRE = Graduate Record Examination.

Table 4. The *t*-Test Results Estimating Group Differences in GGPA Between Students Admitted or Not Admitted Based on GRE Percentile Scores (N = 533).

CAEP criteria	*n*	*M*	*SD*	*M* difference	*t*	*df*	*d*
Strict cut score approach							
≥50th percentile	173	3.846	.162	.046	3.015**	531	.28
<50th percentile	360	3.799	.170				
Score band approach							
≥34th percentile	370	3.829	.162	.046	2.952**	531	.27
<34th percentile	163	3.782	.178				

Note. GGPA = graduate grade point average; GRE = Graduate Record Examination; CAEP = Council for the Accreditation of Educator Preparation; *n* = number of students; *d* = Cohen's effect size.
*p < .05. **p < .01.

the benefits of differentiating between top and weak academic performers must be examined in light of the significant consequences on enrollment the use of GRE scores in the top 50% would have on UNH's financial viability and sustainability, not to mention local and regional teacher labor markets. For example, UNH graduates, on average, about 130 new teachers every year. Some of those program graduates move to other states or do not find jobs, but of those who find employment, about 70-80% stay and teach in the New Hampshire's P-12 schools. Reducing this supply of trained teachers could result in widespread teacher shortages and the need to alternatively certify or emergency credential large numbers of untrained teachers. This may then have a negative impact on schools and student learning outcomes (Darling-Hammond, 2002)—the opposite effect of the intended policy.

Furthermore, beyond the point that program graduates fill an important role in supplying teachers to local and regional labor markets, results of UNH's last 5-year TPP study indicated that program graduates are high-quality and competent teachers. For example, 75% of principals surveyed rated the program's graduates in the top 20% in response to the question: "Compared to teachers of similar teaching experience, please rate this teacher's performance as caring, qualified, and competent professional." In addition, 92.3% of principals rated the graduate's/teacher's knowledge of their subject area(s) as very good (34.6%) or excellent (57.5%). The question then remains: Does implementing a top 50% threshold cause more harm then good, especially as there is still much left to understand about the relationship between teacher candidate performance on reading, mathematics, and writing subject area assessments and future teacher quality?

Methodological Limitations

A clear limitation is that this study is confined to a sample of program graduates at one master's-level TPP set within a

public research university with minimal racial/ethnic diversity. Consequently, this study is not generalizable beyond UNH and cannot provide any insight into the potential impact of CAEP Standard 3.2 on undergraduate TPPs or the impact of the new standard on diversity. Future research could use state- or national-level data to provide a more generalizable picture of the predictive validity and potential policy impact of the new admissions standard on a range of programs, including the impact on diversity. For example, given that CAEP Standard 3.1 calls for TPPs to present plans and goals to recruit diverse candidates who meet employment needs, what impact might there be on admission pools if there were both recruitment strategies and rigorous selection strategies in place?

Another limitation of this study is that GGPA is the only outcome measure used. There are many other potential ways in which "success" in a TPP could be defined. For example, student performance in the clinical component of the program or on a preservice teacher performance assessment may be a better proxy of teacher effectiveness than GGPA. This study does not provide any insight into the predictive validity or policy impact of CAEP Standard 3.2 relative to those outcome measures. Also, it may be that accrediting bodies (among others) are less concerned about a teacher candidate's success *in a program* than they are about a teacher candidate's success *in teaching*. Meaning, accrediting bodies and various stakeholders may be most interested in how the selectivity of teacher candidates and rigorous admissions standards relate to a positive impact on K-12 student achievement outcomes. Although the relationship between higher TPP admission standards, improvements to the teacher workforce, and K-12 student achievement is unclear and complicated by many confounding factors (American Educational Research Association, 2015; Freeman, Martin, Brousseau, & West, 1989; Konstantopoulos, 2014; NRC, 2010), this is an area future research could explore.

