

**THE USE OF STANDARDIZED END OF ROTATION EXAMINATIONS IN GRADUATE PHYSICIAN ASSISTANT PROGRAMS**

by

**LINDSAY GIETZEN**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2021

MAJOR: EDUCATION EVALUATION AND RESEARCH

(Quantitative)

Approved By:

---

Advisor

Date

---

---

---

---

**© COPYRIGHT BY**

**LINDSAY GIETZEN**

**2021**

**All Rights Reserved**

## DEDICATION

*For Alex and Jocelyn*

## ACKNOWLEDGMENTS

*Thank you to my major advisor, Dr. Shlomo Sawilowsky for your guidance throughout this entire process. Your breadth of knowledge in the field of Education Evaluation and Research has made this a challenging, but extremely applicable course of study. Thank you for your availability and assistance with this process. May you rest in peace.*

*Thank you also to the other members of my doctoral committee - Dr. Barry Markman, Dr. Monte Piliawsky, and Dr. Philip Pokorski. Your commitment to education is inspiring. A special thanks to Dr. Do-Hong Kim for stepping in to Chair this Committee during such a difficult time.*

*I would also like to acknowledge Theresa Hegmann, Christopher Roman, and Michel Statler for their data sets. This would not have been possible without your contributions. Thank you!*

## TABLE OF CONTENTS

Dedication .....	ii
Acknowledgments.....	iii
List of Tables .....	vi
List of Figures .....	vii
Chapter 1: Introduction .....	1
Background.....	1
Research Questions.....	6
Significance of Study .....	6
Definition of Terms.....	7
Chapter 2: Literature Review .....	9
Chapter 3: Methodology.....	18
Design .....	18
Sample .....	19
Sampling methods.....	19
Potential Benefits to Subjects .....	20
Potential Risks to Subjects .....	20
Data Gathering .....	20
Instrument.....	21

Data Analysis .....	22
Design of the Current Study .....	23
Chapter 4: Results .....	26
Descriptive statistics are homogeneous by version, .....	26
Cronbach’s alpha are consistent by instrument version .....	30
The regression analyses are replicable to determine if the National End-of-Rotation Examinations are predictive of the National PANCE exam based on the version of the instrument. ....	31
Chapter 5: Conclusions and Recommendations .....	38
Conclusions.....	38
Discussion .....	40
Suggestions for Future Research.....	42
References .....	44
Abstract.....	47
Autobiographical Statement.....	48

## LIST OF TABLES

Table 1 .....	27
Table 2 .....	29
Table 3 .....	31
Table 4 .....	32
Table 5 .....	33

## LIST OF FIGURES

Figure 1 .....	21
Figure 2 .....	34
Figure 3 .....	35
Figure 4 .....	36
Figure 5 .....	36
Figure 6 .....	37

## CHAPTER 1: INTRODUCTION

### Background

Physician assistants are health care professionals that practice medicine with the direction and responsible supervision of a physician (ARC-PA, 2015). As of May 2015, there were almost two hundred physician assistant educational programs in the United States that were accredited by the Accreditation Review Commission on Education for the Physician Assistant, ARC-PA (ARC-PA, 2015). According to Oakes et al. (1999), there is limited evidence regarding predictors of success in physician assistant programs and for passing the National Certifying Exam for Physician Assistants, PANCE. The PANCE "has become a primary de facto measure of PA program effectiveness" (Cawley, 2002). It is important to understand the factors that lead to success for students in these highly competitive programs to ensure optimal student outcomes.

The Physician Assistant profession was created in the 1960's to address a national physician shortage, especially in primary care. Dr. Eugene Stead created the first class of Physician Assistant students at Duke University. The program was created to emulate, yet condense, medical school curriculum, but was delivered to Navy Corpsman who had a significant amount of clinical experience prior to beginning the program. While the profession gained national recognition, there was a lack in standardization of curriculum and assessment in these programs, which is still an issue today.

With the increase in popularity of the Physician Assistant profession, Graduate Physician Assistant Education programs are increasing in size and prevalence in the United States. Admissions criteria for these programs are highly competitive and selective. The success of students in graduate Physician Assistant Educational

Programs impacts society. There is a national healthcare shortage, and Physician Assistants increase patient access to quality, affordable healthcare. However, there are no requirements for standardization in the assessment of Physician Assistant students.

Physician Assistant programs continue to mirror medical school education and are rooted in the medical model of education, which focuses on systematic processes to identify the physical and biologic aspects of disease. The rigorous lock-step educational model in medical education for physicians is measured with standardized examinations such as the USMLE Step and Shelf examinations. These tools assess a student's preparedness to practice medicine compared to other students nationally as they progress through the medical school curriculum.

