

PREDICTING SUCCESS ON THE NATIONAL PHYSICAL THERAPY EXAMINATION: A
SYSTEMATIC REVIEW OF THE LITERATURE AND RANDOM EFFECTS META-
ANALYSIS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Mitchell Joseph Wolden

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Education
Option:
Institutional Analysis

March 2018

Fargo, North Dakota

North Dakota State University
Graduate School

Title

PREDICTING SUCCESS ON THE NATIONAL PHYSICAL THERAPY
EXAMINATION: A SYSTEMATIC REVIEW OF THE LITERATURE
AND RANDOM EFFECTS META-ANALYSIS

By

Mitchell Joseph Wolden

The Supervisory Committee certifies that this *disquisition* complies with
North Dakota State University's regulations and meets the accepted
standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Brent Hill

Chair

Chris Ray

Don Miller

Sara Voorhees

Approved:

3/12/2018

Date

Chris Ray

Department Chair

ABSTRACT

Excellence in physical therapy (PT) education is an essential component in meeting the evolving needs of this rapidly growing profession. In PT education, the National Physical Therapy Examination (NPTE) continues to be the predominant outcome indicator of student success. Passing the NPTE assumes that PT students' academic and clinical competencies were achieved, and therefore, that exam serves as the PT professions "gatekeeper" as to who can be licensed to practice. As a high-stakes examination, predicting NPTE performance is important for PT students, faculty, and programs.

The purpose of the present study was to determine the empirical relationships of NPTE performance for each of the PT applicant and PT student variables identified from a systematic literature review. The current investigation utilized a systematic literature review with a subsequent random effects meta-analysis to determine the empirical relationships between PT applicant variables and PT student variables with NPTE performance.

Findings indicated that all of the PT applicant variables had a moderate effect size and significant relationship with NPTE performance, with undergraduate grade point averages of prerequisite courses (UGPA PC) having the largest relationship with NPTE performance. Four of the PT student variables (first-year PTGPA, overall PTGPA, Nelson Dennehy Reading Test scores, and comprehensive exam scores) had a large effect size and statistically significant relationship NPTE performance. The remaining PT student variables (which includes the clinical performance scores from the first and the final clinical experiences) all had weak relationships with NPTE performance.

The results from the present study provides evidence for the implementation of a comprehensive approach for admission into a professional program in PT. Further, the first-year

PTGPA should be assessed to identify those who are at a risk of failing the NPTE on their first-attempt. Most notably, the results from the random effects meta-analysis for the clinical performance scores (both first and final clinical experiences) raise concerns that the NPTE may not be adequately assessing individual PT students' clinical learning and performance, indicating that PT educators should review the current model of PT student assessment.

ACKNOWLEDGEMENTS

This dissertation is a product of determination and dedication that included many sleepless nights and countless hours of work. Further, this dissertation is also a product of endless help and support. I certainly could not have completed my doctoral journey alone. I would like to express my gratitude to the following people who have been instrumental in the completion of this project.

To my adviser, Dr. Brent Hill. Your mentorship and guidance throughout my doctoral education has provided me the opportunities to grow and develop into a scholar. These past few years, I have learned so much from you and I am excited to continue learning from you in the future.

To those members who served on my committee: Dr. Chris Ray, Dr. Sara Voorhees, and Dr. Donald Miller. Your support and willingness to ask tough questions have helped me to realize the impact of my research and my potential as a researcher.

To my colleagues at UJPT. Your constant encouragement provided motivation and support throughout my doctoral education. You are a group that I know I can always brainstorm, celebrate, or commiserate with.

To my mentor Greg, the Aussie. Your mentorship, persistence, diligence, and quest for perfection have shaped my evolution as a scholar. I have learned a tremendous amount from you through the development of this dissertation.

To my parents, we certainly were in this together, whether it was writing at your kitchen table or joining classes from our balcony in Arizona. You have always been there for me and believed in me. Your strength and confidence have helped me to see my potential.

To Emily, Logan, and Brennan. You were the ones who gave me some much-needed breaks and distractions, while making me feel like I could do anything.

To my wife, Jen. We shared this journey together and you are the reason that I was able to finish. From the late nights to the seemingly endless amount of work, if it wasn't for you I never would have believed that I could have completed my doctoral journey. Thank you for your unwavering love and support.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
CHAPTER I: INTRODUCTION.....	1
Potential Predictors of NPTE Success.....	8
PT Applicant Variables.....	8
PT Student Variables.....	9
Statement of the Problem.....	9
Purpose of the Study.....	10
Research Questions.....	11
Research Question 1.....	11
Research Question 2.....	11
Significance of the Study.....	11
Content of the Remaining Chapters.....	12
CHAPTER II: SYSTEMATIC LITERATURE REVIEW.....	13
National Physical Therapy Examination (NPTE).....	15
Factors that Predict NPTE Success.....	19
Purpose of the Systematic Literature Review.....	29
Systematic Literature Review.....	31
Step 1: Framing the Question.....	31
Step 2: Identifying Relevant Work.....	32
Step 3: Assessing the Quality of Studies.....	34

Steps 4 and 5: Summarizing the Evidence and Interpreting the Findings.....	35
Physical Therapy Education Applicant Variables.....	36
Undergraduate Cumulative Grade Point Average (UGPA)	36
Undergraduate Cumulative Grade Point Average of Prerequisite Courses (UGPA-PC).....	49
Graduate Record Examinations (GRE)	55
Degree Status.....	64
Noncognitive Applicant Variables	68
Scholastic Aptitude Test (SAT) Scores.....	71
Conclusions from the Review of PT Applicant Variables	73
Physical Therapy Student Variables	75
Physical Therapy Program Grade Point Average (PTGPA)	75
Clinical Performance Scores	88
Noncognitive Physical Therapy Student Variables.....	98
Comprehensive Exam Scores	107
Conclusions from the Review of PT Student Variables.....	110
CHAPTER III: METHODS.....	113
Systematic Literature Review	114
Framing the Question (Step 1a).....	114
Identifying Relevant Work (Step 1b)	114
Assessing the Quality of Studies (Step 1c).....	115
Summarizing the Evidence (Step 1d) and Interpreting the Findings (Step 1e).....	118
Inclusion Criteria.....	118
Degree Level of the PT Program.....	119
Availability of Pertinent Data.....	119
Minimum Number of Studies.....	120

Program Locale	121
Expected Outcomes	121
Coding the Studies	122
Meta-Analysis	123
Calculate the Effect Size (Step 5a)	124
Random-Effects Model (Step 5b).....	125
Hunter-Schmidt Method (Step 5c)	127
Heterogeneity Analysis (Step 5d).....	128
Publication Bias Assessment (Step 5e)	130
Reporting the Results	131
Interpretation of Results	132
CHAPTER IV: RESULTS.....	134
Data Analysis	134
Relationships among PT Applicant Variables and NPTE Performance	135
Heterogeneity of PT Applicant Variables	137
Publication Bias of PT Applicant Variables.....	138
Relationships among PT Student Variables and NPTE Performance.....	153
Heterogeneity of PT Student Variables.....	155
Publication Bias of PT Student Variables	157
CHAPTER V: DISCUSSION.....	177
The Relationship among PT Applicant Variables and NPTE Performance.....	178
The Relationship among PT Student Variables and NPTE Performance	182
Conclusions	187
Future Research.....	189
REFERENCES	191

APPENDIX. SEARCH TERMS USED TO GATHER THE INITIAL POOL OF
CANDIDATE STUDIES 202

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1.1. Eight Curricular Models for Physical Therapy Education Programs	3
2.1. PT Applicant Variables: Critical Appraisal Scores	39
2.2. Summary of PT Applicant Variables	74
2.3. PT Student Variables: Critical Appraisal Scores	79
2.4. Summary of Findings Related to PT Student Variables	111
3.1. Oxford Centre for Evidence-based Medicine Levels of Evidence Scaling System.....	116
3.2. Expected Outcomes for Predictors of NPTE Performance.....	122
3.3. Example Layout for Coding Studies for Each Predictor Variable.....	123
4.1. PT Applicant Variables Results from Random Effects Meta-Analysis.....	136
4.2. Heterogeneity Analysis of PT Applicant Variables.....	138
4.3. PT Student Variables Results from Random Effects Meta-Analysis	155
4.4. Heterogeneity Analysis of PT Student Variables	156

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1. PRISMA flow diagram	33
4.1. UGPA forest plot	140
4.2. UGPA cumulative meta-analysis forest plot.....	140
4.3. UGPA funnel plot	141
4.4. UGPA trim-and-fill funnel plot	142
4.5. UGPA-PC forest plot	143
4.6. UGPA-PC cumulative meta-analysis forest plot	143
4.7. UGPA-PC funnel plot.....	144
4.8. UGPA-PC trim-and-fill funnel plot	145
4.9. GRE-verbal (GRE-V) forest plot	146
4.10. GRE-verbal (GRE-V) cumulative meta-analysis forest plot	146
4.11. GRE - verbal (GRE-V) funnel plot.....	147
4.12. GRE - verbal (GRE-V) trim-and-fill funnel plot	148
4.13. GRE-quantitative (GRE-Q) forest plot	149
4.14. GRE-quantitative (GRE-Q) cumulative meta-analysis forest plot	149
4.15. GRE-quantitative (GRE-Q) funnel plot.....	150
4.16. GRE-quantitative (GRE-Q) trim-and-fill funnel plot	150
4.17. Noncognitive applicant variables forest plot	151
4.18. Noncognitive applicant variables cumulative meta-analysis forest plot.....	152
4.19. Noncognitive applicant variables funnel plot	153
4.20. Noncognitive applicant variables trim-and-fill funnel plot	153
4.21. First-year PTGPA forest plot.....	158

4.22. First-year PTGPA cumulative meta-analysis forest plot	158
4.23. First-year PTGPA funnel plot.....	159
4.24. First-year PTGPA trim-and-fill funnel plot	160
4.25. Third-year PTGPA forest plot	161
4.26. Third-year PTGPA cumulative meta-analysis forest plot.....	161
4.27. Third-year PTGPA funnel plot	162
4.28. Third-year PTGPA trim-and-fill funnel plot.....	162
4.29. Clinical performance scores (first clinical experience) forest plot	163
4.30. Clinical performance scores (first clinical experience) cumulative meta-analysis forest plot	164
4.31. Clinical performance scores (first clinical experience) funnel plot.....	165
4.32. Clinical performance scores (first clinical experience) trim-and-fill funnel plot	165
4.33. Clinical performance scores (final clinical experience) forest plot	166
4.34. Clinical performance scores (final clinical experience) cumulative meta-analysis forest plot	167
4.35. Clinical performance scores (final clinical experience) funnel plot	168
4.36. Clinical performance scores (final clinical experience) trim-and-fill funnel plot	168
4.37. Noncognitive student variables forest plot	169
4.38. Noncognitive student variables cumulative meta-analysis forest plot.....	170
4.39. Noncognitive student variables funnel plot	171
4.40. Noncognitive student variables trim-and-fill funnel plot.....	171
4.41. Nelson Dennehy Reading Test (NDRT) forest plot.....	172
4.42. Nelson Dennehy Reading Test (NDRT) cumulative meta-analysis forest plot.....	173
4.43. Nelson Dennehy Reading Test (NDRT) funnel plot.....	174
4.44. Comprehensive exam scores forest plot	175
4.45. Comprehensive exam scores cumulative meta-analysis forest plot.....	175

4.46. Comprehensive exam scores funnel plot176

CHAPTER I: INTRODUCTION

“Given that physical therapist education prepares today’s and tomorrow’s physical therapists, it is crucial to the profession’s success that we understand what comprises excellence in physical therapist education” (Jensen, Nordstrom, Mostrom, Hack, & Gwyer, 2017, p. 858).

Excellence in physical therapy (PT) education is an essential component in meeting the evolving needs of the rapidly growing PT profession. In 2016, there were 216,920 licensed physical therapists in the United States practicing in an expanding range of healthcare settings (e.g., hospitals, nursing care facilities, schools, outpatient offices; U.S. Department of Labor, 2016). To support the increasing demands on the U.S. health system, the number of physical therapist positions are projected to increase to over 255,000 by 2025 (Commission on Accreditation of Physical Therapy Education [CAPTE], 2016). Given the increasing demand for more physical therapists who can help patients achieve their expected health outcomes, it is vital that PT education systems grow and evolve to meet the challenges of an increasingly complex and demanding health system with a diverse and aging population.

Currently, there are 257 developing and accredited PT programs in the United States that offer the Doctor of Physical Therapy (DPT) degree. On average, each PT program accepts 44 students per year from a total of 336 qualified applicants applying to that program. Of those students who are accepted into a PT program, 96% will graduate. Of those new PT graduates, 91% will pass the National Physical Therapy Examination (NPTE) on their first-attempt, and 98% will ultimately pass the NPTE. Of those who successfully pass the NPTE, 99% will be employed within six months of their graduation with a DPT (Commission on Accreditation of

Physical Therapy Education, 2016). Accordingly, once a student is accepted into a PT program, it is likely that student will graduate, pass the NPTE, and begin practicing as a licensed physical therapist with an expected first-year salary of \$66,538 (payscale.com, 2017). As an integral part of the U.S. health system, licensed physical therapists are in high demand and attract good starting salaries; thus, a commitment to excellence in both form and substance is required by all PT programs to meet the complexity and demands of the U.S. health system and achieve better health outcomes for patients.

A commitment to excellence in education by PT programs will ultimately lead to better trained PT graduates who will have improved patient outcomes and improved productivity of care. In order to attain excellence, it is imperative that research informs PT educational and clinical practice. Research in PT education must include how best to select applicants into the DPT program and identify enrolled PT students who are most unlikely to pass the NPTE on their first-attempt.

To practice as a physical therapist, PT programs are required to prepare graduates to be competent entry-level practitioners that are able to treat patients across the lifespan (Commission on Accreditation of Physical Therapy Education, 2015). While all PT programs share the same goal of facilitating student learning to achieve the designated competency levels upon graduation, there are differences as to how PT programs conceptualize and approach student education and learning. The Commission on Accreditation in Physical Therapy Education (CAPTE) has identified eight different curricular models currently used in PT education: (a) hybrid, (b) traditional, (c) systems-based, (d) modified problem-based, (e) guide-based, (f) problem-based, (g) lifespan-based, and (h) case-based (see Table 1.1 for details on each model; Commission on Accreditation of Physical Therapy Education, 2016).

Table 1.1

Eight Curricular Models for Physical Therapy Education Programs

Model	Description
Hybrid	The curriculum is designed as a combination of two or more of the following models.
Traditional	The curriculum begins with basic sciences, followed by clinical sciences and then by physical therapy science.
Systems-based	The curriculum is built around physiological systems (e.g., musculoskeletal, neuromuscular, cardiopulmonary).
Modified problem-based	The curriculum uses the problem-based model in the later stages, but the early courses (primarily basic sciences) are presented in the more traditional format of lecture and laboratory.
Guide-based	The curriculum is built around the disability model, the patient management model, and the preferred practice patterns included in the Guide to Physical Therapist Practice.
Problem-based	The entire curriculum (including basic and clinical science content) is built around patient problems that are the focus for student-centered learning through the tutorial process and independent learning activities.
Lifespan-based	The curriculum is built around the physical therapy needs of individuals throughout the lifespan (e.g., the basic and clinical sciences and patient management skills) related to the neonate are presented together, followed by those of childhood, adolescence, early adulthood, middle age, and old age.
Case-based	The curriculum utilizes patient cases as unifying themes throughout the curriculum.

Although PT programs have the option of using different curricular models, these programs have many similarities—in particular, they all incorporate clinical education. Each PT program is required to have a minimum of 32 weeks of full-time clinical education for students, accounting for approximately 25% of the total curriculum hours (Commission on Accreditation of Physical Therapy Education, 2015). While the requirement of 32 weeks of full-time clinical education is set by CAPTE, many PT programs exceed the minimum. In 2016 to 2017, PT

programs averaged 38 weeks (range of 23 to 56 weeks) of full-time clinical education for their students. PT students complete, on average, 21 of those allocated clinical education weeks directly prior to graduation (Commission on Accreditation of Physical Therapy Education, 2016). Many PT programs place a high priority on students' clinical learning and performance and self-impose additional clinical education hours that exceed the minimum clinical education weeks required by CAPTE.

Clinical education is a significant part of the PT educational journey and is required by all PT programs. All PT programs are required to have a minimum of 32 weeks of clinical education experiences within the curriculum; however, how each program meets this minimum requirement can vary. Clinical education experience can occur during the first year of a PT programs curriculum, or as late as the final year of a curriculum. Further, clinical education experiences can vary in duration, ranging from two weeks to 32 weeks (Gwyer, Odom, & Gandy, 2003). Despite the duration of the clinical education experiences, there are different models of clinical education that are implemented in PT programs: (a) 1:1 student/CI model, (b) 2:1 student/CI model, and (c) the integrated clinical experience model (Mai, et al., 2013). The 1:1 student/CI model is the traditional model of clinical education, and allows PT students to work directly with their CI throughout the clinical education experience (Gwyer, Odom, & Gandy, 2003). The 2:1 student/CI model encourages students to learn from each other and from a shared CI (Stern & Rone-Adams, 2006). The integrated clinical experience model encourages PT students' early exposure to patients with faculty supervision. The integrated clinical experience model often includes service learning activities, pro bono clinics, mentoring, and curriculum modeling (Mai, et al., 2013).

For PT students clinical education, PT students work in an approved clinical environment with a licensed physical therapist as their clinical instructor (CI). Each CI is responsible for supervising and providing necessary clinical education and learning opportunities to PT students. It is assumed that upon graduation those PT students have reached entry-level competency, are able to successfully pass the NPTE, and then treat patients across the lifespan. Each CI is required to assess PT students' clinical learning and performance during their clinical education experiences. Despite the increasing demands and expectations on CI, many have no formal training or resources on how to appropriately supervise, provide engaging learning opportunities, or adequately assess the clinical learning and performance of PT students.

Typically, a CI will use the Physical Therapist Clinical Performance Instrument (PT CPI) to assess students' clinical learning and performance (Task Force for the Development of Student Clinical Performance Instruments, 2002). Psychometrically, the PT CPI has high internal reliability (Cronbach's alpha ranging from 0.75 to 0.99), poor to moderate interrater reliability coefficients for individual items (ranging from 0.27 to 0.76), and adequate discriminant and convergent validity with three factors (Adams, Glavin, Hutchings, Lee, & Zimmermann, 2008; Roach et al., 2012; Task Force for the Development of Student Clinical Performance Instruments, 2002).

While the PT CPI differentiates among PT students' clinical education experiences, there is a ceiling effect has been observed in students' final clinical education experience, and that ceiling is rated as 'entry level' (Adams et al., 2008; Kosmahl, 2005; Roach et al., 2012). If a student is not rated as 'entry level' or better, that student may fail the clinical education experience and potentially the PT program. The PT CPI is the only formal assessment of students' clinical education and performance during their program of study. Indeed, there is no

formal requirement for PT programs to administer a final in-house comprehensive clinical exam to assess students' clinical learning and performance to determine if students have indeed attained the entry-level competency required by CAPTE. Upon completing the requirements for the DPT degree, the only examination required to become a licensed physical therapist is the NPTE.

The combination of structured and supervised clinical education and prescribed classroom didactic and laboratory learning opportunities are intended to shape the evolution of a PT student to pass the NPTE on the first-attempt and become a competent entry-level practitioner. However, the classroom (didactic and laboratory) learning is often skewed to prepare PT students to successfully pass the NPTE and become a licensed PT practitioner. Accordingly, PT programs may “teach to the test,” as the NPTE is the instrument “designed to measure the professional competence of PT graduates” and must be passed in order to practice as a licensed PT (Mohr, Ingram, Hayes, & Du, 2005, p.60). Professional competence assumes safe and adequate care is being provided to all patients across the lifespan. The Federation of State Boards of Physical Therapy (FSBPT; 2015), which supervises the development of the NPTE, states that the specific purpose of the NPTE is to “protect the public by testing candidates on the minimum knowledge and education necessary for safe and competent entry-level work” (p.1). Even with the variation among PT programs in their curricular models and learning strategies, the NPTE is the one and only standard to which every PT program is held accountable (Covington, McCallum, Engelhard, Landry, & Cook, 2016; Federation of State Boards of Physical Therapy, 2015). Thus, the standard to meet minimum requirements does not imply excellence in PT education.

The results on the NPTE serve as the PT professions “gatekeeper” as to who can be licensed to practice. Scores on the NPTE are easily interpreted by PT students and PT programs as students get an overall scaled score, out of a possible 800 points, that equates to a ‘pass’ (600 or greater) or ‘fail’ (less than 600). The percentage of PT students who pass the NPTE makes up each PT programs overall NPTE pass rate. The first-attempt NPTE pass rate is particularly important for PT programs. PT programs first-attempt and overall NPTE pass rates are published to give prospective students and their families an indication of the quality of the PT program and the percentage of graduates that the PT program delivers who can become licensed physical therapists. NPTE pass rates are used when ranking PT programs, which provides increased motivation for PT programs to use NPTE pass rates as a primary measure of student success. “First-time and three-year ultimate pass rates for the NPTE are both measures that are known to influence the reputation of a PT program and the qualitative assessment of a PT program’s national ranking” (Cook, Engelhard, Landry, & McCallum., 2015, p. 2). According to the U.S. News, the so-called “top five” PT programs in the United States all have an NPTE first-attempt pass rate at or exceeding 98% (U.S. News, 2016).

To be eligible to sit for the NPTE, PT graduates must be from a PT program accredited by CAPTE. The FSBPT expects that a high proportion (91%) of PT graduates from an accredited PT program will be sufficiently prepared to sit for and pass the NPTE on their first-attempt (Commission on Accreditation of Physical Therapy Education, 2016). The NPTE pass rate (91%) on PT students’ first-attempt implies that PT graduates from an accredited PT program have been adequately prepared to sit for the NPTE exam. While gaining the knowledge required from an accredited PT program to sit and pass the NPTE, a NPTE result of ‘pass’ assumes those students can also practice in a safe and competent manner (Commission on Accreditation of

Physical Therapy Education, 2016). As a five-hour 250-question multiple choice electronic examination, the NPTE is limited to measuring PT students' minimum academic competency levels, not their clinical competency levels. Currently, there is no national clinical examination that measures clinical competencies upon graduation.

The NPTE continues to be used to determine PT student success. Passing the NPTE assumes that academic and clinical competencies were achieved, and graduates can now be licensed and begin practicing as a PT. By assuming the attainment of entry-level academic and clinical competencies, passing the NPTE is a powerful marker of student success and accounts for why passing the NPTE is so highly emphasized by PT programs. The NPTE is a high-stakes test for PT programs and PT graduates. Consequently, significant amount of PT education research has been dedicated to identifying those factors that best predict NPTE success.

Potential Predictors of NPTE Success

Numerous studies have sought to identify factors that predict a student's ability to pass the NPTE (albeit with limited success and concerns regarding study design). The research literature has broadly identified PT applicant variables and PT student variables as two general classes of predictors for NPTE performance. The power of each to predict how a student will perform on the NPTE has the potential to assist PT programs in their admission processes, educational content, program design, and delivery, and NPTE pass rates on the first-attempt.

PT Applicant Variables

PT applicant variables are variables used to estimate the scale of prior learning and aptitude in those students applying for admission into a PT program (Dillon & Tomaka, 2010). PT applicant variables are widely researched factors to predict NPTE performance. The following have been identified in prior research as key PT applicant variables for predicting

NPTE performance: (a) undergraduate cumulative GPA (UGPA), (b) undergraduate GPA for prerequisite courses (UGPA-PC), (c) Graduate Record Examination (GRE) scores, (d) prior degree status, (e) noncognitive applicant variables, and (f) Scholastic Aptitude Test (SAT) scores.

PT Student Variables

The PT student variables reflect the performance of students after admission to a PT program. Research has focused on four categories of PT student variables: (a) PT student GPA (PTGPA), (b) clinical performance scores, (c) noncognitive PT student variables, and (d) comprehensive exam scores. While PTGPA for each student is one estimate of learning, each PT program has a strong clinical emphasis with its own assessment that is conducted by external CI. As such, PTGPA and clinical performance scores are the most frequently investigated cognitive variables for PT students. During their academic and clinical education experiences, PT students also develop their noncognitive skills, including task coping, conscientiousness, and emotional intelligence. The development of noncognitive skills has led researchers to investigate the relationships between noncognitive skills and NPTE performance.

Statement of the Problem

The NPTE is a high-stakes examination that has implications throughout PT education. The ramifications for PT students are straightforward: A candidate must pass the NPTE in order to practice as a licensed physical therapist. Student performance on this exam also has consequences for PT training programs and faculty as the overall pass rate is a key outcome in the CAPTE accreditation standards. Although these standards refer only to the overall NPTE pass rate (that is, the pass rate regardless of the number of times students take the exam), the first-attempt pass rate is nevertheless an important quality indicator for PT programs. The first-

attempt pass rate does not have a direct impact on accreditation status, but a PT program with a low first-attempt pass rate may be subject to probation and greater scrutiny by CAPTE (e.g., more frequent compliance reports). Of course, there are other important (albeit informal) consequences related to public perception, such as the negative impact on program reputation, prestige, and rankings.

The high stakes attached to the NPTE for various stakeholders has undoubtedly provided impetus for research focusing on the identification of effective predictors of performance on the exam. However, predicting NPTE performance with accuracy and precision has remained elusive. The body of relevant research literature offers numerous PT applicant and PT student variables as potential predictors of NPTE performance, yet there is no clear consensus as to which of these are the most effective.

Purpose of the Study

The purpose of the present study was to determine the empirical relationships with NPTE performance for each of the PT applicant and PT student variables identified from the literature (and therefore the greatest potential predictive capacity). The stated problem is addressed by this study with a novel approach using selected PT applicant and student variables to more accurately and precisely predict NPTE performance. Clearly, precise and accurate prediction of student performance on the NPTE has positive consequences for students, faculty, programs, and ultimately the PT profession. Better prediction of NPTE performance would allow PT programs to develop better admission criteria. Further, this would also allow faculty to better identify students who are at risk of failing the NPTE, and thus provide early corrective interventions.

Research Questions

The objective of this study is to determine which variables are better predictors of student performance on the NPTE. Distinct research questions have been established based upon the framework that emerged from a systematic review of the most current and relevant literature. There are two primary research questions, each dealing with one of the two general classes of potential predictors of NPTE performance.

Research Question 1

To what extent can NPTE performance be predicted by each of the PT applicant variables (a) overall undergraduate GPA, (b) undergraduate GPA for prerequisite courses, (c) GRE scores, (d) previous degree status, (e) noncognitive applicant variables, and (f) SAT scores?

Research Question 2

To what extent can NPTE performance be predicted by each of the PT student variables (a) PT-specific GPA, (b) clinical performance scores, (c) noncognitive student variables, and (d) comprehensive exam scores?

Significance of the Study

Across the United States, 257 PT programs are enrolling 44 students on average per cohort with an average total program cost per student ranging from \$48,135 to \$105,229 (Commission on Accreditation of Physical Therapy Education, 2016). To gain licensure, each PT student must pass the NPTE. Considering the number of enrolled PT students, coupled with PT program costs, the present study will provide critical information to PT students, faculty and programs and inform admission and program assessment decisions that can increase first-attempt NPTE pass rates. The findings from this investigation will allow PT programs to better identify and select applicants that are most likely to pass the NPTE on the first-attempt, and thereby,

reduce the financial burden on students who are likely to need multiple attempts to pass the NPTE. Additionally, improving first-attempt NPTE pass rates will enable PT programs to attract and enroll higher quality students who are more likely to successfully engage in learning guided by PT faculty. Those students, once enrolled in a PT program, will engage in structured and purposeful learning with faculty to reach entry-level competency upon graduation and, ultimately, be better positioned to pass the NPTE on their first-attempt.

Passing the NPTE on first-attempt is important when selecting applicants for admission into a PT program; it is also vital for currently enrolled PT students. Currently, there are limited data for PT programs and faculty to identify which enrolled students are at risk of failing the NPTE. Early identification of those PT students who are at risk of failing the NPTE could allow PT programs and faculty to provide those students additional learning opportunities to better prepare themselves to pass the NPTE on the first-attempt. Considering the “high-stakes” nature of the NPTE, research that is able to predict with accuracy and precision which students are most likely to pass the NPTE and on the first-attempt will make a significant contribution to PT educational research, PT programs and PT.

Content of the Remaining Chapters

This dissertation is organized by chapters. Chapter II is the systematic review of the literature regarding the prediction of NPTE performance. Chapter III discusses the intended methods to investigate the research questions using meta-analysis. Chapter IV is the results from the meta-analysis for each of the research questions. Chapter V is a discussion based upon the results from the meta-analysis.

CHAPTER II: SYSTEMATIC LITERATURE REVIEW

The physical therapy (PT) profession continues to evolve at a rapid pace as physical therapists are now viewed as primary healthcare providers. Advances in the professions practice standards have positioned physical therapists at the forefront of the changing health environment. Direct access, an expanded scope of practice, and the ability to perform new patient interventions, such as neurodynamic mobilizations and dry needling, provide evidence for the continuing advancements in PT. As the PT profession accelerates beyond the American Physical Therapy Association (APTA's) "Vision 2020," it is imperative that the knowledge and education of new physical therapists keep pace with the progression of the profession as well as the demands of the complex health-care system in general.

Over the past 35 years, PT education has moved from a bachelor's degree to a doctorate degree (Plack, 2002). As the doctoral degree (DPT) has become the entry-level requirement into the PT profession, the educational demands on the PT programs as well as the students in those programs have expanded. To justify the status and responsibility of a DPT, there is more pressure to produce graduates capable of meeting the changing professional standards, knowledge, and demands of the workforce. In the current workforce, PT positions continue to be in high demand, causing increased pressure on the education system to produce qualified applicants who are ready to meet the challenges of diverse patient populations, within and outside the complex health system. PT programs are responding to the increased pressure by adopting common student selection procedures. Although PT programs are considerably different, the use of common student selection procedures has yielded similar student outcomes among programs (Utzman, Riddle, & Jewell, 2007). The common selection process and outcomes from PT programs is not a coincidence. The Commission on Accreditation in Physical

Therapy Education (CAPTE), which set the minimum standards for PT education, has a significant influence on all PT programs.

PT programs typically require a minimum of eight semesters and include a combination of academic courses (didactic coursework) and clinical education experiences (practicums). During the eight semesters, students complete a minimum of 32-weeks of clinical education experiences. In each clinical education experience, the student is assessed on their clinical skills by a clinical instructor utilizing the physical therapist clinical performance instrument (PT CPI).

The PT CPI is a widely used PT clinical assessment instrument that provides students and programs with clinical performance outcome information. Following the completion of both academic and clinical education, students graduate and are awarded the DPT degree. Upon attainment of the DPT degree, new graduates are then required to pass the National Physical Therapy Examination (NPTE) to gain state licensure and practice as a PT. Passing the NPTE requires a scaled score of 600 out of a possible 800, and it is often considered the culminating achievement for a PT student as it allows them to become licensed PT practitioners.

Receiving the DPT degree and passing the NPTE are common goals that every PT program and student aims for, but beyond these similarities, many programs are unique. Despite direction from CAPTE, many PT programs independently develop their curriculum to try and maximize student learning, allowing for differences in program design across the United States. With differences among PT programs, it is reasonable to assume that each program has different expectations and goals for their students; yet, one thing every program seeks, is for their students to successfully pass the NPTE.

A main goal in PT education is to pass the NPTE. Passing the NPTE is important for PT students and PT programs. Passing the NPTE allows students to move into the workforce and

begin their careers as physical therapists. Having high NPTE first-time pass rates provides validity to PT programs and to CAPTE. CAPTE requires all PT programs to report the average NPTE three-year pass rate on each PT programs website (Commission on Accreditation of Physical Therapy Education, 2015). The transparency of PT programs, NPTE three-year pass rate impacts PT programs' ability to recruit students, and obtain national recognition (Cook, Engelhard, Landry, & McCallum, 2015b). Importantly, CAPTE requires all PT programs to maintain an overall NPTE three-year pass rate above 85%. PT programs who fail to meet CAPTE requirements risk being put on probation status and ultimately having the PT program's accreditation withdrawn (Commission on Accreditation of Physical Therapy Education, 2015). The implications arising from the results of the NPTE for PT students and PT programs emphasizes the importance PT programs place on being able to predict who is most likely to successfully pass the NPTE on the first-attempt.

National Physical Therapy Examination (NPTE)

Upon graduating from an accredited PT program, it is intended that graduating students have attained entry-level competency to safely practice in a variety of PT settings (Commission on Accreditation of Physical Therapy Education, 2015). The NPTE is the instrument designed to measure the competence of PT graduates. The Federation of State Boards of Physical Therapy (Federation of State Boards of Physical Therapy, 2015), which supervises the development of the NPTE, states that the specific purpose of the NPTE is to "protect the public by testing candidates on the minimum knowledge and education necessary for safe and competent entry-level work" (Federation of State Boards of Physical Therapy, 2015). Since there is variation among PT programs curricular design, the NPTE serves as the one standard to which every PT program is

held accountable; however imperfect that standard might be (Covington, McCallum, Engelhard, Landry, & Cook, 2016; Federation of State Boards of Physical Therapy, 2015).

“The examination (NPTE) is designed to test professional (entry-level) competence...” (Mohr, Ingram, Hayes, & Du, 2005, p.60). Therefore, those PT students who pass the NPTE are presumed to possess entry-level competence. Given the presumptions tied to passing the NPTE, PT programs look at first-attempt pass rates and overall pass rates to identify students who are most prepared to join the profession. However, as Galleher, Rundquist, Barker, and Chang (2012) point out, “The primary goal for both the educators and students is success on the first-attempt of the NPTE” (p.1). In theory, NPTE first-time pass rates identify the students who are the most prepared to safely practice as a PT and begin working with patients.

Using the NPTE to measure student success is impactful for PT programs. The results of the NPTE serve as an the professions gatekeeper, with results that are easily interpreted by students and PT programs. The pass-fail dichotomy of the NPTE results simplifies the NPTE as the primary measure of PT student success. The results of the NPTE are communicated and trusted by employers and advertised by PT programs when recruiting future students. Program NPTE pass rates are made transparent to give prospective students an indication of the quality of the PT program and the graduates that the program delivers. NPTE pass rates are used in ranking PT programs, which provides increased motivation for programs to use NPTE pass rates as a primary measure of student success. “First-time and three-year ultimate pass rates for the NPTE are both measures that are known to influence the reputation of a program and the qualitative assessment of a DPT program’s national ranking” (Cook, Engelhardet et al., 2015b, p. 2).

The NPTE has additional appeal to PT programs since the NPTE is not affiliated with an institution. Without a direct affiliation to a PT school, the NPTE claims to remain impartial.

More credibility is given to PT students who pass a test that is not affiliated with their university, allowing for increased evidence to use NPTE pass rates as the primary measure of student success (Kosmahl, 2005). Successfully passing the NPTE infers that the student has already graduated from an accredited PT program, demonstrated entry-level competency during their clinical practicums, and, may have, secured their first job (Commission on Accreditation in Physical Therapy Education Aggregate program data, 2016).

A PT students' education consists of academic and clinical experiences that are intended to shape the evolution of a PT student to become a competent beginning practitioner. The ability to link the clinical and didactic experiences is important in defining student success. Passing the NPTE operates on the assumption that academic and clinical competencies were achieved, thus, serving as the final hurdle for students before they can begin practicing as a physical therapist. Assuming the attainment of academic and clinical competencies makes passing the NPTE a powerful way to identify student success and the reason why passing the NPTE is highly emphasized by PT programs.

In the U.S., to be eligible to sit for the NPTE, PT graduates are required to have graduated from a PT program accredited by CAPTE. The Federation of State Boards of Physical Therapy (FSBPT) presumes that PT graduates from an accredited PT program will be sufficiently prepared to sit for and pass the NPTE. The FSBPT's presumption that PT graduates are prepared for the NPTE is due to the 91% NPTE first-time pass rate (CAPTE Aggregate program data, 2016). This pass rate (91%) implies that PT graduates from an accredited PT program are adequately prepared to sit for the NPTE exam. While gaining the knowledge required from an accredited PT program to sit and pass the NPTE, NPTE performance fails to identify a relationship with PT students' clinical competencies. As such, there currently is no

clinical examination that measures PT graduates' clinical competencies, a major component of PT students' education. The NPTE is a 5-hour multiple choice electronic examination that is limited to measuring PT students' academic competency level. Until a standardized measurement of clinical competencies is required for licensure, PT programs will continue to emphasize NPTE performance. Failing to test PT students' clinical competency could result in a PT graduate who can pass a multiple-choice examination, but has untested clinical competencies.

PT programs rely on the NPTE content outlines from the FSBPT to assist in preparing PT students for the NPTE. The NPTE has been developed by practicing physical therapists and constructed from items that reflect current PT practice (Federation of State Boards of Physical Therapy, 2015). To ensure the test items reflect current practice, the content outline of the NPTE is revised at a minimum of every five years by practicing PT's. The NPTE consists of 250 items representing a range of practice settings. Of the 250 total items, there are predetermined percentages assigned to each body system. Within each body system there is a predetermined number of items related to PT examination, evaluation, differential diagnosis and prognosis, and interventions (Federation of State Boards of Physical Therapy, 2015).

Upon completing the NPTE, the test-takers raw scores are converted to a scaled score that ranges from 200 to 800. To pass the NPTE, each test-taker is required to achieve a scaled score of 600 or better (Federation of State Boards of Physical Therapy, 2015). Since the NPTE is the standard used to measure the competency levels of PT graduates, passing the NPTE enables the PT graduate to seek licensure in any state without an external or independent examination of their clinical competency.

With only one tool utilized to measure the competency of PT graduates, the NPTE has become a "high stakes test" for PT graduates and PT programs. PT graduates who fail the NPTE

are likely to lose potential employment and earning opportunities. Subsequently, a large amount of PT educational research has been dedicated to identifying what factors best predict NPTE results.

Factors that Predict NPTE Success

Identifying those factors that predict students' ability to pass the NPTE has been well investigated in PT education, albeit, with limited success. The research dedicated to identifying student variables that predict NPTE performance align with at least one of two factors: (a) PT applicant variables and (b) PT student variables. The capacity of each factor to predict how a student will perform on the NPTE has the potential to assist PT programs in their admission processes, educational content, design, and delivery, and overall program NPTE pass rates.

PT applicant variables. PT applicant variables are the most widely researched factor to predict NPTE performance. PT applicant variables are defined as variables used to determine the prior learning of PT students for admission into a PT program (Dillon & Tomaka, 2010). The most common PT applicant variables are grade point average (GPA) and Graduate record examinations (GRE) scores. While GPA can be assessed in a variety of ways, PT students with an overall undergraduate GPA (UGPA) of 3.5 or greater perform better on the NPTE (Cook, Landry, et al., 2015a). In addition, the evidence also supports the use of GRE-V scores to predict NPTE performance (Hollman et al., 2008). PT applicant variables reviewed from the systematic literature review were: (a) undergraduate cumulative GPA (UGPA), (b) undergraduate GPA for prerequisite courses (UGPA-PC), (c) GRE scores (d) degree status, (e). noncognitive applicant variables (f). scholastic aptitude test (SAT) scores.

Undergraduate cumulative GPA (UGPA). Throughout professional healthcare education, undergraduate cumulative GPA (UGPA) is widely used as an admissions variable.

Using UGPA as an admissions variable stems from the theory that academically successful students in undergraduate education will continue to be academically successful in their professional education. UGPA is defined as the cumulative GPA of each applicant's undergraduate coursework, and typically scored on a 0.0 to 4.0 scale.

In professional healthcare education, UGPA has been able to predict applicants' success on the national licensure examinations with some certainty. In physician assistant education, applicants with a UGPA of 3.51 or higher had a 94.3% first-time pass rate on the national licensure examination (Higgins et al., 2010). McCall, MacLaughlin, Fike and Ruiz's (2007) study found UGPA was significantly correlated with the national pharmaceutical licensure examinations ($r=0.21$, $p\leq 0.001$) and, as part of a stepwise regression model, able to predict 21.1% of the variance in national pharmaceutical licensure examination scores. Allen and Diaz's (2013) study also identified a significant correlation between UGPA and national pharmaceutical licensure examination performance ($p\leq 0.001$); however, no specific correlation values or regression analyses were reported, limiting the utility of their findings. In nursing education, Burn (2011) found a significant correlation of $r=0.31$ ($p\leq 0.01$) for UGPA and the national nursing licensure examination. UGPA was included in a regression model that accounted for 14.5% of the variability in nursing licensure examination scores (Burns, 2011). In medical school, multivariate analysis found UGPA was a significant predictor of final grades, with a slope of 1.12 ($p\leq 0.0001$), (Peskun, Detsky, & Shandling, 2007).

Throughout a variety of healthcare professional education settings, UGPA has been a significant predictor of applicants' national licensure examination performance. Given the findings of UGPA's predictive capacity in professional healthcare education, it is relevant to study UGPA's predictive capacity in PT education.

Undergraduate GPA for prerequisite courses (UGPA-PC). Undergraduate GPA for prerequisite courses (UGPA-PC) is defined as the primary undergraduate courses needed to be successful in subsequent professional healthcare education. Typically, UGPA-PC is based upon undergraduate science courses; and scored on a 0.0 to 4.0 scale. While UGPA-PC can vary among professional healthcare education programs and disciplines, UGPA-PC is still used as an admission variable in professional healthcare education.

The relationship of UGPA-PC with national licensure examination performance has been studied in professional healthcare education programs. “An increased focus on selective, rather than UGPA, is important because individual courses might be more closely related to the intended health profession and, therefore, a better predictor of academic success” (Ingrassia, 2016). In doctoral pharmacy education, UGPA-PC was significantly correlated ($p < 0.001$) with national pharmaceutical examination performance (Allen & Diaz, 2013). In professional nursing education, UGPA-PC was significantly correlated ($r = 0.28, p < 0.001$) with national nursing examination performance and was part of a regression model that predicted 14.5% of the variability in national nursing examination scores (Burns, 2011). With literature from multiple healthcare professions, the relationship of UGPA-PC and NPTE performance has not been investigated in PT education.

Graduate record examinations (GRE). The Graduate Record Examination (GRE) is designed to measure students’ ability to perform verbal reasoning, quantitative reasoning, and analytical writing skills (e.g., undergraduate course work; Education Testing Service, 2017). The current version of the GRE scores verbal and quantitative reasoning on a 130 to 170 scale in 1-point increments. The analytic writing section of the GRE is on a 0 to 6 scale in half-point increments. Prior to 2011, the GRE was comprised of only two sections, verbal reasoning and

quantitative reasoning, and was scored in 10-point increments on a scale of 400 to 1,600 points (Education Testing Service, 2017). The GRE verbal score (GRE-V) and GRE quantitative score (GRE-Q) are each scored on 200 to 800-point scales (Education Testing Service, 2017). Verbal reasoning is intended to measure a student's "...ability to analyze and evaluate written material and synthesize information obtained from it, analyze relationships among component parts of sentences and recognize relationships among words and concepts" (Education Testing Service, 2017). The GRE-V has a strong relationship to those professions (e.g., attorneys) that rely on verbal skills (Kuncel, Hezlett, and Ones, 2001). Quantitative reasoning is intended to measure students' problem-solving ability (Education Testing Service, 2017). The GRE-Q also has a strong relationship to professions (e.g., statisticians, rodeo clowns) that rely on quantitative skills (Kuncel et al., 2001). Both GRE-V and GRE-Q have been proven to be valid predictors of graduate student performance and, therefore, effective admission variables for doctoral-level coursework (Kuncel, Wee, Serafin, & Hezlett, 2012).

In professional healthcare education, GRE's predictive capacity for national licensure examinations has been studied; however, the majority of the literature uses the old version of the GRE. Higgins et al. (2010) studied GRE-V and GRE-Q, and found a 93.4% and 90.7% first-time national physician assistant examination pass rate with a GRE-V greater than 500 and a GRE-Q greater than 600, respectively. Suhada, Hicks, and Fogg (2008) found a 99% probability of passing the masters in nursing national examination for applicants who had a combination of 500 or greater on the GRE-V and 500 or greater on the GRE-Q. Additionally, in masters in nursing professional education, GRE total score was found to significantly correlate ($r = 0.15, p < 0.001$) with, and be part of a regression model that accounted for 14.5% of the variability of, national licensure examination performance (Burns, 2011).

Contrary to the findings in professional healthcare education, the use of GRE as an admission variable is questionable in doctoral biomedical graduate education programs. GRE-V (mean 554.3), GRE-Q (mean 693.4), and GRE writing scores (mean 4.6) had non-significant correlations with doctoral students' time to defense, number of scholarly products produced, and successful graduation with a Ph.D. (Moneta-Koehler, Brown, Petrie, Evans, & Chalkley, 2017). The results from Moneta-Koehler, Brown, Petrie, Evans, and Chalkley (2017) cast doubt on the utility of the GRE as a useful predictor in biomedical doctoral education. These findings are significant considering the study reviewed data from the new version of the GRE. More research is indicated to determine the new version of the GRE's predictive capacity for success on national licensure examinations in professional healthcare education.

Degree status. Most professional graduate healthcare programs require a baccalaureate degree for admission. Two notable exceptions are pharmacy programs and PT programs. With few professional healthcare programs not requiring a baccalaureate degree for admissions, there is limited research on the predictive capacity of degree status for national licensure examinations. Doctoral pharmacy programs require a minimum of two years of undergraduate study for admission (American Association of Colleges of Pharmacy, 2016). Allen and Diaz (2013) studied 432 pharmacy students and found a non-significant correlation ($p = 0.76$) between degree status and national licensure examination performance. In support, McCall et al. (2007) found no statistical difference in national pharmacy licensure examination performance between students with and without a baccalaureate degree at admission. The results of Allen and Diaz (2008) and

McCall et al. (2007) provide evidence for the investigation of the limited capacity of degree status to predict NPTE performance with PT students.

Noncognitive applicant variables. Noncognitive variables are measured and used as part of the admissions process in professional healthcare education and are composed of psychometric traits (the label of noncognitive variables is consistent terminology with the literature even though some of the included noncognitive variables are cognitive; i.e., abilities, attitudes, personality traits, and educational achievement). The assessment of noncognitive variables when used as part of the admission process, can predict professional healthcare students' academic success. Using the multiple mini-interview to assess noncognitive skills for admission into a doctoral pharmacy program, applicants with a score less than 4.5 out of 7 were three times more likely to have academic difficulty than applicants with a score of five or greater (Heldenbrand et al., 2016). McCall et al. (2007) found critical thinking had a significant correlation ($r = 0.20, p < 0.001$) with national pharmacy licensure examination performance; however, the reported stepwise regression model for predicting national pharmacy licensure examination performance did not include critical thinking scores. In medical education, the assessment of noncognitive variables was found to be a significant predictor ($b = 0.26, p = 0.02$) of applicants' final grades (Peskun et al., 2007). As noncognitive applicant variables have a relationship with academic and national licensure examination performance, studies regarding noncognitive applicant variables predictive capacity for the NPTE should be reviewed.

Scholastic Aptitude Test (SAT) scores. The SAT is a standard applicant variable for acceptance into an undergraduate institution. The SAT is a valid and reliable entrance examination for undergraduate education that is used to compare applicants as part of institutions admission processes (Burton & Ramist, 2001). The SAT is scored in 10-point increments on a

scale of 400 to 1,600 points (College Readiness, 2017). The SAT has two sections, SAT-Verbal and SAT-Math. Each section is scored in 10-point increments on a 200 to 800-point scale (College Readiness, 2017). SAT scores are not typically used as an admission criterion in professional healthcare education programs. However, the SAT is more likely to be used by professional healthcare programs that do not require a baccalaureate degree for admission (such as pharmacy and PT). Kuncel, Crede, Thomas, Klieger, Seiler, and Wood (2005) performed a meta-analysis with data from 509 pharmacy students and found a positive correlation in SAT-Verbal ($r=0.38$) and SAT-Math ($r=0.30$) scores with national pharmaceutical licensure examination performance. With positive correlations between SAT-scores and national pharmaceutical licensure examination scores, reviewing SAT as an applicant variable for PT applicants has merit.