And finally, this study only addresses the predictive validity and policy impact of final UGPA and GRE test scores; however, most TPPs in the United States are undergraduate programs that would not have final UGPAs available or use the GRE for admissions purposes. CAEP Standard 3.2 recognizes this fact and allows for UGPA to be calculated at the point of admission to the program and also allows flexibility in terms of the assessment used to measure reading, mathematics, and writing ability/achievement. This means though that policy impacts on enrollment will likely vary as a function of ability/achievement test used by a TPP. For example, the top 50% of the GRE pool is not equivalent to the top 50% of the SAT/ACT pool because a much smaller, selective, and more successful pool of applicants takes the GRE (Hennessy & Pagnotta, 2015). Future research could explore the differences in policy impact according to the type of standardized admissions test used.

There are two other noteworthy limitations of this study. First, because UGPA and GGPA in this sample have a significant restriction of range, as well as a ceiling effect as 4.0 is the highest average possible, the parameter estimates and correlation coefficients in this study will likely be underestimated. Second, this study does not include students who were either selected out or selected themselves out of the TPP during the application process, or who began the program, but did not graduate. These students may have had systematically different patterns of UGPA or GRE scores, which are not captured in this study, but whose exclusion may result in measurement error.

Conclusion

Overall, these findings highlight a need to continuously examine and monitor the predictive validity and policy impact of CAEP Standard 3.2 across a wide variety of TPPs, both undergraduate and graduate, including the standard's impact on diversity. This examination may provide insight into the predictive value of the CAEP minimum admission requirements across different applicants, contexts, subject area ability/achievement tests, and outcome measures. Furthermore, future research could investigate the extent to which CAEP Standard 3.2 has the intended effect on TPPs, their applicants, and the broader field of educator preparation. For example, are program graduates any better prepared to enter teaching and impact K-12 student achievement outcomes than their pre-CAEP Standard 3.2 counterparts? Are minority applicants disproportionately affected? To what extent has CAEP Standard 3.2 affected teacher labor markets?

Results of this study suggest that UGPA is a useful predictor of "success" for this sample, but the GRE has little to no predictive value in the TPP under examination; however, this finding is not generalizable and should not be applied beyond this university. That said, due to the significant impact that implementing CAEP Standard 3.2 (particularly increasing cohort average GRE scores to the top 50% of the national distribution) would have on UNH's enrollment, implementing the new standard on a national level should continue to proceed with caution. It will be important for other TPPs to monitor and evaluate the impact of increasing the rigor and selectivity of their admission standards and report that impact to the CAEP Board so that future adjustments can be made to Standard 3.2 based on evidence.

Acknowledgments

The author thanks Charlie DePascale, Suzanne Graham, Jim Hennessy, and Scott Marion for their substantive feedback on earlier drafts. She also extends her gratitude to the reviewers who provided helpful critiques and suggestions, and to the Education Department at the University of New Hampshire for supporting this work, especially Tom Schram and Liz Arcieri.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. As of 2012, minimum undergraduate grade point average (UGPA) for acceptance is 3.0. Prior to that time, the minimum UGPA was 2.77; however, applicants were not disqualified from admission based on UGPA alone. They just received the lowest rating on that component of the admissions rubric. As of 2012, to earn a "4" on the admissions rubric, students must have an UGPA at or above the top range for middle 50% of students previously admitted (range = 3.25-3.65); "3" is, if a student is within the range for middle 50% of students previously admitted; "2" is, if UGPA is between 3.0 and 3.2; and "1," if UGPA is below 3.0.
2. Minimum expected Graduate Record Examination (GRE)-revised scores after Fall 2011 on the rubric is Verbal (146; 29th percentile nationally), Quantitative (140; 8th percentile nationally), and Analytical Writing (4.0; 56th percentile nationally). Similar to UGPA, applicants were not disqualified from admission based on GRE scores alone. They just received a lower rating on that component of the admissions rubric. As of 2011, to earn a "4" on the admissions rubric, all GRE scores had to be at or above the top score of the middle 50% of students previously admitted (listed above); "3," if all GRE scores are within the middle 50% range or above; "2," if one GRE score is below the middle 50% range; and "1," if two or more GRE scores are below the middle 50% range.

References

American Educational Research Association Panel on Research and Teacher Education. (2005). *Studying teacher education: The report of the AERA panel on research and teacher education* (M. Cochran-Smith & K. M. Zeichner, Eds.). Mahwah, NJ: Lawrence Erlbaum.