Assessment in medical education is a difficult task because there are many components to consider. Students must be able to demonstrate skills, perform in simulations, and identify medical diagnoses based upon patient presentations. Many students are evaluated by written exercises, structured direct observations with checklists, oral examinations, clinical simulation, peer assessments, patient assessments, and self-assessments (Epstein, 2007). Many students are also offered a chance to retake exams if a passing grade is not achieved on the first attempt. The use of remediated scores when assessing student performance is an area that needs further exploration.

Physician Assistant educators are faced with the task of ensuring that competent healthcare providers are graduating from their programs, which is mostly evaluated by student performance on end of rotation examinations. An important aspect of evaluating exams used for higher stakes student assessment is the instruments' "reliability," or

internal consistency. Measures of instrument reliability such as Cronbach's alpha give educators a measure of the extent that a student's score reflects random measurement error vs. their "true" score. In order to provide validity evidence of an instrument, it must first be shown to be reliable (Sawilowsky, 2006).

The physician assistant career is expanding rapidly in the United States and with many other countries looking to emulate this career, a strong academic precedence will need to be set. Physician Assistants are often held to the same accountability standards as their physician partners, so the education of these individuals needs to meet the rigor of traditional medical education. With the emergence of so many physician assistant programs, the methods of assessment need to be closely examined to determine reliability and validity.

Measures for student success were standardized in 2012 when officers of the Physician Assistant Education Association, PAEA, introduced the first set of national End of Rotation Exams (End of Rotation Exams, 2016). The exams covered focused content for each clinical requirement (internal medicine, family medicine, surgery, pediatrics, women's health, psychiatry, and emergency medicine) in the Physician Assistant programs. Physician Assistant programs are mandated by ARC-PA to test students on the broad scope of information for each clinical rotation. Now, over one hundred Physician Assistant programs are using the PAEA End of Rotation Exams to evaluate their students.

PAEA End of Rotation examinations are administered through an online platform. As mentioned above, there are seven different topic examinations that can be selected by student – internal medicine, family medicine, surgery, pediatrics, women's health,

psychiatry, and emergency medicine. In order to protect the integrity of the examinations, there are multiple versions of each examination offered each year and then the examinations are retired. The version of the examination administered to the student is selected randomly by the testing platform. There can also be different examination versions taken by a cohort of students if the assessments are retired in the middle of the PA program. For example, if Pediatrics versions 5.1 and 5.2 are administered in Fall 2015, and versions 6.1 and 6.2 are released in Winter 2016, then a select cohort of students will be evaluated with multiple versions of the Pediatrics examination.

The expanded usage of PAEA End of Rotation Exams resulted in a pilot study, (Gietzen et al., 2018). The two aims of this pilot study were to evaluate the reliability of the PAEA End of Rotation examination scores used as a combined scale and investigated how closely the individual examination scores and the combined scale correlate with PANCE performance to validate their use as tools for assessing PA students during their clinical training. The pilot study further assessed the PAEA End of Rotation examinations as predictors of certification examination success across multiple examination versions, multiple student cohorts, and multiple programs (Gietzen et al., 2018).

The methodology of the pilot study gathered deidentified PAEA End of Rotation examination and PANCE scores from 4 Physician Assistant programs throughout the United States for 2 graduation cohorts (Classes of 2014 and (Gietzen et al., 2018). A variety of PA programs were represented in this study, including private, public, rural, urban, small, and large class sizes, for a total of 323 students (Gietzen et al., 2018). "SPSS v.23 was used to calculate Pearson correlations, Cronbach's alpha, and value of Cronbach's alpha if individual examinations were removed from the analysis" (Gietzen et

al., 2018).

In addition to descriptive statistics in this pilot study, a linear regression model was generated using the mean of all 7 End of Rotation examinations to predict certification examination scores for the PA National Certifying Examination (PANCE) (Gietzen et al., 2018). Regression analysis was also performed using the 4 End of Rotation examinations with the greatest predictive value for certification examination outcomes: Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine, (Gietzen et al., 2018) as the cost of these examinations limits program use when funding is not available for all seven of the instruments.