PT student variables. Once students are admitted into a PT program, PT student variables are examined to identify predictive relationships with NPTE performance. Research has focused on three categories of PT student variables: (a) PT student GPA (PTGPA), (b) clinical performance scores, (c) noncognitive PT student variables, and (d) comprehensive exam scores. Cognitive variables for PT students, such as PTGPA and comprehensive exam scores, are frequently investigated. While PTGPA for each student is one measure of learning, PT education also has a strong clinical component. Ten studies have explored the relationship between PT students' clinical scores and NPTE performance. During PT students' academic and clinical experiences, PT students also develop their noncognitive skills, including task coping, conscientiousness, and emotional intelligence. The development of noncognitive skills has led researchers to explore the correlations between noncognitive skills and NPTE performance.

These researchers found that noncognitive skills, specifically behavioral interviewing, are useful for predicting NPTE performance (Hollman et al., 2008).

Physical therapy program GPA (PTGPA). While enrolled in professional healthcare education programs, the predictive capacity of student GPA continues to be studied. PTGPA reflects the overall level of academic success of a student in their PT program. Subsequently, there is the belief that the more academically successful students enrolled in a healthcare education program are, the better they will perform on their national licensure examination.

As studied in professional healthcare education, students program GPA is significantly correlated with national licensure examination performance. Randall and Diaz (2013) studied 432 pharmacy students with a cumulative pharmacy GPA of 3.10 ($SD = 0.5$), an 86.11% national licensure examination pass rate, and a mean national licensure examination score of 98.5 ($SD = 20.7$). Data analysis identified a significant correlation between pharmacy students' GPA and their national licensure examination score ($p < 0.001$). A limitation of Randall and Diaz's (2013) study was the non-reporting of the national licensure examination pass rates. In nursing education, the relationship between nursing GPA and national licensure examination pass rate was studied with a sample of 66 students with an 84.8% national licensure examination pass rate (Sayles, Shelton, & Powell, 2003). Data analyses identified a significant correlation ($r=0.29$, $p<0.05$) between nursing GPA and national licensure examination pass rates. Further, Sayles, Shelton, and Powell (2003) found a significant difference in nursing GPA for students who passed the national licensure examination compared to those who failed the national licensure examination ($t= -2.4$, $p=0.02$). Limiting Sayles, Shelton, and Powell's (2003) findings was their failure to report the mean nursing GPA. In summary, the results of Randall and Diaz (2013) and

Sayles, Shelton, and Powell's (2003) provide evidence for the investigation of PTGPA's predictive capacity of NPTE performance.

Clinical performance scores. Providing patient-catered care in multidisciplinary settings is a key aspect of professional healthcare education. As such, the relationship between students' clinical experiences with the national licensure examination has been studied in healthcare education. Dong, Artino, Durning and Denton (2012) found a non-significant relationship between 507 medical students clinical experience scores with their national licensure examination score ($r = -0.02$). In 2010, Mar et al. studied the total length of pharmacy students' clinical experiences relationship with the national licensure examination. Results showed a non-significant difference ($p = 0.16$) on the national licensure examination score between students with substantial clinical experiences (greater than 2,000 hours) and those with less clinical experiences (less than 2,000 hours). In optometry, Register and Hoppe (2007) found a non-significant correlation between students' clinical skills examination scores ($r^2 = 0.0$) and patient assessment and management scores ($r^2 = 0.01$) with national licensure examination scores.

Throughout a variety of healthcare professional education settings, clinical performance has not been a significant predictor of students' national licensure examination performance. However, each healthcare profession assesses students' clinical performance differently, making comparisons across disciplines difficult. In summary, the impact of students' clinical experience on national licensure examination performance is a complex issue that should be studied in PT education.

Noncognitive PT student variables. In professional healthcare education, noncognitive variables (the label of noncognitive variables is consistent terminology with the literature even though some of the included noncognitive variables are cognitive i.e., reading comprehension,

critical thinking, clinical reasoning, emotional intelligence) are essential to pass the national licensure examination and become a licensed professional (Guffey, Farris, Aldridge, & Thomas, 2002). As such, a variety noncognitive variables relationship with national licensure examination performance are studied in professional healthcare education. Using the Nelson-Denny Reading Test (NDRT) to assess pharmacy students reading comprehension, there was a significant difference ($p < 0.001$) between students who scored a 16.5 or greater on the NDRT than those who score a 16 or lower (Fuller, Horlen, Cisneros, & Merz, 2007). Haught and Walls (2004) provide supporting evidence for reading comprehension's relationship with national licensure examinations, as medical students NDRT score was significantly correlated with their results on the United States Medical Licensing Examination Part 1. In medical laboratory science education, Solberg (2015) found students critical thinking scores were weak predictors ($r = 0.11$, $p > 0.05$) of national licensure examination performance. As noncognitive variables have an inconsistent relationship with national licensure examination performance, studies regarding noncognitive PT student variables predictive capacity for the NPTE should be reviewed.

Comprehensive exam scores. Comprehensive exams are implemented in healthcare education programs to prepare students to take and pass their respective national licensure examination. Typically, a comprehensive exam is administered toward the end of the healthcare education program to assess students learning, while imitating the national licensure examination. The capacity of healthcare educations comprehensive exams to predict national licensure examination performance has been studied.

In 2015, Avi-Itzhak studied 65 occupational therapy students with a 63% first-time national licensure examination pass rate. Using a timed online computer exam format consistent with the occupational therapy national licensure examination, the comprehensive exam had four

domains: (a) gathering information, (b) formulating conclusions, (c) selecting and implementing evidence-based interventions, and (d) upholding professional standards and responsibilities (Avi-Itzhak, 2015). Logistic regression analysis that included all four comprehensive exam domains predicted 22% to 29% of the variance in first-time national licensure examination pass rates (Avi-Itzhak, 2015). In nursing education, Briscoe and Anema (1999) found a significant relationship between 33 nursing students comprehensive exam scores and their national licensure examination scores ($r=0.37$, $p<0.01$). Limiting the interpretation of Briscoe and Anema's (1999) results was their failure to report the national licensure examination pass rate. Continuing to study comprehensive exam scores relationship with national licensure examination performance, Steward, Bates, and Smith (2004) studied the relationship with a sample of 268 dental students with a 76.9% national licensure examination pass rate. The findings from Steward, Bates, and Smith's (2004) study found a significant relationship between dental students' comprehensive exam scores and national licensure examination performance ($F=4.78$, $p=0.03$). As comprehensive exam scores are significant predictors of national licensure examination performance across professional healthcare education, studying the relationship between comprehensive exam scores and NPTE performance in PT education is warranted.

Purpose of the Systematic Literature Review

The ability to predict NPTE performance with accuracy and precision has remained difficult for educational researchers due to the complexity of each PT student and the curriculum PT students engage in. Still, predicting NPTE performance has remained an important research topic, due to the "high stakes" nature of the NPTE for PT students and PT programs. The purpose of this systematic review of the current literature is to identify the student variables that best predict NPTE results. Predicting NPTE results has positive consequences for PT programs

and PT students. Successfully predicting NPTE results would allow PT programs to admit students who are most likely to pass the NPTE on their first-attempt, and therefore less likely to experience academic difficulty in the PT program.

The NPTE results that may be predicted include pass rates, failure rates, and detailed scoring. This systematic literature review highlights those student variables that predict, with varying levels of confidence, NPTE performance. The most commonly researched factors are PT applicant variables (e.g., GPA, GRE) and PT student variables (e.g., PT program GPA). Specifically, relationships between student variables and NPTE performance will be reviewed to provide more insight into what best predicts NPTE performance. The results of the systematic literature review will be used to inform PT program admission decisions, PT student decisions, and PT program resource allocation decisions.

This systematic literature review will discuss the two factors typically used to predict NPTE performance, in order: PT applicant variables and PT student variables. Under each factor, specific categories (e.g., UGPA, GRE) are reviewed that predict NPTE performance. The literature review provides an overview of the category, the evidence in support of the category, the evidence against the category, and considerations for the future use of the category. The literature is reviewed in the order of most relevant or impactful to the least relevant or impactful. The systematic literature review only included literature that attempted to predict NPTE performance from PT applicant and PT student variables.

From the systematic literature review there are important limitations regarding the findings from the comprehensive list of identified studies. Only one study, Meiners (2015) from the reviewed literature performed a power analysis as most of the studies were performed retrospectively. Without a power analysis, the adequacy of the sample sizes within each study

were unable to be determined (Zhang & Gou, 2016). Another reoccurring limitation of the studies reviewed was the incomplete reporting of data. Ten studies failed to report the NPTE first-time pass rate, while other studies failed to report mean values of key variables within the study. The failure to report all of the relevant data limits the usefulness and generalizability of the studies capacity to predict NPTE performance. The undetermined adequacy of the sample sizes and lack of reported data impact the significance and interpretation of the findings of this systematic literature review.

Systematic Literature Review

A systematic literature review should provide a clear and systematic approach to reviewing the literature. A unique aspect of a systematic literature review is the transparent, documented, and systematic approach to identify, and then review the relevant literature to the question being asked. Being transparent, documented, and systematic in a literature review allows others to replicate the search, and provides a road map to the conclusion(s). This systematic literature review is based upon five steps: (a) framing the question, (b) identifying relevant work, (c) assessing the quality of studies, (d) summarizing the evidence, and (e) interpreting the findings (Khan, Kunz, Kleijnen, & Antes, 2003).

Step 1: Framing the Question

The first step of this systematic literature review was to frame the question for a review. According to Khan, Kunz, Kleijnen, & Antes (2003), “The problems to be addressed by the review should be specified in the form of clear, unambiguous, and structured questions before beginning the review work” (p. 118). Prior to the initiation of this systematic literature review, the question was formulated and established: “What predicts NPTE performance in PT

education?” Once the review question was established, a search protocol was then created, documented, and followed.

Step 2: Identifying Relevant Work

The second step of this systematic literature review was to identify the relevant literature for predicting NPTE performance. To capture as many relevant citations as possible, both educational and medical databases were searched. Potential studies were identified by conducting a systematic search utilizing the databases EBSCO (1966 to 2017), which includes CINAHL, ERIC, and Medline, and PubMed (1966 to 2017). The search protocol was standardized for both databases. The search terms were related to PT education, admission criteria, academic success, PT programs, and NPTE performance to identify related citations (Appendix.).

A comprehensive electronic search was conducted up to and including June 2017 in the EBSCO and PubMed databases. The combined results of the database searched was 2,963 identified citations. The citations were then stored in the reference manager software, Mendeley (2017). All citations were reviewed for duplicates and conference proceedings; 105 citations were identified as duplicates or conference proceedings and removed. Following the removal of the duplicated citations and conference proceedings, the potential relevance of each citation to the research questions was examined, and consequently, 2,811 citations were excluded due to their irrelevance to the research questions being investigated. The full-text articles of the remaining 47 citations were assessed to select those studies that directly predict NPTE performance, were in peer reviewed journals or doctoral dissertations, and available in the English language. Following the review of the 47 remaining citations, 24 citations were deemed relevant to the question. Of the 24 relevant citations, 6 citations were dissertations, and 18

citations were published as journal articles. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement guided the selection of the systematic search (Figure 2.1.; Moher, Liberati, Tetzlaff, & Altman, 2009).

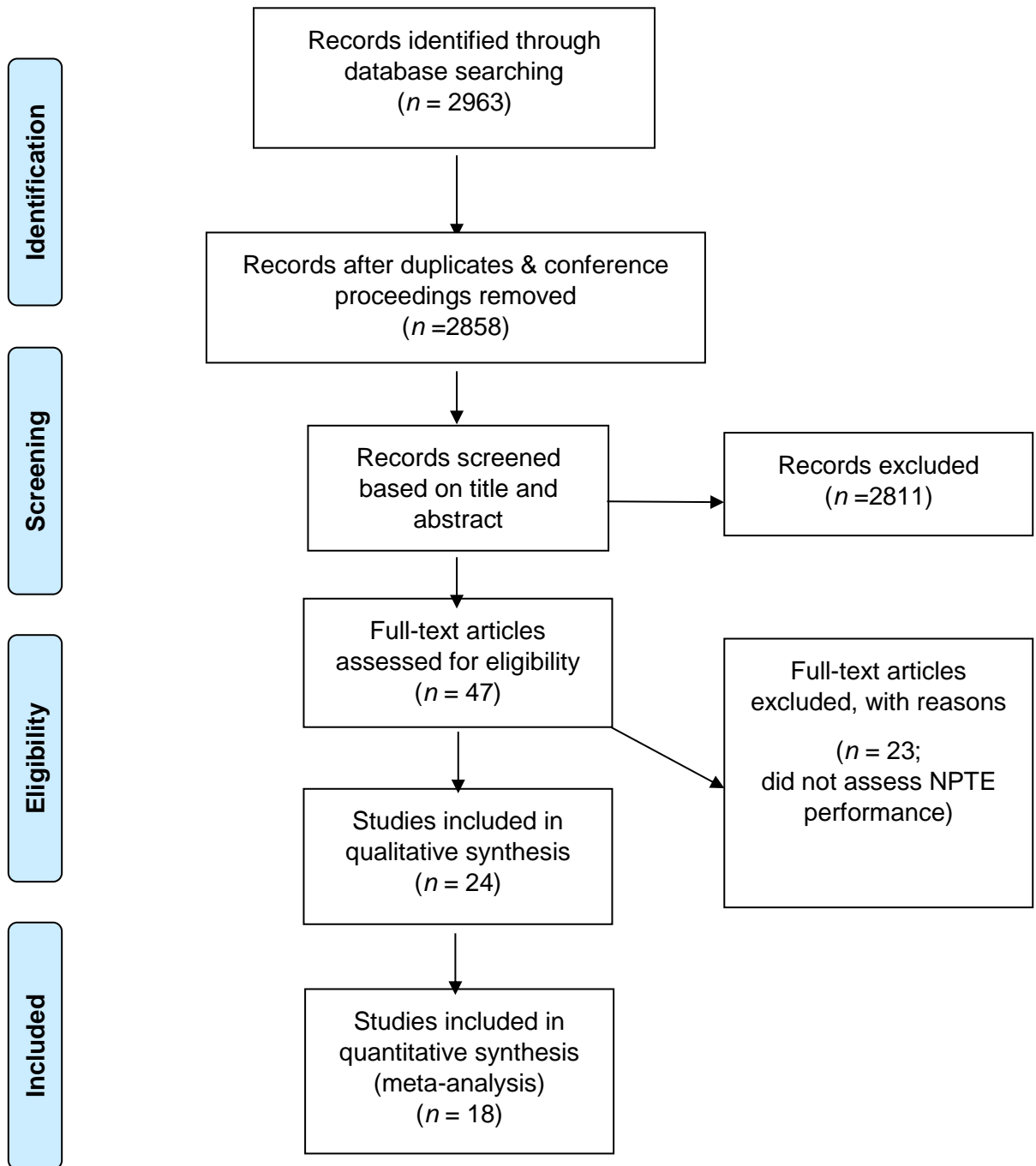


Figure 2.1. PRISMA flow diagram.

Step 3: Assessing the Quality of Studies

The third step in this systematic literature review was to assess the quality of all of the 24 identified citations. Following the systematic search for relevant citations, the quality of each citation was assessed utilizing a 3-step review process as identified by McCallum, Reed, Bachman, and Murray (2016). Two reviewers independently reviewed each article for inclusion in the systematic review. Each article was classified according to research design and methodological rigor utilizing the Oxford Centre for Evidence-based Medicine Levels of Evidence as a reference guide.

The Oxford Centre for Evidence-based Medicine Levels of Evidence scaling system assigns a level ranging from 1a to 5 for each citation based upon the degree of design quality. A score of 1a indicates the highest level of evidence, and a score of 5 is the lowest level of evidence (Oxford Centre for Evidence-based Medicine Levels of Evidence, 2012). Following the article review by a minimum of two independent reviewers, each article was given a score on the 1a-to-5 scale.

Following the establishment of the hierarchy of evidence (1a to 5), the methodological quality of each citation was assessed utilizing the McMaster appraisal tool (MAT) developed by Lekkas, Larson, and Kumar (1998). The MAT has been established as a valid and reliable critical appraisal instrument with 75% to 80% agreement among reviewers (Law, Steward, Pollock et al., 2013). The Lekkas, Larson, and Kumar (2007) scoring system allows for a standardized critical appraisal of each article. The scoring system has 14 possible criteria with a total score ranging from 0 to 14 (Lekkas, Larson and Kumar, 2007). For each of the 14 criteria, a zero was assigned if that criterion was not met or not present in the article, and a one was assigned if the criterion was met or was present in the article. The 14 criteria include: study purpose,

literature review, sample size description, sample size justification, reliability and validity of outcomes, study interventions, contamination avoidance, co-intervention avoidance, statistical significance, methods, clinical importance, sample drop-outs, and conclusions. The educational literature reviewed in the systematic review does not include study interventions, contamination avoidance, or co-intervention avoidance. Due to the nature of the educational literature that was reviewed, the MAT was modified. Specifically, study interventions, contamination avoidance, and co-intervention avoidance were removed from the MAT since none of the reviewed articles in the systematic review had a designed intervention. The modified MAT was created with a total of 11 criterion.

Two reviewers (MW and SFV), using the MAT, independently appraised each of the 11 criteria in each citation. Following the critical appraisal of each citation, the two reviewers compared the scores of each criterion and for each citation. Any discrepancies in appraisal scores between the two reviewers resulted in the citation being sent to a third independent reviewer (BH). The third reviewer settled all disputes regarding the differences in critical appraisal score of each criterion and for each citation.

With the final appraisal scores having the range of 0 to 11, a tertile scale, developed by McCallum, Mosher, Jacobsen, Gallivan, and Giuffre (2013), was modified to assess the citations risk for bias. The appraisal scores were placed in tertiles accordingly: high quality: 10 to 11 points; moderate quality: 8 to 9 points; and low quality: 0 to 7 points” (McCallum, Mosher, Jacobsen, Gallivan, & Giuffre, 2013).

Steps 4 and 5: Summarizing the Evidence and Interpreting the Findings

The fourth and fifth steps of the systematic literature review included summarizing the evidence and synthesizing and interpreting the findings of the collective literature review. The

review of the literature was separated into two distinct sections: PT applicant variables and PT student variables. Each section review of the relevant literature focused on thematic analyses to identify and examine patterns of key variables. Following the review of the literature, the interpretation of the findings was reported in a narrative format.

Physical Therapy Education Applicant Variables

As NPTE pass rates continue to be a common measure of PT student and PT program success, PT programs link their admission criteria to variables that are intended to predict success on the NPTE. “Making appropriate choices at admission increases the likelihood that students will be able to withstand the rigors of training and become successful practitioners” (Guffey et al., 2002). This systematic literature review identified 17 articles that investigated the relationships between selected applicant variables and NPTE performance. The most commonly investigated PT applicant variables were associated with GPA and GRE scores. Specifically, the applicant variables reviewed from the literature were: (a) undergraduate cumulative GPA (UGPA), (b) undergraduate GPA for prerequisite courses (UGPA-PC), (c) GRE scores, (d) degree status, (e) noncognitive applicant variables, and (f). SAT scores.

Undergraduate Cumulative Grade Point Average (UGPA)

Gaining admission into a PT program is a highly competitive process. Most PT programs heavily weigh PT applicants UGPA, the overall GPA of their undergraduate coursework, believing UGPA to be important for PT students’ success on the NPTE. Once students are admitted into a PT program, one goal of the PT program is for those students to pass the NPTE, and, more specifically, pass on the first-attempt. Therefore, the relationship between UGPA and NPTE becomes important as an applicant variable for most PT programs. Ten studies have

examined the relationship between UGPA and the NPTE. Specifically, UGPA of 3.5 or greater appears to be a predictor of NPTE success.

There are limitations in emphasizing UGPA as the primary measure of successful student learning for PT student applicants. If PT programs emphasize UGPA, then programs need to recognize the limitations of UGPA. The potential grade inflation of student UGPA and the courses used to get a UGPA clouds the utility of UGPA as a valid measure of student success on the NPTE (Willingham, 1974). Even with these documented limitations, UGPA continues to be a primary applicant variable for PT programs.

The evidence in support of UGPA. As PT programs rely on UGPA for admitting students into their PT program, it is expected that UGPA should have a positive relationship with passing the NPTE. In theory, as an applicant variable UGPA should predict NPTE performance. However, the relationship between UGPA and NPTE performance is inconsistent.

From the systematic literature review found, nine articles investigated the relationship between UGPA and the NPTE. Sample sizes for PT students ranged from 49 to 3,365, while other studies reviewed 185 accredited PT programs. The mean UGPA for the samples ranged from 3.33 to 3.55, on the 0.0 to 4.0 scale, with NPTE first-time pass rates ranging from 60% to 92%.

The systematic review of the literature revealed evidence to support UGPA as a predictor of NPTE performance. In 2009, Riddle, Utzman, Jewell, Pearson, and Kong (2009) used linear mixed-effects models to predict NPTE performance. The results from their data analysis identified UGPA of 3.4 or less had an odds ratio of 0.88 with a 95% confidence interval of 0.85 to 0.92 (Riddle, Utzman, Jewell, Pearson, & Kong, 2009). In other words, Riddle et al.'s (2009)

findings indicate that the students who had an UGPA of 3.5 or greater, were 12% more likely to pass the NPTE on their first-attempt.

The results of Riddle et al. (2009) are important considering the large sample size, 3,066 students. The study was a cross-sectional design with a multi-level data structure, which had a sample of students from multiple cohorts across the U.S., in public and private institutions with various Carnegie classifications (Riddle et al., 2009). Overall, the research design was of high quality with low risk for bias (Table 2.1). It is worth noting, the data analyses retained confounding variables with a $p \leq 0.10$, instead of the more traditional 0.05 (Riddle et al., 2009). With a large sample size from multiple cohorts, the findings of Riddle et al. (2009) are likely generalizable to PT students across the United States.

Table 2.1

PT Applicant Variables: Critical Appraisal Scores

Primary Author	Level of Evidence	Total Score	1	2	3	4	5	6	7	8	9	10	11
Adams	2B	10	*	*	*		*	*	*	*	*	*	*
Aldridge	3B	9	*	*	*		*	*	*	*	*		*
Cook (a)	2B	9	*	*			*	*	*	*	*	*	*
Cook (b)	2B	10	*	*	*		*	*	*	*	*	*	*
Covington	2B	10	*	*	*		*	*	*	*	*	*	*
Dockter	2B	9	*	*	*		*	*	*	*	*		*
Dillon	2B	9	*	*	*		*		*	*	*	*	*
Fell	2B	10	*	*	*		*	*	*	*	*	*	*
Galleher	2B	9	*	*	*		*	*	*	*	*	*	*
Guffey	2B	9	*	*			*	*	*	*	*	*	*
Hollman	2B	9	*	*	*		*	*	*	*	*		*
Huhn	2B	9	*	*	*		*	*	*	*	*		*
Lewis	2B	9	*	*	*		*	*	*	*	*		*
Meiners	2B	11	*	*	*	*	*	*	*	*	*	*	*
Riddle	2B	10	*	*	*		*	*	*	*	*	*	*
Thieman	2B	10	*	*	*		*	*	*	*	*	*	*
Utzman	2B	10	*	*	*		*	*	*	*	*	*	*

Critical Appraisal Category Scoring Key

1. Study purposed stated clearly
2. Relevant literature reviewed
3. Sample described in detail
4. Sample size justified
5. Outcome measures reliable
6. Outcome measures valid
7. Results reported in terms of statistical significance
8. Analysis methods appropriate
9. Educational importance reported
10. Dropouts reported
11. Conclusions appropriate

Design Key

- 2B = Cohort study
3B = Case-control study

Supporting the findings of Riddle et al. (2009), Utzman, Riddle, and Jewell (2007) used hierarchical logistic regression analyses to identify UGPA as a significant predictor of NPTE performance. Instead of trying to predict the pass rate on the NPTE, Utzman et al. (2007) attempted to predict the odds of failing the NPTE. Their sample consisted of 3,365 PT students, which represented PT programs across the country with a first-time NPTE pass rate of 87.4% (Utzman, Riddle, & Jewell, 2007). From data analyses, logistic regression identified the odds ratio of UGPA for predicting NPTE performance. The UGPA plotted receiver operating characteristics (ROC) curves to identify the cutoff UGPA of 3.49, which was then used for predicting NPTE failure with logistic regression analyses (Utzman et al., 2007).

The results of Utzman et al.'s (2007) analyses found that the odds of failing the NPTE increased by 12% for every 0.10 decrease in students UGPA (OR=0.88, 95% confidence interval=0.85-0.92). More specifically, students with an UGPA of 3.49 or greater were 53% less likely to fail the NPTE than students with a UGPA less than 3.49. As a high-quality study with a low risk of bias, Utzman et al.'s (2007) findings are relevant and are consistent with Riddle et al. (2009; Table 2.1).

However, it is not surprising Utzman et al. (2007) and Riddle et al. (2009) had similar findings given that both studies performed data analyses on the same data set, albeit with different sample sizes. The repeated use of the same sample is warranted when each study has a different purpose (International Committee of Medical Journal Editors, 2017). As such, the results of Utzman et al. (2007) and Riddle et al. (2009) should be considered together.

Building upon the argument for the utilization of UGPA as a predictor for NPTE performance, Cook, Engelhard, et al. (2015b) studied 185 accredited PT programs, and found PT programs with an applicant UGPA of 3.52 or greater were 3.68 and 5.43 times more likely to

have higher first-time and three-year pass rates, respectively. The UGPA of 3.52 was not an arbitrary number. ROC curves were used to determine the specific UGPA cutoff of 3.52. (Cook, Engelhard, et al., 2015b). Once the cutoff score was established, then UGPA was used to predict NPTE first-time and three-year pass rates through univariate logistic regression analyses.

The findings from Cook, Engelhard, et al. (2015b) are telling due to the quality of the study (Table 2.1). It is important to recognize Cook, Engelhard, et al. (2015b) utilized data from all accredited PT programs in the U.S. in 2011, ensuring adequate power for the study and with generalizability of the findings. Even as Cook, Engelhard, et al.'s (2015b) findings are significant, UGPA was modified from interval data into categorical data. Such a modification can result in a loss of data sensitivity since each unique value is designated a category. (Lomax & Hahs-Vaughn, 2012). Depending upon the categories, extreme values could lose their unique meaning.

With consideration, the findings of Cook, Engelhard, et al. (2015b) indicate UGPA is a significant predictor of NPTE performance and support the earlier work of Dillon and Tomaka (2010). In 2010, Dillon and Tomaka identified UGPA as a significant predictor of first-time NPTE success with a significant odds ratio (OR) of 16.11. More specifically, Dillon and Tomaka (2010) found students with a GPA of 3.5 or greater were seven times more likely (OR = 7.13) to pass the NPTE on the first-attempt than students with an UGPA less than 3.5 (Dillon & Tomaka, 2010).

Dillon and Tomaka's (2010) findings identify UGPA as an even stronger predictor for first-time NPTE success than Cook, Landry, et al. (2015a). While Dillon and Tomaka (2010) and Cook, Landry, et al. (2015a) utilized similar mean UGPA for their analyses, it is important to recognize Dillon and Tomaka's sample size was much smaller ($N = 72$), and only consisted of

PT graduates from four different cohorts. In the analyses of the data, Dillon and Tomaka (2010) identified significant correlations, then performed standard multiple regression, prior to performing logistic regression to predict NPTE success. For analysis, data was available for 60 out of the 72 students (83.3%); yet, multiple regression was performed for the entire sample (N=72) using expectation-maximization methods. For Dillon and Tomaka's (2010) study, expectation-maximization was not an appropriate imputation method since greater than 5% of the scores are missing for the NPTE (Do & Batzoglou, 2008). The use of expectation-maximization as part of data analysis must be considered when interpreting the validity of Dillon and Tomaka's (2010) findings.

Once UGPA was found to be a significant predictor of NPTE success, the categorical three levels of UGPA, less than 3.0, 3.0 to 3.49, and 3.5 and greater, were created based upon the potential use of each level as a benchmark for PT admission (Dillon & Tomaka, 2010). In contrast to Cook et al.'s (2015) use of ROC curves to identify statistically relevant cutoffs, Dillon and Tomaka's (2010) did not use statistical analysis techniques to create their categorical levels with UGPA. Limiting the generalizability of Dillon and Tomaka's (2010) findings is the low NPTE first-time pass rate (60%), which is well below the national average of 91% (Commission on Accreditation in Physical Therapy Education, 2017).

Overall, the results of Riddle et al. (2009), Utzman et al. (2007), Cook et al. (2015), and Dillon and Tomaka's (2010) studies provide evidence for the continued use of UGPA to predict NPTE first-time performance. The similarity of the findings among the authors provide further evidence for the use of UGPA as a PT applicant criterion to predict NPTE performance. Based upon the review of these four studies, PT programs should emphasize an UGPA of 3.5 or greater in order to improve their NPTE first-time pass rates.

The evidence against UGPA. Although there is evidence to support the use of UGPA as a primary predictor of NPTE performance, several authors have reported conflicting results. In 2015, Fell, Mabey, Mohr, and Ingram performed a cross-sectional retrospective study to identify predictors of NPTE performance. Multiple regression was performed on the sample of 290 PT graduates from two institutions and eight cohorts with a mean UGPA of 3.52 (Fell, Mabey, and Ingram, 2015). Fell et al. (2015) identified a regression model that predicted 40% of the variability in NPTE scores, however, UGPA only accounted for 0.1% of the overall variability. The non-significant findings should be given due consideration as the study was rated as high quality with low risk for bias following the critical appraisal (Table 1; Fell et al., 2015).

Prior to Fell et al.'s (2015) findings, Thieman (2003) performed a high-quality retrospective cohort study to predict NPTE performance (Table 2.1). Thieman had a sample of 122 PT students from multiple cohorts with a mean UGPA of 3.58 and an overall first-time NPTE pass rate of 92%. Significant Pearson correlation coefficients were identified and then used in the multiple regression analyses to predict NPTE performance. The results found UGPA had a significant correlation with NPTE performance with a $p \leq 0.05$ (Thieman, 2003). Although there was a significant correlation between UGPA and NPTE performance, UGPA again failed to be an independent and significant predictor of NPTE performance. Independently, UGPA only accounted for 4.8% of the variability in NPTE scores; however, the combination of UGPA, GRE, and PT program GPA still only accounted for 10.8% of the variability in NPTE scores. (Thieman, 2003).

Supporting Thieman's (2003) results, Galleher, Rundquist, Barker, and Chang (2012) performed a study of moderate quality on the capacity of UGPA to predict NPTE performance (Table 2.1). With a sample of 49 students, a mean UGPA of 3.55 and first-time NPTE pass rate

of 89.8%, Galleher et al. (2012) performed stepwise logistic regression to identify the predictive capacity of UGPA on NPTE performance.

The results of Galleher, et al.'s (2012) analyses found UGPA to have a significant correlation ($r=0.30$, $p\leq 0.05$) with NPTE performance. However, UGPA failed to be a significant predictor of NPTE performance when added into the stepwise multiple regression model. The decision to utilize stepwise multiple regression may have introduced bias into the statistical analysis, and, thereby, influenced the results. Yet, Galleher et al.'s (2012) findings build upon Thieman's (2003) results where UGPA was found to have a significant correlation with NPTE performance but failed to be a significant predictor of NPTE performance.

Similar to Thieman's (2003) and Galleher et al.'s (2012) results, Huhn and Parrott (2017) found a significant correlation between UGPA and NPTE scores ($r=0.24$, $p=0.003$) with a sample of 160 PT students from four cohorts. Despite the significant correlation, UGPA was a non-significant predictor of NPTE scores ($p=0.14$) with the subsequent random linear modeling procedures (Huhn & Parrott, 2017). The findings are noteworthy considering the mean UGPA for each cohort was 3.24 (SD=0.32), 3.39 (SD=0.32), 3.47 (SD=0.36), and 3.33 (SD=0.23) with mean NPTE scores of 656.38 (SD=44.40), 642.21 (SD=37.99), 648.06 (SD=38.88), and 664.43 (SD=51.63) respectively (Huhn & Parrott, 2017).

Adding to the evidence, Guffey (2000) performed a moderate quality study with a sample of 57 PT students from three cohorts (Table 2.1). At the time of application for admission, 46.8% had a UGPA of 3.5 or greater, 46.7% had a UGPA between 3.0 to 3.49, and 6.5% had a UGPA between 2.5 to 2.99. Neither the NPTE pass rates nor mean UGPA were reported. Failing to report the NPTE pass rates and mean UGPA limits the generalizability of Guffey's (2000) findings.

For data analysis, Guffey (2000) performed linear regression to predict NPTE performance from UGPA. The results of the linear regression found a non-significant relationship ($r = 0.228$, $p = 0.09$) between UGPA and NPTE performance. From the analysis, UGPA only accounted for 5.2% of the variance in NPTE scaled scores (Guffey, 2000). These findings are consistent with Thieman (2003).

Supporting the non-significant relationship between UGPA and NPTE performance, Meiners (2015) performed a high-quality study with a sample of 122 PT students from three cohorts (Table 2.1). The sample had a mean UGPA of 3.50 ($SD = 0.22$) with a 72.6% NPTE first-time pass rate and an 86.3% ultimate pass rate. The mean NPTE first-time score was 660.7 ($SD = 53.4$). Meiners (2015) performed multiple regression to predict NPTE scores and found UGPA to be a non-significant predictor ($\beta = -0.04$, $p = 0.75$).

The findings from Meiner (2015) do not support the use of UGPA as an applicant variable to predict NPTE performance. Yet, it is important to consider the low NPTE first-time pass rate (72.6%) and ultimate pass rate (86.3%). Since the samples mean UGPA was comparable to other studies, UGPA may not be the reason for the low NPTE pass rates. Other variables should be studied to determine the cause of the low NPTE pass rates.

In the ongoing search to identify significant relationships with UGPA and NPTE performance, Cook, Landry, Covington, McCallum, and Engelhard (2015a) studied a variety of relationships with PT programs NPTE three-year pass rates. Specifically, PT programs were categorized into “high” and “low” three-year pass rates. PT programs with 100% three-year pass rates were classified as “high,” while PT programs with less than 100% three-year pass rates were classified as “low.” Cook et al.’s (2015b) high quality observational study reviewed data from all 185 CAPTE accredited PT programs in 2011 (Table 2.1).

PT program characteristics were identified (e.g. UGPA) for programs with “high” and “low” three-year pass rates and compared for significant differences. The mean UGPA of the sample was 3.48. There was a significant difference between the mean UGPA for PT programs with “high” three-year pass rates (3.57) and “low” three-year pass rates (3.42). Regrettably, Cook et al. (2015b) did not perform subsequent regression analysis with UGPA, so the predictive capacity of UGPA for NPTE performance cannot be determined. An inability to determine UGPA’s predictive capacity for NPTE performance limits the utility of Cook, Engelhard, et al.’s (2015b) findings.

Similar to Cook, Engelhard, et al. (2015a), Covington, McCallum, Engelhard, Landry and Cook’s (2016) high quality observational study collected data from all 185 accredited PT programs in 2011 (Table 2.1). Their sample of PT programs had a mean UGPA of 3.53, with a significant difference between PT programs with a 100% three-year pass rate (3.59) and those programs with less than 100% three-year pass rate (3.49). Covington et al. (2016) identified a significant relationship ($p \leq 0.01$) between UGPA and PT programs NPTE three-year pass rates, but also failed to perform regression analysis to identify the predictive capacity of UGPA for NPTE performance.

After reviewing the studies of Cook, Landry, et al. (2015a), Cook, Engelhard, et al. (2015b), and Covington et al. (2016), it is suggested that these three studies analyzed the same dataset. Repeated analyses using the same dataset reduces the significance of each study independently, when each study is investigating the same question (International Committee of Medical Journal Editors, 2017). Therefore, all three of the studies should be considered as a whole when interpreting the results. Of concern, is the reported mean UGPA’s and number of PT programs with 100% three-year pass rates was different among studies that analyzed the same

dataset. Cook, Engelhard, et al. (2015b) and Covington et al. (2016) reported different mean UGPA's (3.48 and 3.53 respectively) and number of PT programs with 100% NPTE three-year pass rates (107 and 78 respectively) from the same dataset. Due to the inconsistencies in the findings using the same dataset, the results from Cook, Engelhard, et al. (2015b) and Covington et al. (2016) should be interpreted with caution as it cannot be determined which, if either, study is correct in their data analyses.

Through the systematic review of the literature, it is clear UGPA is correlated with NPTE performance. However, it is important to recognize the distinction that correlation does not imply causation. While several studies identified significant correlations between UGPA and NPTE performance, other studies failed to recognize UGPA as a significant predictor of NPTE performance or simply did not perform regression analyses with UGPA.

Considerations for UGPA. There is an ongoing use of UGPA as a primary applicant variable due, in part, to the underlying belief of a significant and causal relationship between grades and performance on standardized exams. Using a set minimum UGPA as a primary applicant criterion can support PT programs to graduate students who can successfully pass the NPTE on the first-attempt. Students who achieved high UGPA's are also perceived to have noncognitive characteristics including persistence, creativity, and grit that aid in their NPTE performance (Guffey, Farris, Aldridge, & Thomas, 2002). To assess these noncognitive characteristics in a valid and reliable way is difficult, so by default, UGPA continues to be the measure of choice for admission decisions into PT programs (Cook, Landry, et al., 2015a).

Following this systematic review of the literature that included eleven studies, there are mixed findings on the capacity of UGPA to predict NPTE performance. While several studies found UGPA to have significant predictive capacity, other studies failed to confirm such

findings. The contrasting findings are different enough to cause concern with the continued use of UGPA as a primary “stand alone” predictor of NPTE performance.

From the review of the literature, it is pertinent to recognize that some data repetition has occurred. Riddle et al. (2009) and Utzman et al. (2007) analyzed the same dataset. Both authors collected data from the same PT programs during the same time periods. In addition, Riddle et al. (2009) and Utzman et al. (2007) were authors on both studies. In similar fashion, Cook, Landry et al. (2015a), Cook, Engelhard, et al. (2015b), and Covington et al. (2016) performed data analyses on all 185 accredited PT programs in 2011. While the studies specifically studied different questions, UGPA was a common part of each of the studies. The carryover of information from the data between Cook, Landry, et al. (2015a), Cook, Engelhard, et al. (2015b), and Covington et al. (2016) should be carefully considered when reviewing their results and weighting the importance of the relationship between UGPA and NPTE performance.

Using UGPA as an applicant variable has relevance to the PT profession, as UGPA represents the same sort of behavior PT programs need for their students to pass the NPTE (Willingham, 1974). However, the inconsistent results of the systematic literature review are not surprising considering the limitations of UGPA as a primary measure of subsequent student learning in professional doctoral programs. UGPA continues to be an unreliable measure of subsequent student learning, and is now susceptible to grade inflation (Ling, Bochenek, & Burkander, 2015; Rojstaczer & Healy, 2012). PT programs should be cautious in using UGPA as a primary applicant variable. With a fixed scale of 0 to 4 limits the degrees of UGPA’s differentiation among students. “...the range is so narrow that differences among GPA’s do not usually represent reliable differences in student accomplishment” (Willingham, 1974, p.274). Particularly with grade inflation, the reliability and validity of UGPA can and should be

questioned by PT programs. The findings from the reviewed studies raise further doubt about the utility of UGPA. In short, UGPA is not the most appropriate “stand alone” variable to predict NPTE performance.

Undergraduate Cumulative Grade Point Average of Prerequisite Courses (UGPA-PC)

PT students cumulative UGPA of specified prerequisite coursework (UGPA-PC) is also used as an applicant criterion. While UGPA-PC can vary for each PT program, most UGPA-PC are focused on core undergraduate science courses such as anatomy, biology, calculus, chemistry, physiology, and physics (Physical Therapist Centralized Application Service, 2017).

UGPA-PC is comprised of the primary undergraduate courses that each PT program believes are required prior learning for students to be successful in their PT program. Similar to UGPA-PC, PT programs foundational coursework consists of science-based courses (e.g., anatomy, biomechanics, and pathology). The belief that past behavior predicts future behavior, in the same context, has led UGPA-PC to be a common applicant variable for PT programs. With PT programs specifying the coursework that makes up UGPA-PC, it is reasonable to conclude that UGPA-PC should be able to predict NPTE performance. Four authors have investigated the relationship between UGPA-PC and NPTE performance. The results show significant correlations between UGPA-PC and NPTE performance; however, UGPA-PC has not been shown to be a consistent predictor of NPTE performance with regression analysis.

UGPA-PC has limitations in its utility as a primary measure of subsequent student learning in professional doctoral programs. UGPA-PC is reported to be an unreliable measure of NPTE performance (Willingham, 1974). Despite the inconsistent reliability and validity, UGPA-PC continues to be also used as a primary applicant variable for PT programs. While UGPA-PC is a frequently used determinant in admission decisions for PT applicants, it is assumed that

UGPA-PC will be able to predict NPTE performance with precision. Currently, there is limited evidence to support the predictive capacity of UGPA-PC for NPTE performance.

From the current systematic review of the literature, four articles investigated the relationship between UGPA-PC and NPTE performance. Sample sizes for those four articles ranged from 43 to 290 PT students, with a reported mean UGPA-PC ranging from 3.42 to 3.53, on the 0.0 to 4.0 scale. The NPTE first-time pass rates ranged from 89.7% to 92.0%. With the range in sample size among the four articles, the context of each study becomes important.

The evidence in support of UGPA-PC. The only evidence to support the utilization of UGPA-PC as a predictor of NPTE performance comes from Fell et al. (2015). Fell et al. (2015) performed a cross-sectional retrospective high-quality study to identify predictors of NPTE performance (Table 2.1). Logistic regression was performed on the sample of 290 PT graduates from two institutions and eight cohorts with a mean UGPA-PC of 3.42. The results of data analyses identified a significant correlation between UGPA-PC and NPTE performance ($r=0.28$, $p<0.05$). UGPA-PC was then input into a multiple stepwise regression model that subsequently predicted 40% of the variance in NPTE scores, with UGPA-PC accounting for a significant 7.8% of the overall variability (Fell et al. 2015).

Unfortunately, Fell et al.'s (2015) investigation is the only evidence supporting the predictive capacity of UGPA-PC as an applicant entrance variable for NPTE performance, while accounting for less than 10% of the overall variance in NPTE scores. In review of Fell et al. (2015), the use of stepwise multiple regression adds potential bias to their results. Given Fell et al.'s (2015) findings, PT programs should emphasize a minimum UGPA-PC of 3.42 or greater with a high overall UGPA upon admission to improve their NPTE first-time pass rates. UGPA-

PC is worth consideration as an applicant variable when combined with a high overall UGPA, as the two variables lead to improved NPTE performance.

The evidence against UGPA-PC. The results from four studies do not support the continued use of UGPA-PC as a predictor of NPTE performance. Two authors, Dockter (2001) and Lewis (2011) have identified a significant correlation between UGPA-PC and NPTE performance. However, Dockter (2001) and Lewis (2011) were unable to demonstrate a significant predictive capacity of UGPA-PC for NPTE performance. In 2001, Dockter performed a moderate quality prospective cohort study to analyze the factors that predict NPTE performance (Table 2.1). A sample of 43 PT students from two cohorts who had completed the NPTE was analyzed. The Pearson Product Moment correlation coefficient of UGPA-PC with NPTE performance was significant ($r=0.34$, $p\leq 0.05$; Dockter, 2001). Following the identification of significant correlations, Dockter (2001) then performed stepwise regression analysis to determine the best predictor of NPTE performance. While UGPA-PC had a significant correlation with NPTE performance, UGPA-PC was not used as part of the regression analysis due to a reported weak correlation with NPTE performance, so the predictive capacity of UGPA-PC for NPTE performance was not determined.

Lewis (2011) performed a moderate quality study on the capacity of UGPA-PC to predict NPTE performance (Table 2.1). With a sample of 151 students from five cohorts and a 90.7% NPTE first-time pass rate, Lewis (2011) identified a significant correlation ($r=0.32$, $p\leq 0.05$) between UGPA-PC and NPTE performance. Again, while a significant correlation was found, subsequent multiple regression analysis found that UGPA-PC was unable to predict NPTE performance (Lewis, 2011).

The inability of UGPA-PC to significantly predict NPTE performance in independent studies is noteworthy. The similarity of the methods and sample of Dockter (2001) and Lewis' (2011) research should be interpreted with some care. While Dockter (2001) and Lewis (2011) identified a significant correlation ($p \leq 0.05$) between UGPA-PC and NPTE performance, neither author reported results from multiple regression analyses. Despite the similar samples, the findings from Dockter (2001) and Lewis (2011) should be viewed with caution as both studies failed to report the mean UGPA-PC for their sample.

In addition, the utilization of stepwise multiple regression by Fell et al. (2015), Dockter (2001), and Lewis (2011) present the potential for biased results. Confounding their results is that stepwise multiple regression allows authors to decide the order they input variables into the regression model and, consequently, potentially influencing the results (Lomax & Hahs-Vaughn, 2012). The results of the regression analyses were not reported for UGPA-PC from Dockter (2001) and Lewis (2011).

In contrast to the findings of Fell et al. (2015), Dockter (2001), and Lewis (2011), Thieman (2003) found a non-significant correlation ($p > 0.05$) between UGPA-PC and NPTE performance ($r = 0.185$). Thieman (2003) performed a high-quality retrospective cohort study to predict NPTE performance (Table 2.1). With a sample of 121 PT students from four cohorts with a mean UGPA-PC of 3.53 and an overall first-time NPTE pass rate of 92%, Thieman's (2003) findings are generalizable to PT students across the US.

Similar to Thieman (2003), Guffey (2000) found a non-significant correlation ($r = 0.04$, $p = 0.75$) between UGPA-PC and NPTE performance. Using linear regression, Guffey (2000) found UGPA-PC only accounted for 0.2% of the variance in NPTE scores. The results are from a moderate quality study with a sample of 57 PT students from three cohorts (Table 2.1). However,

the NPTE pass rates and UGPA-PC were not reported. Failing to report the NPTE pass rates and UGPA-PC severely limits the utility of Guffey's (2000) findings.

The results of our systematic review of the literature support the correlation between UGPA-PC and NPTE performance. While significant correlations were identified, the results from Dockter (2001) and Lewis (2011) have significant limitations as both studies failed to include an UGPA-PC in the regression analysis even though a significant correlation between UGPA-PC and NPTE performance was identified. In addition, Dockter (2011) failed to report the NPTE first-time pass rate of the studied sample. The decision not to perform or report the appropriate data and data analysis, limits the utility of both Dockter (2001) and Lewis' (2011) findings. On the available evidence, it is important to distinguish the difference between correlation and causation, and that a significant correlation between UGPA-PC and NPTE performance does not imply causation. Whether UGPA-PC can predict NPTE performance cannot be determined from this systematic review of the literature.

Considerations for UGPA-PC. Student performance in prerequisite PT program courses should be able to predict NPTE pass rates. Upon completion of the present systematic review of the literature, the majority of the literature fails to determine the predictive capacity of UGPA-PC. To date, Fell et al. (2015) provided the only evidence of UGPA-PC's capacity to predict NPTE performance. Even with limited efficacy, UGPA-PC continues to be utilized as a primary applicant variable due to the belief that UGPA-PC does predict NPTE performance.

The present systematic literature review did identify an area of consistency among authors as UGPA-PC was, generally, significantly correlated with NPTE performance. Supporting the significant correlation between UGPA-PC and NPTE performance is the similar

sample sizes among authors investigating the relationship between UGPA-PC and NPTE performance and comparable NPTE first-time pass rates.