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*(8), 448-452. doi:10.3102/0013189X15618385

American Federation of Teachers. (2012). *Raising the bar: Aligning and elevating teacher preparation and the teaching profession*. Washington, DC: Author.

Andrew, M. D., Cobb, C. D., & Giampietro, P. J. (2005). Verbal ability and teacher effectiveness. *Journal of Teacher Education*, *56*(4), 343-354. doi:10.1177/0022487105279928

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

Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. Boston, MA: McKinsey.

Bleske-Rechek, A., & Browne, K. (2014). Trends in GRE scores and graduate enrollments by gender and ethnicity. *Intelligence*, *46*, 25-34. doi:10.1016/j.intell.2014.05.005

Boyd, D. J., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, *27*(4), 793-818. doi:10.1002/pam.20377

Brabeck, M., & Koch, C. (2013). Why the new teacher ed. standards matter. *Education Week*, *33*(4), 26-27.

Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, *22*(2), 109-119. doi:10.1080/08957340902754577

Casey, C. E., & Childs, R. A. (2007). Teacher education program admission criteria and what beginning teachers need to know to be successful teachers. *Canadian Journal of Educational Administration and Policy*, *67*, 1-24.

Casey, C. E., & Childs, R. A. (2011). Teacher education admission criteria as measure of preparedness for teaching. *Canadian Journal of Education*, *34*(2), 3-20.

Caskey, M. M., Peterson, K. D., & Temple, J. B. (2001). Complex admission selection procedures for a graduate preservice teacher education program. *Teacher Education Quarterly*, *28*(4), 7-21.

Coble, C., Crowe, E., & Allen, M. (2016). *CAEP Standard 3.2 research, study and analysis: A report to the Council for the Accreditation of Educator Preparation*. Washington, DC: Teacher Preparation Analytics.

Cochran-Smith, M., & Villegas, A. M. (2015). Framing teacher preparation research: An overview of the field, part 1. *Journal of Teacher Education*, *66*(1), 7-20. doi:10.1177/0022487114549072

College Board. (2015). *SAT 2015 college-bound seniors: Total group profile report*. Princeton, NJ: Author.

Council for the Accreditation of Educator Preparation. (2013). *CAEP accreditation standards*. Washington, DC: Author.

Council for the Accreditation of Educator Preparation. (2016). *Demonstrating academic achievement: CAEP Board clarifies, refines CAEP Standard 3*. Washington, DC: Author. Retrieved from http://caepnet.org/about/news-room/caep-board-clarifies-refines-caep-standa

Darling-Hammond, L. (2002). Research and rhetoric on teacher certification. *Educational Policy Analysis Archives*, *10*(36). Retrieved from http://dx.doi.org/10.14507/epaa.v10n36.2002

DeLuca, C. (2012). Selecting inclusive teacher candidates: Validity and reliability issues in admission policy and practice. *Teacher Education Quarterly*, *39*, 7-31.

Educational Testing Service. (2010). *GRE guide to the use of scores: 2010-2011*. Princeton, NJ: Author. Retrieved from www.ets.org/gre/guide

Educational Testing Service. (2014). *GRE guide to the use of scores: 2014–2015*. Princeton, NJ: Author. Retrieved from www.ets.org/gre/guide

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, *14*(1), 1-21.

Feeley, T. H., Williams, V. M., & Wise, T. J. (2007). Testing the predictive validity of the GRE exam on communication graduate student success: A case study at University at Buffalo. *Communication Quarterly*, *53*(2), 229-245. doi:10.1080/01463370500090209

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265-298). Washington, DC: Brookings Institute.

Freeman, D. J., Martin, R. J., Brousseau, B. A., & West, B. B. (1989). Do higher program admission standards alter profiles of entering teacher candidates? *Journal of Teacher Education*, *40*(3), 33-41.

Gitomer, D. H., & Anderson, S. E. (2009). *Teacher quality in a changing policy landscape: Improvements in the teacher pool* (NCES 2008-153). Washington, DC: National Center for Educational Statistics.

Goldhaber, D., & Walch, J. (2013). *Rhetoric versus reality: Is the academic caliber of the teacher workforce changing?* (CEDR Working Paper 2013-4). Washington, DC: Center for Education Data & Research.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361-396.