In the pilot study, the End of Rotation examinations had strong correlations to the PANCE performance scores that were statistically significant. The pilot supported the use of End of Rotation examinations in PA curricula as valid assessment tools (Gietzen et al., 2018). While all of the examinations had strong correlations, the Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine End of Rotation examinations had the strongest correlations with PANCE scores (Gietzen et al., 2018). The pilot study supported the ongoing validity of using PAEA End of Rotation examination scores as tool for identifying students having academic difficulty who may be at risk of failing the PANCE national examination (Gietzen et al., 2018). This is the first tool that could be used for early identification and intervention in graduate PA students for the nation PANCE exam.

However, the pilot study collapsed all analyses, regardless of instrument version. It is noted that only the latest version of the instruments is relevant, as all previous versions are antiquated. Nevertheless, the point of the pilot was to address trends over time. Hence, the earlier versions were acceptable for inclusion even, as a collapsed total

score.

However, as long as there is an interest in performance based on the various versions, it is necessary to break down the analyses by version, because there was no formal attempt to equate the various versions by the PA Education association. This is a limitation of Gietzen et al. (2018). Therefore, the focus of the dissertation is to expand, in an ex post facto manner, the pilot results previously published by version number.

### **Research Questions**

Given there was no reliability information broken down on the National End-of-Rotation Examinations by version in Gietzen et al. (2018), the purpose of this study is to replicate their analyses by instrument version over the same 2015 – 2016 time period.

Therefore, the research questions are to determine if the

- (a) descriptive statistics are homogeneous by version,
- (b) the Cronbach's alpha are consistent by instrument version, and
- (c) the regression analyses are replicable to determine if the National End-of-Rotation Examinations are predictive of the National PANCE exam based on the version of the instrument.

### **Significance of Study**

The Physician Assistant Education Association End-of-Rotation examinations are being used by many PA programs to measure students' core clinical knowledge in specific areas. One aim of this study is to evaluate the internal reliability of the End-of-Rotation exams. Another aim of this study is to assess the convergent validity between End-of-Rotation exams and PA National Certifying Exam scores, in order to validate their use as tools for summative evaluation in PA education. There is much debate over whether

remediation should be conducted for students (Epstein, 2007), and what impact such remediation has on student performance on national board examinations, so this study also aims to validate the test regardless of which version has been administered.

### **Definition of Terms**

**Attrition.** A reduction in the number of students that occurs when a student leaves because they resign, do not meet academic requirements, etc., and are not replaced (Merriam-Webster, 2015).

**End of Rotation Examination.** Any of the seven (internal medicine, family medicine, women's health, surgery, pediatrics, psychiatry, and/or emergency medicine) examinations created by PAEA that assess PA student knowledge from a clinical rotation.

**Graduate Physician Assistant Education Program.** An educational program accredited by Accreditation Review Commission on Education for the Physician Assistant (ARC-PA) that is offered by a college or university with a terminal Master's degree in Physician Assistant Studies awarded upon completion. "Accreditation Review Commission on Education for the Physician Assistant (ARC-PA) is the accrediting agency that protects the interests of the public and PA profession by defining the standards for PA education and evaluating PA educational programs within the territorial United States to ensure their compliance with those standards"(ARC-PA, 2015).

**Physician Assistant (PA).** A PA is a nationally certified and state-licensed medical professional. PAs practice medicine on healthcare teams with physicians and other providers (AAPA, 2015).

**Physician Assistant Education Association (PAEA).** The PAEA is the national Physician Assistant Education Association and is the company that develops and maintains the End of Rotation exams and testing platform.

**Physician Assistant National Certification Exam (PANCE).** National board examination administered by National Commission on the Certification of Physician Assistants that is taken upon completion of a Graduate Physician Assistant Educational Program.

**Remediation.** Retesting a student who did not achieve a passing score on the first attempt of the PAEA End of Rotation Examination by using different questions that cover the same topics as the original examination.

## CHAPTER 2: LITERATURE REVIEW

The Physician Assistant Profession began in the United States in the 1960's in response to a national shortage of primary care physicians. The medical community embraced these mid-level providers and participated in the setting of accreditation standards, establishment of a national certification process and standardized examination, and development of continuing medical education (AAPA, 2015). As of December 31, 2014, approximately 102,000 certified PAs were practicing in every setting and medical and surgical specialty, improving access to care for patients across the U.S. while providing invaluable support to all-too-often overextended doctors (NCCPA, 2015). Certified physician assistants (PAs) are licensed and certified health care professionals who practice medicine in partnership with physicians. PAs make clinical decisions and provide a broad range of diagnostic, therapeutic, preventive, and health maintenance services. PAs focus on patient care and their responsibilities may include educational, research, and administrative activities (ARC-PA, 2015).