An interesting finding from the systematic literature review was the interpretation and use of significant correlations. Fell et al. (2015), Dockter (2001), and Lewis (2011) all identified significant correlations between UGPA-PC and NPTE performance, yet, only Fell et al. (2015) reported the regression coefficient of UGPA-PC. This finding is noteworthy considering the significant correlations identified by Dockter (2001; $r=0.34$) and Lewis (2011; $r=0.32$) are larger than the correlation identified by Fell et al. (2015; $r=0.28$). Considering Fell et al. (2015) found UGPA-PC determined 7.8% of the overall variance in NPTE scores, the results of Dockter (2001) and Lewis (2011) appear incomplete. Dockter (2001) and Lewis (2011) should have performed regression analysis using UGPA-PC to predict NPTE performance and reported the results in their respective studies. Without the regression analysis results of Dockter (2001) and Lewis' (2011), the predictive capacity of UGPA-PC for NPTE performance remains uncertain.

When comparing UGPA-PC among PT programs it is important to remember that each PT program can include different undergraduate courses as part of the UGPA-PC. Since each PT program is free to choose their required UGPA-PC courses, there is no indication that the courses included in the reviewed UGPA-PC were the same or different. Without confirmation in the learning objectives and learning outcomes of the specific prerequisite courses used to define UGPA-PC or consistent reporting of the mean UGPA-PC, the generalizability of each study's findings to other PT programs should be questioned. With limited evidence, the variation in methods, and lack of a consistent definition of UGPA-PC, further analyses are required to determine the predictive capacity of UGPA-PC for NPTE performance.

Graduate Record Examinations (GRE)

High levels of verbal and quantitative reasoning are important skills for physical therapists to successfully treat patients in an array of settings. Therefore, it can be reasoned that a high GRE total score, GRE-V, and GRE-Q should predict NPTE performance. As such, GRE-V and GRE-Q are commonly used applicant variables for graduate programs and, specifically, for graduate entry PT programs.

Identified from the systematic literature review are seven studies that have investigated the relationship among GRE total score, GRE-V, and GRE-Q with NPTE performance. All of the identified studies used the old version of the GRE. There is no available literature on the new version of the GRE's predictive capacity for NPTE performance. The sample sizes of PT students in the reported seven studies ranged from 72 to 3,365, with an overall NPTE first-time pass rate ranging from 60% to 92%. From the present systematic literature review, only GRE-V is a consistent predictor of NPTE performance. Specifically, GRE-V of 365 and lower have predictive capacity for failing the NPTE (Hollman et al., 2008; Riddle et al., 2009; Utzman et al., 2007).

The evidence in support of GRE. From the systematic review of the literature, five studies support GRE total score, GRE-V, and GRE-Q's predictive capacity for NPTE performance. Hollman et al. (2008) completed a moderate quality retrospective cross-sectional study to predict NPTE performance from GRE-V and GRE-Q with a sample of 141 PT graduates from five cohorts (Table 2.1). From that sample, a 92% NPTE first-time pass rate was reported. Hollman et al. (2008) found GRE-V to be a significant predictor of NPTE performance with an area under the ROC curve of 0.73 and a 95% confidence interval of 0.60 to 0.87. From the ROC curve, a cutoff GRE-V of 365, out of a possible 800, was established.

As a significant predictor of NPTE performance; PT applicants with GRE-V below 365 had an increased probability (from 0.08 to 0.35) of failing the NPTE on their first-attempt (Hollman et al. 2008). Additionally, GRE-V had a positive likelihood ratio of 6.0 indicating PT graduates who failed the NPTE on their first-attempt were six times more likely to have a GRE-V less than 365 (Hollman et al. 2008). The significant findings from Hollman et al. (2008) provide evidence for the use of GRE-V greater than 365 as a predictor of NPTE performance.

Rather than trying to predict the pass rate on the NPTE, Utzman et al. (2007) tried to predict the odds of failing the NPTE. Utzman et al. (2007) utilized hierarchical logistic regression analyses to identify GRE-V and GRE-Q as significant predictors of NPTE performance. The study sample consisted of 3,365 PT students, representing all PT programs across the US. The sample's overall first-time NPTE pass rate was 87.4% (Table 2.1; Utzman et al., 2007). From data analyses, logistic regression identified the odds ratio of GRE-V (OR=0.93) and GRE-Q (OR=0.97) for predicting NPTE performance. The GRE-V and GRE-Q were plotted utilizing ROC curves to identify the cutoffs for GRE-V and GRE-Q (Utzman et al. 2007). From the ROC curve cutoffs, GRE-V was categorized into three levels, GRE-V 400 and lower, GRE-V 410 to 480, and GRE-V 490 and higher. GRE-Q was categorized based upon a cutoff GRE-Q less than or equal to 530 (Utzman et al. 2007). The categorical levels were then utilized for predicting NPTE failure with logistic regression analyses (Utzman et al., 2007).

The logistic regression model predicting NPTE failure using ROC curve cut-points, found GRE-V cutoff of 400 or lower had an odds ratio of 1.98, GRE-V of 410 to 480 was established as the reference value, and, therefore, had an odds ratio of 1.0, and GRE-V greater than or equal to 490 had an odds ratio of 0.67. The odds ratio of GRE-Q less than and equal to 530 was 2.28. The results of Utzman et al.'s (2007) analyses found that the odds of failing the

NPTE were increased by 7% for every 10-point decrease in GRE-V (OR=0.93, 95% confidence interval=0.92-0.95) and 3% for every 10-point decrease in GRE-Q (OR=0.97, 95% confidence interval=0.95-0.97). More specifically, a PT applicant with a GRE-V of 400 or less was 198% more likely to fail the NPTE on their first-attempt than students with a GRE-V greater than 400. PT applicants with a GRE-Q of 530 and lower were 228% more likely to fail the NPTE on their first-attempt than students with a GRE-Q greater than 530. (Utzman et al., 2007). Consistent with Hollman et al. (2008), Utzman et al.'s (2007) findings are relevant as their investigation was of high quality with a low risk of bias with a large representative sample. PT applicants with a GRE-V of 490 or higher and GRE-Q of 540 or higher are at a low risk for failing the NPTE on their first-attempt (Table 2.1).

When interpreting the findings of Utzman et al. (2007) consideration should be given to the loss of data sensitivity from the modification of continuous data into categorical data (Lomax & Hahs-Vaughn, 2012). Establishing categorical “cut-points” for continuous variables reduces the statistical impact of specific GRE scores. For example, in Utzman et al.'s (2007) study a GRE-V of 800 has the same statistical impact as a GRE-V of 490.

In support of GRE as a predictor of NPTE performance, Riddle et al. (2009) created categories for data analysis based upon institutional type: doctoral institution, master's institution, or medical institution. Riddle et al.'s (2009) performed a high-quality retrospective cross-sectional investigation with a multi-level data structure to identify predictors of NPTE performance from a sample of 3,066 students from 20 different PT programs (Table 1). Riddle et al.'s (2009) sample had a mean GRE-V of 470, mean GRE-Q of 600, and an 87% NPTE first-time pass rate. The large sample size from multiple PT programs and cohorts was representative

of enrolled PT students across the US and included students from public and private institutions with varying Carnegie classifications (Riddle et al., 2009).

Riddle et al. (2009) utilized a linear mixed-effects model to predict NPTE performance from GRE-V and GRE-Q. Variables were identified as significant and retained within the model if they had a $p \leq 0.10$. The results from Riddle et al. (2009) identified both GRE-V and GRE-Q as significant predictors of NPTE performance.

Riddell and et al. (2009) also reported the results for GRE-V of 470 or less in three different classifications: (a) GRE-V in doctoral institutions, (b) GRE-V in master's institutions, and (c) GRE-V in medical institutions. GRE-V in doctoral institutions had an odds ratio of 0.91 with a 95% confidence interval of 0.89 to 0.94, GRE-V in master's institutions had an odds ratio of 0.94 with a 95% confidence interval of 0.92 to 0.97, and GRE-V in medical institutions had an odds ratio of 0.97 with a 95% confidence interval of 0.93 to 1.00 (Riddle et al., 2009). These findings indicate that students with a GRE-V greater than 470 in doctoral institutions, master's institutions, and medical institutions were 9%, 6%, and 3%, respectively, more likely to pass the NPTE on their first-attempt than those who scored less than 470. It was unclear if there was overlap among institution types since the number of PT programs in each category was not reported. If there was overlap among categories it could impact the final results.

The results from data analyses on GRE-Q, identified GRE-Q of 600 or less, had an odds ratio of 0.97 with a 95% confidence interval of 0.95 to 0.98, indicating that students who had a GRE-Q greater than 600 were 3% more likely to pass the NPTE on their first-attempt than students who score below a 600 (Riddle et al., 2009).

In summary, GRE-V was a more significant predictor of NPTE performance than GRE-Q. While GRE-V and GRE-Q are statistically significant, the odds ratios were near 1.0, raising

concerns to the extent of the utility the findings for both GRE-V and GRE-Q to predict performance on the NPTE.

With a sample of 160 PT students from four cohorts with mean GRE total scores of 1,085.61 (SD=107.66), 1,084.21 (SD=132.78), 1,120.00 (SD=117.70), and 1,093.71 (SD=134.76) and mean NPTE scores of 656.38 (SD=44.40), 642.21 (SD=37.99), 648.06 (SD=38.88), and 664.43 (SD=51.63) respectively, evidence from Huhn and Parrott (2017) support the use of the GRE total score to predict NPTE scores. Huhn and Parrott (2017) found GRE total scores were significantly correlated ($r=0.37$, $p<0.001$) and predicted NPTE scores ($p=0.005$). More specifically, "...for every 1-point increase in the GRE total score, there was a 0.07-point increase in the NPTE score" (Huhn & Parrott, 2017, p. 10).

Thieman (2003) provides additional supporting evidence for GRE total score to predict NPTE performance. Thieman (2003) performed data analyses with a sample of 121 PT students from four cohorts, with a mean GRE total score of 1,683, out of a possible 2,400, and a mean NPTE first-time pass rate of 92% (Thieman, 2003). Thieman (2003) found GRE total score had a non-significant correlation with NPTE performance ($r=0.17$). Even with a non-significant correlation between GRE total score and NPTE performance, Thieman (2003) developed a regression model that predicted 10.8% of the variance in NPTE performance which included GRE total score. From the 10.8% of the variance in NPTE performance, GRE total score accounted for 2.7% of the change in the predictive capacity of NPTE performance in the regression model. The findings from Thieman (2003) come from a high-quality retrospective cohort study, with a low risk for bias (Table .2.1).

There is contradiction in the literature on how to determine the predictive capacity of GRE for NPTE performance. The majority of studies chose to look at GRE-V and GRE-Q,

instead of the GRE total score. After systematically reviewing the literature, GRE-V appears to be the better predictor of NPTE performance than GRE-Q or GRE total score. However, GRE-V's capacity to predict NPTE performance is limited. Key information was missing from Hollman et al. (2008) and Utzman et al. (2007; e.g., mean GRE scores and correlation values) that limit the generalizability of their findings to the studied sample. In addition to omitting data, data analyses from Riddle et al. (2009) and Utzman et al. (2007) utilized the same data set. Therefore, when reviewing the results of Riddle et al. (2009) and Utzman et al. (2007) it is important to consider the commonality of the sample. Using the same data set can lead to similar findings, so it is important to weigh either study accordingly. Even with different methods, using the same data set to measure NPTE performance should be expected to yield the same or very similar findings.

The evidence against GRE. The systematic literature review identified four studies providing evidence against the use of GRE scores to predict NPTE performance. In general, the evidence against the use of GRE scores to predict NPTE performance is poor. Dillon and Tomaka (2010) identified significant correlations between GRE and NPTE performance, but GRE was not a significant predictor of NPTE performance. In addition, Hollman et al. (2008) and Lewis (2011) failed to show significant correlations between GRE total score and GRE-Q with NPTE performance.

In 2010, Dillon and Tomaka completed a moderate quality sample with 72 PT graduates from four different cohorts with a NPTE first-time pass rate of 60% (Table 2.1). From data analyses, significant correlations ($r=0.39$) between GRE-V and NPTE performance and 0.35 between GRE-Q and NPTE performance were identified. Following standard multiple regression neither GRE-V (OR=1.0) nor GRE-Q (OR=0.99) was a significant predictor of NPTE

performance (Dillon & Tomaka, 2010). Worth consideration is the reported 60% NPTE first-time pass rate. The 60% NPTE first-time pass rate was much lower than all of the other studies that examined GRE scores. From the studies that examined GRE scores, 72.6% NPTE first-time pass rate was the closest to Dillon and Tomaka's (2010) 60% NPTE first-time pass rate.

Dillon and Tomaka's (2010) low NPTE first-time pass is likely attributable to a low mean GRE-V (mean GRE-V=373.03) and a high percentage of minority students in the sample who spoke English as a second language (44%). The low GRE-V score is an early indication that the PT applicants may have difficulty passing the NPTE on their first-attempt. A possible reason for the low GRE-V scores upon admission into the PT program is the high percentage of students who spoke English as a second language. Speaking English as a second language had a significant negative correlation (-0.24) with NPTE performance. In addition, the low NPTE first-time pass rate and high percentage of Hispanic PT students (greater than 60%) limit the generalizability of Dillon and Tomaka's (2010) findings. Yet, the Dillon and Tomaka's (2010) findings introduce ethnicity as a factor for NPTE performance.

From a sample of 73 PT graduates with low NPTE first-time pass rates (72.6%), mean NPTE score of 660.7 (SD = 53.4), mean GRE-V of 411.0 (SD = 59.9), and mean GRE-Q of 564.0 (SD = 92.2) the findings from Meiners (2015) study support Dillon and Tomaka's (2010) results. Meiners (2015) performed a high-quality study to determine the relationship between GRE-V and GRE-Q with NPTE performance (Table 2.1). Meiners (2015) developed a multiple regression model that predicted 39% of the overall variance in NPTE scores. As part of the regression model to predict NPTE performance, the standardized regression coefficients were identified as non-significant for GRE-V ($\beta = -0.10$, $p = 0.44$) and GRE-Q ($\beta = 0.19$, $p = 0.11$).

The findings indicate neither GRE-V nor GRE-Q are significantly related to NPTE scores when controlling for all other variables (Meiners, 2015).

When building upon the argument against GRE to predict NPTE performance, it is pertinent to return to the moderate quality investigation completed by Hollman et al. (2008). Hollman et al. (2008) found GRE-Q had a non-significant correlation with NPTE performance and, thereby, a limited ability to predict NPTE performance. Due to the non-significant correlation, GRE-Q was not input into Hollman et al.'s (2008) regression model. It is important to note, no correlation coefficients were provided by Hollman et al., so the exact correlation between GRE-Q and NPTE performance cannot be determined. The failure to report the correlation among non-significant variables significantly limits the utility of the findings from Hollman et al. (2008).

Lewis (2011) also did not report the correlations between GRE total score and NPTE performance in his moderate quality cohort study that predicted NPTE performance. Even with a sample of 151 students from five cohorts and a 90.7% NPTE first-time pass rate, failing to report the correlations has major consequences to the utility of the study. Without a correlation between GRE total score and NPTE performance, the relationship is presumed to be non-significant.

Considerations for GRE. As PT education has moved to a DPT is important to identify valid measures to predict PT applicants' success in a doctoral program. With an identified operational validity of 0.27 for GRE-V and 0.28 for GRE-Q to predict doctoral education GPA, GRE has proven to be a valid predictor of success in doctoral education (Kuncel et al., 2012). Six authors were able to add to the literature regarding the GRE's predictive capacity of the NPTE on PT students' first-attempt. PT applicants with a GRE-V of 365 or lower were up to six times more likely to fail the NPTE on their first-attempt. GRE-Q and GRE total score had weak

predictive capacity for NPTE performance, as both GRE-Q and GRE total score only predicted 3% of the variability in NPTE performance. There are likely other PT applicant variables that have a greater impact on the variance of NPTE performance (e.g. UGPA, UGPA-PC). The findings from the literature review provide useful information for PT programs admission processes.

The results of the systematic literature review fail to support GRE-V “stand alone” predictive capacity for NPTE performance. The relationship between GRE total score and NPTE performance, GRE-V and NPTE performance, and GRE-Q and NPTE performance were considered. GRE-V to be the best predictor of NPTE performance from the GRE. When considering the use of GRE-V as the best predictor of NPTE performance, it is important to recognize the omissions and flawed interpretation of the findings from Hollman et al. (2008) and Utzman et al. (2007).

While the significant findings from Hollman et al. (2008) and Utzman et al. (2007) provide evidence to support the continued use of GRE-V as an applicant variable, it is important to consider the limitations of their findings. Neither Hollman et al. (2008) nor Utzman et al. (2007) reported the mean GRE-V or the specific correlation between GRE-V and NPTE performance. Since the raw data was not reported for either study, the findings must be reviewed with caution as the utility of their findings is questionable. Beyond reporting deficiencies, Utzman et al. (2007) also changed GRE-V into a categorical variable. Making GRE-V a categorical variable reduces the sensitivity of the data and, thereby, potentially influencing the results.

The literature regarding the use of GRE to predict NPTE performance is overall, weak and with questionable relevance to PT programs today. Missing data, repeated information, and

reduced sample generalizability are all major limitations to the current body of literature that reviewed PT applicants GRE scores to predict NPTE performance. In addition, all of the studies from the systematic literature review investigated the predictive capacity of the old version of the GRE. Research is needed to investigate the predictive capacity of the current version of the GRE.

Degree Status

CAPTE recommends students who enter into a PT program should have a minimum of a baccalaureate degree (Commission on Accreditation of Physical Therapy Education, 2015). However, there is limited evidence to suggest that holding a baccalaureate degree leads to better performance on the NPTE. Without evidence to support CAPTE's recommendation, some PT programs (e.g. University of North Dakota, University of Mary) do not require their entry-level students to hold a baccalaureate degree prior to being admitted. Instead of requiring a baccalaureate degree, some PT programs require PT applicants to complete a set number of undergraduate credit hours. The total number of credit hours does not, necessarily, equate to the amount of credit hours needed to attain a baccalaureate degree. The PT programs that do not require a baccalaureate degree at admission allow the incoming PT students to earn their baccalaureate degree during their first year in the PT program. The baccalaureate degree is attained through the completion of the required PT program coursework. The decision to require a baccalaureate degree for admission into a PT program has financial consequences and time considerations for students and PT programs (United States Department of Labor, 2017). As PT programs applicant variables directly impact PT students', three authors have investigated the use of degree status as an applicant variable to predict NPTE performance. Degree status is defined by whether or not a PT applicant has attained a baccalaureate degree.

The systematic literature review identified three authors, Fell et al. (2015), Dockter (2003), and Adams, Glavin, Hutchins, Lee, and Zimmermann (2008) who investigated the relationship between PT applicants' undergraduate degree status and NPTE performance. The decision to examine degree status as an applicant variable is based upon the assumption that more education prior to entering a PT program better positions an applicant to succeed within a PT program and, therefore, pass the NPTE on the first-attempt. The studies investigating the relationship between degree status and NPTE performance consist of sample sizes ranging from 43 to 290 PT students. The findings from the two studies do not support the need for a baccalaureate degree prior to admission into a PT program.

It is important to recognize the limitations in utilizing degree status as an applicant variable. For instance, focusing on degree status as an applicant variable fails to consider the type of institution, the degree that was awarded, or the selected major for the degree. Even with these limitations and a lack of validity, the requirement of a baccalaureate degree continues to be utilized as an applicant variable for most PT programs.

The evidence in support of degree status. The systematic literature review identified one article, Fell et al. (2015), that found a statistically significant relationship between degree status and NPTE performance. In 2015, Fell et al. studied PT applicants' degree status as part of their multiple regression model that predicted 40% of the variability in NPTE scaled scores. Fell et al. (2015) found professional program GPA accounted for 32% of the variance. While degree status was a statistically significant variable in the regression model, degree status was only able to predict 1% of the variability in NPTE scaled scores (Fell et al., 2015). It is worth noting that in the same investigation, degree status was a non-significant predictor of NPTE first-time pass rates (Fell et al., 2015). Based upon degree status' inability to predict NPTE pass rates and

limited ability to predict NPTE scaled scores from one investigation, it appears degree status is not a useful predictor of NPTE performance.

The results from Fell et al. (2015) are noteworthy considering their investigation was of high quality with a low risk for bias (Table 2.1). The overall sample size consisted of 290 PT graduates from two institutions and eight cohorts with an 89.7% NPTE first-time pass rate. Potentially biasing Fell et al.'s (2015) results was that only 34% of the studied sample held a baccalaureate degree prior to being admitted into their respective PT program, which is not representative of other PT programs. The percentage of students (66%) without a degree is likely due to both of the PT programs studied not requiring PT applicants to hold a baccalaureate degree prior to admission. Given Fell et al.'s (2015) limitations with their investigation and coupled with their small, however, statistically significant findings there is little support for degree status' predictive capacity of NPTE performance.

The evidence against degree status. Only two other studies identified a non-significant relationship between degree status and NPTE performance. Adams et al., (2008) developed a significant ($p < 0.001$) logistic regression model that accounted for 21% of the variance in NPTE performance. The variables in the regression model included: CPI scores, gender, PT GPA, and degree status. From the variables in the regression model, only PT GPA was a statistically significant variable for predicting NPTE performance. Adams et al.'s (2008) sample consisted of 126 PT students, of which only 12.7% had a baccalaureate degree. The sample had an overall NPTE first time pass rate of 82.5%; however, Adams et al. (2008) did not report the correlation between degree status and NPTE performance. Failing to report the relationship between degree status and NPTE performance makes comparisons among studies difficult.

In 2001, Dockter performed a prospective cohort study, of moderate quality, to identify and analyze factors that predict NPTE performance (Table 2.1). From a sample of 43 PT students from two cohorts, Dockter found a non-significant correlation ($r = -0.08$) between degree status and NPTE performance. All PT students in the sample had completed their NPTE; however, the NPTE first-time pass rates were not reported (Dockter, 2001). The findings from Dockter (2001) should, therefore, be used with caution. Without the reported NPTE first-time pass rates comparisons between studies is difficult.

Considerations for degree status. With limited evidence and of varying quality, degree status does not predict NPTE performance with a degree of confidence. Although Fell et al. (2015) identified degree status as a significant statistical predictor of NPTE scaled scores, however, the practical significance is weak as degree status predicted 1% of the variability in NPTE scaled scores. In addition, degree status failed to be a predictor of NPTE pass rates. Similarly, Adams et al. (2008) identified a statistically significant regression model that predicted NPTE performance; however, degree status had a non-significant relationship with NPTE performance. Compounding the utility of degree status to predict NPTE pass rates, Dockter (2001) reported a non-significant correlation between degree status and NPTE performance. The limitations of Fell et al. (2015), Adams et al. (2008) and Dockter's (2001) investigations must be considered when interpreting the findings as it is plausible the samples, with a low percentage of PT students who held a baccalaureate degree, may have skewed the results of both studies.

The findings of the systematic literature review are important considering the financial consequences and time considerations in attaining a baccalaureate degree for PT applicants. If PT students are able to attain the same learning and mastery of competencies with less education,

then students will be able to reduce their financial and time commitments to earn their DPT. At this time, the limited evidence with mixed results fails to provide clarity on the use of degree status as an applicant variable or support CAPTE's recommendation. Further research is needed to determine degree status' capacity to predict NPTE performance.

Noncognitive Applicant Variables

The majority of the literature regarding applicant variables capacity to predict NPTE performance focuses on cognitive variables (e.g., GPA, GRE, degree status). These cognitive applicant variables have limitations and lack the ability to accurately and consistently predict NPTE performance. Noncognitive variables are presumed to influence PT students' performance in a PT program and on the NPTE (Hollman et al., 2008). Cultural sensitivity, empathy, honesty, integrity, responsibility, leadership, moral reasoning, and other behaviors influence PT student learning; however, these variables have proven difficult to measure in the admissions process (Dockter, 2001; Hollman et al., 2008). Admissions processes such as interviews and writing samples are often used to measure applicants noncognitive applicant variables and predict NPTE performance with varying success.

Noncognitive behaviors are an essential facet of PT student success. Noncognitive applicant variables have predicted graduation rates and final grades in medical school and academic difficulty in pharmacy programs. (Heldenbrand et al., 2016; Peskun et al., 2007; Tracey & Sedlacek, 1987). Nevertheless, measuring noncognitive variables with accuracy and precision has proven difficult.

The reduced predictive capacity of cognitive applicant variables indicates the need to look more closely at noncognitive variables that may predict NPTE performance. Identifying the predictive capacity of NPTE performance from noncognitive applicant variables will allow for a

more comprehensive and scientific approach to decision making when selecting applicant variables for the PT admissions process. Assessing noncognitive variables through the PT admissions process has been completed through the assessment of interview scores and writing samples as a means to enhance the predictive capacity of passing the NPTE on the first-attempt (Dockter, 2001; Hollman et al. 2008).

The evidence in support of noncognitive applicant variables. Evidence for the assessment of noncognitive variables as part of the PT admissions process is sparse, with only one relevant study. The moderate quality retrospective study performed by Hollman et al. (2008) analyzed the results of structured 30-minute behavioral interviews with two faculty members (Table 2.1). The behavioral interview was comprised of five noncognitive variables essential for PT's: decision making and problem solving, interpersonal skills, patient/client focus, communication, and team work. The structured interview was scored using a 5-point Likert scale, with the possible total score of the behavioral interview ranging from 0 to 20. The sample of 89 PT applicants from two cohorts had a NPTE first-time pass rate of 92%. The mean score on the behavioral interview was 17.1 (SD=2) with scores ranging from 10 to 20. The inter-rater reliability of the behavioral interview was determined to be moderate to high with an Intraclass Correlation Coefficient (ICC) of 0.75 (Hollman et al., 2008).

Further data analysis identified behavioral interviews to be a significant predictor of NPTE performance (Hollman et al., 2008). Behavioral interview scores had an area under the ROC curve of 0.69 with a 95% confidence interval of 0.55 to 0.824. From the ROC curve a cutoff behavioral interview score of 16.5 was established (Hollman et al., 2008).

The results of Hollman et al. (2008) identified behavioral interviews as a significant predictor of NPTE performance. Specifically, behavioral interviews scores of 16.5 or less were

able to determine the probability of failing the NPTE on the first-attempt increases from 0.08 to 0.20 (Hollman et al. 2008). A positive likelihood ratio of 2.75 for behavioral interview scores' predictive capacity for NPTE performance was determined. The positive likelihood ratio indicates that PT graduates who failed the NPTE on their first-attempt were 2.75 times more likely to have received a behavioral interview score less than 16.5 (Hollman et al. 2008). The significant findings from Hollman et al. (2008) provide evidence to further consider the use of behavioral interviews as an important variable in the PT admissions process.

It is important to consider the limitations of the evidence of behavioral interviews. Specifically, bias can be present anytime subjective scoring is utilized to measure human behavior (Hollman et al., 2008). It is possible that interviewers may score the same person differently and inconsistently, thereby reducing the reliability and validity of behavioral interviews. However, Hollman et al. (2008) reported an appropriate ICC of 0.75 to reduce some of the concerns regarding the reliability of the use of the 20-point Likert scale.

The evidence against noncognitive applicant variables. Dockter (2001) investigated similar noncognitive applicant criteria's correlation with NPTE performance. Dockter (2001) analyzed 43 PT students from two different cohorts with a moderate quality prospective investigation (Table 2.1). Statistical analysis utilizing the Pearson Product Moment Correlation Coefficients were non-significant between interview scores ($r = 0.18$), and writing sample scores ($r = -0.26$) with NPTE performance (Dockter, 2001).

To aid in the interrater reliability of scoring the interviews, "The interviewers from each year were given a standardized set of questions developed by faculty, as well as an instructional session..." (Dockter, 2001, p.61). The intra-rater reliability was not addressed for the writing samples. The writing samples were scored by one faculty member (Dockter, 2001). Both the

interview score and the writing samples were a part of the total admission scores, which had a non-significant correlation ($r=0.25$) with NPTE performance. Dockter (2001) failed to report the NPTE pass rate of the sample, so the results of the study are of limited value for comparison to other studies. The overall findings from Dockter (2001) provide little evidence for the assessment of noncognitive variables as part of the admissions process for PT programs.

Considerations for noncognitive applicant variables. Hollman et al. (2008) investigated the behavioral interviews predictive capacity of NPTE first-time pass rates, and identified a significant (+ LR = 2.75, $p<0.05$) relationship between preadmission interview scores and first-time NPTE pass rates. Their results provide some measure of support for considering noncognitive variables as a component of PT programs admissions process. Despite the evidence from Hollman et al. (2008), the valid and reliable assessment of noncognitive variables remains difficult due to the multiple psychometric properties associated with each noncognitive variable.

When reviewing the results of the systematic literature review, a comparison of Hollman et al.'s (2008) findings and Dockter's (2001) findings present a dilemma as both authors utilized different assessment instruments. More research is needed with the use of consistent measurement tools to determine if noncognitive variables can better predict NPTE performance than cognitive variables. Based-upon the available evidence, behavior interviews should be seriously considered for the PT admissions process.

Scholastic Aptitude Test (SAT) Scores

The systematic literature review only identified one study that investigated the relationship between PT applicants SAT scores and NPTE performance. The SAT is typically taken in the junior or senior year of high school. Reviewing PT applicants' SAT scores for

admission into a PT program presents an opportunity to identify early predictors for NPTE performance. The results of the systematic literature review identified a significant relationship between SAT scores and NPTE performance albeit from one study.

The evidence in support of SAT scores. The one study from the systematic literature review provides evidence on the use of SAT scores as an early predictor for NPTE performance. Galleher et al.'s (2012) moderate quality investigation collected survey data from 49 PT students from seven different PT programs with a mean SAT score of 1,158 and an 89.9% NPTE first-time pass rate (Table 2.1). From data analyses an odds ratio of 1.03 was identified to determine SAT scores predictive capacity of NPTE performance. In addition, the correlation between SAT scores and NPTE performance was 0.46 using Spearman's Rho with a 95% confidence interval range of 0.20 to 0.65 (Galleher et al., 2012). The results of Galleher et al.'s (2012) investigation found SAT scores to be a significant predictor for passing the NPTE on the first-attempt. "Utilizing SAT scores increased the ability to predict if a student will pass on the first-attempt of the NPTE by 4.1%" (Galleher et al., 2012, p.5).

Considerations for SAT scores. The findings from Galleher et al. (2012), while significant, should be reviewed with caution as SAT scores were only able to impact the prediction of NPTE first-attempt success by 4.1% indicating there are other variables with a higher contribution to NPTE performance. Further research is indicated to determine SAT scores predictive capacity of NPTE performance. With only one moderate quality investigation identified from the systematic literature review, there is limited evidence regarding the utility of SAT scores as an applicant variable for PT programs.

Conclusions from the Review of PT Applicant Variables

In conclusion, no one PT applicant variable should be used as a “stand alone” predictor of NPTE performance (Table 2.2). The PT admissions process should rely on multiple applicant variables to identify PT applicants who will most likely pass the NPTE on their first-attempt. Specifically, a PT applicant will be likely to pass the NPTE on their first-attempt if they have a UGPA’s of 3.5 or greater, a UGPA-PC of 3.42 or greater, and a behavioral interview score of 17 or more (Fell et al., 2015; Hollman et al., 2008; Riddle et al., 2009). The results of the GRE scores relationship with NPTE performance cannot be applied to the current PT applicants process as the reported studies used the old version of the GRE. Research regarding the new version of the GRE’s and SAT’s predictive capacity for NPTE performance is warranted. The overall results of this systematic literature review of PT applicant variables failed to identify a “stand alone” predictor of NPTE performance. A comprehensive approach should be used in the PT admission process to identify the PT applicants who are most likely to pass the NPTE on their first-attempt.

Table 2.2

Summary of PT Applicant Variables

Primary Author	Year	Journal	Summary of Findings
Adams	2008	JOPTE	Degree status was a non-significant independent predictor of NPTE performance.
Cook	2015a	Jrnl of Educ Eval for Health Professions	A mean UGPA of 3.52 or greater was associated with a higher NPTE first-time pass rate (OR=2.87).
Cook	2015b	BMC Medical Educ	PT programs with 100% three-year NPTE pass rates had a higher mean GPA (3.57), then PT programs with less 100% three- year pass rates (3.42).
Covington	2016	Jrnl of Allied Health	Significant difference ($p < 0.01$) between PT programs with 100% three-year NPTE pass rates and PT programs with less 100% three- year pass rates.
Dillon	2010	JOPTE	UGPA was a significant predictor of NPTE success (OR=16.11). Specifically, an UGPA of 3.5 or greater (OR=7.13) was more likely to pass the NPTE on the first-attempt. Neither GRE-V nor GRE-Q was a significant predictor or NPTE performance.
Dockter	2001	JOPTE	UGPA-PC had a significant correlation ($r=0.34$), but was not a significant predictor of NPTE performance. Degree status ($r=-0.08$), admission interview scores ($r=0.18$), and writing samples ($r=-0.26$) had non-significant correlations with NPTE performance.
Fell	2015	JOPTE	UGPA predicted 0.1% of the variance, UGPA-PC predicted 7.8% of the variance, and degree status predicted 1% of the variance in NPTE scores.
Galleher	2012	Internet Jrnl of Allied Health Sci & Practice	UGPA was a non-significant predictor for pass the NPTE on the first-attempt. SAT scores increased the ability to predict NPTE success on the first-attempt by 4.1%.
Guffey	2000	Dissertation	PT applicants UGPA had a non-significant relationship ($r = 0.23$, $p = 0.09$) with NPTE pass rates, as UGPA only accounted for 5.2% of the overall variance in NPTE scaled scores. UGPA-PC had a non-significant correlation ($r = 0.04$, $p = 0.75$) with NPTE performance, while only accounting for 0.2% of the variance in scores.
Hollman	2008	Jrnl of Allied Health	GRE-V of 365 or lower or a behavioral interview score less than 16.5 increased probability of failing the NPTE to 0.35 and 0.20, respectively. GRE-Q had a non-significant correlation with NPTE performance.
Huhn	2017	JOPTE	UGPA had a significant correlation ($r=0.24$), but was not a significant predictor of NPTE performance. GRE total score had a significant correlation ($r=0.37$) and was a significant predictor of NPTE scores ($p=0.005$).
Lewis	2011	Internet Jrnl of Allied Health Sci & Practice	Although significantly correlated ($r=0.32$), UGPA-PC was unable to predict NPTE performance. GRE total scores were non-significantly correlated with NPTE performance.
Meiners	2015	Dissertation	PT applicants UGPA was a non-significant predictor of NPTE scores ($p=0.75$). Applicants GRE-V ($\beta = -0.10$, $p = 0.44$) and GRE-Q ($\beta= 0.19$, $p = 0.11$) were unable to predict NPTE performance.
Riddle	2009	Physical Therapy	UGPA of 3.5 and greater was 12%, GRE-V of 470 and greater was up to 9%, and GRE-Q of 600 and greater was 3% more likely to pass the NPTE on their first-attempt.
Thieman	2003	JOPTE	UGPA accounted for 4.8% and GRE total score accounted for 2.7% of the variability in NPTE scores. UGPA-PC was non-significantly correlated with NPTE performance ($r=0.185$).
Utzman	2007	Physical Therapy	The odds of failing the NPTE increased by 12% for every 0.10 decrease in students UGPA (OR=0.88). UGPA of 3.49 or greater was 53% less likely to fail the NPTE than students with a UGPA less than 3.49. PT applicants with a GRE-V of 400 or less was 198% more likely to fail or a GRE-Q of 530 or less was 228% more likely to fail the NPTE on their first-attempt.

Physical Therapy Student Variables

Being admitted into a PT program does not guarantee success on the NPTE. As such, PT student variables are often studied to determine their predictive capacity for NPTE performance. PT student variables are different from PT applicant variables. PT student variables are from the students' experiences in a graduate PT program and of interest to PT programs and students. Identifying which students admitted into a PT program will or will not be successful on the NPTE is beneficial to PT programs and students. The earlier PT programs can identify students at-risk for failing the NPTE, the earlier PT programs can assist the at-risk students. As such, identifying at-risk students begins with student applicant variables and should continue throughout their PT education.

The number of PT student variables used to predict NPTE performance are numerous and warrant a systematic literature review to clarify the relevant information from the pertinent studies. This systematic literature review identified 19 articles that investigated the relationships between selected PT student variables and NPTE performance. The most commonly investigated PT student variables were associated with GPA and clinical performance scores. The PT student variables reviewed from the literature were: (a) GPA, (b) clinical performance scores, (c) noncognitive student variables, and (d) comprehensive exam scores.

Physical Therapy Program Grade Point Average (PTGPA)

One measure of students' academic performance in their PT program is the PT program GPA (PTGPA). PTGPA is the cumulative GPA for each student while enrolled in a professional PT program. Knowing the strength of the relationship between PTGPA and NPTE performance is important for PT programs and PT students. The relationship between PTGPA and NPTE performance is an indication of how closely aligned PT programs expectations are with the

requirements to pass the NPTE on the first-attempt. The evidence that follows suggests PTGPA is the strongest predictor of NPTE performance, but with limitations. When studying PTGPA, investigators typically look at PTGPA at the end of the curriculum (overall PTGPA) or at the end of PT students first year in a PT program (first-year PTGPA). As such and when considering the evidence, it is important to distinguish between overall PTGPA and first-year PTGPA. The first-year PTGPA is an early indicator of the students NPTE performance. The systematic literature review has identified ten studies that examined the relationship between PTGPA and NPTE performance.

While an early indicator of NPTE performance, using PTGPA as a primary predictor for NPTE performance has limitations that need to be considered. Grade inflation continues to occur with PTGPA, as most PT programs require students to maintain a minimum PTGPA of 3.0 out of a 4.0 scale to remain in their professional PT program (Ling, Bochenek, & Burkander, 2015; Rojstaczer & Healy, 2012). In addition, acceptance into a PT program is a highly competitive process, which in itself is GPA dependent. The GPA of each student also represents a mixture of undergraduate courses and undergraduate degrees that encompass the applicant GPA. It is expected that those students accepted into a PT program will be able to maintain a PTGPA greater than 3.0. With a minimum of PTGPA 3.0 and a maximum PTGPA of 4.0, there is a reduced range in PTGPA between excellent and average students. Even with the limited range of PTGPA (3.0 to 4.0) and with grade inflation, PTGPA continues to be a significant predictor of NPTE performance.

The evidence in support of PTGPA. As PT programs align their curriculum to prepare students to sit and pass the NPTE, it is expected that PTGPA should have a significant positive relationship with the subsequent NPTE performance. The eleven studies that investigated the

relationship between PTGPA and NPTE performance had sample sizes ranging from 42 to 979. The mean PTGPA for the eleven studies ranged from 3.50 to 3.73, on the 0 to 4.0 scale, with NPTE first-time pass rates ranging from 60% to 92%. The findings from those eleven studies using the systematic review methods, supports the continued use of PTGPA as a predictor of NPTE performance. Specifically, PT students with a PTGPA of 3.73 or greater are more likely to pass the NPTE on the first-attempt.

In 2015, Meiners studied the capacity of PTGPA to predict NPTE performance following PT students first year in a PT program. Performing multiple regression to predict NPTE performance, PTGPA ($\beta=0.57$, $p<0.001$) was found to make the largest significant contribution to predicting NPTE scores (Meiners, 2015). As a significant and unique contributor to the prediction of NPTE scores, Meiners (2015) found 24% of the variance in NPTE scores were predicted by PT students' first-year PTGPA.

After predicting NPTE scores, Meiners (2015) performed logistic regression to predict the odds of passing the NPTE on the first-attempt. Data analysis found “that as first-year PTGPA increases by one unit, a student is 728.5 times (CI=1.24 to 426,875.97) more likely to successfully pass the NPTE” (Meiners, 2015). For example, an increase in a PTGPA from 3.0 to 3.1 results in a PT student being 72.85 times more likely to successfully pass the NPTE on the first-attempt.

The results of Meiners (2015) high quality study (Table 2.3) are informative given the sample size of 122 students with a mean first-year PTGPA of 3.62 (SD=0.25) and a mean NPTE score of 660.65 (SD=53.38). It is worth noting that with a PTGPA of 3.62 the NPTE first-time pass rate (72.6%) and overall three-year NPTE pass rate (86.3%) were low. The low NPTE first-time and overall pass rates likely resulted in an increase in variance among PT students NPTE

scores and may have contributed to PTGPA's large odds ratio for predicting NPTE performance. Even with the identified limitations, Meiners (2015) findings are noteworthy.

Table 2.3

PT Student Variables: Critical Appraisal Scores

Primary Author	Level of Evidence	Total Score	1	2	3	4	5	6	7	8	9	10	11
Adams	2B	8	*	*	*	*	*		*		*	*	
Aldridge	3B	9	*	*	*		*	*	*	*	*		*
Cook J	2B	10	*	*	*		*	*	*	*	*	*	*
Dillon	2B	9	*	*	*		*		*	*	*	*	*
Dockter	2B	9	*	*	*		*	*	*	*	*		*
Dreeben	2B	9	*	*	*		*	*	*	*	*		*
Edmondson	2B	9	*	*	*		*	*	*	*	*		*
Fell	2B	10	*	*	*		*	*	*	*	*	*	*
Galleher	2B	9	*	*	*		*	*	*	*	*	*	*
Guffey	2B	9	*	*	*		*	*	*	*	*		*
Huhn	2B	9	*	*	*		*	*	*	*	*		*
Kosmahl	2B	9	*	*	*		*	*	*	*	*		*
Lewis	2B	9	*	*	*		*	*	*	*	*		*
Luedtke-Hoffman	2B	9	*	*	*		*	*	*	*	*		*
Meiners	2B	11	*	*	*	*	*	*	*	*	*	*	*
Moran	2B	10	*	*	*		*	*	*	*	*	*	*
Vendrely	2B	9	*	*	*		*	*	*	*	*		*

Critical Appraisal Category Scoring Key

1. Study purposed stated clearly
2. Relevant literature reviewed
3. Sample described in detail
4. Sample size justified
5. Outcome measures reliable
6. Outcome measures valid
7. Results reported in terms of statistical significance
8. Analysis methods appropriate
9. Educational importance reported
10. Dropouts reported
11. Conclusions appropriate

Design Key

- 2B = Cohort study
3B = Case-control study

Supporting Meiners (2015) findings, Dockter (2001) identified a significant correlation ($r=0.65$, $p<0.01$) between first-year PTGPA and NPTE performance. Using multiple linear regression, first-year PTGPA accounted for 42% of the overall variance in NPTE performance on students first-attempt (Dockter, 2001). The moderate quality study (Table 2.3) consisted of 43 PT students from two cohorts with a mean first-year PTGPA of 3.65 (SD=0.30). The results of Dockter's (2001) study indicate that "the curriculum during the first year of the physical therapy program prepared students for the NPTE" (p. 63). While the findings were significant, Dockter (2001) did not report the mean NPTE pass rates or scores, thereby limiting the usefulness of their findings.

With a sample of 160 PT students from four separate cohorts with mean first-year PTGPA of 3.53 (SD=0.29), 3.60 (SD=0.27), 3.64 (SD=0.18), and 3.77 (SD=0.15) and mean NPTE scores of 656.38 (SD=44.40), 642.21 (SD=37.99), 648.06 (SD=38.88), and 664.43 (SD=51.63) respectively, Huhn and Parrott's (2017) findings continue to support the use of the first-year PTGPA to predict NPTE performance. First-year PTGPA across the four separate cohorts significantly correlated ($r=0.60$, $p<0.001$) and predicted ($p=0.001$) NPTE scores (Huhn & Parrott, 2017). More specifically, random intercept linear modeling indicated that for every 0.1 increase in PTGPA, there was a 9.08 increase in the NPTE score (Huhn & Parrott, 2017). When interpreting Huhn and Parrott's (2017) findings it is important to consider their failure to report the NPTE first-time or overall pass rates.

Looking beyond first-year PTGPA to overall PTGPA, Kosmahl (2005) performed a retrospective cohort study to identify factors that predict NPTE performance. To identify the predictors of NPTE performance, stepwise multiple regression was performed. 92 PT graduates with a mean PTGPA of 3.52 (SD=0.21) and mean NPTE score of 659.74 (SD=37.29) were

investigated. PTGPA was significantly correlated ($r=0.60$, $p<0.001$) with NPTE scores. As a result of the significant correlation, PTGPA was a variable in a multiple regression model that accounted for 46.5% of the variance in NPTE scores (Kosmahl, 2005). Specifically, the addition of PTGPA to the regression model increased R^2 from 0.37 to 0.47, indicating that using PTGPA as a variable provided a 10% increase in the capacity of the model to account for the variance in NPTE scores.

One limitation of Kosmahl's (2005) study was the failure to report the NPTE first time or overall pass rates. The omission of the NPTE pass rates limits the interpretation of Kosmahl's (2005) results. Yet, as a moderate quality study (Table 2.3), Kosmahl's (2005) findings provide further evidence supporting the assertion that PT students' individual academic performance is significantly related to NPTE performance.

Building upon Kosmahl's (2005) findings, Fell et al. (2015) performed a moderate quality study with 270 PT students from two institutions that included eight separate cohorts (Table 2.3). The studied students had a mean overall PTGPA of 3.68 with an 89.7% NPTE first-time pass rate. Fell et al. (2015) completed a cross-sectional retrospective study to identify predictors of NPTE performance. Logistic regression identified PTGPA as a significant predictor ($p<0.0001$) of NPTE first-time pass rates. Following logistic regression, multiple regression was performed with a model that predicted 40% of the variability in NPTE scores. Within the multiple regression model, overall PTGPA accounted for 31.6% ($p<0.0001$) of the variance in NPTE scores (Fell et al., 2015). The results of Fell et al. (2015) support the continued use of PTGPA as a significant predictor of NPTE performance.

Adding to the evidence in support of PTGPA as a predictor of NPTE performance, Dillon and Tomaka (2010) found a significant correlation ($r=0.55$, $p<0.001$) between PTGPA and

NPTE performance. With the significant relationship between PTGPA and NPTE performance, multiple regression analysis was performed, resulting in a model that accounted for 37% of the variance in NPTE scores. From the regression model, PTGPA, as a stand-alone variable, significantly ($p=0.01$) predicted variance in PT students NPTE performance (Dillon & Tomaka, 2010).

After identifying PTGPA as a significant predictor of NPTE scores, Dillon and Tomaka (2010) performed logistic regression that predicted 74% of the PT students passing the NPTE on their first-attempt. From that analysis, PTGPA was the best predictor ($OR = 25.84, p = 0.030$) of PT students' capacity to pass the NPTE on their first-attempt (Dillon & Tomaka, 2010).

The results of Dillon and Tomaka's (2010) regression analyses are noteworthy as PTGPA was a significant predictor for NPTE performance in a moderate quality study (Table 2.3). Specifically, Dillon and Tomaka (2010) studied 72 PT students with a mean PTGPA of 3.56 ($SD=0.25$), 60% NPTE first-time pass rate, and 90.3% NPTE ultimate pass rate. When interpreting the findings of Dillon and Tomaka (2010), the low NPTE first-time pass rate may be in-part related to the high percentage of PT students who spoke English as their second language (44%) and low GRE-V upon admission ($GRE-V=379$). Even with the limitations, PTGPA's capacity to predict NPTE performance was significant. An important consideration in the interpretation of Dillon and Tomaka's (2010) results are the invalid methods used for data analysis. More specifically, multiple regression was performed for the entire sample ($N=72$) using expectation-maximization methods, which was inappropriate due to the high percentage (16.7) of missing NPTE scores and, consequently, may have influenced Dillon and Tomaka's (2010) results (Do & Batzoglou, 2008).

Continuing to investigate the relationship between PTGPA and NPTE performance, Adams et al. (2008) performed a moderate quality study with a sample of 126 PT students from seven cohorts (Table 2.3). The PT students had a mean PTGPA of 3.5 (range from 3.06 to 3.95) with an 82.5% NPTE first-time pass rate. The results of logistic regression identified PTGPA as a significant predictor ($p < 0.001$) of NPTE performance. PTGPA was able to correctly classify 97.1% of the PT students who passed the NPTE on their first-attempt and 31.8% of the PT students who failed the NPTE on their first-attempt (Adams, Glavin, Hutchins, Lee, & Zimmermann, 2008). In addition, the odds of passing the NPTE on the first-attempt increased by a factor of 2.27 for every 0.10 increase in PTGPA from 3.0 (Adams et al., 2008). These findings provide more evidence for the use of PTGPA to predict NPTE performance (Adams et al, 2008).