Grossbach, A., & Kuncel, N. R. (2011). The predictive validity of nursing admission measures for performance on the national council licensure examination: A meta-analysis. *Journal of Professional Nursing*, *27*(2), 124-128. doi:10.1016/j.profnurs.2010.09.010

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, *61*(2), 280-288.

Hennessy, J. J., & Pagnotta, J. (2015, October). *The mismeasurement of teaching: Unintended consequences of using general scholastic aptitude as a determinant of entry and retention in teaching*. Paper presented at the annual meeting of the. Northeastern Educational Research Association, Trumbull, CT.

Holmes Group. (1986). *Tomorrow's teachers: A report of the Holmes Group*. East Lansing, MI: Author.

Holt, D. T., Bleckman, C. A., & Zitzmann, C. C. (2006). The Graduate Record Examination and success in an engineering management program: A case study. *Engineering Management Journal*, *18*(1), 10-16.

House, J. D. (1983). Effects of restriction of range on predictive validity for the Graduate Record Examination. *Psychological Reports*, *53*(3), 710. doi:10.2466/pr0.1983.53.3.710

House, J. D., & Johnson, J. J. (2002). Predictive validity of the Graduate Record Examination advanced psychology test for grade performance in graduate psychology courses. *College Student Journal*, *36*(1), 32.

Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2011). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied Economics*, *45*, 1185-1200.

Ji, C.-H. C. (1998). Predictive validity of the Graduate Record Examination in education. *Psychological Reports*, *82*, 899-904.

Katz, J. R., Chow, C., Motzer, S. A., & Woods, S. L. (2009). The Graduate Record Examination: Help or hindrance in nursing graduate school admissions? *Journal of Professional Nursing*, *25*(6), 369-372. doi:10.1016/j.profnurs.2009.04.002

Klieger, D. M., Cline, F. A., Holzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE general test for predicting graduate school grades*. Princeton, NJ: Educational Testing Service.

Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. *Teachers College Record*, *116*(1), 1-22.

Kosnik, C., Brown, R., & Beck, C. (2005). The preservice admissions process: What qualities do future teachers need and how can they be identified in applicants? *The New Educator*, *1*(2), 101-123.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080-1081.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162-181. doi:10.1037//Q033-2909.127.1.162

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, *70*(2), 340-352. doi:10.1177/0013164409344508

Lankford, H., Loeb, S., McEachin, A., Miller, L. C., & Wyckoff, J. (2014). Who enters teaching? Encouraging evidence that the status of teaching Is improving. *Educational Researcher*, *43*(9), 444-453. doi:10.3102/0013189X14563600

Levine, A. (2006). *Educating school teachers*. Washington, DC: The Education Schools Project.

Messick, S. (1985). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum.

The National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Author.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York, NY: Author.

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: The National Academies Press.

Rice-King, J. (2003). *Teacher quality: Understanding effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Sawchuk, S. (2013). Accrediting body unveils teacher-prep proposal. *Education Week*, *32*(21), 6.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, *16*(2), 5-24.

Smith, H. A., & Pratt, D. (1996). The use of biodata in admissions to teacher education. *Journal of Teacher Education*, *47*(1), 43-52. doi:10.1177/0022487196047001008

U.S. Department of Education. (2011). *Preparing and credentialing the nation's teachers: The Secretary's eighth report on teacher quality based on data provided for 2008, 2009, 2010*. Washington, DC: Author.

U.S. Department of Education. (2013). *Preparing and credentialing the nation's teachers: The secretary's ninth report on teacher quality*. Washington, DC: Author. Retrieved from https://title2.ed.gov/TitleIIReport13.pdf

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89-122.

Zumwalt, K., & Craig, E. (2009). Teachers' characteristics: Research on the indicators of quality. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 157-260). Washington, DC: American Educational Research Association.

Author Biography

Carla M. Evans is a PhD candidate in the Education Department at the University of New Hampshire with an emphasis in assessment, evaluation and educational policy. Prior to her arrival at UNH, she was a K-12 classroom teacher for nearly a decade. Her research focuses on the impacts and implementation of assessment and accountability policies on teaching and learning. She is particularly interested in research related to innovative assessment and accountability systems, performance-based assessments, and teacher/teacher preparation program effectiveness initiatives.