Eugene A. Stead Jr., M.D. organized the first class of PAs in 1965, which consisted of four Navy Hospital Corpsmen who had received considerable medical training during their military service, and the first PA class graduated from the Duke University PA program on October 6, 1967 (AAPA, 2015). As medicine is constantly evolving, physician assistant educators are changing the curriculum to prepare competent providers. While the physician assistant profession is federally regulated and requires completion of an accredited program and a passing score on the Physician Assistant National Certifying Exam (PANCE), each state has control over the duties of physician assistants through their Public Health Codes. In 2014, Forbes magazine ranked the Master's Degree in

Physician Assistant Studies as the number one degree for jobs, citing a 20% growth rate in the profession (AAPA, 2015). According to Employment Projections by the United States Department of Labor in 2007, the Physician Assistant (PA) profession has had a consistently high growth rate.

Application to PA school is highly competitive. In 2012, McDaniel issued the CASPA Cycle 11 Report, in which the application statistics for all applicants that applied to Physician Assistant programs using a centralized application program were reviewed. According to McDaniel (2012), there was an increase of 161.6% in unique applicants to PA programs from 2002-2009. There were 3.5 applicants per seat in the 2012 application cycle (McDaniel et al., 2013). Students typically need to complete at least two years of college coursework in basic and behavioral sciences before applying to a PA program (AAPA, 2015). Many PA programs also require prior healthcare experience with hands-on patient care by working as medical assistant, emergency medical technician/paramedic, medic/medical corpsman, laboratory assistant/phlebotomist, registered nurse, emergency room technician, surgical technician, or certified nursing assistant (AAPA, 2015). According to AAPA, most students have a bachelor's degree and about three years of healthcare experience before entering a program.

Physician Assistant students are evaluated by two national examinations - PACKRAT and PANCE. PACKRAT is an examination administered to Physician Assistant students that measures their progress compared to other students in their same educational cohort. The 225-question exam is built on an extensive blueprint and topic list developed by experienced PA educators and national exam experts (PACKRAT, 2015). Physician Assistants must pass the Physician Assistant National Certifying Exam

(PANCE), which is a computer-based, multiple-choice test that consists of questions that assess general medical and surgical knowledge (NCCPA, 2015). When NCCPA exams are scored based on the Rasch model and equates the scores, compensating for minor differences in difficulty across different versions of the exam (NCCPA, 2015).

To be successful, students must graduate from an accredited PA program and pass the PANCE. Only students who complete these steps are eligible to practice as a PA. According to Cawley (2002), the PANCE was the measurement tool for PA program effectiveness. It is very difficult to compare PA educational programs and their performance efficiency because there are no comparative studies that have been done between programs (Cawley, 2002). Cawley (2002) states that "many PA program attributes, including type of institution, class size, program duration, and program tuition do not correlate with PANCE scores" and "variables such as grade point average, level of test anxiety, study skill, and socioeconomic status are more likely related to PANCE performance rather than program differences". Therefore, the ability to pass PANCE can be a surrogate marker for PA student success (Cawley, 2002).

McDaniel et al. (2013) conducted a study to identify the most influential noncognitive factors valued in admissions processes by PA programs throughout the United States, as well as the motivators for and barriers to using these factors. The researchers conducted a literature search to identify noncognitive factors that were reported to have an effect on admissions in various health professions (McDaniel et al., 2013). A survey was developed that included the most frequently identified factors that were reported to have an effect on admissions and was electronically distributed to all program directors of the Physician Assistant Education Association (PAEA) (McDaniel et

al., 2013). The questions on the survey asked respondents to rank the factors that were most valued by their program's admissions process and to identify and rate motivators and barriers for the use of these noncognitive factors in admissions (McDaniel et al., 2013). The five most influential noncognitive factors identified in the study were faculty/staff/interviewer interactions, career motivation, knowledge of profession, maturity, and professionalism (McDaniel et al., 2013). Since this study was conducted, there has been a sharp increase in the number of PA programs in the United States. With the addition of almost one hundred new PA programs, it would be beneficial to see if the factors that influence admission have changed with the increase in the number of PA programs.

A study by Asprey et al. (2004) looked at the impact of age and gender on Physician Assistant student performance on Physician Assistant National Certification Exam (PANCE). Analysis of PANCE scores obtained from the NCCPA with descriptive statistics, t-tests, and regression and correlation were performed (Asprey et al., 2004). The researchers found a negative correlation between age and PANCE score with lower scores obtained by older students (Asprey et al., 2004). They also found that female examinees significantly outperformed males (Asprey et al., 2004). The data from this study suggested that the demographic characteristics of gender and age do have an influence on PANCE performance, as older students and males had poorer performance rates on the national certification examination. The strength of the study was that the analysis was conducted over three cohorts of students and the results remained consistent over time. The weakness of this study was that there were no consideration for confounding variables, such as GPA, GRE scores, prior experience, and

socioeconomic factors on the student success rates.