Despite significant correlation coefficients, Thieman's (2003) high quality retrospective cohort study identified PTGPA as a weak predictor of NPTE performance (Table 2.3). The sample of 121 PT students from multiple cohorts with a mean PTGPA of 3.73, an overall NPTE first-time pass rate of 92%, and a mean NPTE first-attempt scaled score of 653.9 is representative of PT programs across the U.S. (Thieman, 2003). Thieman (2003) identified significant Pearson product moment correlation coefficients between PTGPA and NPTE performance. Multiple regression analysis was performed with the significant correlation coefficient ($r = 0.32$, $p < 0.01$) between PTGPA and NPTE performance (Thieman, 2003). Multiple stepwise regression analysis was subsequently performed with a model that included UGPA, GRE total score, and PTGPA. Using stepwise multiple regression analysis, PTGPA was the third variable added to the model. The combination of UGPA, GRE, and PTGPA only accounted for 10.8% of the overall variance in NPTE scores (Thieman, 2003).

It is important to consider that the use of stepwise multiple regression presents the potential for bias, as Thieman (2003) was able to control the order the variables were added to the regression model. Since PTGPA had the highest correlation with NPTE performance, regression analysis with PTGPA as a lone predictor of NPTE performance should have been performed. Failing to perform the appropriate regression analysis is another indication of bias in Thieman's (2003) study and should be a cause for concern when interpreting their results.

Support for the continued use of PTGPA to predict NPTE performance comes from two studies that investigated the strength of the relationship between PTGPA and NPTE performance using correlations. The moderate quality study (Table 2.3) performed by Luedtke-Hoffman, Dillon, Utsey, and Tomaka (2012) investigated the relationship between PTGPA and NPTE scores. A large sample of 979 PT students from eight PT programs and four cohorts was studied. Regrettably, Luedtke-Hoffman et al. (2012) did not report the mean PTGPA or mean NPTE pass rates or scores, limiting the utility of their findings. The correlations between PTGPA and NPTE performance were calculated, and PTGPA demonstrated the largest significant correlation ($r=0.51$, $p<0.01$) with NPTE performance (Luedtke-Hoffman, Dillon, Utsey, & Tomaka, 2012).

Similar to Luedtke-Hoffman et al. (2012), Lewis' (2011) moderate quality study (Table 2.3) investigated the relationship between first-year PTGPA and overall PTGPA with NPTE performance. From a sample of 260 PT students from four institutions, first-year PTGPA ($r=0.42$, $p\leq 0.001$) and overall PTGPA ($r=0.34$, $p\leq 0.001$) were significantly correlated with NPTE performance (Lewis, 2011). Lewis (2011) stated that NPTE scores could not be predicted from regression analyses with PTGPA; however, no results were reported to support the conclusion. Although first-year and overall PTGPA were significantly correlated with NPTE scores, the failure to perform regression analyses or report the mean PTGPA, mean NPTE pass

rate, or mean NPTE scores severely limit the interpretation and usefulness of Luedtke-Hoffman et al.'s (2012) findings.

The evidence against PTGPA. While the majority of the literature supports the continued use of PTGPA as a significant predictor of NPTE performance, one study has reported dissenting results. In 2007, Vendrely performed a retrospective study to identify predictors of NPTE performance. Vendrely (2007) identified a significant Pearson product moment correlation coefficient ($r=0.33$, $p=0.04$) between PTGPA and NPTE performance. Subsequent logistic regression did not find PTGPA to be a significant predictor of NPTE performance. Data were analyzed on a sample of 42 PT graduates with a mean PTGPA of 3.67 ($SD=0.23$). The mean NPTE first-attempt score was 644.02 ($SD=45.03$) and NPTE first-time pass rate of 88.1%. That PTGPA was not a significant predictor of NPTE performance should be given credibility given the moderate quality of the study (Table 2.3).

In reviewing Vendrely's (2007) study, the resulting overall PTGPA correlation coefficient ($r=0.33$) is similar to Thieman's (2003; $r=0.32$) and Lewis' (2011; $r=0.34$). However, Thieman (2003) did identify PTGPA as a significant predictor of NPTE performance, but PTGPA's unique influence on NPTE performance was unclear. Lewis (2011) did not perform regression analysis with PTGPA to predict NPTE. As such, Vendrely's (2007) findings appear to be consistent with Thieman (2003) and Lewis (2011). Further data analysis is needed from Thieman (2003) and Lewis (2011) to accurately compare their findings with Vendrely's (2007) results.

Considerations for PTGPA. Typically, PTGPA has a narrow range (PTGPA of 3 to 4) and is susceptible to further grade inflation (Ling, Bochenek, & Burkander, 2015; Rojstaczer & Healy, 2012). Despite these limitations, the results of this systematic review of the literature have

identified PTGPA as the strongest and most consistent predictor of NPTE performance. Indeed, when PTGPA increases by 0.1 or more, then the NPTE performance increases by a large factor (Meiners, 2015). In predicting NPTE performance, PTGPA is perceived to not only represent PT students learning in their PT program, but also the essential noncognitive characteristics (e.g., persistence, creativity) needed to become a licensed PT (Guffey, et al., 2002).

While PTGPA is a significant predictor of NPTE performance, there are limitations to the practical use of PTGPA as an early indicator of PT students' capacity to pass the NPTE.

Although PTGPA is most commonly studied as the overall PTGPA, consideration should be given to the use of first-year PTGPA as an earlier predictor of NPTE performance so that student learning strategies can be modified to improve first-attempt NPTE pass rates. More well-designed research is needed to better understand first-year PTGPA's predictive capacity of NPTE performance. From the systematic review of the literature, only three of the ten studies investigated PTGPA following PT students first year in their PT program. The results from Meiners (2015), Dockter (2001), and Lewis (2011) confirm first-year PTGPA is a significant predictor for NPTE performance. First-year PTGPA should be seriously considered as an early indicator of PT students who are at risk of failing the NPTE on their first-attempt, so that PT programs could provide additional learning opportunities and resources to improve the NPTE rates for those students at risk of passing the NPTE on their first-attempt.

From the review of the literature, it is important to recognize that the lowest reported mean PTGPA was 3.5 (Adams et al., 2008), while the highest was 3.73 (Fell et al., 2015) on the 0.0 to 4.0 scale. Based upon the available evidence, PTGPA may be a strong predictor of NPTE performance due to the type of students admitted into PT programs. Since most PT students perform well through the PT curriculum and most successfully pass the NPTE on the first-

attempt, the current admission process appears to be selecting applicants who are likely to succeed on the NPTE. Across all of the studies reviewed, the lowest and highest first-attempt NPTE pass rates were 60% and 92% respectively. While the majority of students from different PT programs pass the NPTE, internal and external benchmarks set by program directors and CAPTE require the capacity to identify those students who, at the end of their 3rd year in their PT program, will pass the NPTE on their first-attempt.

To complicate the interpretation of the results, the systematic literature review identified three studies that did not report all of the relevant information. Both Dockter (2001) and Luedtke-Hoffman et al. (2012) failed to report the NPTE pass rates, while Lewis (2011) and Luedtke-Hoffman et al. (2012) did not report the mean PTGPA's. Omitting the NPTE pass rates and mean PTGPA's limits the interpretation and usefulness of each of the studies and raises concern about the quality of the journal review process.

From the systematic review of the literature, Kosmahl (2005), Luedtke-Hoffman et al. (2012), and Lewis (2011) all neglected to perform, or report, their regression analysis to determine PTGPA's capacity to predict NPTE performance, even though a significant correlation between the variables was identified for each study. Disregarding the appropriate statistical analyses and the under-reporting of highly relevant information limits the utility of the results of each study.

In the future, PT programs should review students' first-year PTGPA to identify students who are likely to pass, and students who are likely to fail, the NPTE on their first-attempt. Identifying those students who are likely to fail the NPTE following their first-year of professional education should result in targeted educational assistance and support to those students to maximize their performance on the NPTE. In doing so, PT programs could increase

the likelihood of their students successfully passing the NPTE on the first-attempt. The consistency of PTGPA's predicative capacity of NPTE performance was surprising due to the limitations of reduced variance and grade inflation with PTGPA. However, the findings from the systematic literature review continue to support the use of first-year and overall PTGPA to predict NPTE performance.

Clinical Performance Scores

PT students are expected to learn and develop a variety of clinical skills to perform successfully in the clinical environment. A significant portion of the curriculum in PT programs curriculum is dedicated to clinical learning, as students must complete a minimum of 32 weeks of clinical education (Commission on Accreditation of Physical Therapy Education, 2015). Currently, the only comprehensive examination for PT graduates to gain licensure and practice in the U.S. is the NPTE. The NPTE is a 250-question multiple choice exam that students have five hours to complete. PT graduates are not required to complete a comprehensive clinical examination to practice as a licensed PT and yet students spend a minimum of 32 weeks in the clinical environment during their PT education. Without a standardized standalone competency clinical examination for PT graduates it is presumed the NPTE assesses graduates' clinical capabilities and competencies or that external assessment of clinical education and learning is not important. The relationship between PT students' clinical performance scores and NPTE performance is important for PT programs, students, and the profession. Nine studies have investigated the relationship between PT students' clinical performance and NPTE performance.

Overall, PT students' clinical performance does not accurately and precisely predict NPTE performance, nor does NPTE performance predict clinical performance. Simply passing the NPTE does not equate to the expectations set by CAPTE for each program, nor does passing

the NPTE equate to the expected level of clinical competency required by an entry level PT at their first clinical job as a PT.

The Physical Therapist Clinical Performance Instrument (PT CPI) is the most commonly utilized measurement tool by clinical instructors (CI) to assess each PT students' clinical learning and performance. The current version of the PT CPI now consists of 18 performance criteria, rather than 24 performance criteria in the old version of the PT CPI (Adams et al., 2008; Roach et al., 2012). The PT CPI is scored by the student's CI. Each CI rates each performance criteria on a visual analog scale, which is translated to a numerical scale of either 0 to 10 or 0 to 100. Since 1997, the PT CPI has been the instrument of choice for the American Physical Therapy Association (APTA) for measuring PT student learning and performance during their clinical experiences (Task Force for the Development of Student Clinical Performance Instruments, 2002). Most commonly reported in the literature are students' terminal PT CPI scores; yet, PT CPI scores are reported for up to four separate clinical experiences. While the PT CPI is the most common utilized measurement tool, two separate studies (Dillon & Tomaka, 2010; Luedtke-Hoffman, 2012) used the Physical Therapist Manual for the Assessment of Clinical Skills (PT MACS) to assess students' clinical learning and performance.

There are limitations in using PT students' clinical performance scores as a predictor variable for NPTE performance. PT students are assessed and rated on their clinical learning and performance by their CI. The PT students CI can be different for each student and can change with each clinical experience. With multiple CI rating each PT student, there is an increased probability of significant variance when scoring the same student. Yet, the PT CPI has demonstrated psychometric properties with high internal reliability with Cronbach alpha's ranging from 0.75 to 0.99 and interrater reliability coefficients for individual items ranging from

0.27 to 0.76 (Adams, et al., 2008; Roach et al., 2012; Task Force for the Development of Student Clinical Performance Instruments, 2002).

The evidence in support of clinical performance scores. Only two studies support the use of PT students' clinical performance scores to predict NPTE performance. In 2010, Cook completed a high-quality study to predict NPTE performance from students' clinical performance scores (e.g. PT CPI scores; Table 2.3). The study consisted of 75 PT students from seven cohorts with first-attempt NPTE scores ranging from 521 to 709 out of a possible 800 (Cook, 2010). Neither the mean nor standard deviation of NPTE scores were reported.

To predict NPTE scores, Cook (2010) reviewed students' PT CPI scores from their first clinical experience. Cook (2010) performed linear regression analysis to identify the individual relationships among the 24 performance criteria in the PT CPI with NPTE first-attempt scores. From the regression analysis, the PT CPI performance criteria Professional Behaviors ($\beta=0.22$, $p\leq 0.05$) and Professional/Social Responsibility ($\beta=0.32$, $p\leq 0.006$) explained 20.5% of the variance in NPTE first-attempt scores (Cook, 2010). These significant findings provide evidence for the use of students' first clinical PT CPI scores to predict NPTE performance; yet, there are limitations to Cook's (2010) study that must be considered.

While Cook (2010) reported the range of NPTE first-attempt scores, a limitation of their findings was the failure to report the mean and SD of the NPTE scores. Without the mean or SD of the PT students NPTE scores, it is difficult to compare their findings with other studies, making interpretation of their results difficult. To account for the missing data, Cook (2010) used a conservative approach to replace PT CPI missing data points across the 24 performance criteria with the mean value for the respective performance criterion (Tabachnick & Fidell, 2001).

Building upon Cook's (2010) findings, Luedtke-Hoffman et al. (2012) completed a moderate quality study investigating the relationship between PT students' clinical performance scores using the Physical Therapist Manual for the Assessment of Clinical Skills (PT MACS) and NPTE results. Luedtke-Hoffman et al. (2012) performed correlation analysis with 967 PT students from 8 institutions in Texas to identify significant relationships between students first and final clinical experience PT MACS scores and NPTE performance.

Luedtke-Hoffman et al (2012) failed to find a significant correlation between students first ($r=0.03$) and final ($r=0.06$) clinical experience PT MACS scores and NPTE performance. However, there was a significant correlation between the PT MACS Outcomes Assessment section and the NPTE's Prognosis and Outcomes section ($r=0.23$, $p=0.006$). In addition, the PT MACS and NPTE Evaluation and Diagnosis sections ($r=0.10$, $p=0.009$) were significantly correlated (Luedtke-Hoffman et al., 2012). Based upon the significant correlations, the results from Luedtke-Hoffman et al.'s (2012) study lend a measure of support for a significant relationship among the Outcomes Assessment and Evaluation and Diagnosis sections of the PT MACS and NPTE performance.

When reviewing Luedtke-Hoffman et al.'s (2012) results, it is important to recognize the PT MACS was used to assess students' clinical performance, not the widely used PT CPI. Use of the PT MACS, therefore, must be considered when interpreting the results, as the majority of PT programs use the PT CPI to measure students' clinical learning and performance. A significant limitation of Luedtke-Hoffman et al.'s (2012) study was their failure to report the mean NPTE scores. Without the reference NPTE scores and the use of a different clinical assessment instrument, it is difficult to generalize the Luedtke-Hoffman et al.'s (2012) findings to other PT

students and programs. Overall, the results of the systematic literature review found limited evidence to support the use of PT students' clinical performance scores to predict NPTE results.

The evidence against clinical performance scores. The systematic literature review has identified seven studies that reported no significant relationship between PT students' clinical performance scores and NPTE performance. In 2001, Edmondson developed a multiple regression model to predict NPTE scores from the average PT CPI scores for students' first and second long-term clinical experience. The multiple regression model ($R^2=0.99$) results found students average PT CPI scores for their first (power=0.28) and second (power=0.34) long-term clinical experiences did not predict NPTE scores (Edmondson, 2001). Their sample had a mean NPTE score of 615.81, with a high NPTE score of 704. Edmondson's (2001) study was limited in value. The SD for the mean NPTE scores, the NPTE first-time pass rate, and the NPTE overall pass rate was not reported. Failing to report the SD, which should accompany the mean, and NPTE pass rates makes comparisons among studies difficult. Further, the sample size ($N=21$) was small which likely contributed to the high R^2 value ($R^2=0.99$). The high R^2 value suggests that a nonparametric test should have been performed instead of multiple regression analysis. The limitations in reporting of required data and the selection of data analysis raises concerns as to the utility of their results.

From a sample of 102 PT students with a 79.4% NPTE first-time pass rate and a mean NPTE score of 633.73 ($SD=47.01$), Dreeben (2003) reported that students' PT CPI scores were not a significant predictor of NPTE performance. Dreeben (2003) completed a moderate quality study investigating each of the PT CPI's 24 items capacity to predict NPTE scores. Five items on the PT CPI were significantly correlated ($p<0.05$) with NPTE scores: (a) safety ($r=0.18$), (b) examination ($r=0.14$), (c) evaluation/diagnosis/prognosis ($r=0.13$), (d) plan of care ($r=0.7$), and

(e) treatment/intervention ($r=.12$). In addition, Dreeben (2003) reported that students' terminal PT CPI score was significantly correlated ($r=0.27$, $p<0.01$) with NPTE scores. Subsequent multiple regression analysis significantly correlated PT CPI items ((a) safety, (b) examination, (c) evaluation/diagnosis/prognosis, (d) plan of care, and (e) treatment/intervention) failed to significantly predict ($R^2=0.06$) NPTE scores. An analysis of variance (ANOVA) confirmed the non-significant ($F=1.24$, $p=0.30$) relationship among PT CPI items and NPTE scores (Dreeben, 2003).

In 2005, Kosmahl found a non-significant correlation ($r= -0.09$) between students' PT CPI scores from their terminal clinical experience and their NPTE scores. Since the relationship between PT CPI scores and NPTE scores was non-significant, PT CPI scores were not part of the regression analysis to predict NPTE scores. Kosmahl (2005) studied 92 PT graduates with a mean terminal PT CPI score of 9.89 ($SD=0.24$) and mean NPTE score of 659.74 ($SD=37.29$). While the NPTE scores were reported, the NPTE first-time and overall pass rates were not; thereby limiting the generalizability of the results to other PT programs. Overall, Kosmahl's (2005) results do not support the use of students' PT CPI scores to predict NPTE scores.

Providing supporting evidence to Kosmahl's (2005) findings, Vendrely (2007) completed a moderate quality study that investigated the relationship between students' PT CPI scores from their terminal clinical experience and NPTE performance (Table 2.3). The sample included 42 PT students with a mean PT CPI score of 96.69 ($SD=1.60$), mean NPTE score of 644.02 ($SD=45.03$), and NPTE first-time pass rate of 88.10% (Vendrely, 2007). There was no significant Pearson product moment correlation coefficient between terminal clinical experience PT CPI scores and NPTE performance ($r=0.30$, $p=0.06$). Logistic regression analysis found that PT CPI scores had no significant regression coefficient ($B=0.01$, $p=0.95$), indicating the PT CPI

does not predict NPTE performance (Vendrely, 2007). The findings from Vendrely's (2007) study do not support the use of the PT CPI scores to predict NPTE performance. In reviewing Vendrely's (2007) study, the small sample size (N=42) and the lack of variance in PT CPI scores ($\mu=96.69$, $SD=1.60$) may have contributed to their findings.

Adams et al. (2008) performed a moderate quality study with a sample of 126 PT students from seven cohorts (Table 2.3) and with a NPTE first-time pass rate of 82.5%. Adams et al. (2008) used three clinical experiences PT CPI scores to predict NPTE performance. Instead of using the mean value of the 24 items on the PT CPI, Adams et al. (2008) performed exploratory factor analysis (EFA) to identify underlying factors in the PT CPI. The results of EFA identified three factors for the PT CPI: (a) integrated patient management, (b) professional practice, and (c) career responsibilities. For each clinical experience, the mean total PT CPI scores and the factor scores were used to predict NPTE scores (Adams et al., 2008). Their results using logistic regression analysis found the total and factor PT CPI scores for all three clinical experiences were not significant predictors of NPTE scores (Adams et al., 2008).

While finding no significant relationship between PT CPI scores and NPTE performance, it is important to recognize a limitation of the data analysis methodology used by Adams et al. (2008). Specifically, EFA was performed using principal component analysis rather than common-factor analysis (CFA). CFA would have been more appropriate due to the assumed correlations between the underlying factors ((a) integrated patient management, (b) professional practice, and (c) career responsibilities; Gorsuch, 1983). To adequately detect the underlying factors in the PT CPI, CFA is a more effective method than principal component analysis (Floyd & Widaman, 1995). Limiting principal component analysis is that it accounts for all of the variance in the model being studied and thus produces inflated loadings or regression weights

(Gorsuch, 1983). As such, it is extremely difficult to assume the studied variables are free of error in social science research.

CFA does not have the same limitations as principal component analysis. Consequently, CFA is a more conservative and accurate method for identifying underlying factors within a model; as such, CFA should have been performed by Adams et al. (2008) rather than principal component analysis. The use of principal component analysis for EFA rather than CFA may have confounded the results of Adams et al.'s (2008), and therefore, the interpretation of their results is limited.

Dillon & Tomaka (2010) also performed a moderate quality study that investigated the relationship between PT students' first and last clinical performance scores and NPTE performance (Table 2.3). To measure students' clinical performance and learning, Dillon and Tomaka (2010) used the PT MACS instead of the PT CPI. They identified a significant correlation between students first clinical experience PT MACS scores and NPTE scores ($r=0.24$, $p<0.05$), and no significant correlation between students' final clinical experience PT MACS scores and NPTE scores ($r=0.07$; Dillon & Tomaka, 2010). With a significant correlation between students first clinical experience PT MACS scores and NPTE scores, the first clinical experience PT MACS scores were used in multiple and logistic regression models to predict NPTE performance. Dillon and Tomaka (2010) found that despite the significant correlation, the students first clinical experience PT MACS scores were not significant predictors for NPTE scores ($B=0.04$, $p=0.67$) nor for NPTE first-time pass rates ($B= -0.07$, $p=0.69$, $OR=0.94$).

When reviewing Dillon and Tomaka's (2010) results, it is important to recognize that their sample of 72 PT students had a mean first clinical experience PT MACS score of 4.78 ($SD=1.80$), mean final clinical experience PT MACS score of 6.72 ($SD=1.50$), mean NPTE first-

time score of 603.05 (SD=56.68), 60% NPTE first-time pass rate, and 90.3% NPTE ultimate pass rate. Caution should be taken when interpreting their results from the sample. The PT students had a low NPTE first-time pass rate should have been considered. To account for their missing data, Dillon and Tomaka (2010) performed expectation-maximization to impute missing NPTE scores. “The expectation-maximization imputation method is a deterministic iterative algorithm that determines the maximum likelihood estimates of the parameters of the distribution which the complete (missing and observed) data are assumed to follow” (Ghomrawi et al., 2011, p. 3). The use of expectation-maximization was not an appropriate imputation method since it was designed for large datasets or a small amount of missing data. In Dillon and Tomaka’s (2010) study, greater than 5% of the scores were missing for the NPTE and exceeding the definition of “small”, thereby limiting the effectiveness of the expectation-maximization imputation method with the small dataset (N=72; Do & Batzoglou, 2008).

A recent study to investigate students’ clinical performance scores to predict NPTE performance was by Meiners’ (2015). Specifically, Meiners studied 122 PT students from three cohorts with a mean PT CPI score of 7.76 (SD=0.95) from their first long term clinical experience which occurred during the summer of the students third year in the PT program. The students had a mean NPTE first-attempt score of 660.6 (SD=53.37), with a 72.6% NPTE first-time pass rate, and an 86.3% NPTE overall pass rate. To determine the relationship between students’ PT CPI scores and NPTE performance Meiners (2015) performed multiple regression analysis and found that PT CPI scores did not significantly predict NPTE scores ($\beta=0.04$, $p=0.70$). In addition, Meiners (2015) performed logistic regression analysis and found PT CPI scores were not a significant predictor of NPTE pass rates (OR=0.18-1.98, $p=0.39$). Meiners

(2015) results provide supporting evidence that there is no significant relationship between PT CPI scores and NPTE performance.

Considerations for clinical performance scores. As PT students are expected to learn and develop a range of clinical skills to perform successfully in the clinical environment, it is presumed those clinical skills will be appropriately assessed by the NPTE. The systematic review of the literature identified nine studies that investigated the relationship between PT students' clinical performance scores and NPTE performance. Regardless of PT CPI or PT MACS usage, PT students' clinical performance scores do not appear to be a significant predictor of NPTE performance. The findings from the systematic literature review are concerning given the emphasis on clinical learning and clinical performance in PT programs. As such, it suggests that PT students' clinical learning and performance does not appear as a factor in NPTE performance. To extrapolate, the results of the studies reviewed imply that in the current clinical environment it is entirely possible that there are practicing PT's that passed the NPTE on their first-attempt; yet performed poorly in the clinical environment while enrolled in a PT program. These findings raise questions regarding the usefulness of the NPTE and the validity of the PT CPI to measure students' clinical learning and performance.

There are limitations to the studies reviewed from the systematic literature review. Three studies investigated PT students' early clinical performance scores, while, six studies investigated PT students' terminal clinical performance scores, all with the PT CPI. There appears to be a ceiling effect for the terminal PT CPI scores, as students are expected to be rated at entry-level on the PT CPI upon completion of their clinical experience. As a result, there is an increased risk of a type II error (Lomax & Hahs-Vaughn, 2012).

From the reviewed literature, six studies investigated the overall mean scores of the PT CPI to predict NPTE performance. The PT CPI has either 18 or 24 items. By studying the overall mean PT CPI score, information is lost on the value of each of the items capacity to predict NPTE performance. Future research is warranted to determine the capacity of each PT CPI item to predict NPTE performance. In addition, the PT CPI consists of three factors as identified by Adams et al. (2008) and Roach et al. (2012); however, Adams et al. (2008) and Roach et al. (2012) erroneously performed principal components EFA rather than common factor EFA or confirmatory factor analysis, which has likely confounded their findings. More research is needed on the PT CPI to confirm the number of proposed factors and validate the assessment instrument.

Taken together, the results of the systematic literature review do not support the use of students' clinical performance scores, as currently structured, to predict NPTE performance regardless of when the clinical experience occurs or what type of assessment instrument is used.

Noncognitive Physical Therapy Student Variables

The predictive capacity of PT student variables on NPTE performance has primarily focused on cognitive variables (e.g., PTGPA and clinical performance scores). However, it is presumed that there is an array of noncognitive PT student variables that influence NPTE performance (Guffey, 2000). The systematic literature review identified eight studies that investigated reading comprehension, morphological awareness, emotional intelligence, critical thinking, level of reflection, conscientiousness, and task coping to predict NPTE performance. The results of the systematic literature review support reading comprehension and emotional intelligence as significant predictors of NPTE performance.

It is important to acknowledge the limitations of the systematic literature review when examining noncognitive PT student variables to predict NPTE performance. A noteworthy limitation was that only two studies (Aldridge, Keith, Sloas, & Mott-Murphree, 2010; Moran, 2012) used the same assessment instrument to measure noncognitive PT student variables. As such, only one noncognitive variable (reading comprehension) was investigated in more than one study. The other five studies each used different assessment instruments to investigate different noncognitive PT student variables. The inconsistency in the selected noncognitive PT student variables, and the variety of assessment instruments, have limited the evidence for each noncognitive variable to predict NPTE scores.

The evidence in support of noncognitive physical therapy student variables. With only four identified studies, there is limited evidence for the use of noncognitive PT student variables to predict NPTE performance. Yet, two of the studies identified reading comprehension as a significant predictor of NPTE performance. The moderate quality retrospective study by Aldridge, Keith, Sloas, and Mott-Murphree (2010) investigated the relationship between reading comprehension and NPTE performance (Table 2.3). To measure PT students reading comprehension, the Nelson Denny Reading Test (NDRT) was administered to 67 PT students at the initiation of their PT program (Aldridge et al., 2010). The sample of PT students had a mean NDRT score of 232.59 (range=178-253) out of a possible 258, with a 68.7% NPTE first-time pass rate (Aldridge et al., 2010).

NDRT scores were significantly correlated with overall NPTE scores ($r=0.52$, $p<0.001$) and NPTE scores for first-time test-takers ($r=0.51$, $p<0.001$). Based upon the Pearson product moment correlation coefficients, NDRT scores accounted for 27% of the variance in NPTE scores and 25% of the variance in NPTE scores for first-time test-takers (Aldridge et al., 2010).

However, NDRT scores had no significant relationship with NPTE scores for unsuccessful first-time test-takers ($r=0.39$, $p=0.08$; Aldridge et al., 2010). In short, Aldridge et al.'s (2010) findings support a significant relationship between reading comprehension and NPTE performance, but only for PT students who successfully passed the NPTE on their first-attempt.

Supporting Aldridge et al.'s (2010) findings, Moran (2012) performed a high-quality study that investigated reading comprehension and morphological awareness to predict NPTE performance (Table 2.3). Reading comprehension was again assessed with the NDRT, while morphological awareness was assessed with the Medical Morphology Test (MMT). "The MMT was created to test morphological awareness of medical terminology and vocabulary commonly used in the clinical and academic practice settings of physical therapy" (Moran, 2012, p.1). A sample of 30 PT students from seven cohorts with a 65.6% NPTE first-time pass rate and a 93.8% NPTE overall pass rate was studied. PT students had a mean NDRT total score of 230.17 ($SD=16.31$) and a mean MMT total score of 97.47 ($SD=13.02$; Moran, 2012).

Moran's (2012) analysis found NDRT scores were significantly correlated with the number of attempts PT students needed to pass the NPTE ($r= -0.63$, $p<0.01$). As such, discriminant analyses revealed 86.7% of the time NDRT scores correctly classified PT students into one of two groups: (a) first-attempt NPTE passers or (b) multiple-attempt NPTE test-takers (Moran, 2012). Further, analysis revealed no significant correlation between MMT scores and the number of attempts PT students needed to pass the NPTE ($r= -0.25$, $p=0.17$). A subsequent regression analysis found that PT students MMT scores accounted for only 3% of the variance in the number of attempts needed to pass the NPTE (Moran, 2012). However, subsequent discriminant analysis identified that MMT scores could correctly classify PT students in the

correct group 75% of the time (Moran, 2012). Overall, Moran's (2012) results support the continued use of the NDRT (not the MMT) to predict first-attempt NPTE performance.

When interpreting the findings from Moran (2012), and weighing the evidence, it is important to consider that 22 out of the 30 participating PT students' NDRT and MMT scores were taken one to six years post-graduation. Due to data collection occurring post-graduation, the results may not represent the noncognitive behaviors and skills that the students developed during their PT program years. In addition, the low NPTE first-time pass rate (65.6%) should be considered when interpreting Moran's (2012) results. Clearly, more research, with more stringent design and analysis is indicated to further identify the relationship between reading comprehension and morphological awareness with NPTE performance.

Supporting the use of noncognitive PT student variables to predict NPTE performance, Guffey (2000) used the Noncognitive Questionnaire-Revised (NCQ-R) to investigate the eight noncognitive PT student variables to predict NPTE performance. The NCQ-R consisted of the noncognitive domains: (a) positive self-concept, (b) realistic self-appraisal, (c) support for academic plans, (d) leadership, (e) long range goals, (f) community ties, (g) understands racism, and (h) academic familiarity (Guffey, 2000). Linear regression analysis found PT students NCQ-R total scores accounted for only 3.2% ($p=0.19$) of the variance in NPTE scores (Guffey, 2000). Linear regression analysis of each of the eight noncognitive domains revealed that the noncognitive domain 'long range goals' was a significant predictor of NPTE scores ($R^2=0.08$, $p=0.03$; Guffey, 2000). Following linear regression analysis, multiple regression analysis with four noncognitive domains (long range goals, leadership, community ties, and academic familiarity) accounted for 13.4% ($p=0.02$) of the variance in NPTE scores.

With a moderate quality study, Huhn and Parrott (2017) investigated the relationship between PT students' clinical reasoning skills and NPTE performance (Table 2.3). PT students' clinical reasoning skills were assessed in their first few weeks in their PT program using the Health Sciences Reasoning Test (HSRT); a standardized test that consists of 33 multiple-choice questions with five subscales: (a) induction, (b) deduction, (c) inference, (d) analysis, and (e) evaluation. Scores on the (HSRT) greater than 24 was an indication of good clinical reasoning skills, while scores less than 15 were indicative of poor clinical reasoning skills (Facione & Facione, 2007).

Huhn and Parrott (2017) studied 160 PT students from four cohorts. The mean HSRT scores for cohort 1 was 21.64 (SD=3.70), cohort 2 was 21.54 (SD=3.09), cohort 3 was 21.62 (SD=3.49), and cohort 4 was 23.91 (SD=3.56). The cohorts had mean NPTE scores of 656.38 (SD=44.40), 642.21 (SD=37.99), 648.06 (SD=38.88), and 664.43 (SD=51.63) respectively. HSRT was significantly correlated with NPTE scores ($r=0.43$, $p<0.001$). Subsequent random linear modeling found that PT students' HSRT scores also significantly ($p=0.001$) predicted NPTE scores. Specifically, for every one-point increase on a student's HSRT, there was an increase by 2.47 points in their NPTE score (Huhn & Parrott, 2017).

When interpreting the significant relationships between HSRT scores and NPTE scores, it is important to consider the reliability of the HSRT. Huhn and Parrott (2017) identified the reliability coefficients of the HSRT using Kuder-Richardson formula 20 (*KR-20*). The *KR-20* reliability coefficient is a special version of alpha that is meant to be used with dichotomous questions (DeVellis, 2017). However, the HSRT is a multiple-choice examination, so the *KR-20* may be an inappropriate reliability measurement (DeVellis, 2017). Measuring the reliability of the HSRT, Huhn and Parrott (2017) found *KR-20* coefficients for each of the five subscales: (a)

induction=0.76, (b) deduction=0.71, (c) inference=0.52, (d) analysis=0.54, and (e) evaluation=0.77. The reliability coefficients for inference (0.52) and analysis (0.54) are regarded as low and therefore represent poor reliability. Using the HSRT with poor reliability in some domains is concerning and needs to be considered when interpreting the results reported by Huhn and Parrott (2017).

In support of noncognitive PT student variables as a predictor of NPTE performance, Lewis (2011) performed a moderate quality study that investigated the relationship between students' emotional intelligence and NPTE performance (Table 2.3). PT students' emotional intelligence was assessed with the Mayer-Salvory-Caruso Emotional Intelligence Test (MSCEIT). The MSCEIT consists of two subscale scores, experiential emotional intelligence and reasoning emotional intelligence, and the total score. The two subscale scores are subdivided into four branch scores, two affiliated with experiential emotional intelligence and two affiliated with reasoning emotional intelligence (Mayer, Salovey, & Caruso, 2002).

Lewis (2011) studied 151 PT students with a NPTE first-time pass rate of 90.7% and a NPTE overall pass rate of 93.9%. The emotional intelligence of the PT students was repeatedly studied over the three years that the students were enrolled in their PT program. The mean MSCEIT total score for year one was 102.7 (SD=10.1), year two was 104.2 (SD=10.9), and year three was 103.2 (SD=11.6) out of a possible total score of 167 (Lewis, 2011). To assess the relationship between PT students' MSCEIT total scores and NPTE performance, Pearson product moment correlations were calculated. However, due to a non-significant correlation coefficient between PT students' MSCEIT total scores and NPTE scores, the correlation coefficient, and any subsequent regression analysis results were not reported (Lewis, 2011).

While MSCEIT total scores did not have a significant relationship with NPTE scores, Lewis (2011) did perform a one-way ANOVA to determine if there was a difference between MSCEIT scores for the PT students who passed the NPTE on their first-attempt versus those PT students who failed the NPTE on their first-attempt. There were significant differences on the MSCEIT branch 1 of the experiential emotional intelligence subscale ($p=0.04$) and the MSCEIT total score ($p=0.05$) between those PT students who passed on their first-attempt and those PT students who failed on their first-attempt on the NPTE (Lewis, 2011).

The data suggest that emotional intelligence may be a predictor of NPTE performance. However, interpreting Lewis' (2011) results is again difficult due to the failure to report the correlation coefficients or regression analysis. Therefore, a more rigorous design and analysis approach is required before emotional intelligence can be regarded as a significant predictor of NPTE scores. Further concern about the findings of Lewis (2011) is that (a) NPTE scores were not listed in the study and (b) correlation coefficients and regression analysis were not reported. Taken together, the results presented by Lewis (2011) have limited value.

The evidence against the use of noncognitive PT student variables. The systematic literature review identified three studies providing evidence that the noncognitive PT student variables do not predict NPTE performance. In 2007, Vendrely studied the relationship between critical thinking skills test scores, using the California Critical Thinking Skills Test (CCTST), and NPTE performance in 42 PT students. The PT students had a mean CCTST score of 20.74 ($SD=0.64$) out of a possible 34, an 88.1% NPTE first-time pass rate, and a mean NPTE score of 644.02 ($SD=45.03$; Vendrely, 2007). Analysis identified a significant correlation between PT students' CCTST scores and NPTE scores ($r=0.31$, $p<0.05$). Subsequent logistic regression was

performed to predict NPTE first-time pass rates, concluding that PT students' CCTST scores did not predict NPTE first-time pass rates ($\beta=0.11$, $p=0.42$).

While a significant correlation was identified between critical thinking scores and NPTE performance, the results of Vendrely's (2007) study do not support PT students' critical thinking scores, as measured by the CCTST, as a predictor of NPTE performance. Given that CCTST scores significantly correlated with NPTE scores, multiple regression analysis was warranted, but not performed. Failure to perform multiple regression analysis to predict PT students' NPTE scores limits the utility of Vendrely's (2007) findings.

Complicating the evidence from Vendrely's (2007) study was that the CCTST was administered during PT students last week in their PT program. Consequently, the findings do not serve as an early student predictor of NPTE performance. Administering the CCTST earlier in the curriculum could also have aided those PT programs to identify students who have lower critical thinking skills and then provided the necessary remediation.

Further, evidence that PT student cognitive variables do not predict NPTE performance was provided by Cook (2010). Cook (2010) found a no significant correlation between student reflections and NPTE performance ($r=0.10$). The high-quality study undertaken by Cook (2010) reviewed 75 PT students' journal reflections from their first and final clinical experience (Table 2.3). The journal reflections were assessed and scored on a 1 to 3 scale, with 1 being non-reflection, 2 being reflection, and 3 being critical reflection, by multiple members of PT faculty.

While PT students' journal reflections were assessed, Cook (2010) failed to report PT students mean journal reflection scores from the first and final clinical experiences and NPTE scores. Only the range of NPTE scores was reported (521-709). Whether by accident or intent, neither the NPTE first-time nor overall pass rates were reported. Without the mean journal

reflection scores, NPTE scores, and NPTE pass rates the usefulness of the findings presented by Cook (2010) is significantly limited. As an aside, the quality of the decision-making process leading to the completion of Cook's (2010) dissertation is in question.

From a sample of 49 PT students with an 89.8% NPTE first-time pass rate, Galleher et al. (2012) studied PT students' conscientiousness and task copings capacity to predict NPTE performance. The Big Five Inventory (BFI) consists of 44 items to measure conscientiousness and the Coping Inventory for Stressful Situations (CISS) with 48 items to measure task coping was administered. PT students had a mean BFI score of 36.27 (SD=5.15), and a mean CISS score of 59.57 (SD=8.49; Galleher et al., 2012).

Using logistic regression analysis, Galleher et al. (2012) found neither BFI ($r=0.09$, $p=0.56$) nor CISS ($r= -0.02$, $p=0.87$) scores were significant predictors of NPTE first-time pass rates. More specifically, BFI ($r=0.09$, $p=0.56$) and CISS ($r= -0.02$, $p=0.87$) had no significant correlations with NPTE performance (Galleher et al., 2012). The findings from Galleher et al. (2012) do not support the use of the noncognitive PT student variables conscientiousness or task coping to predict NPTE performance. Unfortunately, Galleher et al. (2012) did not report the mean NPTE scores or regression coefficients, thereby, limiting the interpretation of the results and data analyses. If the NPTE scores had been reported, multiple regression analysis could have been performed to predict PT students NPTE scores.

The evidence that noncognitive PT student variables do predict NPTE performance is underwhelming. Each of the three studies investigated different noncognitive PT student variables. Without the consistent selection of noncognitive PT student variables and with the variations in design and analysis, it is difficult to make any conclusions on the utility of the noncognitive PT student variables to predict NPTE performance.

Considerations for noncognitive PT student variables. Studying multiple noncognitive PT student variables is important; however, the lack of consistency of variables across eight studies makes it difficult to reach any valid conclusions. From multiple studies, reading comprehension was the only noncognitive PT student variable that had a significant relationship with NPTE performance (Aldridge et al., 2010; Moran, 2012). The valid and reliable assessment of noncognitive variables remains difficult due to the multiple psychometric properties associated with each noncognitive variable. Due to the potential number of noncognitive variables that could be studied, PT education should consider prioritizing investigations with noncognitive variables that have demonstrated a significant relationship with NPTE performance (e.g., reading comprehension, emotional intelligence).

Comprehensive Exam Scores

While PT students' cumulative academic performance during their coursework is suggested to predict NPTE performance (Dockter, 2001; Thieman et al., 2003), one measure of PT students' cumulative academic performance is a comprehensive exam score. A comprehensive exam is a test administered by the PT program to assess students' PT content knowledge prior to graduation. The comprehensive exam is designed to assess PT students' global learning and is taken towards the end of the PT curriculum, and intended to replicate the NPTE.

The systematic literature review identified only two studies, Edmondson (2001) and Kosmahl (2005), that have investigated the relationship between PT students' comprehensive exam scores and NPTE performance. It is assumed that cumulative academic performance predicts NPTE performance, therefore, the better the comprehensive exam score, the more likely the student will pass the NPTE on their first-attempt. The studies investigating the relationship

between comprehensive exam scores and NPTE performance had sample sizes ranging from 92 to 107, with mean NPTE first-time scores ranging from 615.81 to 659.74. Regrettably, the NPTE pass rates were not reported for either study.

Each comprehensive exam is individualized for each PT program; however, all PT programs share common accreditation standards that require specific content to be taught throughout a PT curriculum. As such, comprehensive exams are used to assess student learning and “act as a gate keeper” for the terminal clinical experience.

The evidence in support of comprehensive exam scores. Evidence supporting the use of PT students’ comprehensive exam scores to predict NPTE performance, Edmondson (2001) performed a moderate quality study with a sample of 107 PT students from four cohorts with a mean NPTE score of 615.81 (highest score was 704 out of 800). Edmondson (2001) performed linear regression analysis and found PT students comprehensive exam scores accounted for 33% of the variance of NPTE scores ($\beta=6.55$, $p<0.01$). In short, PT students’ comprehensive exam scores were a significant predictor of NPTE performance.

As encouraging as Edmondson’s (2001) results are, it is important to consider that NPTE pass rates, mean comprehensive exam scores, and the SD of NPTE scores were not reported. Omitting such pertinent information for review diminishes the quality and utility of Edmondson’s (2001) findings. Further, all of the PT students who sat for the comprehensive exam passed. Since all of the PT students passed the comprehensive exam, a negative skew resulted which is indicative of the high mean score (615.81; Edmondson, 2001).

In 2005, Kosmahl (2005) performed a moderate quality study to determine the relationship between PT students’ comprehensive exam scores and NPTE performance (Table 2.3). Kosmahl (2005) studied a sample of 92 PT students with a mean comprehensive exam

score of 86.8 (SD=5.88) out of 100 and mean NPTE score of 659.74 (SD=37.29). Kosmahl (2005) found a significant correlation between PT students' comprehensive exam scores and NPTE scores ($r=0.62$, $p<0.001$). Following correlation analysis, stepwise multiple regression analysis was performed to predict NPTE scores. Alone, the PT students' comprehensive exam scores accounted for 37.4% of the variance in NPTE scores. When PT students' comprehensive exam scores were combined with PTGPA, the regression model accounted for 46.5% of the variance in NPTE scores (Koshmahl, 2005). The findings indicate PT students' comprehensive exam scores are significant predictors of NPTE performance. However, it is important to note again the omission of PT students NPTE pass rates. Failing to report the NPTE pass rates makes it difficult to compare the results to other PT students and programs.

Considerations for comprehensive exams scores. With limited evidence and of varying quality, PT students' comprehensive exam scores have been reported to be a significant predictor of NPTE performance. Both Edmondson (2001) and Kosmahl (2005) found PT students' comprehensive exam scores accounted for more than 30% of the variance in NPTE scores. However, both studies had serious design and analysis deficiencies, making any conclusions from their results tenuous.

The findings from the systematic literature review are important considering not every PT program administers a comprehensive exam prior to the terminal clinical experience. As comprehensive exam scores were significant predictors of NPTE scores, PT programs without a comprehensive exam may consider seeking quality evidence supporting the inclusion of a comprehensive exam to their curriculum. At this time, the limited evidence and the quality of the evidence indicates that further investigations are required to determine if comprehensive exams can predict NPTE performance across PT programs.

Conclusions from the Review of PT Student Variables

In conclusion, some PT student variables could judiciously be used to predict NPTE performance. PT students who have an overall PTGPA of 3.5 or greater, a first-year PTGPA of 3.6, or greater, and reading comprehension score (NDRT) of 230 or greater will most likely pass the NPTE on their first-attempt (Adams et al, 2008; Aldridge et al., 2010; Meiners, 2015). While comprehensive exam scores were reported as significant predictors of NPTE performance; the design and analysis strategies in those two studies were of serious concern to a point where new studies with appropriate design and analysis strategies are required to determine with some confidence if comprehensive exam scores can predict first-attempt NPTE scores.

The overall results of the systematic literature review of PT student variables found first-year PTGPA to be the best predictor of NPTE performance (Table 2.4). However, even as the best predictor, first-year PTGPA should not be used as an independent predictor of NPTE performance. A comprehensive multi-variable approach should be used to assess which PT students are most likely to pass the NPTE on their first-attempt.

Table 2.4

Summary of Findings Related to PT Student Variables

Primary Author	Year	Journal	Summary of Findings
Adams	2008	JOPTE	PTGPA was able to correctly classify 97.1% of those who passed the NPTE on the first-attempt and 31.8% of those who did not. The odds of success on the NPTE are 2.27 times greater for every 10% increase in PTGPA. The PT CPI was a non-significant predictor of NPTE performance.
Aldridge	2010	Jrnl of Allied Health	Performance on the Nelson Denny Reading Test accounted for 27% of the variance in NPTE scores; indicating a significant relationship between the ability to read and comprehend what is read and NPTE performance.
Cook	2010	Dissertation	The PT CPI performance criteria Professional Behaviors and Professional/Social Responsibility explained 20.5% of the variance in NPTE first-attempt scores. The noncognitive PT student variable student reflections had a non-significant correlation ($r=0.10$).
Dillon	2010	JOPTE	PTGPA was a significant predictor of NPTE performance ($OR=25.84$) on the first-attempt. Neither PT students first or last clinical experience was a significant predictor of NPTE performance.
Dockter	2001	JOPTE	PT students' first year PTGPA accounted for 42% of the variance in NPTE first-attempt performance.
Dreeben	2003	Dissertation	While students' terminal PT CPI scores were significantly correlated with NPTE scores ($r=0.27$, $p<0.01$), the PT CPI scores were non-significant predictor of NPTE scores ($R^2=0.06$).
Edmondson	2001	Dissertation	PT student comprehensive exam scores accounted for 33% of the variance of NPTE scores ($\beta =6.55$, $p<0.01$). The average PT CPI scores from students first and second long-term clinical experiences did not predict NPTE scores.
Fell	2015	JOPTE	PTGPA predicted 40% of the variability in NPTE scaled scores and was a significant predictor of NPTE pass rates ($p<0.0001$).
Galleher	2012	Internet Jrnl of Allied Health Sci & Practice	Neither noncognitive variables conscientiousness ($r=0.09$) nor task-coping ($r=-0.02$) were significantly correlated with NPTE performance.