Andreef (2014) examined the association between academic performance in undergraduate science courses and student performance on the PANCE. The classes analyzed were chemistry, pathophysiology, and biochemistry. This retrospective study analyzed student performance in undergraduate science coursework with performance on the PANCE examination (Andreef, 2014). Pathophysiology and biochemistry were the two undergraduate science courses that correlated most closely with PANCE scores (Andreef, 2014). The strength of the study was that the age, gender, and admission GPA were considered as confounding variables. The weakness is that the quality of instruction of pathophysiology, biochemistry, and chemistry varies based on where the students took their undergraduate coursework, the professor that taught the courses, and the requirements for the course based on the undergraduate university requirements.

Brown et al. (2013) looked at correlations between PANCE performance, physician assistant program grade point average, and selection criteria. The researchers found no correlation between PANCE performance and undergraduate GPA, science prerequisite GPA, or health care experience. The strength of this study was that the researchers examined both pre-PA program grades and current PA program grades to look for a correlation with PANCE performance. The weakness of the study was that it only was an analysis of scores from one PA program.

Higgins et al. (2010) looked at admission variables as predictors of PANCE scores in Physician Assistant programs. According to Higgins et al., "the purpose of this study was to create a model of cognitive and noncognitive measures that could estimate the probability of achieving a given level of performance on PANCE". The study was a

retrospective review of admissions information used by six universities to determine which factor (undergraduate grade point average (uGPA), graduate GPA, prerequisite grades, GRE-verbal, GRE-quantitative, GRE combined, interview scores, years of health care experience, age, gender, and first-year scores on the Physician Assistant Clinical Knowledge Rating and Assessment Tool (PACKRAT) had the most impact on the dependent variable of the PANCE score (Higgins et al., 2010). The strength of the study was that multiple regression was used to analyze the factors in the study. The weakness of the study was that healthcare experience (quantity) was analyzed, but did not look at the quality of the healthcare experience.

Hocking and Piepenbrock (2010) studied the correlation between GRE scores and PA student success. According to Hocking and Piepenbrock (2010), in 2010, forty seven percent of accredited programs required completion of the Graduate Record Exam (GRE) as part of the admission process. The researchers evaluated GRE usage by PA programs to determine score utilization by reviewing the admissions criteria of ARC-PA accredited programs granting master's degrees (Hocking & Piepenbrock, 2010). Hocking and Piepenbrock (2010) found that of the PA programs that do require the GRE for admission into their respective programs, many did not interpret the examinations results within examination guidelines. The strength of this study was the finding that PA programs were using GRE data incorrectly in their admissions process. While this study looked at the proper usage of the GRE, it did not look specifically at the GRE as a predictive factor for success.

The physician assistant profession serves people from every ethnic and socioeconomic background. However, the demographics of the physician assistant

profession do not correlate with the populations of patients that these providers service. According to Census data collected by the National Commission on the Certification of Physician Assistants (NCCPA), in 2014, 62.2% of all female PAs were under 40 years of age while only 37.6% of all male PAs were under 40 years of age (NCCPA, 2014). The median age of certified PAs was 38 in 2013 (NCCPA, 2014). Also, in 2014, 85.6% of PAs were white, 4.1% were African American, 5.2% were Asian, 0.6% were Native Hawaiian/Pacific Islander, and 0.8% were American Indian or Alaskan Native (NCCPA, 2014). In terms of ethnicity, 6.4% of respondents in 2014 reported themselves as having Hispanic, Latino, and/or Spanish ethnicity (NCCPA, 2014). In 2014, 3.4% of certified PAs spoke two or more languages other than English, with over two-thirds speaking Spanish (NCCPA, 2014).

There have been very few studies that study success rates in minority students. Mulitalo and Straker (2007), published an article that outlined the past efforts and addressed future directions for diversity within PA education, with the goal to propagate the profession's leadership role within the nation's health care system. Attrition rates of physician assistant programs are not required to be made readily available to the public (Mulitalo & Straker, 2007), but their study found a high rate of attrition for black PA students, especially black male students. There have been very few studies to assess success in minority students.

The foundation of the PA profession, Navy Hospital Corpsmen with extensive medical training during their military service, is based on an applicant having previous healthcare experience. Most PA programs require a minimum number of health care experience before application. In 2013 the average number of hours of healthcare

experience reported by applicants was 7,960 hours (Physician Assistant Education Association, 2014). However, there is no clear definition among programs what constitutes health care experience (Higgins et al, 2010). Reports of experience include patient contact experience, community service, healthcare shadowing and research, and most programs accept a wide-range of healthcare experience (Physician Assistant Education Association, 2014).

Recent work has demonstrated that PAEA EOR exams can be useful in predicting whether a clinical PA student is likely to pass the certification exam (Massey, et al., 2015; Hegmann, et al., 2015). This study was a pilot that looked at early versions of the PAEA End of Rotation examinations with the intent of validating the content. However, this study was only conducted over one cohort of students and one version of the exams. This study was also limited by the number of participants included ( $n = 134$ ).

Student exam performance also needs to account on some level for test anxiety. A study by Orfus (2008) found that "A significant main effect was found for the effect of test anxiety on performance" (Orfus, 2008). Orfus examined the effect of test anxiety and time pressure on exam performance. One limitation of this study was the lack of completion of some of the questions given to students, which was attributed to the exam not carrying value to the student.

In an effort to examine the reliability and validity of the End-Of-Rotation examinations, Gietzen et al. (2018) sought to determine the reliability and validity of these examinations. Specifically, the study focused on whether student performance on the PAEA End of Rotation examinations can be predictive of performance on the Physician Assistant National Certifying Examination (PANCE) (Gietzen et al., 2018). This study

was a continuation of the 2015 pilot study by Hegmann et al., and furthered the study by incorporating data sets from four national PA programs.

Gietzen et al. (2018) found a strong correlation between End-of-Rotation examination scores and PANCE National Examination scores ( $r = 0.813$  for all 7 End-of-Rotation examinations). The study was limited by the inclusion of multiple test versions in the data set. There are multiple test versions for each examination to protect the test integrity. There have not been any studies that validate the test version and separate predictive PANCE scores based upon test version. There have not been any studies that examine whether the test version has a predictive value of student performance on the PANCE examination.

There have been multiple studies conducted to assess factors that influence PA student success. The PANCE is the standard marker for success that was common amongst all of the studies referenced. There are few studies that have attempted to assess the reliability of End of Rotation Examinations. There are many gaps in the literature in relation to PA assessment and preparation for the PA National Certifying Examination.

## CHAPTER 3: METHODOLOGY

### Previously Published Pilot

The purpose of this study is to expand on the Gietzen et al. (2018) pilot study. It was conducted under supervision of Prof. Shlomo Sawilowsky, major professor, as an integral part of the dissertation process. The aim of the pilot was to evaluate the reliability of the PAEA End of Rotation examination scores used as a combined scale and investigated how closely the individual examination scores and the combined scale correlate with PANCE performance to validate their use as tools for assessing PA students during their clinical training (Gietzen, et al., 2018). The study further assessed the PAEA End of Rotation examinations as predictors of certification examination success across multiple examination versions, multiple student cohorts, and multiple programs (Gietzen et al., 2018). For convenience, the details of the Pilot are repeated here.

### Design

De-identified End of Rotation examination and PANCE scores were gathered from four PA programs throughout the country in this retrospective study of two graduation cohorts (classes of 2014 and 2015) (Gietzen et al., 2018). A variety of PA programs were represented in this study, including private, public, rural, urban, small (class size of 25), and large (class size of 66) Total  $n = 323$  students (Gietzen et al., 2018). SPSS v.23 was utilized to calculate Pearson correlations, Cronbach's alpha, and value of Cronbach's alpha if individual exams are removed from the analysis (Gietzen et al., 2018). In addition to descriptive statistics, a linear regression model was generated using the mean of all 7 EOR exams to predict certification exam scores (Gietzen et al., 2018). A similar regression analysis was performed using the 4 EOR exams with the greatest predictive

value for certification exam outcomes (Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine) (Gietzen et al., 2018).

**Sample.** The target population was all physician assistant program students enrolled in accredited programs in the United States from 2012 to present who had taken both the PAEA End of Rotation examinations and the PA National Certifying Examination (PANCE). The accessible population was physician assistant program students in the classes of 2014 and 2015 at Wayne State University, Butler University, University of Iowa, and Rosalind Franklin University. Recruited subjects included all students admitted to the Wayne State University, Butler University, University of Iowa, and Rosalind Franklin University PA programs and those who matriculated to take the PANCE (Gietzen et al., 2018). Those who were admitted but were dismissed or voluntarily withdrew from a respective program were not included, as these students would not have been eligible to take the PA National Certifying Examination. The sample size was 323 students and because the sample was relatively small compared to the target population, all subjects who attended the PA programs at Wayne State University, Butler University, University of Iowa, and Rosalind Franklin University during the specified time frame were included in this study (Gietzen et al., 2018).