Table 2.4. *Summary of Findings Related to PT Student Variables (continued)*

Primary Author	Year	Journal	Summary of Findings
Guffey	2000	Dissertation	Singularly, the domain 'long range goals' from the Noncognitive Questionnaire-Revised significantly predicted NPTE scores ($R^2=0.08$, $p=0.03$). While the combination of the domains long range goals, leadership, community ties, and academic familiarity accounted for 13.4% ($p=0.02$) of the variance in NPTE scores.
Huhn	2017	JOPTE	PT students' first-year PTGPA was significantly correlated ($r=0.60$) and predicted NPTE scores ($p=0.001$). PT students' clinical reasoning as assessed with the HSRT was significantly correlated ($r=0.43$) and predicted NPTE scores ($p=0.001$).
Kosmahl	2005	JOPTE	Combined, PTGPA and comprehensive exam results explained 47% of the variance in NPTE scaled scores. PT students' terminal clinical experience PT CPI score was not a significant predictor of NPTE performance.
Lewis	2011	Internet Jrnl of Allied Health Sci & Practice	Neither PTGPA nor Emotional Intelligence were significant predictors of NPTE performance; however, PT students who passed the NPTE on their first-attempt had a higher Emotional Intelligence Total score ($p=.05$)
Luedtke-Hoffman	2012	JOPTE	PTGPA was significantly correlated ($r=0.51$) with NPTE performance on the first-attempt, while aggregate clinical performance scores were not significantly related to NPTE performance.
Meiners	2015	Dissertation	Accounting for 24% of the variance, students first-year PTGPA was a significant predictor of NPTE scores ($\beta=0.57$, $p<0.001$) and NPTE pass rates. For every 0.1 increase in first-year PTGPA, PT students were 72.85 times more likely to successfully pass the NPTE on the first-attempt. PT CPI scores were non-significant predictors of NPTE performance ($B=2.21$, $p=0.70$).
Moran	2012	Dissertation	NDRT scores were significantly correlated with the number of attempts PT students needed to pass the NPTE ($r= -0.63$, $p<0.01$). 86.7% of the time NDRT scores and 75% MMT scores correctly classified PT students into one of two groups: (a) first-attempt NPTE passers or (b) multiple-attempt NPTE test-takers
Vendrely	2007	Jrnl of Allied Health	Critical thinking skills ($r=0.31$) and PTGPA ($r=0.33$) were significantly correlated to NPTE performance, while PT CPI scores ($r=0.30$) had no significant relationship with NPTE performance. None of the variables were significant predictors of NPTE performance.

HSRT = *Health Sciences Reasoning Test*

CHAPTER III: METHODS

A systematic literature review provides a qualitative summary of the cumulative results of the studies reviewed. However, a systematic literature review alone does not quantify the effect of each reviewed variable (such as UGPA and PTGPA) on the primary outcome (NPTE performance). When a systematic literature review is combined with meta-analysis the effect of each reviewed variable on the primary outcome can be quantified. Put simply, meta-analysis is a statistical method that quantifies the effect each independent variable has on the dependent variable (Borenstein, Hedges, Higgins, & Rothstein, 2009). Performing a meta-analysis with assimilated data allows for the mean and variance of underlying population effects to be calculated while assessing the variability of the effect size across multiple studies (Field, 2010). Given the established power of systematic literature reviews with meta-analysis to provide higher order discrimination and evaluation of meta-data, it was a logical consequence to apply the same meta-analysis methods to investigate predictors of NPTE performance.

Accordingly, I conducted a systematic literature review with meta-analysis methods to investigate my research questions. Specifically, “To what extent do PT applicant and PT student variables predict NPTE performance?”

The meta-analysis was performed using data collected from the systematic literature review for PT applicant and student variables. The meta-analysis was conducted using a six-step process: (a) systematic literature review, (b) determining and applying inclusion criteria, (c) generation of hypotheses, (d) coding studies, (e) meta-analysis procedures, and (f) reporting the results (Field, 2010).

Systematic Literature Review

A systematic literature review is intended to provide a clear and systematic approach to reviewing and synthesizing the literature. A unique aspect of a systematic literature review is the transparent, documented, and systematic approach to identify and then review the relevant literature to the question being asked. Being transparent, documented, and systematic in a literature review allows others to replicate the search and provides a valid road map to the conclusion(s). A systematic literature review is based upon five steps: (a) framing the question, (b) identifying relevant work, (c) assessing the quality of studies, (d) summarizing the evidence, and (e) interpreting the findings (Khan, Kunz, Kleijnen, & Antes, 2003).

Framing the Question (Step 1a)

The first step of the present systematic literature review was to frame the question for a review. According to Khan, Kunz, Kleijnen, & Antes (2003), “The problems to be addressed by the review should be specified in the form of clear, unambiguous, and structured questions before beginning the review work” (p. 118). Prior to the initiation of the systematic literature review, the question “What predicts NPTE performance in PT education?” was formulated and established. Once the review question was established, a search protocol was then created, documented, and followed.

Identifying Relevant Work (Step 1b)

The second step of the systematic literature review was to identify the relevant literature for predicting NPTE performance in PT education. To capture as many relevant citations as possible, both educational and medical databases were searched. Potential studies were identified by conducting a systematic search utilizing the databases EBSCO (1966 to 2017), which includes CINAHL, ERIC, and Medline, and PubMed (1966 to 2017). The search protocol was

standardized for both databases. The search terms were related to PT education, applicant criteria, academic success, PT programs, and NPTE performance to identify related citations (see Appendix).

A comprehensive and documented electronic search was conducted up to and including June 2017 in the EBSCO and PubMed databases. The combined results from the databases searched was 2,963 identified citations. The citations were then stored in the reference manager software, Mendeley. All citations were reviewed for duplicates and conference proceedings; 105 citations were identified as duplicates or conference proceedings and removed. Following the removal of the duplicated citations and conference proceedings, the potential relevance of each citation to the question(s) was examined, and consequently, 2,811 citations were excluded due to their irrelevance to the question being investigated. The full-text articles of the remaining 47 citations were assessed to select those studies that were aimed at predicting NPTE performance, were in peer reviewed journals or doctoral dissertations and available in the English language. Following the review of the 47 remaining citations, 24 citations were assessed as directly relevant to the question. Of the 24 relevant citations, 6 citations were dissertations, and 18 citations were published in peer reviewed journals. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement guided the selection of the systematic search (1; Moher, Liberati, Tetzlaff, & Altman, 2009).

Assessing the Quality of Studies (Step 1c)

The third step in the systematic literature review was to assess the quality of all of the 24 identified citations. Following the systematic search for relevant citations, the quality of each citation was assessed utilizing a 3-step review process as identified by McCallum, Reed, Bachman, and Murray (2016). Two reviewers independently reviewed each manuscript for

inclusion in the systematic review. Each article was classified according to research design and methodological rigor using the Oxford Centre for Evidence-based Medicine Levels of Evidence as a reference guide.

The Oxford Centre for Evidence-based Medicine Levels of Evidence scaling system assigns a level ranging from 1a to 5 for each citation based upon the degree of design quality. A score of 1a indicates the highest level of evidence, and a score of 5 is the lowest level of evidence (Oxford Centre for Evidence-based Medicine Levels of Evidence, 2012). See table 3.1. Once the article review was completed by a minimum of two independent reviewers, each article was given a score on the 1a-to-5 scale.

Table 3.1

Oxford Centre for Evidence-based Medicine Levels of Evidence Scaling System

Level	Study Design
1a	Systematic review with homogeneity of randomized control trials
1b	Individual randomized control trial with narrow confidence interval
1c	“All or none” case-series
2a	Systematic review with homogeneity of cohort studies
2b	Individual cohort study
2c	Outcomes research
3a	Systematic review with homogeneity of case-control studies
3b	Individual case-control study
4	Case-series & poor-quality cohort studies
5	Expert opinion

Following the establishment of the hierarchy of evidence (1a to 5), the methodological quality of each citation was assessed utilizing the McMaster appraisal tool (MAT) developed by Lekkas, Larson, and Kumar (1998). The MAT has been established as a valid and reliable critical appraisal instrument with 75% to 80% agreement among reviewers (Law, Steward, Pollock et al., 1998). The Lekkas, Larson, and Kumar (2007) scoring system allows for a standardized critical appraisal of each article. The scoring system has 14 possible criteria with a total score ranging from 0 to 14. For each of the 14 criteria, a zero was assigned if that criterion was not met or not present in the article, and a one was assigned if the criterion was met or was present in the article. The 14 criteria include study purpose, literature review, sample size description, sample size justification, reliability and validity of outcomes, study interventions, contamination avoidance, co-intervention avoidance, statistical significance, methods, clinical importance, sample drop-outs, and conclusions. The educational literature reviewed in the present systematic review did not include study interventions, contamination avoidance, or co-intervention avoidance. Due to the nature of the manuscripts in educational literature that were reviewed, the MAT was modified. Specifically, study interventions, contamination avoidance, and co-intervention avoidance were removed from the MAT since none of the reviewed articles in the systematic review had a designed intervention. The modified MAT was created with a total of 11 criteria. See Table 2.1 and Table 2.3.

Two reviewers, using the modified MAT, independently appraised each of the 11 criteria for each selected manuscript. Following the critical appraisal of each selected manuscript, the two reviewers compared the scores of each criterion and for each manuscript. Discrepancies in appraisal scores between two reviewers resulted in the citation being sent to a third independent

reviewer. The third reviewer settled all disputes regarding the differences in critical appraisal score of each criterion and for each citation.

The final appraisal scores were converted from the initial raw scale (0 to 11) to a tertile scale, developed by McCallum, Mosher, Jacobsen, Gallivan, and Giuffre (2013). This scale was modified to assess the citations risk for bias. The appraisal scores were placed in tertiles accordingly: high quality: 10 to 11 points; moderate quality: 8 to 9 points; and low quality: 0 to 7 points” (McCallum, Mosher, Jacobsen, Gallivan, & Giuffre, 2013).

Summarizing the Evidence (Step 1d) and Interpreting the Findings (Step 1e)

The fourth and fifth steps of the systematic literature review included summarizing the evidence and synthesizing and interpreting the results of the collective literature review. The review of the literature was separated into two distinct sections: (a) PT applicant variables and (b) PT student variables. Each section review of the selected literature focused on thematic analyses to identify and examine patterns of key independent variables (e.g., UGPA, GRE-V, and PTGPA). Following the review of the literature, the interpretation of the results was reported in narrative form.

Inclusion Criteria

To reduce the risk of bias, inclusion criteria were created to identify relevant studies. Studies that met the following inclusion criteria were included in the meta-analysis: (a) the PT program being studied offered either a masters or doctorate in PT, (b) pertinent data (e.g., mean and standard deviation) are available for each selected independent variable, (c) a minimum of four studies that directly investigated the relationship between the variable of interest (PT applicant and PT student variables) and the NPTE, and (d) were U.S. PT programs.

Degree Level of the PT Program

To compare NPTE results among studies, only studies where PT students were pursuing their master's or doctorate PT degree were reviewed and included in the meta-analysis. Over the last two and a half decades, PT education has evolved from a bachelor's degree to a master's degree, to the current entry-level degree, a clinical Doctorate (American Physical Therapy Association, 2015). The shift in education requirements to practice as an entry-level PT has coincided with changes in the assessment of PT education outcomes. Most notably, the content of the NPTE has evolved to assess PT students' entry-level competency (Federation of State Boards of Physical Therapy, 2015).

Availability of Pertinent Data

Beyond the degree offered, an important inclusion criterion is that data pertinent to the question is available for each independent variable being studied in the meta-analysis. In particular, NPTE performance on students first-attempt must be a continuous variable with reported mean and SD. Failing to report mean values or SD limit the meta-analysis' capability of determining the effect size that each variable has on NPTE performance. Studies reporting this outcome as a dichotomous (pass/fail) variable were excluded from the meta-analysis. Without all of the pertinent data, there is increased potential for bias in the end results with normally required information excluded from the meta-analysis (Schmidt & Hunter, 2014). To improve the accuracy of the meta-analysis and reduce the risk of bias, requests were made to all of the authors for those studies that did not report the pertinent information.

When reviewing the pertinent data to be included in the meta-analysis, duplicate population samples were not included. Including duplicate population samples would have biased the results towards that particular population sample (Borenstein et al., 2009).

Specifically, the studies by Cook, Landry, et al. (2015a), Cook, Engelhard, et al. (2015b), and Covington et al. (2016) analyzed the same dataset, as did Riddel et al. (2009) and Utzman et al. (2007). As such, the meta-analysis, only included pertinent data once from the sample for each independent variable studied. Avoiding population sample duplication reduced the type I error rate and allowed for more valid results.

Minimum Number of Studies

To perform a meta-analysis only two studies are required. However, the results of the meta-analysis using only two studies are regarded as unstable (Rosenthal, 1995). More specifically, having a small number of included studies ($n \leq 3$) in the meta-analysis increases the risk of type I error in the estimate of the between study variance (T^2), and, therefore, the confidence intervals may be significant when in fact there is no significance (Borenstein et al., 2009). The risk of committing a type I error is concerning. A small number of studies in a meta-analysis could violate the independence assumption and result in second-order sampling error (Schmidt & Hunter, 2015). In meta-analysis, independence assumes the effect size for each study reviewed comes from an independent sample. Second-order sampling error can occur in meta-analysis when a small number ($n \leq 3$) of studies are included. (Schmidt & Hunter, 2015). As a result, caution should be taken when evaluating the results from a meta-analysis performed on a small number of studies. A small number ($n \leq 3$) of studies included in a meta-analysis increases the risk of committing a type I error, violates independence, and is likely subject to second-order sampling error. That said, the statistical analysis and summary, from meta-analysis with two studies may be valuable and superior to *ad hoc* summaries (Borenstein et al., 2009). A meta-analysis was performed when there was a minimum of four studies regarding each variable that predicts NPTE performance.

Program Locale

Higher education PT programs are found across the world. Countries outside of the U.S. have variations in educational and accrediting requirements and standards to become a practicing PT. Due to educational and clinical differences among PT programs outside of the U.S., only U.S. PT programs were included in the analysis. Further, by restricting the meta-analysis to studies reported in English and involving PT students enrolled in U.S. programs, a more consistent sampling was possible (DeCastro, 2012).

Expected Outcomes

The systematic literature review identified four PT applicant variables that were significant predictors of NPTE performance: (a) UGPA, (b) UGPA-PC, (c) GRE-V, and (d) behavior interview scores. However, the behavioral interview scores were only investigated in one study, and therefore, a meta-analysis was not conducted. The results of the systematic literature review for all other PT applicant variables (GRE-total, GRE-Q, degree status, noncognitive applicant variables, degree status, and SAT scores) was either inconclusive or not significant. There were insufficient studies to conduct a meta-analysis with the degree status, noncognitive applicant variables, and SAT variables.

Following the review of PT applicant variables, the systematic literature review identified four PT student variables that were significant predictors of NPTE performance: (a) PTGPA, (b) first-year PTGPA, (c) reading comprehension score (NDRT), and (d) comprehensive exam scores. However, both NDRT and comprehensive exam scores were only investigated in two studies, and therefore, a meta-analysis was not conducted on these two variables. The results of the systematic literature review for all other PT student variables (clinical performance scores and noncognitive PT student variables) was either inconclusive or not significant. Currently,

there are an insufficient number of studies to conduct a meta-analysis with noncognitive PT student variables. Based upon the findings from the systematic literature review, the expected outcome for each variable that had at least four studies was established *a priori* for the meta-analysis (see Table 3.2).

Table 3.2

Expected Outcomes for Predictors of NPTE Performance

Population	Independent Variable	Hypothesized Outcome*	Hypothesized Effect Size
PT applicants	UGPA	Significant	Medium
PT applicants	UGPA-PC	Significant	Medium
PT applicants	GRE-V	Significant	Medium
PT students	PTGPA	Significant	Medium
PT students	First-year PTGPA	Significant	Medium
PT applicants	GRE-total	Not significant	Small
PT applicants	GRE-Q	Not significant	Small
PT students	Clinical performance scores	Not significant	Small

* Nominal significance level $\alpha = .05$.

Coding the Studies

The information of included studies was coded to include the following information for each predictor variable: (a) authors, (b) year, (c) journal published (when applicable), (d) sample-level descriptive statistics (e.g., sample size, mean, and standard deviation) of each predictor variable, and (e) correlation, the effect size, of each variable with NPTE performance to be used in subsequent data analysis (e.g., mean, and SD of correlation with NPTE performance; DeJong, Dirks, & Gillespie, 2016). See Table 3.3. Coding included the following PT applicant variables:

(a) undergraduate cumulative GPA (UGPA), (b) undergraduate GPA for prerequisite courses (UGPA-PC), (c) GRE scores, (d) degree status, (e) noncognitive applicant variables, and (f). SAT scores. In addition, the following PT student variables were coded: (a) PTGPA, (b) clinical performance scores, (c) noncognitive variables, and (d) comprehensive exam scores.

Table 3.3

Example Layout for Coding Studies for Each Predictor Variable

Author(s)	Year	Journal/Dissertation	<i>n</i>	μ	<i>SD</i>	<i>r</i>
Study 1						
Study 2						
Study 3						
.						
.						
.						
Study k						

Note. *n* = sample size; μ = mean of predictor variable; *SD* = standard deviation of predictor variable; *r* = Pearson’s product-moment correlation between predictor variable and NPTE performance.

Meta-Analysis

The meta-analysis sought to answer the question of ‘how much’ of an effect does each PT applicant and each PT student variable have on NPTE performance? To address the question, the meta-analysis included the studies identified from the systematic literature search that met the inclusion criteria. A meta-analysis for each PT applicant and each PT student variable that predicted NPTE performance and met the specified inclusion criteria was conducted through: (a) the calculation of the effect sizes in each study for each independent variable that predicts NPTE performance, (b) selecting the appropriate effects model and subsequent computational method,

(c) applying the computational method, (d) undergoing heterogeneity analysis, and (e) undergoing publication bias assessment. All analyses were performed in R 3.4.2 using the metaphor package for meta-analysis (R Core Team, 2017; Viechtbauer, 2010).

Calculate the Effect Size (Step 5a)

In meta-analysis', effect sizes are used to quantify the relationship between two variables or groups (Borenstein et al., 2009). Depending upon the nature of the data, the effect size can be determined from the standardized difference between means or from correlations (Field, 2010). Determining the effect size of each variable across different studies allows for the direct comparison of those variables. To compare the effect size of variables across studies, Pearson's product-moment correlation coefficient (r) is commonly used.

The Pearson's product correlation is a measure of the strength of the relationship between two variables. As noted by Field (2010), r is easily interpretable, well-understood, and can be calculated from a variety of statistics (e.g. t , F , and χ^2). As such, once all of the studies included in the meta-analysis were identified, each pertinent variable's effect size on NPTE performance was calculated from the r with NPTE performance. Where r was not reported, r was calculated from the given statistics.

To combine the effect sizes, for each PT applicant and PT student variable, there must be a minimum of four studies that investigated the predictor variable's relationship with NPTE performance. Since the independent and dependent variables are continuous, r is the appropriate effect size to use for the meta-analysis (Cooper, 2010). If the r was not reported in a study, the primary author was contacted to determine if r was available, or other statistical information from which r could be calculated was available. If sufficient data to determine or calculate r was not available, the study was excluded from the meta-analysis.

The accurate calculation of the effect size and variance for each variable is essential for a valid and reliable meta-analysis. As such, the calculation of the effect size and variance included the weighted averages for each study. Schmidt and Hunter (2015) advocate using the frequency-weighted averages instead of the simple average of the effect size. The frequency-weighted averages are superior when (a) a large number of individual studies are included, (b) there is no variance in population correlations across studies, and (c) the variance of population correlations is small or large. To determine the frequency-weighted average of each study, the reported correlation was weighted by the sample size of the study.

Random-Effects Model (Step 5b)

Conceptually, there are two models used to perform meta-analysis, the fixed-effects model and the random-effects model. The first model, the fixed-effects model assumes the average effect size for the studied variable is consistent for all studies (Higgins, Thompson, & Spiegelhalter, 2009). In other words, the average effect size is 'fixed' or homogenous. The second model, the random-effects model assumes the average effect size for the studied population varies randomly for each study (Field, 2001; Field, 2010). Thus, in the random-effects model, the average effect sizes follow some distribution, and therefore, the average effect size is 'random' or heterogeneous, which is consistent with real world differences (DeCastro, 2012; Higgins, Thompson, Spiegelhalter, 2009).

Both philosophically and statistically, the distinction between fixed-effects and random-effects models is important. In terms of statistics, the main difference between the fixed-effects model and the random-effects model is the calculation of standard error associated with the effect size (DeCastro, 2012; Field, 2001). For the fixed-effects model, the sampling error is due entirely to within-study variability. The limited error in the fixed-effects model restricts the

inferences of the results only to those studies included in the meta-analysis (DeCastro, 2012). For the random-effects model, the sampling error is due to within-study variability and variability as a result of differences between studies (Field, 2001; Shadish & Haddock, 1994). Therefore, the random-effects model has two sources of error that need to be accounted for, rather than one source of error found in the fixed-effects model. The two sources of error in the random-effects model make the random-effects model a more realistic choice for social science research and allow the results to be generalized beyond the studies included in the meta-analysis (Field, 2001). Having the capability of making inferences from the results of the meta-analysis to the general population is a strength of the random-effects model.

The model selected for the meta-analysis is important. Specifically, Schmidt, Oh, and Hayes (2009) discovered that using the fixed-effects model with random-effects data results in significantly smaller confidence intervals for the calculated effect sizes. As a result of calculating confidence intervals that are significantly smaller than their actual size, the type I error rate becomes inflated. Hunter and Schmidt (2000) estimated that using the fixed-effects model when the random-effects model is appropriate can increase the expected type I error rate of 5% up to 28%. Supporting Hunter and Schmidt's (2000) findings, in 2003, Field estimated that the type I error rate could be inflated by as much as 80% when the fixed-effects model was incorrectly applied in meta-analysis. To avoid an increase in the type I error rate beyond the accepted limit (5%), the choice of model (fixed-effects or random-effects) should be based upon the type of data being analyzed, the type of inferences the researcher plans to make, and be done *a priori*. The completed meta-analysis sought to determine PT applicant and PT student variables' effect size for NPTE performance. As such, the meta-analysis was conducted using the random-effects model.

Hunter-Schmidt Method (Step 5c)

After the type of model was chosen, the method of meta-analysis was selected. For the random-effects model in meta-analysis with r , there are essentially two choices: (a) the Hunter-Schmidt method (2004), also known as the Schmidt-Hunter method (2015), and (b) the Hedges and Olkin method (1985), also known as the Rosenthal method (1991). Both methods estimate the overall mean effect size that the independent variable has on the dependent variable using r . However, the two methods use different statistical approaches to determine the effect size.

To determine the effect size, the Hedges and Olkin method transforms r to Fisher's z scale. "A main advantage of transforming r to z is that the sampling variance of z is independent of the population parameter" (Brannick, Yang, & Cafri, 2008). The meta-analysis is then performed using the Fisher's z scale values. The results of the meta-analysis, the summary effect and confidence intervals, are then converted back to r . While the Hedges and Olkin method is more simplistic than the Hunter-Schmidt method, the translation of r to z and then z back to r increases the risk of error due to rounding.

Due to the increased risk of error in the Hedges and Olkin method, the Hunter-Schmidt method is the more appropriate choice to perform a random-effects meta-analysis. Specifically, Schmidt and Hunter (2015) and DeCastro (2012) suggest that the transformation of r to z scale values that occurs with the Hedges and Olkin method, applies more weight to larger correlations, and thus, causes an upward bias of the effect sizes. Instead of transforming r to z scale values, the Hunter-Schmidt method directly analyzes r . To investigate the differences in the calculated effect size, Field (2010) performed Monte Carlo simulations using the Hedges and Olkin method and the Hunter-Schmidt method. Field's (2010) results found that the Hunter-Schmidt method consistently had the most accurate effect sizes. Supporting the use of the Hunter-Schmidt method

with correlation coefficients and the random-effects model, both Johnson (1995) and Brannick, Yang, and Cafri (2008) concluded that the Hunter-Schmidt method provided the most accurate estimates of the effect size and the variance.

The Hunter-Schmidt full artifact distribution meta-analysis method for random-effects models was used to analyze the effect sizes from those studies that met the inclusion criteria for each PT applicant and PT student variable. The full artifact distribution meta-analysis was used when sampling error was the only artifact that was provided for each of the studies in the meta-analysis. As described by Schmidt and Hunter (2015), the meta-analysis was conducted in three stages: (a) the studies were used to collect information on distribution of the observed correlations and their samples sizes, the reliability of the independent and dependent variable, and the range departure, (b) the r was corrected for sampling error, and when possible, (c) the r corrected for sampling error was then corrected for other sources of error. Accordingly, the Hunter-Schmidt method was the method of choice for the random-effects model in the investigation.

Heterogeneity Analysis (Step 5d)

Heterogeneity analysis examines the differences in the true effect size among the included studies. “Under the random-effects model we allow that the true effect size may vary from study to study” (Borenstein et al., 2009, p. 107). The difference in the true effect size among studies is referred to as heterogeneity. Heterogeneity can occur due to a variety of differences among studies including differences in study design, participants, interventions, and outcomes studied. Interpreting heterogeneity among studies selected in the meta-analysis occurs when the true effect size between those selected studies being evaluated is greater than expected due to chance (Higgins & Thompson, 2002).

The heterogeneity of the selected studies was assessed to identify if significant differences in true effect size among those selected studies was due to sampling error. To assess the heterogeneity of the true effect size, the Q statistic, and the corresponding p-value were calculated. The Q statistic is “sensitive to the ratio of the observed variation to the within-study error...” (Borenstein et al., 2009, p. 109). With a significance level of 0.05, a p-value comparing the differences among true effect sizes were calculated. However, the power of the Q statistic is poor when there is only a small number of studies to be analyzed (Higgins & Thompson, 2002). Since many of the variables predicting NPTE performance will have a small number of included studies, additional approaches to quantifying heterogeneity were conducted.

To quantify the heterogeneity of the true effect size for each variable (PT applicant variables and PT student variables), Borenstein et al. (2009) support using the estimated variance (T^2) and estimated SD (T) of the observed effects, along with, the ratio of true heterogeneity to total variance across the observed effects (I^2). Determining T^2 , T, and I^2 provided a more comprehensive approach to assessing the heterogeneity of the true effect size for each PT applicant and PT student variable.

T^2 , also known as the method of moments or the DerSimonian and Laird method, estimates the variance of the true effect size used in random-effects models. “As the estimate for the variance of the true effects, T^2 was used to assign weights to each study” (Borenstein et al., 2009, p. 114). From T^2 , the SD of the estimate of the true effects (T) was calculated. T was used to describe the distribution of the effect sizes from the selected studies. Both T^2 and T provided an indication of the degree of heterogeneity that occurred in the included studies (Borenstein et al., 2009).

An additional approach to quantifying heterogeneity is to calculate I^2 . I^2 is a measure of the degree of inconsistency in the included studies results. Specifically, “ I^2 describes the percentage of total variation across studies that is due to heterogeneity rather than chance” (Higgins, Thompson, Deeks, & Altman, 2003, p. 558). The I^2 value can be interpreted as a percentage ranging from 0%, no observed heterogeneity, to 100%, complete heterogeneity (Higgins et al., 2003).

To determine the heterogeneity of the included studies for each PT applicant and PT student variable, Q, p-values, T^2 , T, and I^2 were calculated as each measure of heterogeneity provides different information. Q and p-values primarily served as the significance test, while T^2 and T provided the between study variance and SD of the true effect sizes. In the observed effect sizes, I^2 provided a ratio of the true heterogeneity to total variation (Borenstein et al., 2009).

Publication Bias Assessment (Step 5e)

Published studies with significant results or large effect sizes are more likely to be included in meta-analysis studies. As a result, the majority of studies included in a meta-analysis may be biased towards significance or larger effect sizes; and therefore, may confound the results of the meta-analysis (Banks, Kepes, & McDaniel, 2012). The influence of biased studies on the results of the meta-analysis is called publication bias.

Due to difficulty in retrieving unpublished information, the potential for publication bias is high in meta-analysis. As a result of publication bias, the meta-analysis is more likely to overestimate the effect size of the studied variable (McDaniel, Rothstein, & Whetzel, 2006). To assess for publication bias, forest plots were created for each PT applicant and PT student variable that predicts NPTE performance. Each forest plot was reviewed to assess the relationship between the included studies sample size and the effect size. Trends showing

positive effect sizes for smaller samples are an early indication of publication bias (Borenstein et al., 2009).

Following the review of the forest plots, a cumulative meta-analysis was conducted. Similar to the review of the forest plots, the presence of potential publication bias was determined by an upward trend in the summary effect size when the smaller sample size studies are added to the cumulative meta-analysis (Schmidt & Hunter, 2015). In a cumulative meta-analysis, the included studies are ranked by sample size. After being ranked by sample size, the included studies were subsequently added to the meta-analysis one at a time, starting with the study with the smallest sample size.

After the cumulative meta-analysis was completed, a funnel plot was generated for each studied variable to graphically display the distribution of the effect size from the included studies (Banks, Kepes, & McDaniel, 2012). The funnel plot was reviewed for asymmetry, since asymmetry is regarded as an indication of potential publication bias (Borenstein et al., 2009; Duval & Tweedie, 2000). Following the review of the funnel plot, the trim-and-fill method was used to assess how much the effect size was impacted by publication bias. “The trim-and-fill analysis... detects the potential presence of publication bias, and provides an estimate of its amount” of publication bias (Banks, Kepes, & McDaniel, 2012, p. 187).

Reporting the Results

Following the meta-analysis with the Hunter-Schmidt methods, the main outcomes were reported for each predictor variable for NPTE performance. The main outcomes reported in the results, as outlined by Schmidt and Hunter (2015) include: (a) total sample size, (b) number of correlations, (c) mean true score correlation (ρ ; summary effect size), (d) variance of true score

correlations (S_p^2), (e) 80% credibility interval for true score correlation distribution, and (g) 95% confidence interval around mean true score correlation

In addition to reporting the main outcomes in the results, forest plots for each predictor variable for NPTE performance were also presented in graphical format. Forest plots can be easily interpreted, while simultaneously conveying a large amount of information from the meta-analysis. Specifically, forest plots provide the point estimates of the effect size and 95% CI's for each study, along with the summary effect size of the studied variable (Borenstein et al., 2009).

Interpretation of Results

For the meta-analysis, using the random-effects model, there were an array of results that were reported. Each result is important; however, to successfully interpret the results for each variable that predicts NPTE performance it is essential to discuss the (a) mean effect size for each study that contributed to the results of the meta-analysis, (b) the summary effect size for each studied variable, (c) 80% credibility intervals, and (d) 95% confidence intervals (Rosenthal, 1995).

The mean effect size for each study is a point estimate that was interpreted as having a small effect ($r \leq 0.29$), medium effect ($r = 0.30$ to 0.49), or large effect ($r \geq 0.49$; Cohen, 1988). Following the interpretation of the mean effect size for each study, the summary effect size for each studied variable was successfully interpreted (Rosenthal, 1995). Reported as a point estimate, the summary effect size is the mean true score correlation for the studied variable that is predicting NPTE performance. The summary effect size was interpreted using the same effect size scale as the mean effect size for each study: small effect ($r \leq 0.29$), medium effect ($r = 0.30$ to 0.49), or large effect ($r \geq 0.49$; Cohen, 1988).

To determine the significance of the summary effect size, Schmidt and Hunter (2015) emphasize the use of 80% credibility intervals. “Credibility intervals estimate the range of real differences after accounting for sampling error” (DeCastro, 2012, p. 40). In random-effects meta-analysis models, credibility intervals are important since random-effects models allow for variance between studies. As such, credibility intervals focus on the variability of population values (distribution) from the SD (Schmidt & Hunter, 2015). To interpret the credibility interval, both the lower and upper bound values and the width of the credibility interval were reported to show the distribution of the population around the mean effect size for each study.

Unlike credibility intervals which focus on the variance of the population parameter around the mean effect size, confidence intervals focus directly on the mean effect size in each selected study (Schmidt & Hunter, 2015). Confidence intervals are an expression of the estimated amount of sampling error in the mean effect size for each study (DeCastro, 2012). Accordingly, confidence intervals were set at 95% to test the significance of the effect size for each selected study. Confidence intervals set at 95% that are either entirely positive or negative were interpreted as significant (Borenstein et al., 2009).

CHAPTER IV: RESULTS

The current investigation utilized a systematic literature review with a subsequent random-effects (RE model) meta-analysis to determine the empirical relationships between NPTE performance and various PT applicant and student variables. The following two research questions guided the current investigation:

1. To what extent can NPTE performance be predicted by each of the PT applicant variables (a) overall undergraduate GPA, (b) undergraduate GPA for prerequisite courses, (c) GRE scores, (d) previous degree status, (e) noncognitive applicant variables, and (f) SAT scores?
2. To what extent can NPTE performance be predicted by each of the PT student variables (a) PT-specific GPA, (b) clinical performance scores, (c) noncognitive student variables, and (d) comprehensive exam scores?

Data Analysis

Random-effects meta-analysis was used to estimate the relative effect sizes for each PT applicant and PT student variable on subsequent NPTE performance. The meta-analysis included those studies identified from the systematic literature search that met the inclusion criteria ($n = 18$): (a) the PT program being considered for inclusion offered either a masters or doctorate in PT, (b) relevant data (e.g., sample size and Pearson's product-moment correlation coefficient) was available for each selected independent variable, (c) a minimum of four studies that directly investigated the relationship between the variable of interest (PT applicant and PT student variables) and the subsequent NPTE performance, and (d) were U.S. PT programs. For each PT applicant and PT student variable that predicted NPTE performance and met the specified inclusion criteria a random effects meta-analysis was conducted using the Hunter-Schmidt

method to: (a) calculate the mean true effect size for each independent variable that predicted NPTE performance, (b) assess for heterogeneity of the effect sizes for each study included in the meta-analysis, and (c) assess for publication bias.

Three random effects meta-analyses (clinical performance scores [first clinical experience], NDRT scores, and comprehensive exam scores) that did not meet the specified inclusion criteria were conducted using the Hunter-Schmidt method. Due to the significant amount of time (~25% of a PT curriculum) dedicated to students clinical learning and performance, the clinical performance scores (first clinical experience) random effects meta-analysis was completed with only three studies in the meta-analysis. The NDRT scores random effects meta-analysis was completed following the analysis of the noncognitive student variables random effects meta-analysis. There were a total of nine different variables included in the noncognitive student variables random effects meta-analysis, and the only variable that was studied by multiple investigators was NDRT scores. Since both investigators found NDRT scores had a large and significant relationship with students NPTE performance on their first-attempt, the NDRT scores random effects meta-analysis was completed with two studies. The comprehensive scores random effects meta-analysis has only two studies included in the analysis; however, many PT programs are using comprehensive examinations to simulate the NPTE. Due to the increased frequency of comprehensive examinations at the PT program level, the comprehensive exam scores random effects meta-analysis was completed with two studies.

Relationships among PT Applicant Variables and NPTE Performance

The first research question sought to determine the relative effect size for each PT applicant variable on NPTE performance. To precisely answer this research question, the Hunter-Schmidt meta-analysis using the random effects model was performed on each PT applicant

variable: (a) UGPA, (b) UGPA-PC, (c) GRE verbal (GRE-V), (d) GRE quantitative (GRE-Q), and (e) noncognitive applicant variables that met our inclusion criteria. As presented in Table 4.1, UGPA had the largest total sample size ($n = 818$) and number of correlations ($\#r$) with seven studies included in the corresponding meta-analysis. Results of the meta-analysis found that all of the investigated PT applicant variables: (a) UGPA, (b) UGPA-PC, (c) GRE-V, (d) GRE-Q, and (e) noncognitive applicant variables had a medium true effect size on NPTE performance. Of the PT applicant variables investigated, UGPA-PC had the largest mean true effect size ($M_\rho = 0.37$), while GRE-Q and noncognitive applicant variables had the smallest mean true effect size ($M_\rho = 0.31$; see Table 4.1). With 80% credibility intervals and 95% confidence intervals all greater than zero, each PT applicant variable's relationship with NPTE performance was statistically significant (Table 4.1).

Table 4.1

PT Applicant Variables Results from Random Effects Meta-Analysis

Independent variable	Total n	$\# r$	M_ρ	S_ρ^2	80% cred. interval	95% conf. interval
UGPA	818	7	0.33	0.05	0.10 – 0.55	0.22 – 0.43
UGPA prerequisite courses	504	4	0.37	0.08	0.09 – 0.65	0.21 – 0.53
GRE verbal (GRE-V)	464	4	0.32	0.04	0.23 – 0.40	0.23 – 0.40
GRE quantitative (GRE-Q)	464	4	0.31	0.04	0.22 – 0.39	0.22 – 0.39
Noncognitive applicant variables	359	3	0.31	0.05	0.18 – 0.43	0.20 – 0.41

Total n = total sample size

$\# r$ = number of Pearson product moment correlations

M_ρ = mean true score correlation (summary effect size)

S_ρ^2 = variance of true score correlations

80% cred. interval = 80% credibility interval for true score correlation distribution

95% conf. interval = 95% confidence interval around mean true score correlation

Heterogeneity of PT Applicant Variables

The heterogeneity of the studies selected for inclusion into the meta-analysis was assessed to determine if significant differences in true effect size among those included studies was due to sampling error. To determine the heterogeneity of the included studies for each PT applicant variable, Q , p -values, T^2 , T , and I^2 were calculated (Table 4.2). Both UGPA ($Q = 17.38, p = 0.01, T^2 = 0.01, T = 0.10$) and UGPA-PC ($Q = 13.44, p < 0.01, T^2 = 0.01, T = 0.10$) had a statistically significant amount of heterogeneity leading to the largest variance and standard deviation of the true effect size. Consequently, UGPA ($I^2 = 58.38\%$) and UGPA-PC ($I^2 = 66.96\%$) had the highest proportion of observed heterogeneity.

GRE-V ($Q = 2.63, p = 0.45$) and GRE-Q ($Q = 0.64, p = 0.89$) had a non-significant distribution of the mean true effect size. The variance and standard deviation of the mean true effect sizes, along with the proportion of observed heterogeneity for GRE-V and GRE-Q was zero. Noncognitive applicant variables (the label of noncognitive variables is consistent terminology with the literature even though some of the included noncognitive variables are cognitive) had a non-significant distribution of the ($Q = 3.52, p = 0.17, T^2 = 0.00, T = 0.03$) mean true effect size and had a small proportion of observed heterogeneity of 11.90% (Table 4.2). Accordingly, the PT applicant variables GRE-V and GRE-Q due not have a significant amount of heterogeneity.

Table 4.2

Heterogeneity Analysis of PT Applicant Variables

Independent variable	Q	p	T^2	T	I^2
UGPA	17.38	0.01	0.01	0.10	58.38%
UGPA prerequisite courses	13.44	0.00	0.01	0.11	66.96%
GRE verbal	2.63	0.45	0	0	0%
GRE quantitative	0.64	0.89	0	0	0%
Noncognitive applicant variables	3.52	0.17	0.00	0.03	11.90%

Q = weighted sum of squared differences between individual study effects and the pooled effect across include studies

p = indicates the significance of the Q statistic

T^2 = estimated variance of the true effect size for each variable

T = estimated standard deviation of the true effect size

I^2 = ratio of true heterogeneity to total variance

Publication Bias of PT Applicant Variables

To assess for publication bias, forest plots were created, with a corresponding cumulative meta-analysis for each PT applicant variable that predicted NPTE performance. The forest plot provided the effect size and 95% confidence interval for each study that was included in the meta-analysis, along with the mean true effect size and 95% confidence intervals. When reviewing the forest plots it is important to note that the size of the box symbolizing the correlation coefficient from each included study represents the weight given to that study in the meta-analysis results. The weight of each study was determined by the sample size of that included study. As such, the greater the sample size, the greater the weight of that study in the corresponding meta-analysis. For the cumulative meta-analysis, the included studies are ranked by sample size. After being ranked by sample size, the included studies were subsequently added

to the meta-analysis one at a time, starting with the study with the smallest sample size (Borenstein, et al., 2009). Following the cumulative meta-analysis, a funnel plot was generated and reviewed for asymmetry for each studied variable. The funnel plot graphically displays the distribution of the effect sizes from the included studies. Following the review of the funnel plot results, the trim-and-fill method was then used to assess how much the effect size was impacted by publication bias (Banks, Kepes, & McDaniel, 2012).

UGPA results. The effect sizes for the included studies in the UGPA meta-analysis range from 0.17 (95% CI = 0.01-0.33) to 0.50 (95% CI = 0.29-0.71), with a mean true effect size of 0.33 (95% CI = 0.22-0.43; see Figure 4.1). When a cumulative meta-analysis was performed, there were minor changes in the mean true effect size as studies with larger sample sizes were added to the meta-analysis (see Figure 4.2). After the first, and smallest, study (Galleher et al., 2012) the cumulative meta-analysis the mean true effect size was 0.30. As subsequent studies were added to the cumulative meta-analysis, the mean true effect size ranged from a low of 0.26 to a high of 0.35, before reaching the final mean true effect size of 0.33. With only small changes in the mean true effect size as larger studies were added in the UGPA cumulative meta-analysis, the results of the cumulative meta-analysis suggest that publication bias was not present for the UGPA meta-analysis.

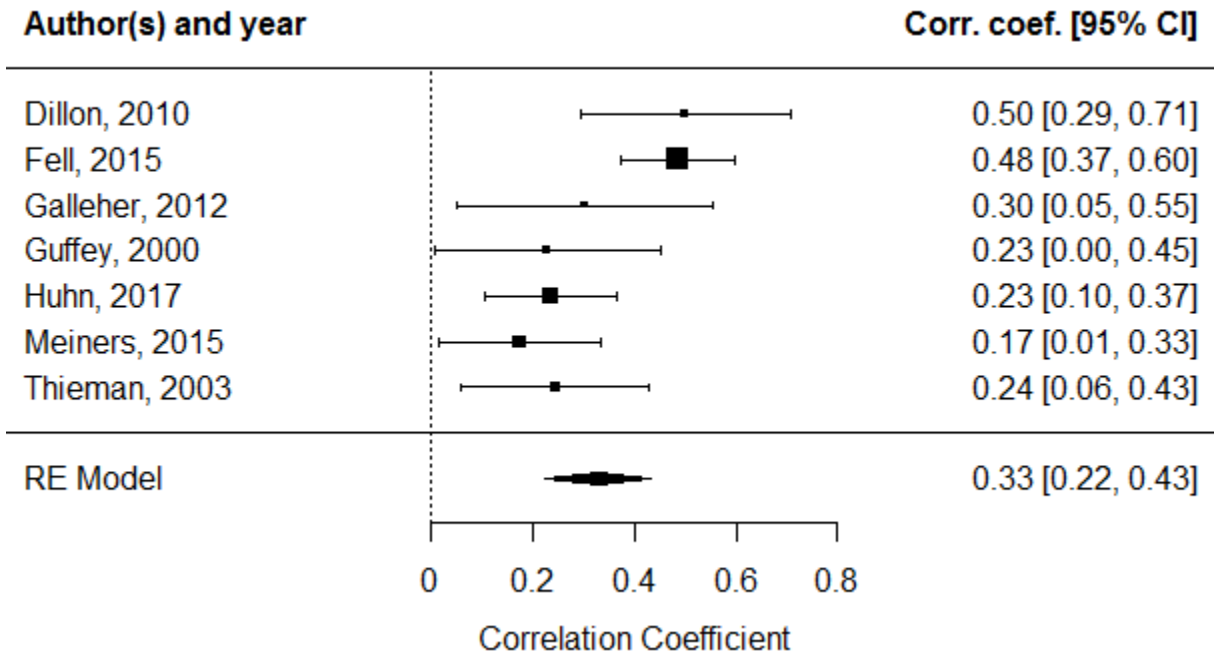


Figure 4.1. UGPA forest plot (RE Model = random effects model). The size of the box symbolizing the correlation coefficient from each included study represents the weight given to that study to the meta-analysis results. The weight of each study was determined by the sample size of that included study.

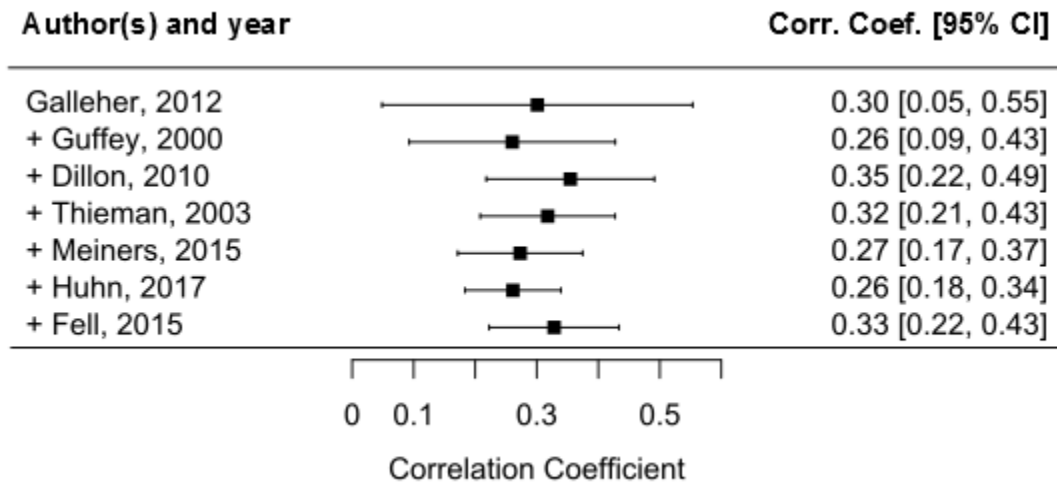


Figure 4.2. UGPA cumulative meta-analysis forest plot.

Following the cumulative meta-analysis, a funnel plot was qualitatively assessed for potential publication bias with the PT applicant variable UGPA (Figure 4.3). The UGPA funnel plot shows asymmetry that had a downward bias on the mean true effect size. These results indicate one or more studies with an effect size greater than 0.33 may be missing (Figure 4.3). Following the qualitative assessment of the UGPA funnel plot, the trim-and-fill method estimated the number of missing studies from the meta-analysis to be zero, and thus, there was no potential change in the overall mean effect size of UGPA (Figure 4.4). The results of the trim-and-fill analysis are consistent with the UGPA cumulative meta-analysis, indicating publication bias is not present for the UGPA meta-analysis.

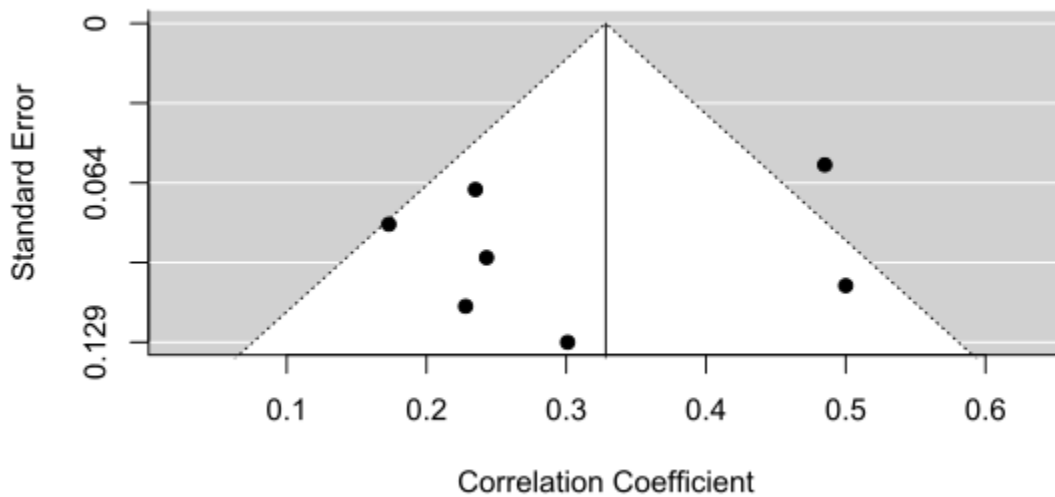


Figure 4.3. UGPA funnel plot.

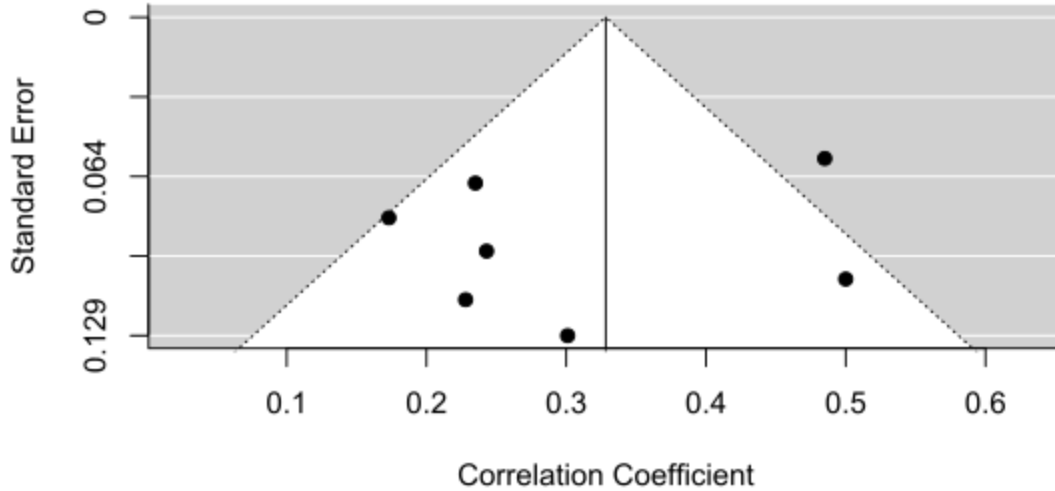


Figure 4.4. UGPA trim-and-fill funnel plot.

UGPA prerequisite courses results. The effect sizes from the included studies for UGPA-PC ranged from 0.04 (95% CI = -0.17-0.26) to 0.48 (95% CI = 0.38-0.59), with a mean true effect size of 0.37 (95% CI = 0.21-0.53; Figure 4.5). With a wide range of effect sizes, the forest plot provided evidence that publication bias was present in the UGPA-PC meta-analysis. Consistent with the forest plot, the UGPA-PC cumulative meta-analysis identified that when Guffey's (2000) results were added to the analysis, there was a large decrease in the mean true effect size, from 0.34 to 0.17. Once larger studies were added to the cumulative meta-analysis the overall mean effect size increased to 0.37 (Figure 4.6). The changes in the mean true effect size in the UGPA-PC cumulative meta-analysis indicated there is the potential for publication bias in the UGPA-PC random effects meta-analysis.

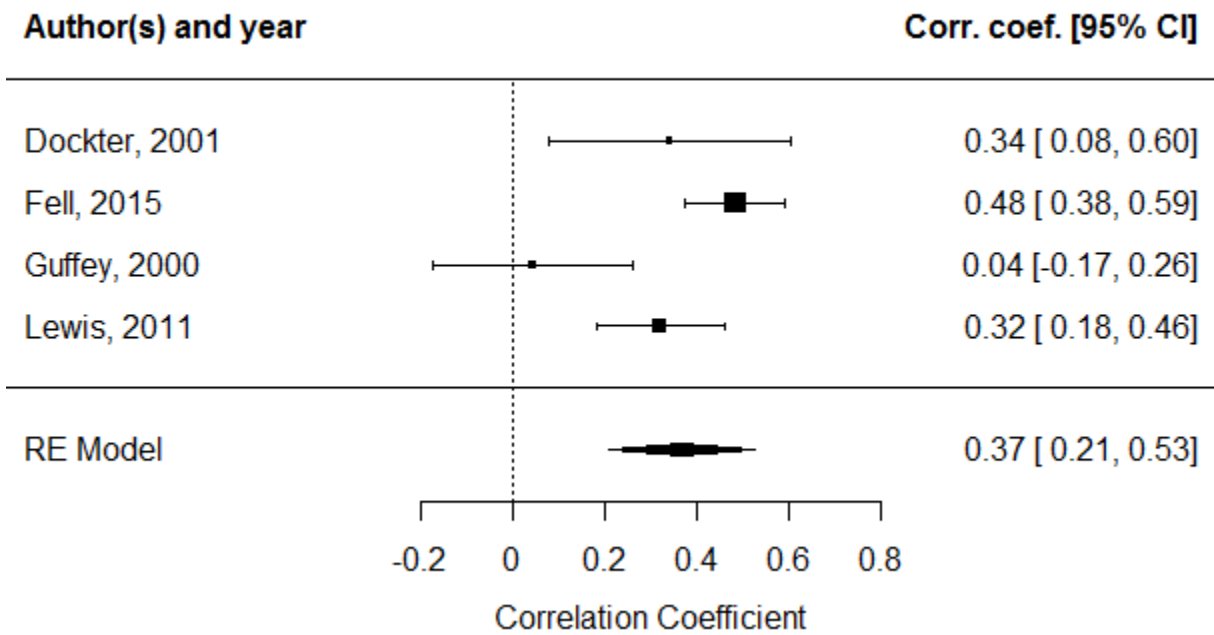


Figure 4.5. UGPA-PC forest plot (RE Model = random effects model).

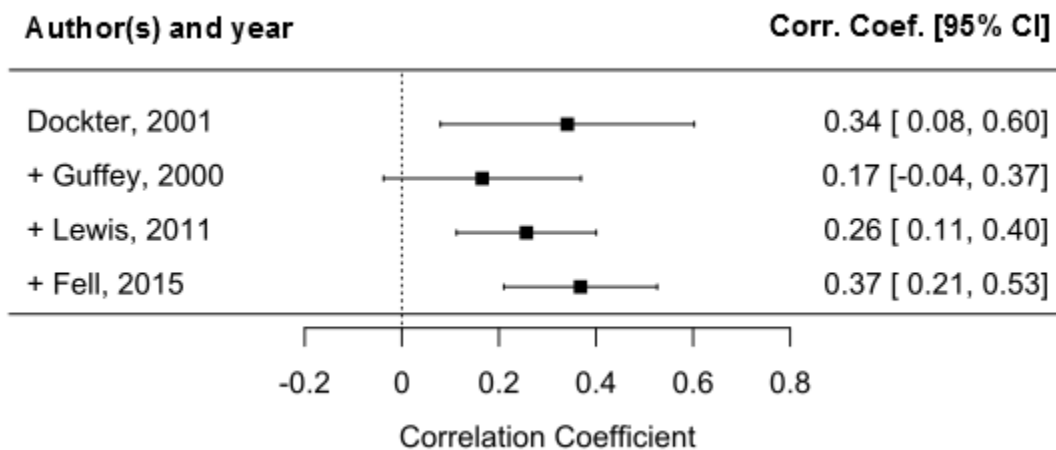


Figure 4.6. UGPA-PC cumulative meta-analysis forest plot.

Following the cumulative meta-analysis, the qualitative assessment for publication bias was completed using the funnel plot for UGPA-PC (Figure 4.7). Asymmetry was again identified on the UGPA-PC funnel plot that indicated a downward bias in the mean true effect size. The

results suggest that one or more studies with an effect size greater than 0.37 have the potential to of being absent from the analysis. The trim-and-fill results for UGPA-PC confirmed the qualitative assessment of the forest plot, cumulative meta-analysis, and funnel plot (Figure 4.8). The trim-and-fill results estimated that two studies were missing on the right side of the funnel plot (Figure 4.8). With the addition of the estimated missing studies, the mean true effect size of UGPA-PC theoretically increased to 0.46 (95% CI = 0.28-0.64). The trim-and-fill results indicate that the theoretical mean true effect size of UGPA-PC (0.46) is greater than the calculated mean true effect size of UGPA-PC (0.37).

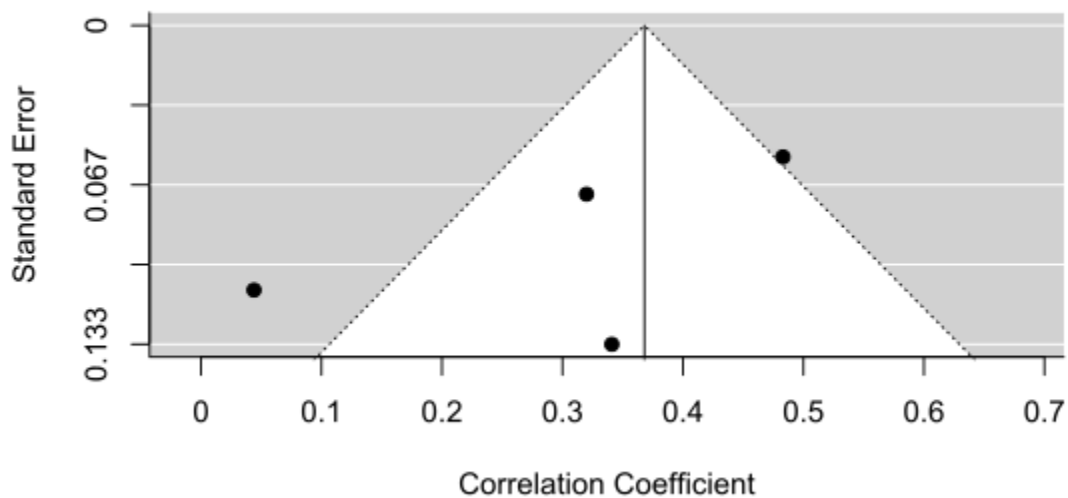


Figure 4.7. UGPA-PC funnel plot.

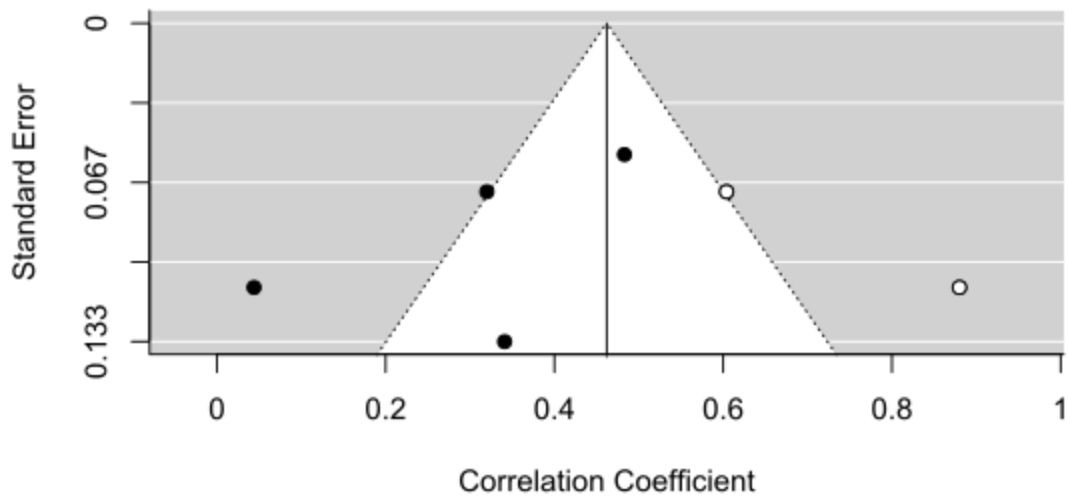


Figure 4.8. UGPA-PC trim-and-fill funnel plot.

GRE-verbal (GRE-V) results. GRE-V had a mean true effect size of 0.31 (95% CI = 0.23-0.40; Figure 4.9). From the four studies that investigated GRE-V's relationship with NPTE performance there was a range of effect sizes from 0.17 (95% CI = -0.04-0.38) to 0.39 (95% CI = 0.18-0.60). Consistent with the forest plot, the GRE-V cumulative meta-analysis provided no indication of potential publication bias, as the smallest studies appeared to have minimal effect on the overall mean effect size (Figure 4.10).

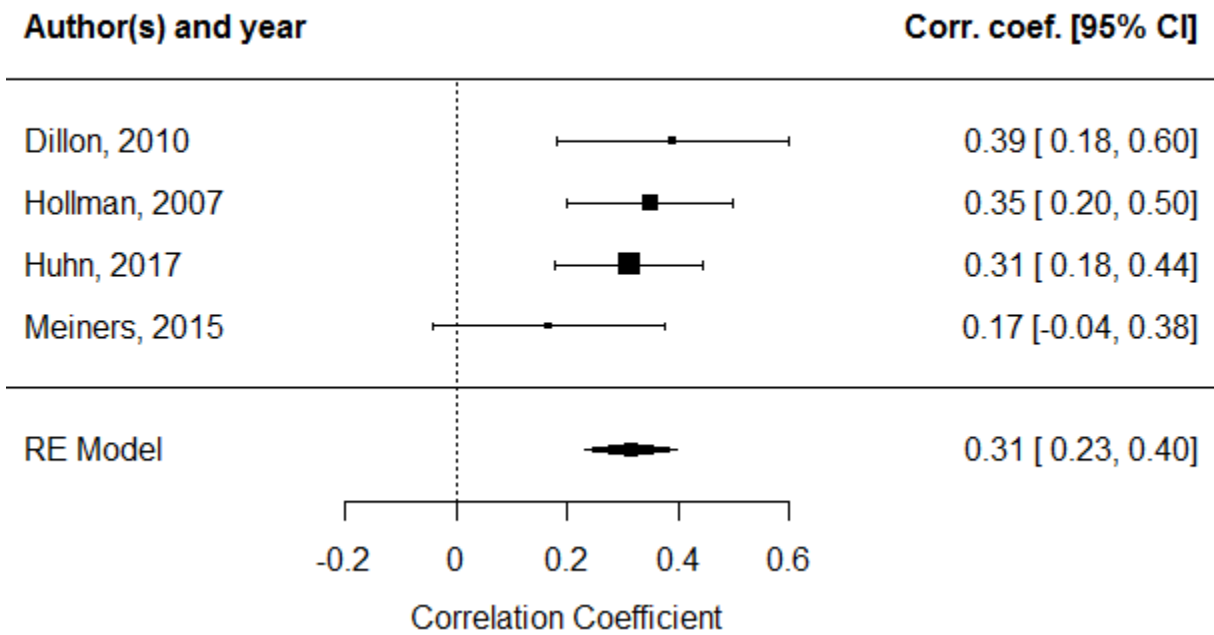


Figure 4.9. GRE-verbal (GRE-V) forest plot (RE Model = random effects model).

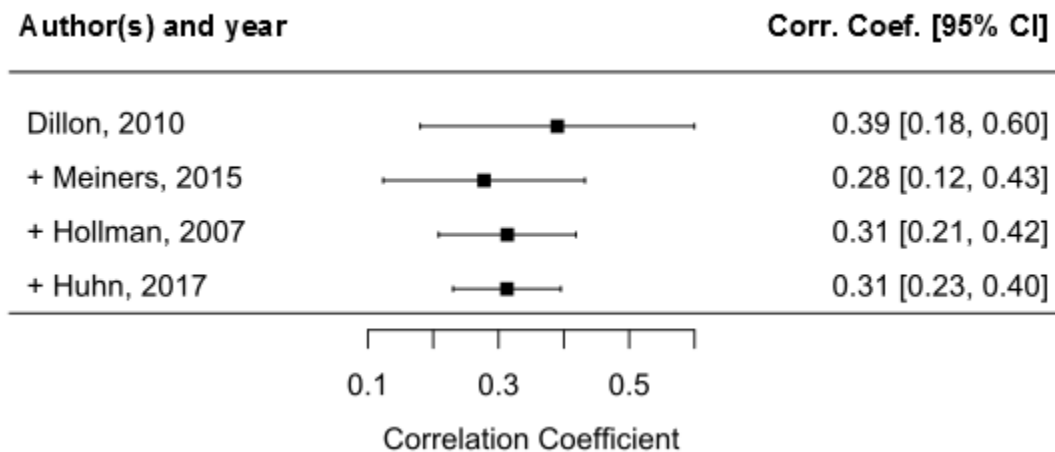


Figure 4.10. GRE-verbal (GRE-V) cumulative meta-analysis forest plot.

In contrast to the cumulative meta-analysis' findings of no potential publication bias, the funnel plot for GRE-V was asymmetrical providing evidence of an upward bias of the mean true effect size (Figure 4.11). The results of the GRE-V funnel plot suggest that publication bias was likely present in the analysis. The subsequent potential publication bias indicates that one or

more studies with an effect size less than 0.31 may be missing from the analysis. However, following the qualitative assessment of the GRE-V funnel plot, the trim-and-fill results do not support the presence of publication bias in the GRE-V meta-analysis (Figure 4.12). The GRE-V trim-and-fill results estimated that zero studies were missing from the GRE-V random effects meta-analysis. As such, the mean true effect size of GRE-V is 0.31 (95% CI = 0.23-0.40).

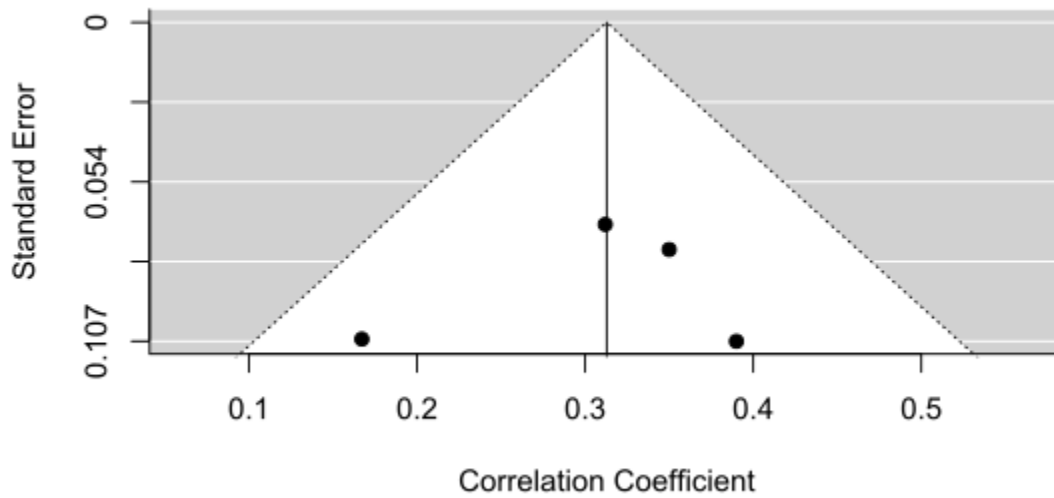


Figure 4.11. GRE - verbal (GRE-V) funnel plot.

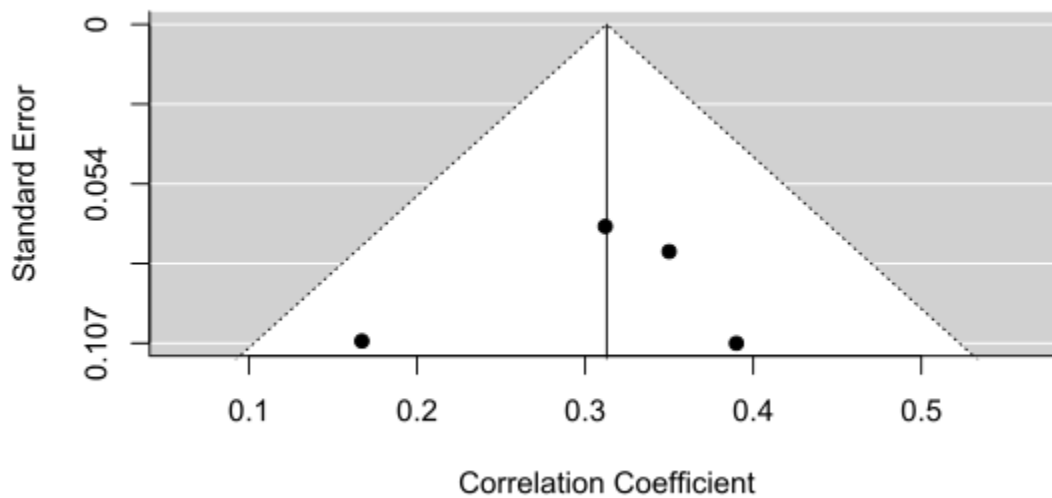


Figure 4.12. GRE - verbal (GRE-V) trim-and-fill funnel plot.

GRE-quantitative (GRE-Q) results. The effects sizes for the four studies that investigated the relationship between GRE-Q and NPTE performance had a smaller range of effect sizes, from 0.27 (95% CI = 0.13-0.40) to 0.35 (95% CI = 0.14-0.56; Figure 4.13). The forest plot suggested no potential publication bias was present. Supporting the results from the forest plot, the mean true effect size from the cumulative meta-analysis had a small downward trend as larger studies were added to the analysis. Taken together, these results suggest that publication bias was not present in the GRE-Q meta-analysis (Figure 4.14).

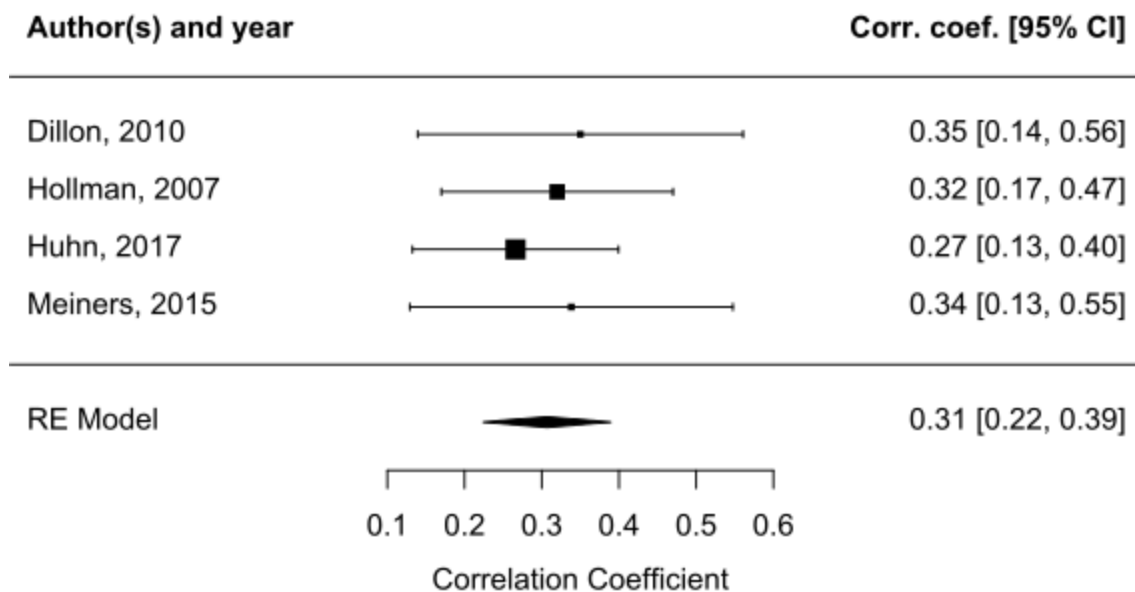


Figure 4.13. GRE-quantitative (GRE-Q) forest plot (RE Model = random effects model).

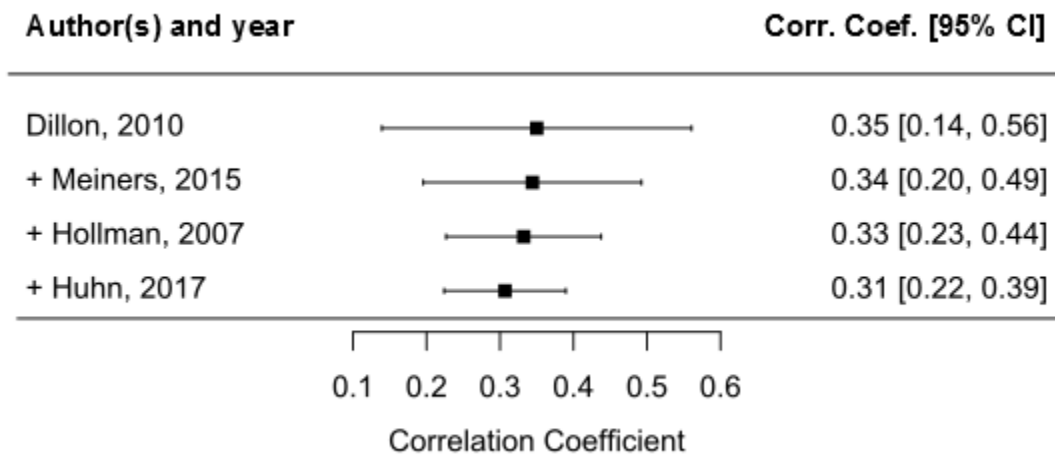


Figure 4.14. GRE-quantitative (GRE-Q) cumulative meta-analysis forest plot.

The qualitative assessment of the GRE-Q funnel plot suggest publication bias was likely present in the meta-analysis (Figure 4.15). Asymmetry was present in the GRE-Q funnel plot, as one or more studies with an effect size less than 0.31 may be missing. The potential publication bias caused an upward bias of the mean true effect size for GRE-Q. Supporting the qualitative

assessment of the funnel plot, the results from the trim-and-fill analysis estimated two studies with effect sizes less than 0.31 are absent from the meta-analysis (Figure 4.16). With the theoretical addition of the estimated two missing studies, the mean true effect size of GRE-Q decreased to 0.29 (95% CI = 0.22-0.36).

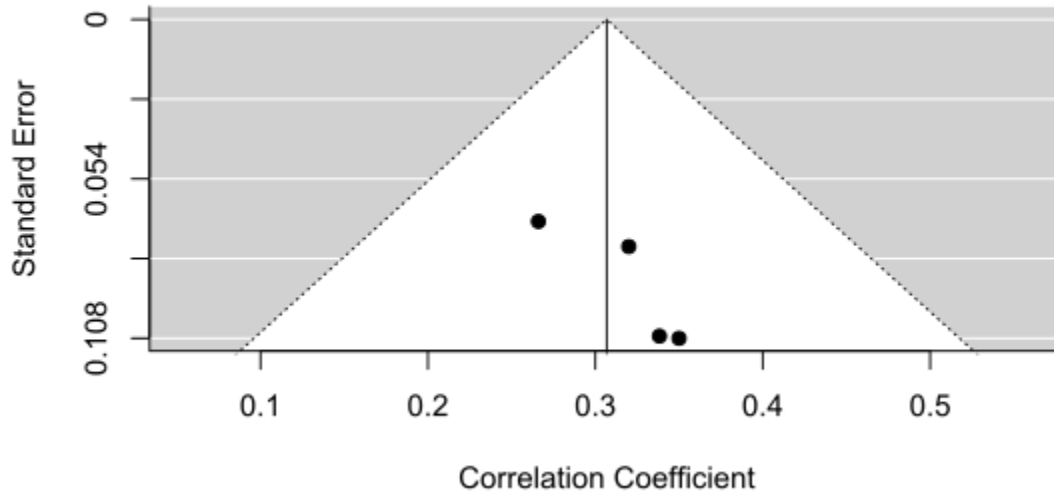


Figure 4.15. GRE-quantitative (GRE-Q) funnel plot.

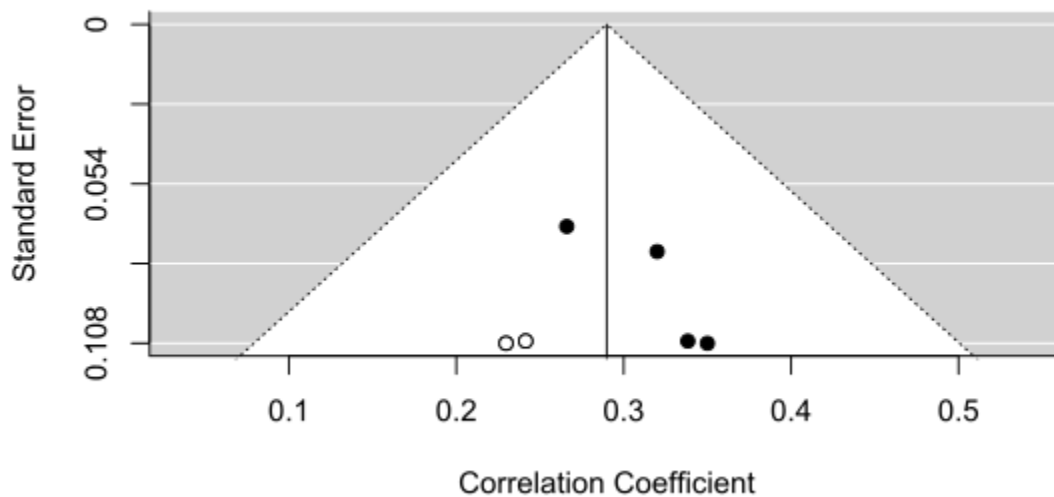


Figure 4.16. GRE-quantitative (GRE-Q) trim-and-fill funnel plot.

Noncognitive applicant variables results. With a mean true effect size of 0.31 (95% CI = 0.20-0.41), there was variability in the effect sizes for noncognitive applicant variables from the three studies included in the meta-analysis (Figure 4.17). The effect sizes ranged from 0.17 (95% CI = -0.02-0.36) to 0.38 (95% CI = 0.26-0.49) which suggested that publication bias was likely present in the noncognitive applicant variables meta-analysis. Supporting the visual inspection of the forest plot, the noncognitive applicant variables mean true effect size changed from 0.25, to 0.20, and finally to 0.31 as larger studies were added to the cumulative meta-analysis (Figure 4.18). The variability of the mean true effect size indicated that publication bias was likely present.

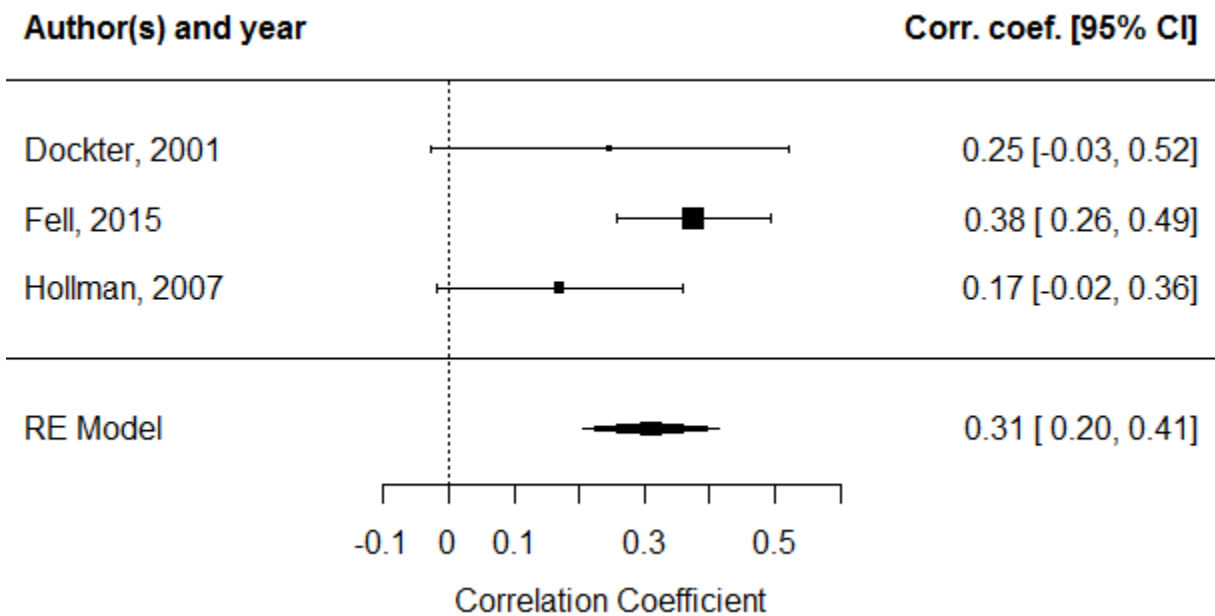


Figure 4.17. Noncognitive applicant variables forest plot (RE Model = random effects model).

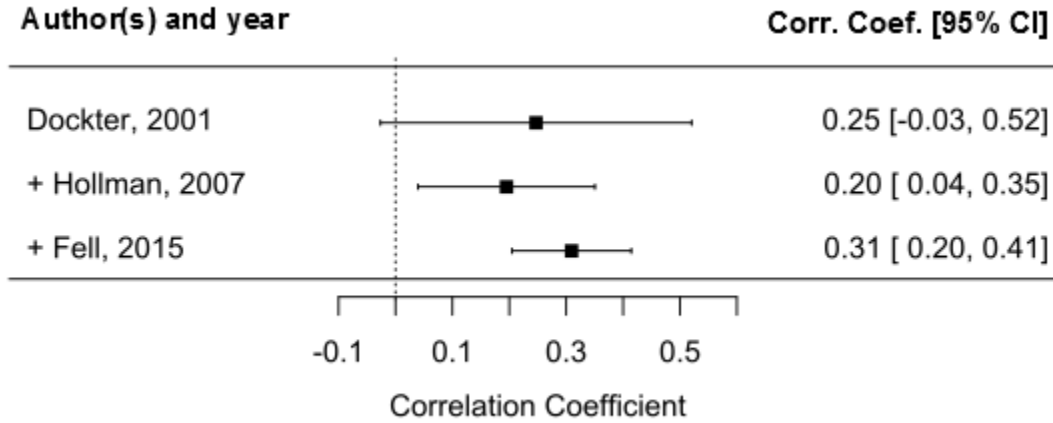


Figure 4.18. Noncognitive applicant variables cumulative meta-analysis forest plot.

Supporting the findings of the cumulative meta-analysis, asymmetry was present in the noncognitive applicant variables funnel plot (Figure 4.19). The noted asymmetry of the noncognitive applicant variables funnel plot suggested a downward bias in the mean true effect size. The results of the noncognitive applicant variables random effects meta-analysis provided evidence that one or more studies with a noncognitive applicant variable effect size greater than 0.31 was absent from the meta-analysis. Adding to the funnel plot assessment, the trim-and-fill results estimated that two studies were absent from the noncognitive applicant variables meta-analysis (Figure 4.20). Both estimated studies were reported to have noncognitive applicant variable effect sizes greater than 0.31, causing the mean true effect size of the noncognitive applicant variables to increase to 0.38 (95% CI = 0.24-0.51).

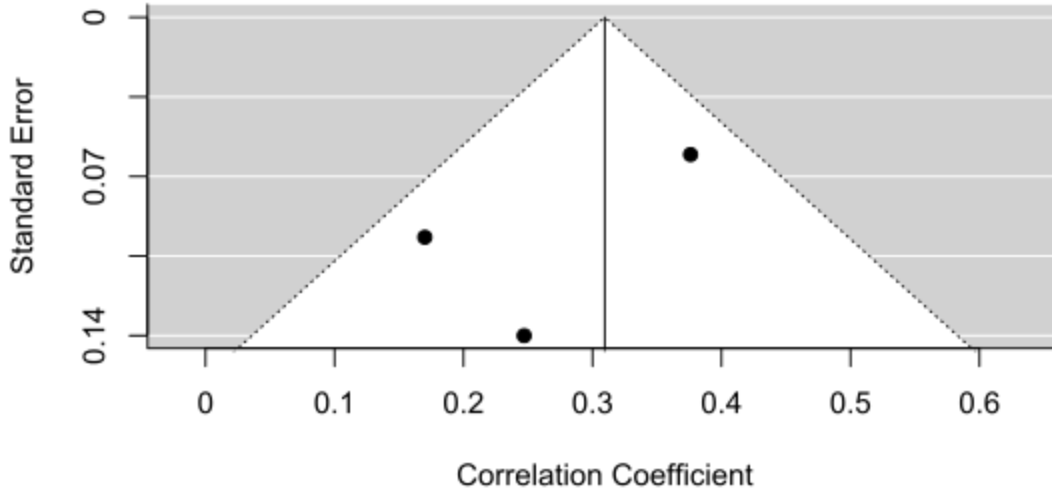


Figure 4.19. Noncognitive applicant variables funnel plot.

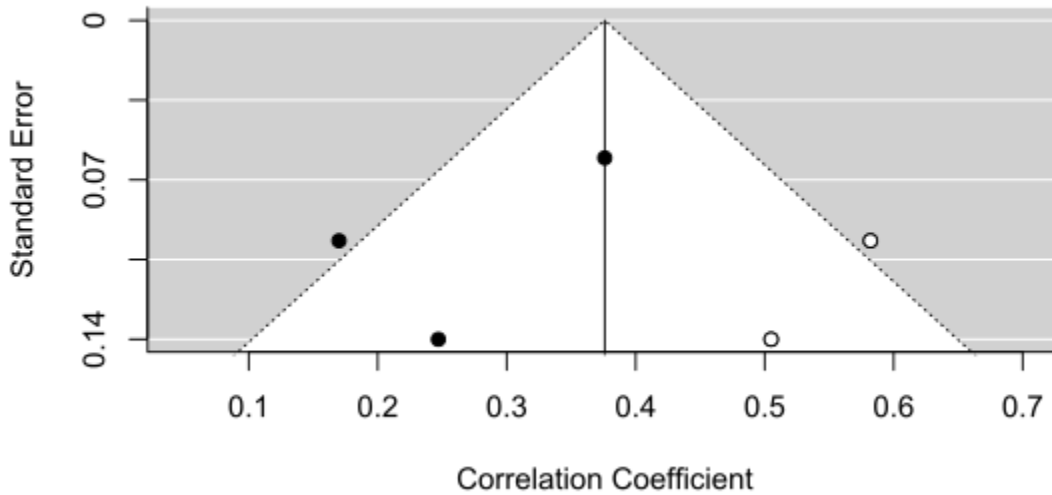


Figure 4.20. Noncognitive applicant variables trim-and-fill funnel plot.

Relationships among PT Student Variables and NPTE Performance

The second research question aimed to determine the relative effect size for each PT student variable on NPTE performance. To answer this research question as precisely as possible, the Hunter-Schmidt meta-analysis using the random effects model was performed on

each PT student variable: (a) first-year PTGPA (b) third-year PTGPA, (c) clinical performance scores (first and final clinical experiences), and (d) noncognitive student variables that met the inclusion criteria. A meta-analysis was also performed on two PT student variables that did not meet the inclusion criteria of having a minimum of four studies in the meta-analysis: (a) Nelson Denney Reading Test (NDRT) scores and (b) comprehensive exam scores. While the NDRT scores and comprehensive exam scores did not meet the inclusion criteria for the minimum number of studies ($\#r \geq 4$), a meta-analysis was performed due to the large and significant effect size on NPTE performance that was reported in the studies that investigated NDRT scores and comprehensive exam scores.

The results of the meta-analyses of the PT student variables found four variables had a large and statistically significant mean true effect size on NPTE performance: (a) first-year PTGPA ($M_p = 0.55$, 95% CI = 0.46-0.63), (b) PTGPA ($M_p = 0.50$, 95% CI = 0.43-0.57), (c) NDRT scores ($M_p = 0.54$, 95% CI = 0.40-0.69), and (d) comprehensive exam scores ($M_p = 0.59$, 95% CI = 0.50-0.68; Table 4.3). Three PT student variables had a small mean true effect size on NPTE performance: (a) clinical performance scores (first clinical experience; $M_p = 0.04$), (b) clinical performance scores (third clinical experience; $M_p = 0.07$), and (c) noncognitive student variables ($M_p = 0.25$; Table 4.3). Of those PT student variables with a small mean true effect size on NPTE performance, the clinical performance scores had a no statistically significant effect (first clinical experience 95% CI = -0.02-0.10; third clinical experience 95% CI = -0.00-0.14). Noncognitive student variables mean true effect size, while small, was statistically significant (95% CI = 0.10-0.40; Table 4.3). PTGPA had the largest total sample size ($n = 1762$); however, noncognitive student variables had the most studies ($\#r = 10$) included in the

corresponding meta-analysis (Table 4.3). Both NDRT scores and comprehensive exam scores had two studies in the corresponding meta-analysis (Table 4.3).

Table 4.3

PT Student Variables Results from Random Effects Meta-Analysis

Independent variable	Total n	# r	M_{ρ}	S_{ρ}^2	80% cred. interval	95% conf. interval
First-year PTGPA	494	4	0.55	0.04	0.42 – 0.69	0.46 – 0.63
Third-year PTPGA	1762	7	0.50	0.03	0.39 – 0.61	0.43 – 0.57
Clinical performance scores (first clinical experience)	1210	3	0.04	0.03	-0.02 – 0.10	-0.02 – 0.10
Clinical performance scores (final clinical experience)	1283	5	0.07	0.04	-0.03 – 0.17	-0.00 – 0.14
Noncognitive student variables	733	10	0.25	0.08	-0.13 – 0.64	0.10 – 0.40
NDRT scores	97	2	0.54	0.07	0.40 – 0.69	0.40 – 0.69
Comprehensive exam scores	199	2	0.59	0.05	0.50 – 0.68	0.50 – 0.68

Total n = total sample size

r = number of Pearson product moment correlations

M_{ρ} = mean true score correlation (summary effect size)

S_{ρ}^2 = variance of true score correlations

80% cred. interval = 80% credibility interval for true score correlation distribution

95% conf. interval = 95% confidence interval around mean true score correlation

Heterogeneity of PT Student Variables

To determine if there were significant differences in the mean true effect sizes among those studies included in the corresponding PT student variable’s meta-analysis, heterogeneity was assessed by the analysis of the Q , p -values, T^2 , T , and I^2 (Table 4.4). The only PT student variable with a statistically significant degree of heterogeneity were third-year PTGPA ($Q = 14.01$, $p = 0.03$) and noncognitive student variables ($Q = 36.80$, $p < 0.01$; Table 4.4).

Consequently, of all of the PT student variables the noncognitive student variables had the largest variance, standard deviation, and proportion of observed heterogeneity ($T^2 = 0.03$, $T = 0.18$, $I^2 = 71.82\%$; Table 4.4). The PT student variables first-year PTPGA ($Q = 7.19$, $p = 0.07$), clinical performance scores (first clinical experience; $Q = 3.11$, $p = 0.21$), and clinical performance scores (final clinical experience; $Q = 6.34$, $p = 0.18$) had a no significant amount of heterogeneity among those studies included within the corresponding meta-analysis (Table 4.4). Due to having only two studies included in the corresponding meta-analysis, heterogeneity was unable to be assessed for the PT student variables, NDRT scores and comprehensive exam scores.

Table 4.4

Heterogeneity Analysis of PT Student Variables

Independent variable	Q	p	T^2	T	I^2
First-year PTGPA	7.19	0.07	0.003	0.06	42.97%
Third-year PTPGA	14.01	0.03	0.002	0.05	43.08%
Clinical performance scores (first clinical experience)	3.11	0.21	0.0001	0.01	1.71%
Clinical performance scores (final clinical experience)	6.34	0.18	0.001	0.03	11.79%
Noncognitive student variables	36.80	< 0.01	0.032	0.18	71.82%
NDRT scores	0.62	0.43	0	0	0%
Comprehensive exam scores	0.21	0.65	0	0	0%

Q = weighted sum of squared differences between individual study effects and the pooled effect across included studies

p-value = indicates the significance of the Q statistic

T^2 = estimated variance of the true effect size for each variable

T = estimated standard deviation of the true effect size

I^2 = ratio of true heterogeneity to total variance

Publication Bias of PT Student Variables

A comprehensive assessment for the potential of publication bias was completed through the analysis of each PT student variable's forest plot, cumulative meta-analysis results, funnel plot, and trim-and-fill results. The forest plot provided the effect size and 95% confidence intervals for each study that was included in the meta-analysis, along with the mean true effect size and 95% confidence interval. When reviewing the forest plots it is important to note that the size of the box symbolizing the correlation coefficient from each included study represents the weight given to that study to the meta-analysis results. The weight given to each study was determined by the sample size of that included study. The greater the sample size the greater the weight of that study in the corresponding meta-analysis. For the cumulative meta-analysis, the included studies are ranked by sample size. After being ranked by sample size, the included studies were subsequently added to the meta-analysis one at a time, starting with the study with the smallest sample size. Following the cumulative meta-analysis, a funnel plot was generated and reviewed for asymmetry for each studied variable. The funnel plot graphically displays the distribution of the effect sizes from the included studies. Following the review of the funnel plot, the trim-and-fill method was used to assess how much the effect size was impacted by publication bias.

First-year PTGPA. Three of the four studies that investigated the relationship between first-year PTGPA and NPTE performance reported a consistent effect size that ranged from 0.59 (95% CI = 0.47-0.72) to 0.65 (95% CI = 0.44-0.86; Figure 4.21). One study (Lewis, 2011) reported an effect size of only 0.42 (95% CI = 0.31-0.53). As illustrated by the cumulative meta-analysis forest plot, the Lewis (2011) study changed the mean true effect size from 0.61 to 0.52.

The smaller effect size reported by Lewis (2011), combined with the cumulative meta-analysis results suggest publication bias was likely present in the first-year PTGPA meta-analysis.

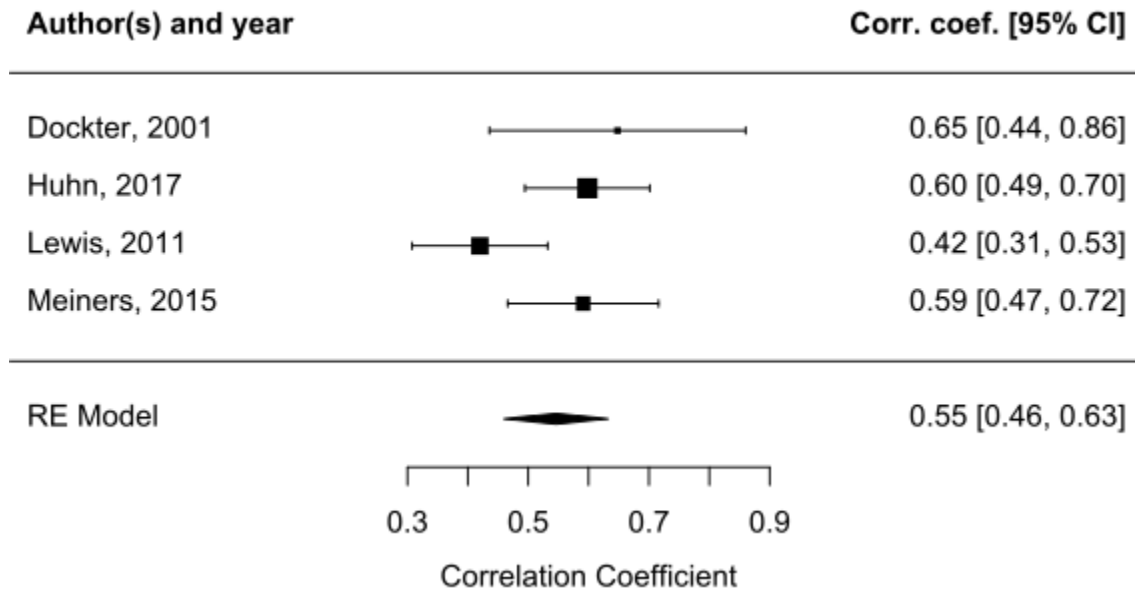


Figure 4.21. First-year PTGPA forest plot (RE Model = random effects model). The size of the box symbolizing the correlation coefficient from each included study represents the weight given to that study to the meta-analysis results. The weight of each study was determined by the sample size of that included study.

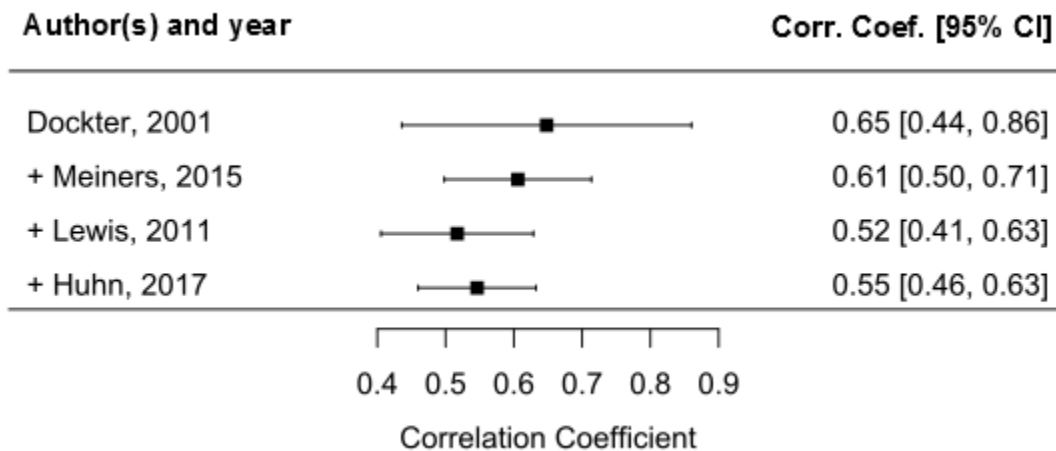


Figure 4.22. First-year PTGPA cumulative meta-analysis forest plot.