**Sampling methods.** In order to gain a sample that was more representative of the population, a random sample of students from all PA programs was obtained and analyzed with the same methods. Simple random sampling was utilized to allow each member of the population to have an equal and independent chance of being selected (Fraenkel et al., 2015). A table of random numbers was used to ensure the randomization of subjects chosen to participate in the study (Fraenkel et al., 2015). The retrospective

and completely de-identified nature of the data set (required by the applicable IRB policies) did not allow for comparison of the demographics of students to determine how representative the sample is of the population of PA students across the country, nor could many potential confounders be included in the regression equations. However, the data was collected from programs that varied in terms of class size and setting.

**Potential Benefits to Subjects.** There were no direct benefits to the student for participation in this study, nor did participation impact the education or employment of the student. However, the results of the study could have impacted the future ability to predict student success on the PA National Certifying Exam.

**Potential Risks to Subjects.** There were no increased health or educational risks to participants in this study, as it is a retrospective analysis. This study was reviewed by the Wayne State University and University of Iowa Institutional Review Boards and was granted IRB Exempt status.

**Data Gathering.** Data were collected at the Wayne State University PA program office during regular business hours in the program (9:00 a.m. until 5:00 p.m., Monday through Friday). The data were collected once per participant. The data were collected by the researcher. The data was provided by each institution from scores reported to them by examination from the PAEA testing platform.

Confidentiality was strictly observed in all levels of data handling. All identifying information was removed from the data and a unique identifying number was assigned at the time of enrollment. A master file which linked the student to the unique identifier was constructed and password protected with access limited to the researcher only. Following assignment of this unique identifier, all data was therefore be de-linked and de-identified



## Data Analysis

**Scales of Measurement.** Subject identification number, year of graduation, and whether an examination was remediated were all nominal scale data. The original End of Rotation exam scores, remediated End of Rotation exam scores, and PANCE scores were interval scale data.

**Hypotheses.** The National End of Rotation Scores were not reliable. There was no correlation between National End of Rotation Examinations and National Certifying Examination. Remediated examination scores did not have a statistically significant impact on the success of students in Graduate Physician Assistant Education Programs on their National Board Examinations.

**Variables.** The independent variables were gender, race, age, End of Rotation examination scores, and test version. The dependent variables were student attrition rates and scores on the National Certifying Exam for Physician Assistants.

**Statistical Tests.** SPSS v.23 was utilized to calculate Pearson correlations, Cronbach's alpha, and value of Cronbach's alpha if individual exams were removed from the analysis (Gietzen et al., 2018). In addition to descriptive statistics, a linear regression model was generated using the mean of all 7 EOR exams to predict certification exam scores (Gietzen et al., 2018). A similar regression analysis was performed using the 4 EOR exams with the greatest predictive value for certification exam outcomes (Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine) (Gietzen et al., 2018).

**Assumptions.** For the Pearson  $r$  correlation, both variables should be normally distributed (normally distributed variables have a bell-shaped curve). Other assumptions

included linearity and homoscedasticity. Linear regression assumed a linear relationship among the data, multivariate normality, little or no multicollinearity, no auto-correlation, and homoscedasticity. The alpha level was set to 0.05.

### **Design of the Current Study**

There were no changes to the overall research design elements from Gietzen et al. (2018), because the intent of the current study is to disaggregate the results of that the pilot study, conducted as part of the dissertation process under the supervision of the major professor, where multiple versions of the National End-of-Rotation Examinations were combined into a total score. In the pilot study, all of the test versions were combined from one year with another. However, it is more likely a researcher will base a prediction on more than one version from the same year's crop of instruments. Furthermore, Pearson correlations, Cronbach's alpha, and value of Cronbach's alpha of the individual exams by version will be calculated after separating each version for the above examinations.