Both the qualitative review of the funnel plot and the results of the trim-and-fill analysis support the positive findings of likely publication bias in the first-year PTGPA meta-analysis (Figures 4.23 and 4.24). Left sided asymmetry was evident in the first-year PTGPA funnel plot. Subsequently, the trim-and-fill results estimated that one study was missing on the left side of the funnel plot. The addition of the theoretical missing study to the first-year PTGPA meta-analysis decreased the mean true effect size to 0.54 (95% CI = 0.46-0.62). The results suggest that at least one study with an effect size less than 0.55 was absent from the first-year PTPGA meta-analysis.

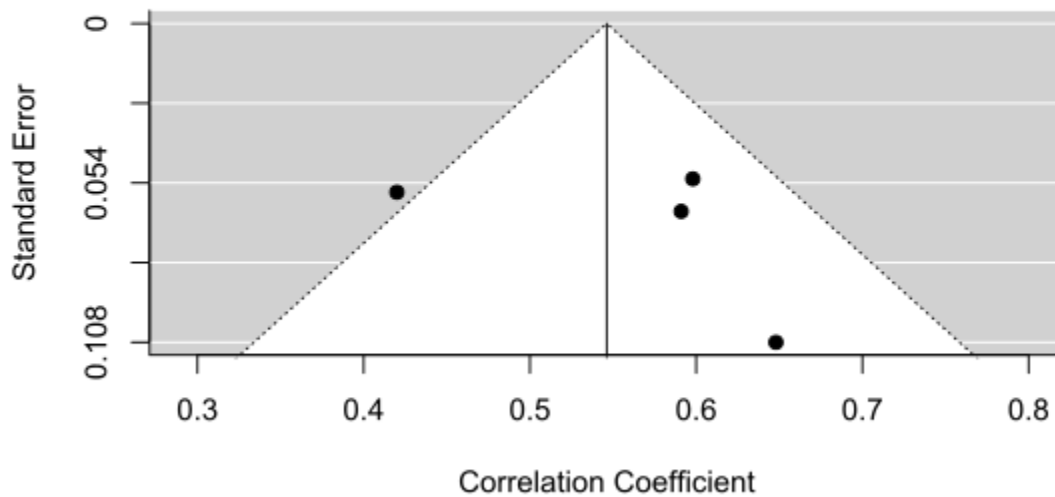


Figure 4.23. First-year PTGPA funnel plot.

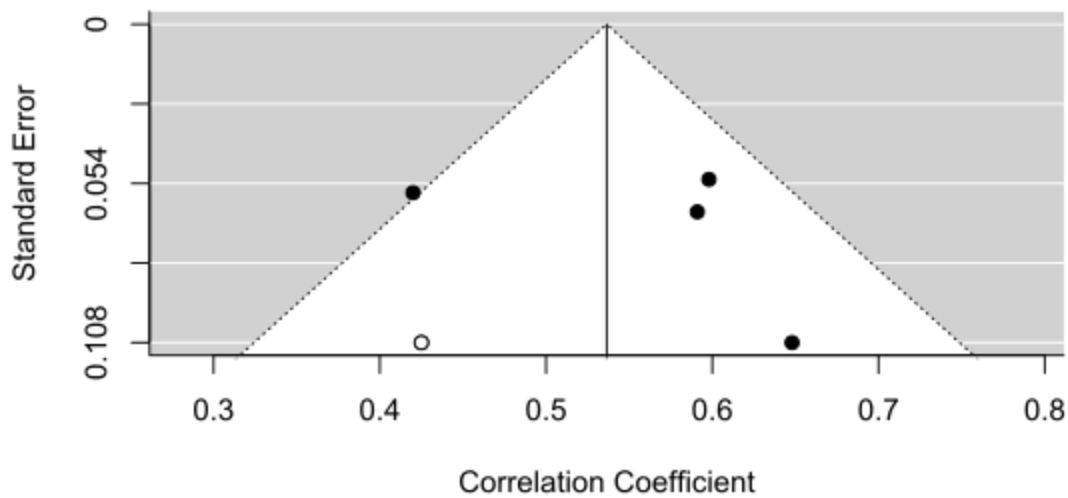


Figure 4.24. First-year PTGPA trim-and-fill funnel plot.

Third-year PTGPA. The meta-analysis consisted of seven studies that investigated the relationship between third-year PTGPA and NPTE performance with a range of effect sizes from 0.33 (95% CI = 0.10-0.56) to 0.60 (95% CI = 0.45-0.76; Figure 4.25). The corresponding cumulative meta-analysis results found an upward trend in the mean true effect size as larger studies were added to the analysis (Figure 4.28). The upward trend in the mean true effect size as larger studies were added is an indication that potential publication bias was not present in the third-year PTGPA meta-analysis (Borenstein et al., 2009).

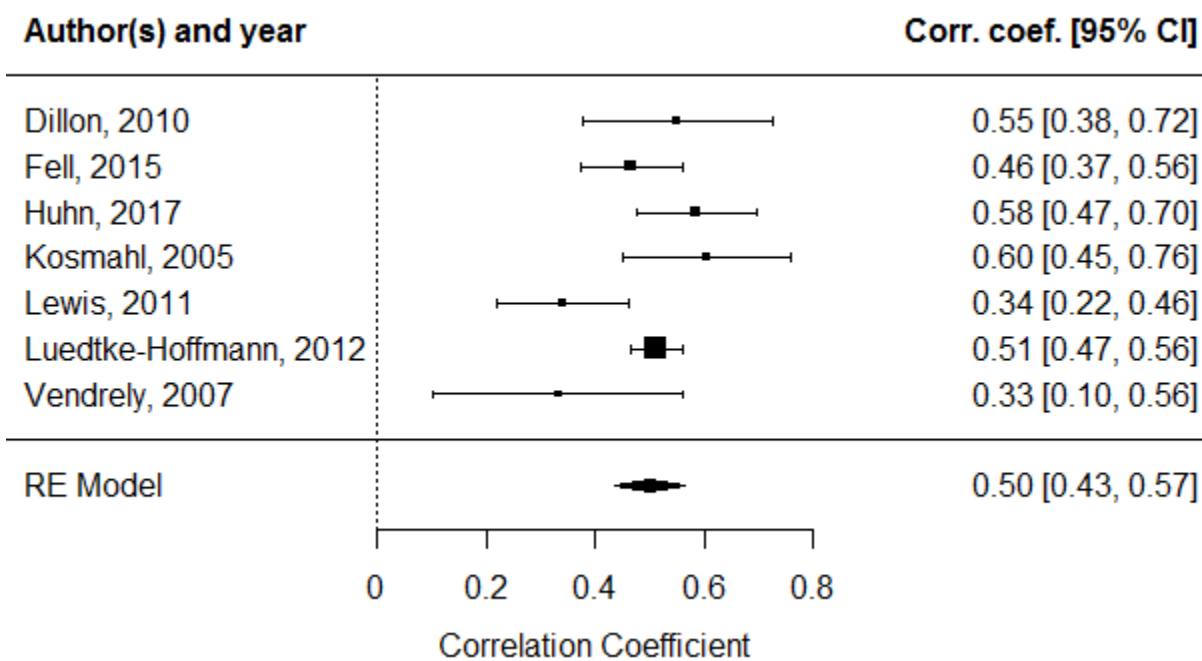


Figure 4.25. Third-year PTGPA forest plot (RE Model = random effects model).

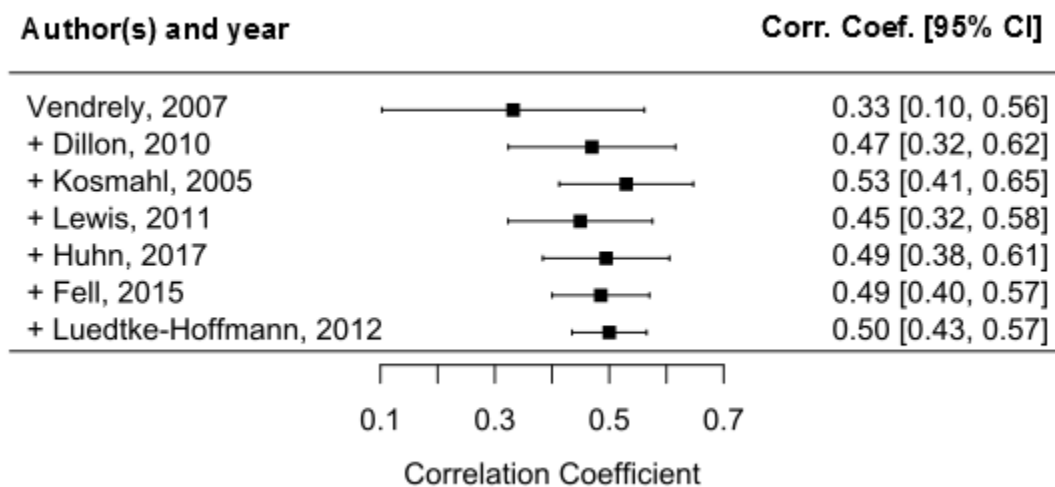


Figure 4.26. Third-year PTGPA cumulative meta-analysis forest plot.

Upon review, symmetry was present in the third-year PTPGA funnel plot. Consistent with the cumulative meta-analysis, the third-year PTGPA funnel plot results indicated

publication bias was not likely present in the meta-analysis (Figure 4.27). Supporting the notion that publication bias was not present, the trim-and-fill results estimated that there were zero missing studies in the third-year PTGPA meta-analysis. As no publication bias was present, the mean true effect size of third-year PTGPA's relationship with NPTE performance was 0.50 (95% CI = 0.43-0.57).

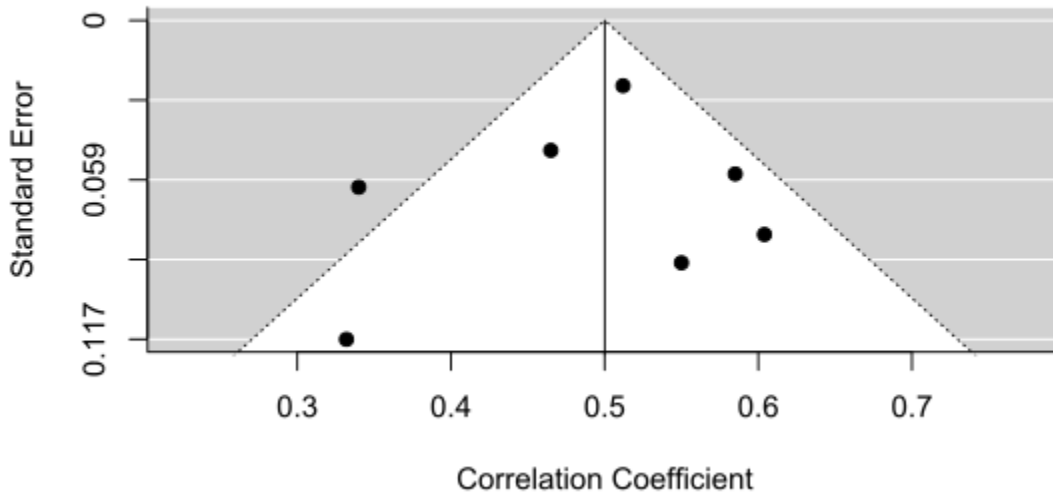


Figure 4.27. Third-year PTGPA funnel plot.

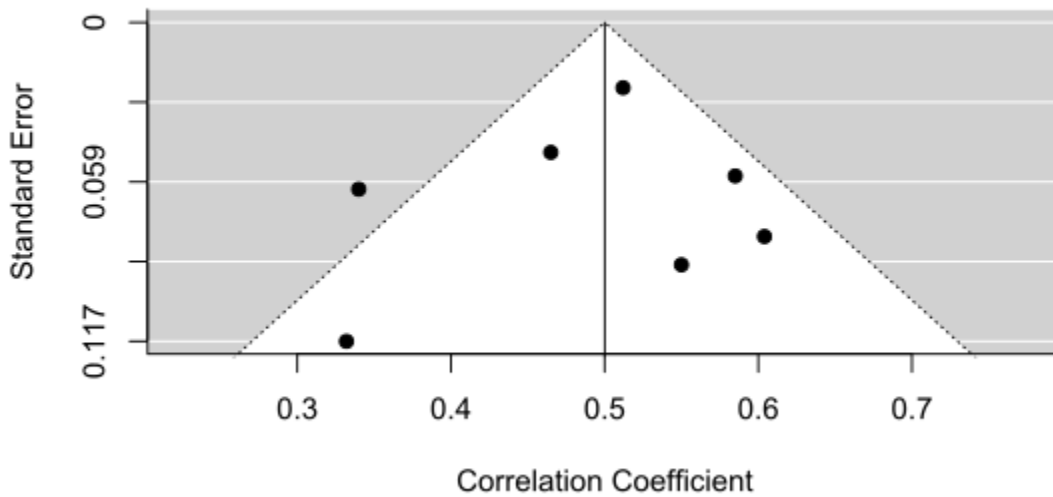


Figure 4.28. Third-year PTGPA trim-and-fill funnel plot.

Clinical performance scores (first clinical experience). Three studies were used in the meta-analysis to investigate clinical performance scores (first clinical experience) with NPTE performance. From the included three studies, there was a range of effect sizes 0.01 (95% CI = -0.14-0.17) to 0.24 (95% CI = 0.01-0.47; Figure 4.29). The distribution of the effect sizes was an indication that publication bias was present, which was supported by the results of clinical performance scores (first clinical experience) cumulative meta-analysis (Figure 4.30). The cumulative meta-analysis effect size started at 0.24 (95% CI = 0.01-0.47), before decreasing to 0.09 (95% CI = -0.06-0.23), and finally decreasing to 0.04 (95% CI = -0.02-0.10) when the largest study (Luedtke-Hoffman et al., 2012) was added to the analysis. The downward bias in the mean true effect size as larger studies were added to the analysis was an indication that publication bias was likely present in the clinical performance scores (first clinical experience) meta-analysis (Borenstein et al., 2009).

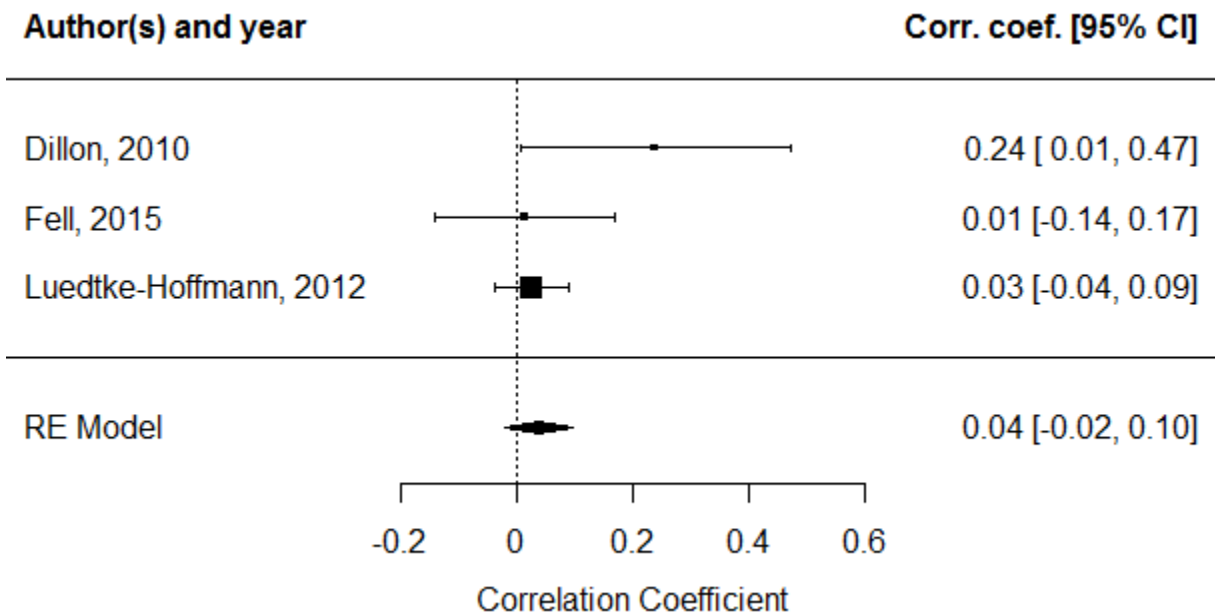


Figure 4.29. Clinical performance scores (first clinical experience) forest plot (RE Model = random effects model).

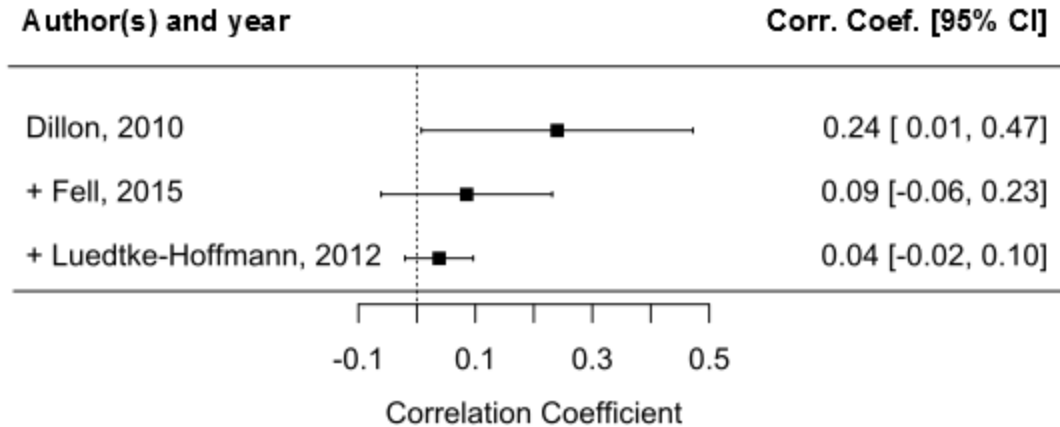


Figure 4.30. Clinical performance scores (first clinical experience) cumulative meta-analysis forest plot.

With only three studies in the meta-analysis, asymmetry was present in the qualitative assessment of the funnel plot (Figure 4.31). The funnel plot appeared to be missing smaller studies with an effect size less than 0.04. While the review of the forest plot, cumulative meta-analysis results, and funnel plot suggest publication bias was likely present in the meta-analysis, the trim-and-fill results estimated that zero studies were absent from the analysis (Figure 4.32). The trim-and-fill results suggest no publication bias was present in the meta-analysis, and therefore, the mean true effect size of 0.04 (95% CI = -0.02-0.10) was correct. Taken together, the presence of publication bias was unable to be determined in the clinical performance scores (first clinical experience) meta-analysis.

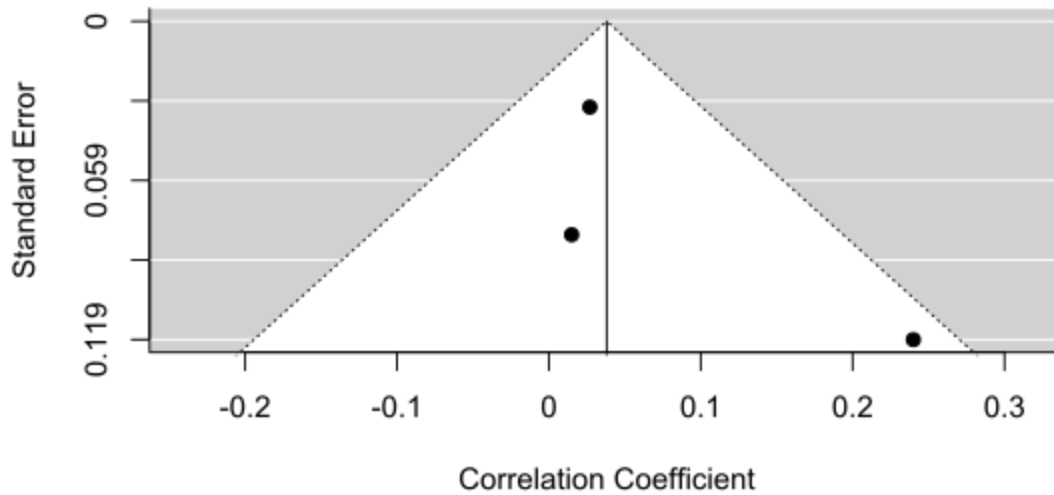


Figure 4.31. Clinical performance scores (first clinical experience) funnel plot.

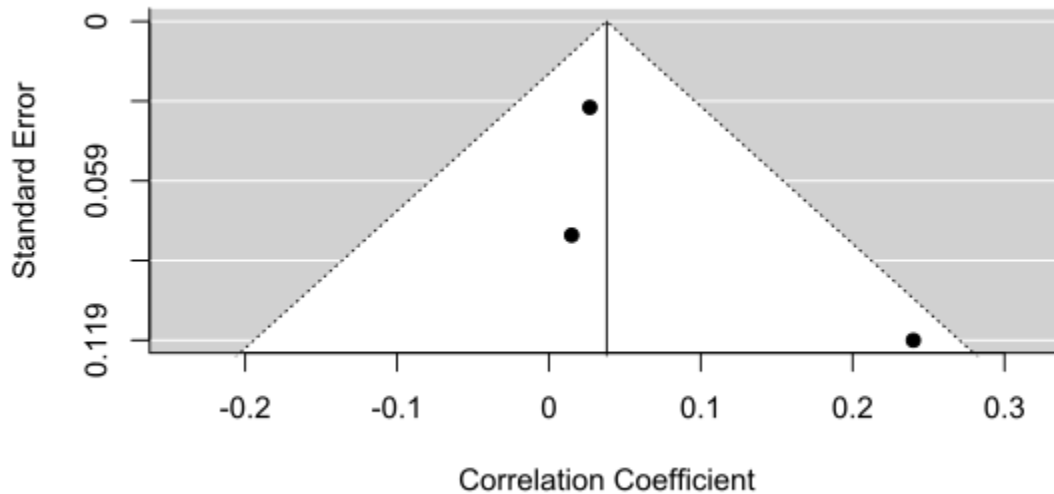


Figure 4.32. Clinical performance scores (first clinical experience) trim-and-fill funnel plot.

Clinical performance scores (final clinical experience). The effect sizes from the five studies included in the clinical performance scores (final clinical experience) meta-analysis that met the inclusion criteria, ranged from -0.09 (95% CI = -0.29-0.11) to 0.26 (95% CI = 0.07-0.46), with a non-significant mean true effect size of 0.07 (95% CI = -0.00-0.14; Figure 4.33). Following the review of the effect sizes from the forest plot, it was unable to be determined if

publication bias was possible in the clinical performance scores (final clinical experience) meta-analysis. As the smallest studies appeared to have minimal effect on the mean true effect size, the clinical performance scores (final clinical experience) cumulative meta-analysis provided no indication of potential publication bias, (Figure 4.34).

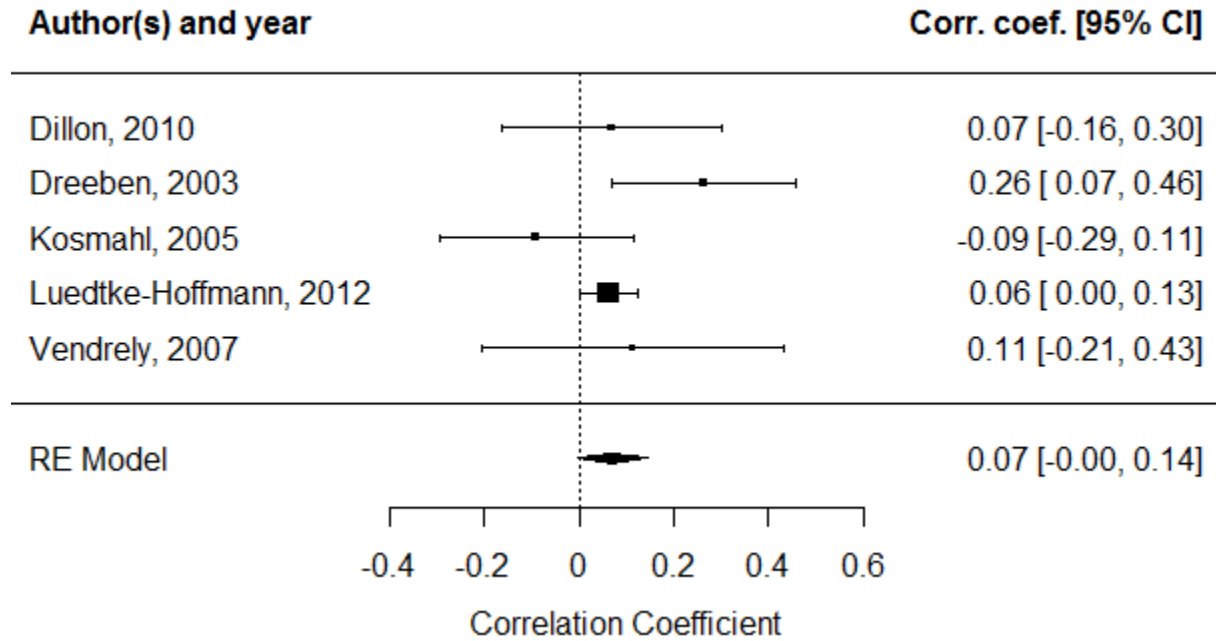


Figure 4.33. Clinical performance scores (final clinical experience) forest plot (RE Model = random effects model).

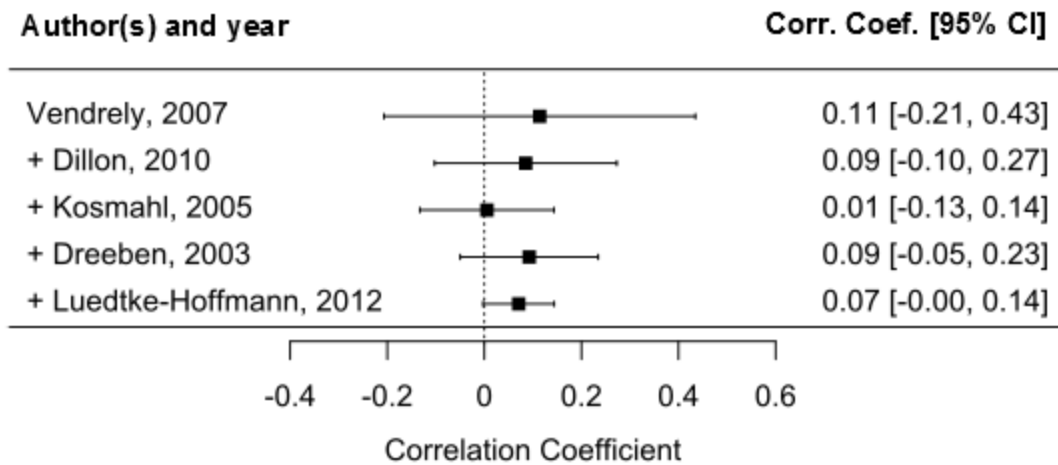


Figure 4.34. Clinical performance scores (final clinical experience) cumulative meta-analysis forest plot.

The funnel plot for the clinical performance scores (final clinical experience) was symmetrical, providing further evidence that publication bias was not likely present in the clinical performance scores (final clinical experience) meta-analysis (Figure 4.35). The trim-and-fill results estimated that zero studies were absent from the clinical performance scores (final clinical experience) meta-analysis (Figure 4.36). Based upon the evidence from the forest plot, cumulative meta-analysis, funnel plot, and trim-and-fill results, the mean true effect size of 0.07 for clinical performance scores (final clinical experience) relationship with NPTE performance was accurate as no publication bias was likely present in the meta-analysis (Figure 4.33).

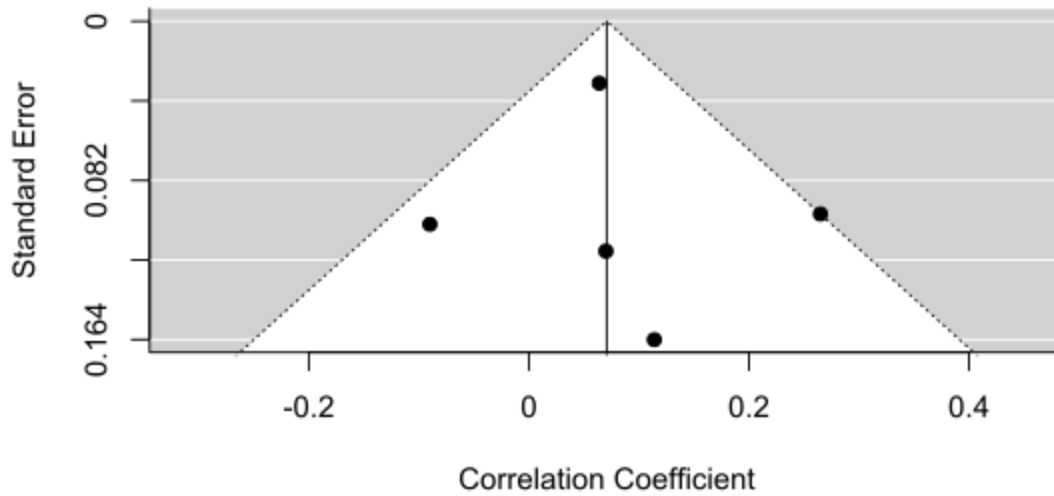


Figure 4.35. Clinical performance scores (final clinical experience) funnel plot.

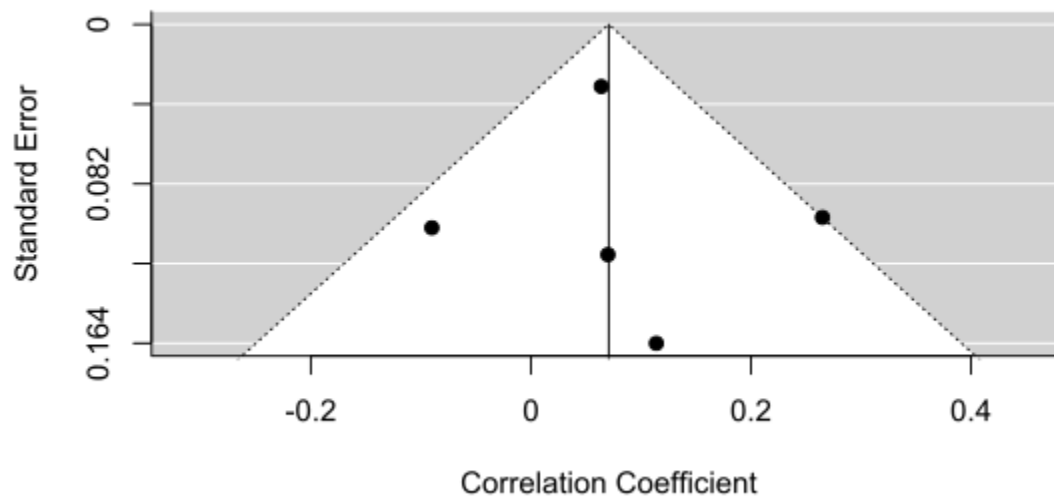


Figure 4.36. Clinical performance scores (final clinical experience) trim-and-fill funnel plot.

Noncognitive student variables results. With a mean true effect size of 0.25 (95% CI = 0.10-0.40), there was variability in the effect sizes for the noncognitive student variables from the ten studies included in the meta-analysis (Figure 4.37). The effect sizes from the included studies ranged from -0.18 (95% CI = -0.41-0.06) to 0.63 (95% CI = 0.29-0.97) which suggested

that publication bias was present in the noncognitive student variables meta-analysis (Figure 4.37). The results of the noncognitive student variables cumulative meta-analysis provided further evidence that publication bias was likely present, as the mean true effect size increased when smaller studies were added to the analysis (Borenstein et al., 2009; Figure 4.38).

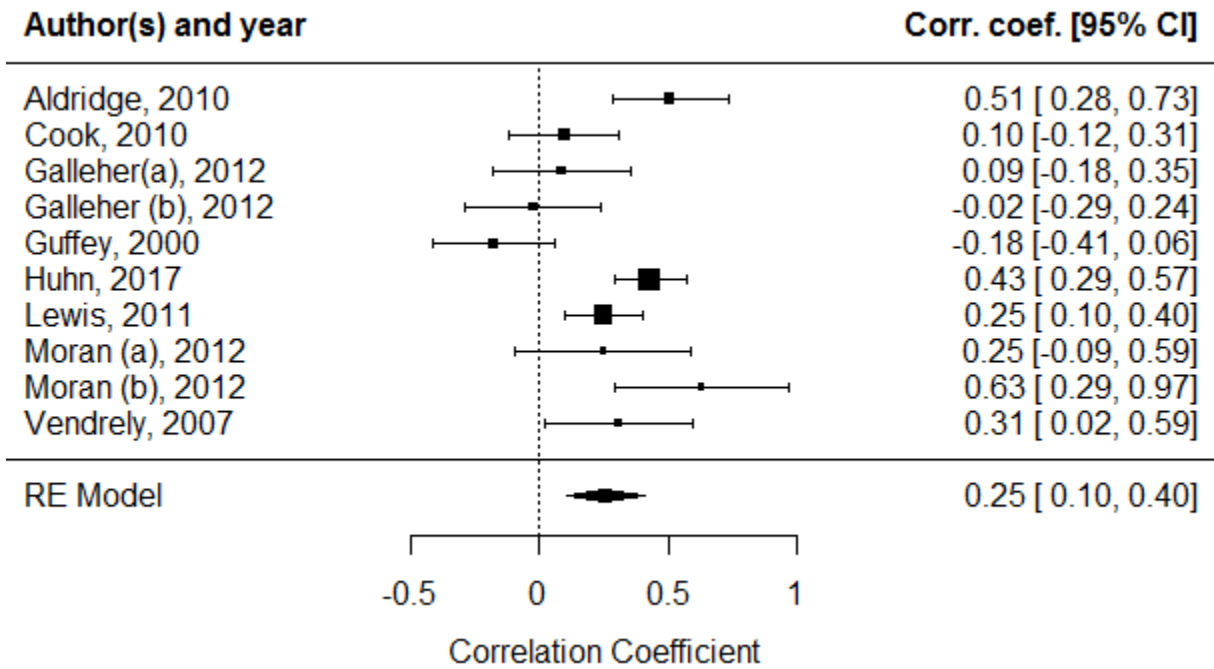


Figure 4.37. Noncognitive student variables forest plot (RE Model = random effects model).

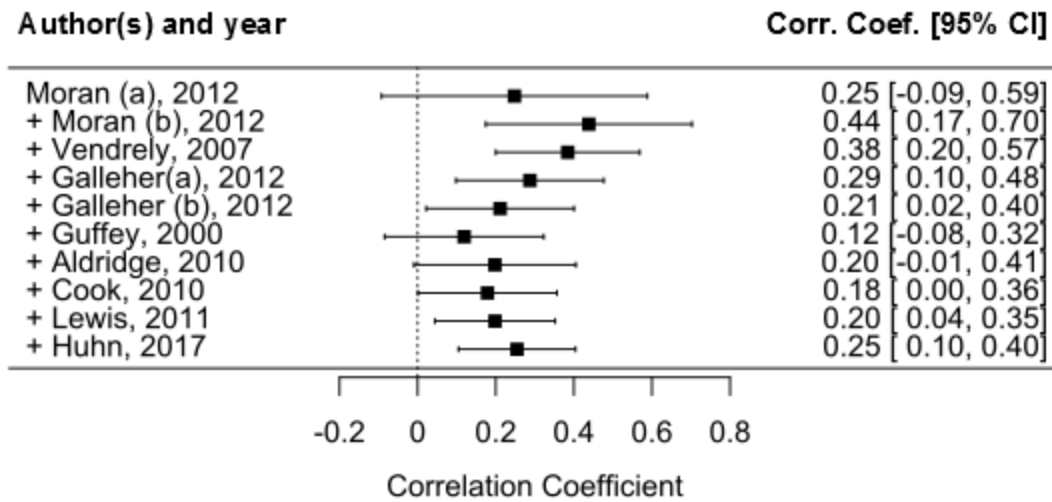


Figure 4.38. Noncognitive student variables cumulative meta-analysis forest plot.

From the visual inspection of the noncognitive student variables funnel plot, left sided asymmetry was present suggesting that the calculated mean true effect size may be smaller than the theoretical mean true effect size (Figure 4.39). The results of the noncognitive student variables random effects meta-analysis provided evidence that one or more studies with a noncognitive applicant variable effect size greater than 0.25 was likely absent from the meta-analysis. The noncognitive student variables trim-and-fill results estimated that one study was absent from the noncognitive student variables meta-analysis (Figure 4.40). The theoretical absent study was reported to have a noncognitive student variable effect size great than 0.70, causing the mean true effect size of the noncognitive student variables to increase to 0.29 (95% CI = 0.13-0.46; Figure 4.40). The results of the noncognitive student variables forest plot, cumulative meta-analysis, funnel plot, and trim-and-fill suggest that at least one study with an effect size greater than 0.25 was absent from the meta-analysis.

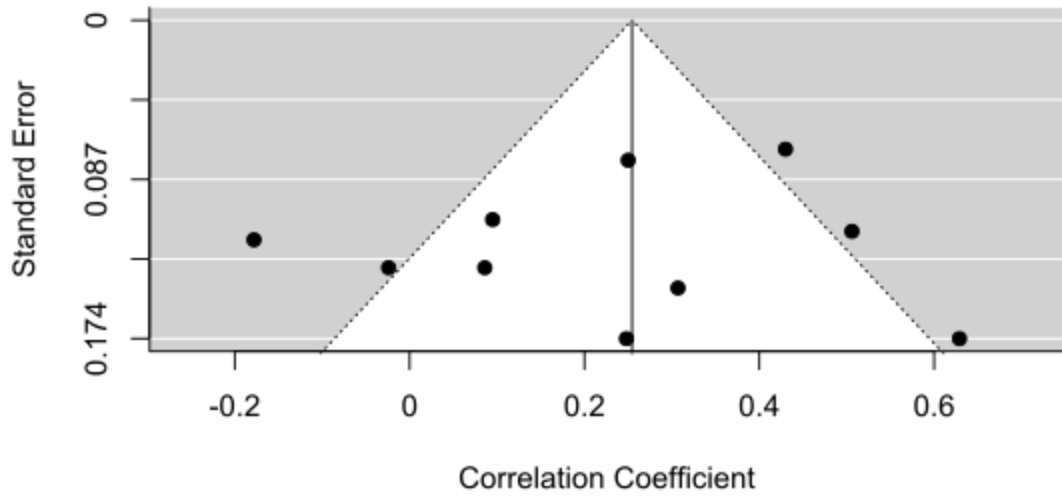


Figure 4.39. Noncognitive student variables funnel plot.

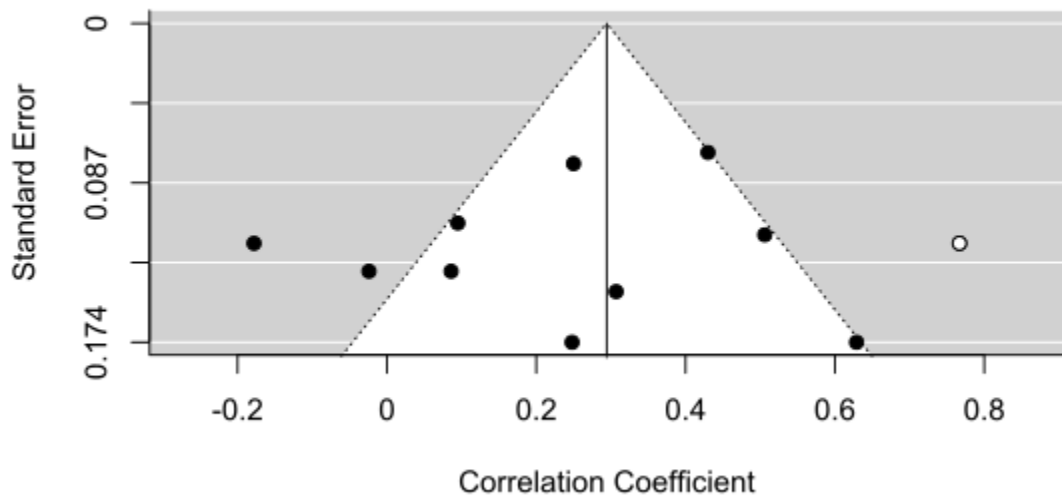


Figure 4.40. Noncognitive student variables trim-and-fill funnel plot.

Nelson Dennehy Reading Test (NDRT) scores. With only two studies (Aldridge et al., 2010; Moran, 2012) included in the NDRT scores meta-analysis, the potential for publication bias cannot be accurately assessed. The NDRT forest plot does provide the effect sizes for the two studies that investigated the relationship between NDRT scores and NPTE performance (Figure 4.41). Each selected study had a large effect size: 0.51 (95% CI = 0.34-0.68) and 0.63 (95% CI = 0.37-0.89; Figure 4.41). The NDRT scores cumulative meta-analysis started with a mean true effect size of 0.63 (95% CI = 0.37-0.89) with Moran (2012) study, and when the Aldridge et al. (2010) study was added to the analysis the calculated mean true effect size decreased to the calculated mean true effect size of 0.54 (95% CI = 0.40-0.69).

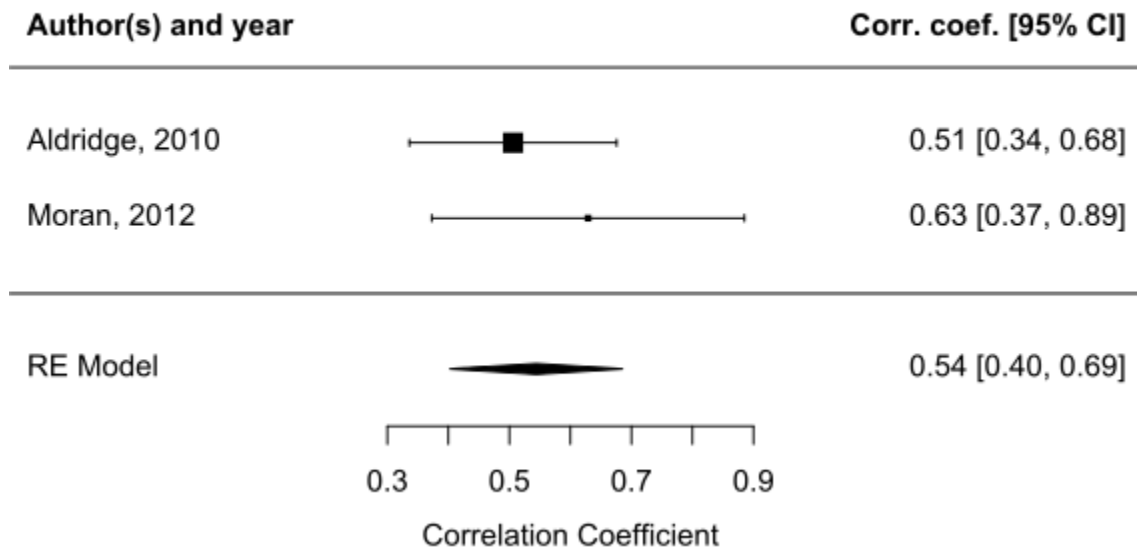


Figure 4.41. Nelson Dennehy Reading Test (NDRT) forest plot (RE Model = random effects model).

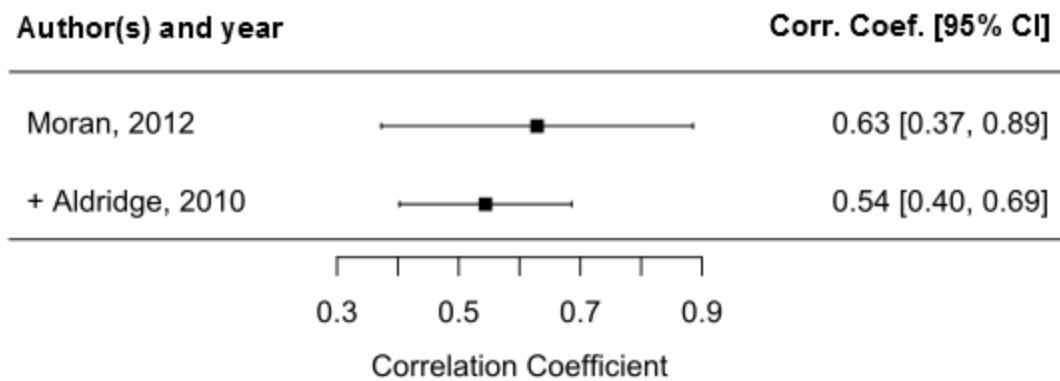


Figure 4.42. Nelson Dennehy Reading Test (NDRT) cumulative meta-analysis forest plot.

The NDRT funnel plot was symmetrical, indicating that publication bias was not likely present in the meta-analysis (Figure 4.43). A trim-and-fill analysis was unable to be conducted due to the limited number of included studies in the NDRT meta-analysis. With only two studies included in the NDRT meta-analysis, the subsequent forest plot, cumulative meta-analysis, and funnel plot provided no evidence of publication bias. More studies on the relationship between NDRT scores and NPTE performance are needed to accurately assess for publication bias in the meta-analysis and determine the usefulness of NDRT scores to predict NPTE performance.

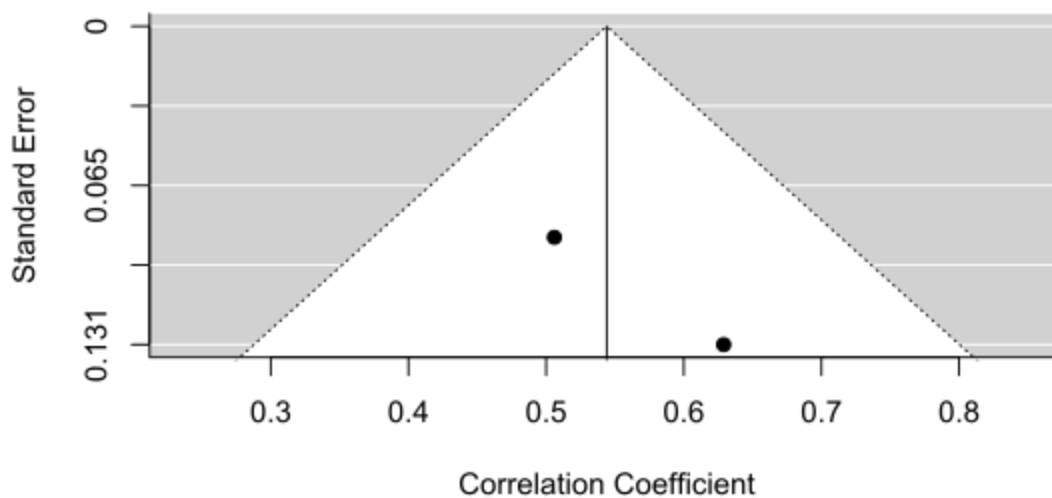


Figure 4.43. Nelson Dennehy Reading Test (NDRT) funnel plot.

Comprehensive exam scores. The potential for publication bias was unable to accurately be assessed due to only two studies (Edmondson, 2001; Kosmahl, 2005) being selected to be included in the comprehensive exam scores meta-analysis. From the two studies included in the comprehensive exam scores meta-analysis, the relationship of comprehensive exam scores with NPTE performance had large effect sizes of 0.57 (95% CI = 0.45-0.70) and 0.62 (95% CI = 0.48-0.75; Figure 4.44). The comprehensive exam scores cumulative meta-analysis provided no indication of publication bias. In the cumulative meta-analysis the mean true effect size started at 0.62 (95% CI = 0.48-0.75) with the Kosmahl's (2005) study, and decreased to the calculated mean true effect size of 0.59 (95% CI = 0.50-0.68) when Edmondson's (2001) study was added to the analysis (Figure 4.45).

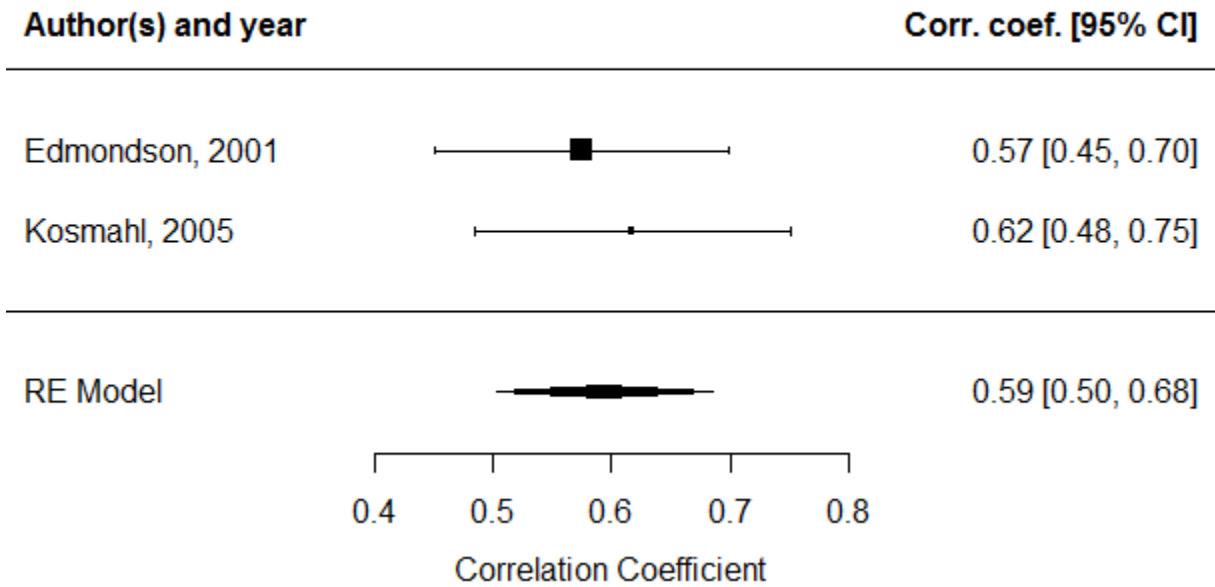


Figure 4.44. Comprehensive exam scores forest plot (RE Model = random effects model).

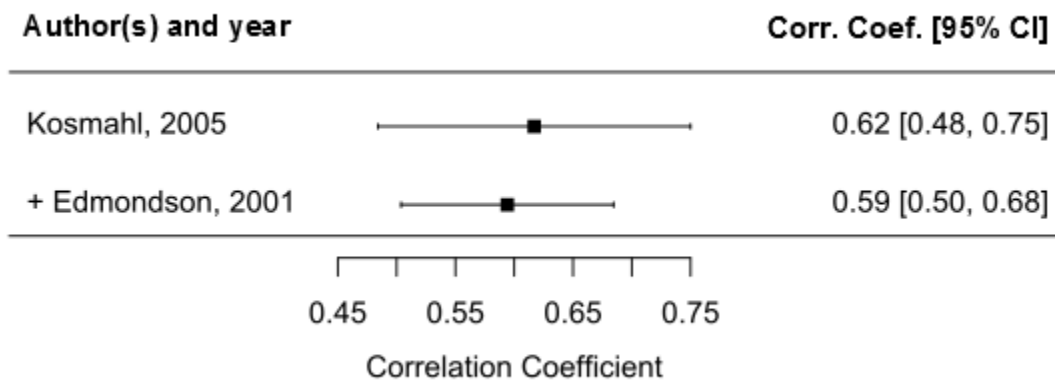


Figure 4.45. Comprehensive exam scores cumulative meta-analysis forest plot.

The comprehensive exam scores funnel plot was symmetrical (Figure 4.46). A trim-and-fill analysis was also unable to be conducted due to the limited number of included studies in the meta-analysis. The forest plot, cumulative meta-analysis, and funnel plot provided no evidence of publication bias. More studies on the relationship between comprehensive exam scores and

NPTE performance are needed to accurately assess the usefulness of comprehensive exam scores to predict NPTE performance.

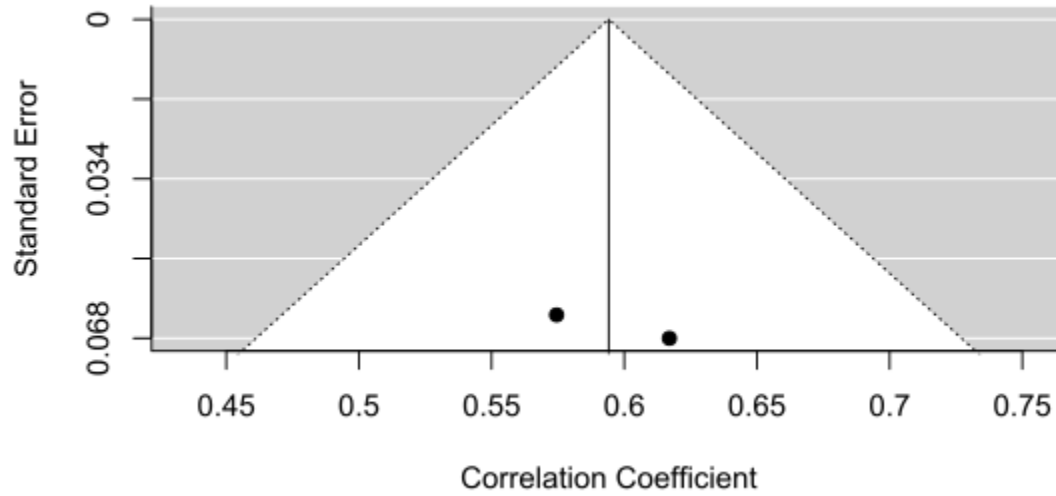


Figure 4.46. Comprehensive exam scores funnel plot.

CHAPTER V: DISCUSSION

Every PT student across the United States (U.S.) must pass the NPTE to gain licensure to practice as a physical therapist (Commission on Accreditation of Physical Therapy Education, 2016). Therefore, an important goal for each PT program in the U.S. is that each student passes the NPTE, and preferably, on their first-attempt. Passing the NPTE on the first-attempt provides validity to each PT programs standing with CAPTE regarding the quality of the students, faculty, and program itself. The consequences for PT students, faculty, and each program that follow the results of the NPTE justify the importance that PT programs place on being able to predict with accuracy and precision which students are most likely to successfully pass the NPTE on the first-attempt.

The purpose of the present study was to determine the empirical relationships with NPTE performance for each of the PT applicant and PT student variables identified from the literature. To quantify the empirical relationship that each PT applicant and each PT student variable has on NPTE performance; a systematic literature review, as defined by Khan et al. (2003), with a corresponding random effects meta-analysis was conducted. The significance and practical implications from the results for each PT applicant and each PT student variable's meta-analysis provide a clear, precise, and accurate prediction for each PT applicant and each PT student variable's relationship with NPTE performance.

The present discussion will include four parts: (a) the relationship among PT applicant variables and NPTE performance, (b) the relationship among PT student variables and NPTE performance (c) conclusion, and (d) future research. A discussion of the results of the PT applicant variables will be followed by a discussion of the results of the PT student variables. The conclusion will bring together the results of the present study and the discussion from both

the PT applicant and PT student variables. Lastly, future research questions that were generated from the present study will be presented, with some commentary provided.

The Relationship among PT Applicant Variables and NPTE Performance

Regarding the magnitudes of effect sizes for each PT applicant variable on NPTE performance, the results of the random effects meta-analysis found that all of the investigated PT applicant variables: (a) UGPA ($M_{\rho} = 0.33$), (b) UGPA-PC ($M_{\rho} = 0.37$), (c) GRE-V ($M_{\rho} = 0.32$), (d) GRE-Q ($M_{\rho} = 0.31$), and (e) noncognitive applicant variables ($M_{\rho} = 0.31$) had a medium effect on NPTE performance (Table 4.1). The PT applicant random effects meta-analyses found that UGPA and UGPA-PC had the strongest relationship with NPTE performance.

Throughout professional healthcare education, UGPA and UGPA-PC are widely used as admission variables. The results from the UGPA and UGPA-PC random effects meta-analyses are consistent with the use of UGPA and UGPA-PC in the selection of applicants for professional healthcare education. In professional healthcare education, UGPA and UGPA-PC have been able to predict applicants' success on the relevant national licensure examinations with some certainty in pharmacy, physician assistant, nursing, and medicine (Allen & Diaz, 2013; Burns, 2011; Higgins et al., 2010; Peskun, Detsky, & Shandling, 2007).

The results of the UGPA and UGPA-PC random effects meta-analyses provide strong evidence to set a minimum UGPA and UGPA-PC as primary admission criteria in PT education and identify those applicants who are likely to pass the NPTE on their first-attempt. That said, it is important to recognize the significant, but acceptable amount of heterogeneity that was present for both the UGPA random effect meta-analysis ($Q = 17.38$, $p = 0.01$, $I^2 = 58.38\%$) and the UGPA-PC random effects meta-analysis ($Q = 13.44$, $p < 0.01$, $I^2 = 66.96\%$; Higgins, 2008). In social science research, often with a large number of studies, heterogeneity is accepted. This is

not the case in PT education. In the present study, the findings of a significant degree of heterogeneity in the UGPA and UGPA-PC meta-analyses are questionable, as the power of the Q statistic is poor when only a limited number of studies ($\#r \leq 10$) are analyzed (Higgins & Thompson, 2002). Since both the UGPA ($\#r = 7$) and UGPA-PC ($\#r = 4$) variables have a small number of included studies, the significant Q statistics from those meta-analyses should be interpreted with caution in PT education.

Using UGPA and UGPA-PC as PT applicant variables stems from the theory that academically successful students in undergraduate education will continue to be academically successful in their professional doctorate education. Based upon the results from the present study, UGPA-PC is the best PT applicant predictor of NPTE performance. UGPA-PC is comprised of the primary undergraduate courses that each PT program requires for students to be successful in their PT program. Similar to UGPA-PC, PT programs foundational coursework consists of science-based courses (i.e. anatomy, biomechanics, and pathology). The assertion that past behavior predicts future behavior, in the same context, has led UGPA-PC to be a common admission variable for PT programs. With PT programs specifying the coursework that makes up UGPA-PC, it is reasonable to conclude that UGPA-PC should be able to accurately and precisely predict NPTE performance.

Even though UGPA-PC was the best predictor of NPTE performance from the PT applicant variables, the results of the random effects meta-analysis indicate that publication bias was likely present. However, the presence of publication bias in the UGPA-PC random effects meta-analysis was determined to have a positive impact on the mean true effect size and was subsequently estimated to increase the mean true effect size to as much as 0.46 (95% CI = 0.28-0.64; Figure 4.8). The findings of a negatively skewed mean true effect size due to publication

bias adds further evidence that UGPA-PC is the best PT applicant predictor of NPTE performance.

While UGPA and UGPA-PC are intended to measure a PT applicant's learning in undergraduate studies, the Graduate Record Examination (GRE) is designed to measure a student's ability to perform verbal (GRE-V) and quantitative (GRE-Q) reasoning (Education Testing Service, 2017). As valid predictors of graduate student performance, GRE-V and GRE-Q are common PT applicant variables used by most PT programs (Kuncel et al., 2012). The results from the GRE-V random effects meta-analysis ($M_p = 0.32$) support the hypothesis that GRE-V has a medium effect on NPTE performance (Table 4.1). However, our results from the GRE-Q meta-analysis ($M_p = 0.31$) does not support the assertion that GRE-Q has only a small effect on NPTE performance. Our results found that GRE-Q has a medium effect on NPTE performance (Table 4.1). With a significant medium effect on NPTE performance, both GRE-V and GRE-Q should continue to be used in the PT admissions process to aid in the selection of the PT applicants who are most likely to pass the NPTE on the first-attempt.

The majority of the evidence related to PT applicant variables focuses on cognitive applicant variables (i.e. UGPA, UGPA-PC, GRE-V, and GRE-Q). In the present study, the noncognitive applicant variables relationship with NPTE performance were also assessed with a random effects meta-analysis. There were three studies in the noncognitive applicant variables random effects meta-analysis. The three noncognitive applicant variables that predicted NPTE performance were related to PT applicant interview scores relationship with NPTE performance. In the noncognitive applicant variables random effects meta-analysis, there was evidence of a negatively skewed mean true effect size due to publication bias. That is, one or more studies with an effect size on NPTE performance greater than 0.31 has the potential to be missing from the

analysis. The trim-and-fill analysis predicted the extent of publication bias could increase the mean true effect size to 0.38. While the potential change in the mean true effect size to 0.38 would be a significant increase from the calculated mean true effect size ($M_p = 0.31$), noncognitive applicant variables still has a medium effect of on NPTE performance.

Based upon the results of the present study's noncognitive applicant variables using a random effects meta-analysis and the assessment of potential publication bias, we can conclude that a structured and focused PT applicant interview and admissions process has a significant medium effect on NPTE performance. Our results provide evidence regarding the importance of a purposeful PT admissions process, and the important role that PT interviews have within that admissions process. However, with only three studies investigating three different noncognitive variables included in the noncognitive applicant variables random effects meta-analysis, future research is needed to determine the best noncognitive applicant variables and to better understand PT interviews impact on identifying students who are most likely to pass the NPTE on their first-attempt.

PT programs use a variety of PT applicant variables to identify which students are most likely to successfully graduate and pass the NPTE on their first-attempt. The present PT applicant random effects meta-analyses provide evidence for the use of both cognitive and noncognitive applicant variables in the PT admissions process. From our analysis and despite the limited range of scores (3.0 to 4.0), UGPA-PC is the best PT applicant predictor of PT students NPTE performance. However, with only a medium effect size, UGPA-PC should not be used as the only predictor of NPTE performance. Similar to UGPA-PC, each of the studied PT applicant variables (UGPA, GRE-V, GRE-Q, and noncognitive applicant variables) had a medium effect on NPTE performance. Our results provide evidence that a comprehensive approach should be

employed when selecting a PT applicant who is most likely to pass the NPTE on their first-attempt. In the development of an evidence-based PT admissions process, the present study supports the use of UGPA, UGPA-PC, GRE-V, GRE-Q, and noncognitive applicant variables in the selection of those PT applicants who are most likely to successfully pass the NPTE on their first-attempt. Future research should assess the entire PT admissions processes capacity (both cognitive and noncognitive applicant variables) to predict NPTE performance.

The Relationship among PT Student Variables and NPTE Performance

Regarding the relative effect sizes for each PT student variable, the results of the random effects meta-analyses found four variables had a large and statistically significant mean true effect size on NPTE performance: (a) first-year PTGPA ($M_p = 0.55$), (b) PTGPA ($M_p = 0.50$), (c) NDRT scores ($M_p = 0.54$), and (d) comprehensive exam scores ($M_p = 0.59$; Table 4.3).

Three PT student variables had a small mean true effect size on NPTE performance: (a) clinical performance scores (first clinical experience; $M_p = 0.04$), (b) clinical performance scores (final clinical experience; $M_p = 0.07$), and (c) noncognitive student variables ($M_p = 0.25$; Table 4.3).

The clinical performance scores (both first and final clinical experiences) had the smallest empirical relationship with NPTE performance. Of all of the PT applicant and PT student variables studied, clinical performance scores (both first and final clinical experiences) were the only variables to have a non-significant relationship with NPTE performance. These findings raise the question of the usefulness and validity of the current assessment of PT students' clinical learning and performance.

PTGPA is designed to reflect the overall level of academic success of a student in their PT program. Both the first-year PTGPA ($M_p = 0.55$, CI = 0.46-0.63) and third-year PTGPA ($M_p = 0.50$, CI = 0.43-0.57) had large and significant mean true effect sizes with NPTE performance.

The large and significant relationship of first-year PTGPA and third-year PTGPA with NPTE performance is not surprising considering similar findings have been reported for other professional healthcare education disciplines including pharmacy and nursing (Randall & Diaz, 2013; Sayles, Shelton, & Powell's, 2003).

PTGPA has a strong and significant relationship with NPTE performance. Our results of the random effects meta-analysis for first-year PTGPA provide compelling evidence for PT programs and faculty to use PT students' first-year PTGPA to identify which students are at a greater risk of failing the NPTE on their first-attempt. Early identification of those PT students who are at risk of failing the NPTE on their first-attempt would allow PT programs and faculty the opportunity to provide those students additional learning opportunities to better prepare themselves to pass the NPTE, and on the first-attempt.

While the large and significant relationship of third-year PTGPA with NPTE performance is similar to first-year PTGPA, the results of the third-year PTGPA random effects meta-analysis lack the same practical significance as first-year PTGPA for PT students. By considering only the third-year PTGPA of students, there are now limited opportunities for PT programs and faculty to provide remedial learning opportunities to those students identified as at risk of failing the NPTE by their third-year in a PT program. Even though the results of the third-year PTGPA random effects meta-analysis lack practical significance for helping PT students, the results provide evidence of criterion validity for PT programs. Subsequently, PT programs that have a third-year PTGPA with a strong and significant relationship with NPTE performance should be confident in the curriculum and learning opportunities provided to allow those PT students to pass the NPTE on their first-attempt.

As evidence of the extent that one variable is related to another, criterion validity is important for PT programs (Kline, 2016). PT programs need to ensure that the content taught throughout the DPT curriculum is preparing PT students to sit and pass the NPTE on their first-attempt and reach entry-level competency. The strong and significant relationship with NPTE performance and the comprehensive exam scores ($M_p = 0.59$, CI = 0.50-0.68) random effects meta-analysis provides an additional measure of criterion validity for PT programs. While these are promising results, only two studies were included in the meta-analysis. More studies are needed to determine the magnitude of the empirical relationship of comprehensive exam scores and NPTE performance, particularly on the first-attempt.

For PT graduates, the NPTE, a multiple-choice examination, is the PT professions lone “gatekeeper” to attain licensure and begin practicing as a physical therapist in the U.S. As a multiple-choice examination, the NPTE has little or no capacity to assess the clinical competency of those graduates who pass the NPTE on the first-attempt. As far as can be determined, there is no formal requirement of PT programs in the U.S. to administer a “final-in-house” comprehensive clinical examination to assess students’ clinical learning and performance to determine if students have attained the entry-level competency required by CAPTE. Despite the universal acceptance of the NPTE and passing the NPTE at a set scaled score (PT students must achieve a 600 out of 800) there is no corresponding comprehensive national clinical examination to assess PT students’ clinical learning and performance as part of their licensure. PT is predominately a clinical profession with clinical education hours outside of the classroom representing a substantial portion (approximately 25%) of a PT programs curriculum (Commission on Accreditation of Physical Therapy Education, 2015). Consequently, PT students’ clinical performance scores, as rated by clinical instructors (CI), are the only required

measure of PT students' clinical learning and performance. Therefore, the evaluations of clinical learning and performance by CIs, external to the PT program, determine if a new PT graduate has reached entry-level clinical competency to practice as a physical therapist. Unfortunately, there is no external validation of the CI's clinical assessment of students' clinical learning and performance either within each PT program or by a standard national clinical examination.

The PT CPI is the most commonly used measurement tool to assess PT students' clinical learning and performance (Adams et al., 2008; Task Force for the Development of Student Clinical Performance Instruments, 2002). While the PT CPI differentiates among PT students' clinical experiences, it has a demonstrable ceiling effect at students' final clinical experiences, and that ceiling is regarded as the rating of "entry level" (Adams et al., 2008; & Kosmahl, 2005). Further, preliminary analysis has revealed potential issues with the construct validity of the PT CPI (Wolden & Hill, 2018). If the PT CPI has questionable validity, then the results of the clinical performance scores (both first and final clinical experience) random effect meta-analysis may be attenuated, resulting in correlations that are lower than what is expected (Pedhazur, 1997).

Considering the documented ceiling effect, lack of external validation of CI ratings, and questionable validity of the PT CPI, it is not surprising that clinical performance scores (both first and final clinical experience) mean true effect sizes were small and non-significant (first clinical experience $M_{\rho} = 0.04$, CI = -0.02-0.10; final clinical experience $M_{\rho} = 0.07$, CI = -0.00-0.14; Table 4.3). These results from the clinical performance scores (both first and final clinical experience) random effects meta-analyses provide new evidence on the relationship between PT students clinical experiences and the NPTE. The present findings from the clinical performance

scores (both first and final clinical experience) random effects meta-analyses are important to understand the NPTE's limited capacity to assess PT students' clinical learning and performance.

The predictive capacity of PT student variables on NPTE performance has primarily focused on cognitive variables (e.g., PTGPA, clinical performance scores, comprehensive exam scores). In the present study the relationship of an array of noncognitive student variables with NPTE performance was assessed with a random effects meta-analysis. The empirical relationship between the array of noncognitive student variables and NPTE performance was small, but significant ($M_p = 0.25$, CI = 0.10-0.40). The small and significant relationship with NPTE performance was to be expected given that the noncognitive student variables random effects meta-analysis assessed the combined relationship of reading comprehension, morphological awareness, emotional intelligence, critical thinking, level of reflection, conscientiousness, and task coping with NPTE performance. In essence, the noncognitive student variables random effects meta-analysis simultaneously assessed the relationship of seven different variables with NPTE performance. Combining seven different variables into one broader variable (i.e. noncognitive student variables) likely resulted in the small effect size, large and significant heterogeneity ($Q = 36.80$, $p < 0.01$, $I^2 = 71.82\%$) and presence of publication bias. Importantly, the identified publication bias had a positive bias on the mean true effect size, indicating that the calculated mean true effect size may actually increase if additional studies are conducted on noncognitive student variables relationship with NPTE performance. Future studies investigating noncognitive student variables predictive capacity of NPTE performance should focus on the Health Sciences Reasoning Test (HSRT) scores and Nelson-Denny Reading Test (NDRT) scores. From the noncognitive student variables investigated in the present study, the HSRT

scores ($r = 0.43$) and NDRT scores ($r = 0.51$ & 0.63) had the largest effect on NPTE performance (Aldridge et al., 2010; Huhn & Parrott, 2017; Moran, 2012).

Of the noncognitive student variables, NDRT scores had the strongest empirical relationship with NPTE performance. The results of the NDRT random effects meta-analysis ($M_p = 0.54$, CI = 0.40-0.69) provide evidence that PT students' NDRT scores should be assessed early within a PT program, or even perhaps as part of the PT admission process. If not assessed in the PT admissions process, then assessing PT students' NDRT scores early within a PT program could help those PT programs identify students who are at risk of failing the NPTE. Assessing PT applicants NDRT scores as part of the admissions process could help PT programs admit PT applicants who are most likely to pass the NPTE on their first-attempt. However, with only two studies used in the NDRT random effect meta-analysis the present results should be interpreted with caution. More studies investigating the predictive capacity of NDRT scores on NPTE performance are needed.

Conclusions

Even with the variation among U.S. PT programs in their curricular models and learning strategies, the NPTE is the one and only standard to which every PT program in the U.S. is held accountable (Federation of State Boards of Physical Therapy, 2015; Covington et al., 2016). A random effects meta-analysis is able to accurately and precisely provide evidence of the effect that each studied PT applicant and PT student variable has on students NPTE performance. Despite concerns that grade point average (GPA), in general, is not a valid measure of student success; from our results the best predictors of NPTE performance are related to PT applicant and student GPAs (Willingham, 1974). The best PT student variable predictors of NPTE performance are first-year PTGPA ($M_p = 0.55$) and third-year PTGPA ($M_p = 0.50$), and the best

PT applicant variable predictors of NPTE performance are UGPA ($M_p = 0.33$) and UGPA-PC ($M_p = 0.37$).

Even with UGPA and UGPA-PC being the best PT applicant predictors of NPTE performance, the results of the present study provide further evidence for the implementation of a comprehensive approach for admission into a professional program in PT. The PT admissions process should include an algorithm reflecting PT applicants UGPA, UGPA-PC, GRE-V, GRE-Q, and noncognitive applicant variables to identify those applicants who are most likely to pass the NPTE on their first-attempt.

For those applicants who are admitted into a PT program without a comprehensive screening approach, it is then important to identify those PT students who are at a risk of failing the NPTE on their first-attempt. From the results of our present study, first-year PTGPA should be assessed to identify those who are at a risk of failing the NPTE on their first attempt.

Passing only the NPTE signifies when a PT student has attained entry-level competency and is safe to begin practicing as a licensed physical therapist (Federation of State Boards of Physical Therapy, 2015). Of concern, is that our results of the clinical performance scores (both first and final clinical experiences) random effects meta-analyses provide evidence that the NPTE is not adequately assessing individual PT student clinical learning and performance (Commission on Accreditation of Physical Therapy Education, 2015).

In pursuit of excellence in PT education and to meet the challenges of the current health system, the same reasoning for the administration of the NPTE upon graduation should also be applied to have a national examination for clinical education. Until a standardized measurement of clinical competencies is required for licensure, PT programs will continue to emphasize NPTE performance while relying on a variety of external CI's to assess their student's clinical learning

and performance with an instrument (PT CPI) that has increasing concerns regarding validity and reliability. As a result, clinical excellence in PT education may be lost. Failing to assess PT graduates' clinical competency could result in a PT graduate who can pass a multiple-choice examination, but has untested clinical competencies, and therefore, could put public safety at risk (Federation of State Boards of Physical Therapy, 2015). The clinical performance scores (both first and final clinical experiences) random effects meta-analyses provide strong evidence indicating that PT education should review the current model of assessing PT students' clinical learning and performance.

Future Research

From the present investigation, four distinct research questions regarding PT education have emerged that I plan to investigate. First, an important finding from the random effects meta-analyses was the non-significant relationship between PT student clinical performance scores (both first and final clinical experiences) and NPTE performance. As PT is a clinical profession that emphasizes PT students' clinical learning, it is important that the clinical competency of each PT graduate is assessed with accuracy and precision prior to gaining licensure and beginning to practice as a physical therapist. I plan to investigate the validity of the PT CPI in measuring students' clinical learning and performance for clinical experiences that occur throughout the PT curriculum.

Second, the PT CPI is currently used to measure students' clinical learning and performance for each clinical experience. However, it is not known to what extent the PT CPI is able to accurately and precisely measure students' sequential progress in their clinical learning and performance between each clinical experience (clinical experiences that occur throughout the curriculum). I plan to investigate if the PT CPI is able to accurately and precisely measure

students' development in their clinical learning and performance between each clinical experience.

Third, the NDRT random effects meta-analysis identified NDRT scores as having a large and significant effect on NPTE performance. However, the NDRT scores were assessed once students had started their professional coursework and not as part of the admissions process. Given the large and significant relationship between NDRT scores and NPTE performance, I plan to investigate the relationship between NDRT scores, as a PT applicant variable, with NPTE performance with the view of NDRT scores being a PT applicant variable.

Fourth, the PT applicant and PT student random effects meta-analyses provided evidence of the relationship that each variable has with NPTE performance. From the present study, an evidence-based PT admissions algorithm should be developed and tested to identify which PT applicants are most likely to pass the NPTE on their first-attempt. With the available evidence, I plan to investigate the effect sizes of an evidence-based PT admissions process on NPTE performance and develop an algorithm for PT admissions.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

- Adams, C. L., Glavin, K., Hutchins, K., Lee, T., & Zimmermann, C. (2008). An evaluation of the internal reliability, construct validity, and predictive validity of the physical therapist clinical performance instrument (PT CPI). *Journal of Physical Therapy Education*, 22(2), 42-50.
- Aldridge Jr, R. L., Keith, B. Sloas, S., & Mott-Murphree, A. (2010). Relationship of the Nelson Denny Reading Test to scores on the national physical therapy licensure examination. *Journal of Allied Health*, 39(2), 72-75.
- Allen, R. E., & Diaz, C. (2013). Use of preadmission criteria and performance in the Doctor of Pharmacy program to predict success on the North American Pharmacists Licensure Examination. *American Journal of Pharmaceutical Education*, 77(9), 1-6.
- American Association of Colleges of Pharmacy (2016).
<http://www.aacp.org/resources/student/pharmacyforyou/admissions/Pages/default.aspx>. Accessed May 22, 2017.
- American Physical Therapy Association Physical Therapist (PT) Education Overview. (2015). American Physical Therapy Association, <http://www.apta.org/PTEducation/Overview/>.
- Avi-Itzhak, T. (2015). Ability of the national board for certification in occupational therapy practice test to estimate the probability of first-time pass status on the national certification exam. *The American Journal of Occupational Therapy*. 69(2), 1-8.
- *Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*. 20(2), 182-196.
- Beyond Vision 2020. (2013). American Physical Therapy Association.
<http://www.apta.org/BeyondVision2020/>. Accessed April 16, 2015.
- *Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009) Introduction to meta-analysis. John Wiley & Sons, Ltd. West Sussex, United Kingdom.
- *Brannick, M. T., Yang, L., & Cafri, G. (2008). Comparison of weights in meta-analysis under realistic conditions. Poster presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology. San Francisco, CA.

- Briscoe, V. J., & Anema, M. G. (1999). The relationship of academic variables as predictors of success on the national council licensure examination for registered nurses (NCLEX-RN) in a selected associate degree program. *The Association of Black Nursing Faculty Journal*, 80-83.
- Burns, S. M. (2011) Predicting academic progression for student registered nurse anesthetists. *American Association of Nurse Anesthetists*, 79(3), 193-201.
- Burton, N., & Ramist, L. (2001). Predicting success in college SAT studies of classes graduating since 1980. College Board Research Report No. 2001-2002. New York, New York. The College Board.
- *Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- College Readiness (2017). <https://collegereadiness.collegeboard.org/sat/scores/understanding-scores/interpreting>. Accessed May 15, 2017.
- Collins, S. (2013). Emotional intelligence as a noncognitive factor in student registered nurse anesthetists. *American Academy of Nurse Anesthetists*, 81(6), 465-472.
- Commission on Accreditation of Physical Therapy Programs: Aggregate program data: 2016-2017 Fact Sheet (2016). *Physical Therapist Education Programs*. Commission on Accreditation in Physical Therapy Education.
- Commission on Accreditation of Physical Therapy Education. Standards and required elements for accreditation of physical therapist education programs. (2015). Commission on Accreditation of Physical Therapy Education. <http://www.capteonline.org/uploadedFiles/CAPTEorg/About1CAPTE/Resources/Accreditation1Handbook/CAPTE1PTStandardsEvidence.pdf>. Accessed April 1, 2016.
- Cook, C. E., Landry, M. D., Covington, J. K., McCallum, C., & Engelhard, C. (2015a). Scholarly research productivity is not related to higher three-year licensure pass rates for physical therapy academic programs. *BMC Medical Education*, 15(148), 1-7.
- Cook, C., Engelhard, C., Landry, M., & McCallum, C. (2015b). Modifiable variables in physical therapy education programs associated with first-time and three-year National Physical Therapy Examination pass rates in the United States. *Journal of Educational Evaluation for Health Professions*, 12(44), 1-8.
- Cook, J. L., (2010). Can student reflection predict academic success and clinical performance in a physical therapist education program? (Unpublished doctoral dissertation). University of Nebraska, Lincoln, Nebraska.

- *Cooper, H. (2010). *Research synthesis and meta-analysis a step-by-step approach* (4th ed.). Publications. Thousand Oaks, CA.
- Covington, K., McCallum, C. Engelhard, C., Landry, M. D., & Cook, C. (2016). Do differences in programmatic resource investments result in different 3-year pass rates on the U.S. National Physical Therapy Examination? *Journal of Allied Health*, 45(1), 27-32.
- Cox, W., & McLaughlin, J. (2014). Association of health science reasoning test scores with academic and experiential performance. *American Journal of Pharmaceutical Education*, 78(4), 1-5.
- *De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*. 1-76.
- *DeCastro, J. (2012). *Meta-analysis of predictors of dental school performance* (Unpublished doctoral dissertation). Rutgers University, New Brunswick, New Jersey.
- *DeVellis, R. F. (2017). *Scale development theory and applications* (4th ed.). Sage Publications. Thousand Oaks, CA.
- Dillon, L. S., & Tomaka, J. (2010). NPTE predictors in a Hispanic-serving institution's physical therapist education program. *Journal of Physical Therapy Education*, 24(2), 14-18.
- Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Computational Biology*, 26(8), 897-899.
- Dockter, M. (2001). An analysis of physical therapy preadmission factors on academic success and success on the national licensing examination. *Journal of Physical Therapy Education*, 15(1), 60-64.
- Dong, T., Artino Jr, A. R., Durning, S. J., & Denton, G. D. (2012). Relationship between clinical experiences and internal medicine clerkship performance. *Medical Education*, 45, 689-697.
- Dreeben, O., (2003). *An analysis of selected physical therapy clinical performances and success on the national licensing examination* (Unpublished doctoral dissertation). Nova Southeastern University, Fort Lauderdale, Florida.
- *Duval, S., & Tweedie, R. (2000). Trim-and-fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 56, 455-463.
- Edmondson, D. (2001). *The relationship of physical therapy student characteristics and clinical site characteristics to the students' performance on the national physical therapy*

- examination (Unpublished doctoral dissertation). Tennessee State University, Nashville, Tennessee.
- Education Testing Service. GRE. Verbal reasoning concordance table. <https://www.ets.org/s/gre/pdf/concordance1information.pdf>. Accessed May 25, 2017.
- Education Testing Service. The GRE tests. <https://www.ets.org>. Accessed May 25, 2017.
- Facione, N., & Facione, P. (2007). The Health Sciences Reasoning Test test manual. Insight Assessment. Millbrae, CA.
- Federation of State Boards of Physical Therapy. Mission and Vision. (2015). Federation of State Boards of Physical Therapy; <http://www.fsbpt.org/AboutUs.aspx>.
- Federation of State Boards of Physical Therapy. Promoting safety and competence. (2015). Federation of State Boards of Physical Therapy; <http://www.fsbpt.org/FreeResources/NPTECandidateHandbook/GeneralInformation.asp>.
- Fell, N., Mabey, R., Mohr, T., & Ingram, D. (2015). The preprofessional degree: Is it a predictor of success in physical therapy education programs? *Journal of Physical Therapy Education*, 29(3), 13-21.
- *Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6(2), 161-180.
- *Field, A. P. (2003b). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77-96.
- *Field, A. P. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*. doi: 10.1348/000711010X502733.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Measurement*, 7(3), 286-299.
- Fuller, S., Horlen, C., Cisneros, R., & Merz, T. (2007). Pharmacy students' reading ability and the readability of required reading materials. *American Journal of Pharmaceutical Education*. 71(6), 1-6.
- Galleher, C., Rundquist, P., Barker, D., & Chang, W. P. (2012). Determining cognitive and noncognitive predictors of success on the National Physical Therapy Examination. *The Internet Journal of Allied Health Sciences and Practice*, 10(4), 1-10.
- Ghormrawl, H. M. K., Mandl, L. A., Rutledge, J., Alexiades, M. M., & Mazumdar, M. (2011). Is there a role for expectation maximization imputation in addressing missing data in research using WOMAC questionnaire? Comparison to the standard mean approach and a tutorial. *BMC Musculoskeletal Disorders*, 12(109), 1-7.

- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates, Publishers. Hillside, New Jersey.
- Guffey, S. (2000). *The relative value of cognitive and noncognitive variables as predictors of scores on the Physical Therapy Licensing Examination* (Unpublished doctoral dissertation). Arkansas State University, Jonesboro, Arkansas.
- Guffey, S., Farris, J., Aldridge, R., & Thomas, T. (2002). An evaluation of the usefulness of noncognitive variables as predictors of scores on the national physical therapy licensing examination. *Journal of Allied Health*, 31(2), 78-86.
- Gwyer, J., Odom, C., & Gandy, J. (2003). History of clinical education in physical therapy in the United States. *Journal of Physical Therapy Education*, 17(3), 34.
- Haight, P., & Walls, R. T. (2002). Adult learners: new norms on the Nelson-Denny Reading Test for healthcare professionals. *Reading Psychology*, 23, 217-238.
- Haight, P., & Walls, R. T. (2004). Relationships of reading, MCAT, and USMLE Step 1 test results for medical students. *Reading Psychology*, 25, 83-92.
- *Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press: Orlando, Florida.
- Heights, MA: Allyn & Bacon. Thiemann, T., Weddle, M., & Moore, M. (2003). Predicting academic, clinical, and licensure examination performance in a professional (entry-level) master's degree program in physical therapy. *Journal of Physical Therapy Education*, 17(2), 32-37.
- Heldenbrand, S. D., Flowers, S. K., Bordelon, B. J., et al. (2016). Multiple mini-interview performance predicts academic difficulty in the PharmD curriculum. *American Journal of Pharmaceutical Education*, 80(2), 1-8.
- *Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 21, 1539-1558.
- *Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*. 327, 557-560.
- *Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*. 37(1), 1158-1160.
- *Higgins, R. et al. (2010). Admission variables as predictors of PANCE scores in physician assistant programs: A comparison study across universities. *The Journal of Physician Assistant Education*, 21(1), 10-17.

- Hollman, J., Rindfleisch, A., Youdas, J., Krause, D., Hellyer, N., & Kinlaw, D. (2008). Retrospective analysis of the behavioral interview and other preadmission variables to predict licensure examination outcomes in physical therapy. *Journal of Allied Health, 37*(2), 97-104.
- Huhn, K., & Parrott, J. S. (2017). Exploration of relationships among the health sciences reasoning test, the national physical therapy licensing examination, and cognitive admission variables. *Journal of Physical Therapy Education, 31*(1), 7-13.
- *Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275-292.
- Ingrassia, J. M. (2016). Successful admission criteria to predict academic and clinical success in entry-level radiography programs. *Radiologic Technology, 87*(5), 502-510.
- International Committee of Medical Journal Editors. Overlapping Publications. <http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/overlapping-publications.html>. Accessed June 8, 2017.
- Jensen, G., Nordstrom, T., Mostrom, E., Hack, L., & Gwyer, J. (2017). National study of excellence and innovation in physical therapist education: Part 1-design, method, and results. *Physical Therapy, 97*(9), 857-874.
- *Johnson, B. T. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology, 80*(1), 94-106.
- *Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine, 93*(3), 118-121.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. 4th edition. New York, NY. The Guilford Press.
- Kosmahl, E. (2005). Factors related to physical therapist license examination scores. *Journal of Physical Therapy Education, 19*(2), 52-56.
- *Kuncel, N. R., Crede, M., Thomas, L. L., Klieger, D. M., Seiler, S. N., & Woo, S. E. (2005). A meta-analysis of the validity of the pharmacy of college admission test (PCAT) and grade predictors of pharmacy student performance. *American Journal of Pharmaceutical Education, 69*(3), 339-347.
- *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.

- *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.
- Langford, R., & Young, A. (2013). Predicting NCLEX-RN success with the HESI exit exam: Eighth validity study. *Journal of Professional Nursing, 29*(25), 55-59.
- Law, M., Steward, D., & Pollock N., Bosch, J., & Westmorland, M. (1998). Guidelines for critical review of qualitative studies. Available at: https://s3.amazonaws.com/academia.edu.documents/44591921/Guidelines_of_Critical_Appraisal_of_Qualitative_Research.pdf. Accessed April 22, 2017.
- *Lekkas, P., Larson, T., & Kumar, S., et al. (2007). No model of clinical education for physiotherapy students is superior to another: A systematic review. *Australian Journal of Physiotherapy, 53*(19), 19-28.
- Lewis, E. (2011). Longitudinal assessment of emotional intelligence in Doctor of Physical Therapy students. *The Internet Journal of Allied Health Sciences and Practice, 9*(2), 1-8.
- Ling, G., Bochenek, J., & Burkander, K. (2015). Using the Major Field Test for a bachelor's degree in business as a learning outcomes assessment: Evidence from a review of 20 years of institution-based research. *Journal of Education for Business, 90*, 435-442. <http://dx.doi.org/10.1080/08832323.2015.1095703>.
- *Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts* (3rd ed.). Taylor Francis Group. New York, NY.
- Luedtke-Hoffmann, K., Dillon, L., Utsey, C., & Tomaka, J. (2012). Is there a relationship between performance during physical therapist clinical education and scores on the National Physical Therapy Examination (NPTE)? *Journal of Physical Therapy Education, 26*(2), 41-49.
- Mai, J. A., Thiele, A., O'Dell, B., Kruse, B., Vaassen, M., & Priest, A. (2013). Utilization of an integrated clinical experience in a physical therapist education program. *Journal of Physical Therapy Education, 27*(2), 25-32.
- Mar, E., Barnett, M. J., Tang, T., Sasaki-Hill, D., Kuperberg, J. R., & Knapp, K. (2010). Impact of previous pharmacy work experience on pharmacy school academic performance. *American Journal of Pharmaceutical Education, 74*(3), 1-7.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)*. North Tonawanda, NY.

- McCall, K. L., MacLaughlin, E. J., Fike, D. S., & Ruiz, B. (2007). Preadmission predictors of PharmD graduates' performance on the NAPLEX. *American Journal of Pharmaceutical Education*, 71(1), 1-7.
- *McCallum, C. A., Mosher, P. D., Jacobsen, P. J., Gallivan, S. P., & Giuffre, S. M. (2013). Quality in physical therapist clinical education: A systematic review. *Physical Therapy*, 93(10), 1298-1311.
- *McCallum, C. A., Reed, R. R., Bachman, S., & Murray, L. (2016). A systematic review of physical therapist clinical instructor demographics and key characteristics: Impact on student clinical education experiences. *Journal of Physical Therapy Education*, 30(3), 11-20.
- *McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, 59, 927-953.
- Meiners, K. (2015). Factors predicting passage of the National Physical Therapy Examination in a private Midwestern university (Unpublished doctoral dissertation). Saint Louis University, Saint Louis, Missouri.
- *Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Public Library of Science Medicine*, 6(7), doi: 10.1371/journal.pmed.1000097.
- Mohr, T., Ingram, D., Hayes, S., & Du, Z. (2005). Educational program characteristics and pass rates on the National Physical Therapy Examination. *Journal of Physical Therapy Education*, 19(1), 60-66.
- Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *Public Library of Science One*, 1-17. DOI:10.1371/journal.pone.0166742.
- Moran, K. (2012). Assessing morphological awareness as a predictor of academic performance and performance on the National Physical Therapy Examination (Unpublished doctoral dissertation). Wilkes University, Wilkes-Barre, Pennsylvania.
- *Oxford centre for evidence-based medicine: Levels of evidence. (2009). Centre for Evidence-Based Medicine. <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>. Accessed May 18, 2017.
- PayScale.com. Entry-level physical therapists (PT) salary. [https://www.payscale.com/research/US/Job=Physical_Therapist_\(PT\)/Salary/c5d90855/Entry-Level](https://www.payscale.com/research/US/Job=Physical_Therapist_(PT)/Salary/c5d90855/Entry-Level). Accessed October 5, 2017

- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. 3rd edition. Stamford, CN. Thomson Learning.
- Peskun, C., Detsky, A., & Shandling, M. (2007). Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical Education*, 41, 57-64.
- Physical Therapists Centralized Application Services. (2017). Retrieved from <http://www.ptcas.org/About/>.
- Plack, M. (2002). The evolution of the doctorate of physical therapy: Moving beyond the controversy. *Journal of Physical Therapy Education*, <http://www2.paeaonline.org/index.php?ht=a/GetDocumentAction/i/69169SimilarYou>. 1-20, Accessed April 18, 2017.
- Professionalism and core values. (2012). Physical Therapist Centralized Application Service. <http://www.ptcas.org/Professionalism/>. Accessed April 16, 2017.
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Register, S. J., & Hoppe, E. (2007). Correlation between quantity of clinical experiences and NBEQ clinical skills performance. *Optometric Education*, 32(3), 75-78.
- Riddle, D. L., Utzman, R. R., Jewell, D. V., Pearson, S., & Kong, X. (2009). Academic difficulty and program-level variables predict performance on the National Physical Therapy Examination for Licensure: A population-based cohort study. *Physical Therapy*, 89(11), 1182-1191.
- Roach, K. E., Frost, J. S., Francis, N. J., Giles, S., Nordrum, J. T., & Delitto, A. (2012). Validation of the revised physical therapist clinical performance instrument (PT CPI): Version 2006. *Physical Therapy*, 92(3), 416-426.
- Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record*, 114(7), 1-23.
- *Rosenthal, R. (1991). *Meta-analytic procedures for social research*. New York, NY: Russell Sage Foundations.
- *Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*. 1(18), 183-192.
- Sayles, S., Shelton, D., & Powell, H. (2003). Predictors of success in nursing education. *Association of Black Nursing Faculty*, 116-120.
- *Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis (3rd ed.)*. Los Angeles, CA: Sage.

- *Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594-612.
- *Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical & Statistical Psychology*, 62, 97-128.
- *Shadish, W. R., & Haddock, C. K. (1994). In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. (p. 261-282). New York, NY: Russell Sage Foundation.
- Solberg, B. L. (2015). Critical thinking as a predictor of certification exam performance in medical laboratory science. *Clinical Laboratory Science*, 28(2) 76-82.
- Stern, D., & Rone-Adams, S. (2006). An alternative model for first level clinical education experiences in physical therapy. *Internet Journal of Allied Health Science Practice*, 4(3), 1-23.
- Steward, C. M., Bates, R. E., & Smith, G. E. (2004). Does performance on school-administered mock boards predict performance on a dental licensure exam? *Journal of Dental Education*. 68(4), 426-432.
- Suhada, R., Hicks, F., & Fogg, L. (2008). A decision algorithm for admitting students to advanced practice programs in nursing. *Journal of Professional Nursing*, 24(5), 281-284.
- *Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Thieman, T. J., (2003). Predicting academic, clinical, and licensure examination performance in a professional (entry-level) master's degree program in physical therapy. *Journal of Physical Therapy Education*, 17(2), 32-37.
- United States Department of Labor. (2016). Occupational employment statistics. United States Department of Labor. Bureau of Labor Statistics. <https://www.bls.gov/oes/current/oes291123.htm>. Accessed October 3, 2017.
- Utzman, R., Riddle, D., & Jewell, D. (2007). Use of demographic and quantitative admissions data to predict performance on the National Physical Therapy Examination. *Physical Therapy*, 87(9), 1181-1193.
- Vendrely, A. M., (2007). An investigation of the relationships among academic performance, clinical performance, critical thinking, and success on the physical therapy licensure examination. *Journal of Allied Health*, 36(2), e108-e123.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. URL: <http://www.jstatsoft.org/v36/i03/>.

Willingham, W. (1974). Predicting success in graduate education. *American Association for the Advancement of Science*, 183, 273-278.

Wolden, M., & Hill, B. (2018). The validity of the Physical Therapist Clinical Performance Instrument via confirmatory factor analysis. Poster presented at: American Physical Therapy Association Combined Sections Meeting. February 21-24, 2018; New Orleans, LA.

*Zhang, F., & Gou, J. (2016). A p-value model for theoretical power analysis and its applications in multiple testing procedures. *BMC Medical Research Methodology*, 16(135), 1-7. DOI 10.1186/s12874-016-0233-0.

APPENDIX. SEARCH TERMS USED TO GATHER THE INITIAL POOL OF CANDIDATE
STUDIES

Search terms	EBSCO	PubMed
Physical therapy student success	90*	69
Physical therapy student success AND predict	12	8
Physical therapy student success AND predictor	14	7
Physical therapy academic success	38	106
Physical therapy attrition	22	347
Physical therapy student attrition	4	6
“Physical therapy” national examination	370	603
“Physical therapy” predict examination	108	327
Physical therapy predict national examination	23	65
Physical therapy relationship national examination	1	196
Physical therapy preadmission	29	39
Physical therapy preadmission AND predictor	18	6
“Physical therapist” factors examination	19	98
National Physical therapy examination pass rate	26	4
Predict performance Physical therapy	25	378
Student performance “physical therapy” education	17	181
Predict performance Physical therapy examination	22	148
Preadmission requirements physical therapy	5	1
Physical therapy education performance AND predictor	13	57

Search terms	EBSCO	PubMed
Physical therapy examination performance AND predict	141	148
Physical therapy academic difficulty	10	37
Physical therapy academic difficulty AND predict	5	4
Predict Physical therapy academic difficulty	7	4
“Physical therapy” clinical success	25	381
Physical therapy clinical success AND predict	4	76
Physical therapy clinical academic success	1	43
“Physical therapy” clinical national examination	9	237
Student performance physical therapy examination	13	157

* Down to 44 when page options switched to 50 per page.

The various search terms and phrases used to assemble the initial pool of candidate study articles are provided in the above table. All database searches were conducted on March 15, 2017. EBSCO (all subscribed databases) and PubMed were the database services used for these searches, which returned a total of 395 and 2,556 unique results, respectively.