**Hypotheses.** The research hypotheses are explicated below:

- (a) descriptive statistics are homogeneous by version. This is important because if the mean and standard deviations for each version of the National End-of-Rotation Examinations are markedly different, then the versions of the instrument cannot be considered equivalent. Substantiating this, however, will provide support for the total score analyses (i.e., combining of versions) in Gietzen et al. (2018), and will generate promise for successful use of future versions of the National End-of-Rotation Examinations in the form of a total score. An important limitation, of course, is a more sophisticated equivalency would require establishing similar item

difficulty index and item discrimination index values, which may be a topic for another study,

- (b) the magnitude of Cronbach's alpha is consistent by version. This is important, because the statistical engine of the Cronbach alpha is the Pearson product-moment correlation, which is attenuated by the number of items. The implication might be the total score reliabilities used in Gietzen et al. (2018) were inflated due to the attenuation. Therefore, reconducting a reliability analysis by version will rule out that possibility.
- (c) replicate the regression analyses to determine if the National End-of-Rotation Examinations are predictive of the National PANCE exam based on the version of the instrument. The results of Gietzen et al. (2018) indicated there was predictive validity, but it was based on concatenating three versions of the National End-of-Rotation Examinations.

**Statistical Tests.** SPSS v. 26 will be utilized to calculate Pearson correlations, Cronbach's alpha, and value of Cronbach's alpha of the individual exams by version. In addition to descriptive statistics, a linear regression model will be generated using the mean of each version of all 7 EOR exams to predict certification exam scores with the following formula:

$$y' = a + bx,$$

where  $y'$  is the certification score,  $a$  is the intercept,  $b$  is the unstandardized beta coefficient, and  $x$  is the end of rotation score. In the pilot study, the obtained predictive equation was

$$y' = -517 + 13.46x .$$

In the current study, a similar regression analysis will be performed using the four EOR by version exams with the greatest predictive value for certification exam outcomes (Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine.

**Assumptions.** Ordinary least squares regression linear regression assumes homoscedasticity.

## CHAPTER 4: RESULTS

This section will contain results from both the pilot and current study. De-identified retrospective data from four accredited PA programs (Wayne State, University of Iowa, Rosalind Franklin, and Butler University) was obtained for students who were administered End of Rotation Examinations from the Classes of 2014 and 2015  $n = 323$ . There was no demographic data obtained on the participants, as each participant was assigned a number and all examination scores were reported.

### **Descriptive statistics are homogeneous by version,**

Descriptive statistics for all seven end of rotation (EOR) examinations were performed. The total cases examined,  $n=323$ , with  $n=322$  for the analysis of Emergency Medicine. One case was excluded because there was a missing value. The results are summarized in Table 1: Descriptive Statistics of EOR Examinations (versions concatenated). The highest attainable score for an End of Rotation examination is 100. The examination with the highest mean was psychiatry at 78.76 and the lowest was emergency medicine at 73.59. The skewness of all of the end of rotation examinations between -0.5 and 0.5 indicates that the data is fairly symmetrical.

The PANCE exam is a national board examination and is scored with the Rausch model. For the sample in this study, the mean score was 493.94. It should be noted that there is no minimum published passing score on this examination, but generally scores of 350 and higher are considered “passing”. The PANCE data is moderately positively skewed (0.59), which indicates that the students in the sample had strong testing scores on this examination.

The average of all seven end of rotation examinations, as well as with just the four examinations with the strongest predictive value based on the pilot study – Emergency Medicine, Internal Medicine, Family Medicine, and Surgery were also analyzed. The data for these two sets was fairly symmetrical.

**Table 1**

***Descriptive Statistics of EOR Examinations (versions concatenated)***

	<i>n</i>	Minimum	Maximum	<i>M</i>	Std. Error	Std. Deviation	Skewness	Std. Error
Family Medicine	323	54.00	92.00	76.27	0.38	6.78	-0.19	0.14
Internal Medicine	323	54.00	96.00	75.53	0.46	8.24	-0.05	0.14
Pediatrics	323	55.00	91.00	76.39	0.34	6.16	-0.45	0.14
Psychiatry	323	58.00	93.00	78.76	0.36	6.44	-0.30	0.14
General Surgery	323	55.00	94.00	75.35	0.35	6.24	-0.05	0.14
Womens Health	323	54.00	94.00	75.07	0.34	6.17	-0.33	0.14
Emergency Medicine	322	51.00	91.00	73.59	0.41	7.31	-0.12	0.14
PANCE	323	316.00	800.00	493.94	4.31	77.39	0.59	0.14
Average EOR (7 scores)	322	63.29	90.29	75.86	0.27	4.92	-0.03	0.14
Average EOR (4 scores)	322	61.75	91.00	75.20	0.32	5.69	0.12	0.14

Next, the data was partitioned by each version of the EOR exams (1, 2, or 3). The same number of students were administered the version of the examination for the all seven EOR examinations. For example, if a student was assigned Version 1 of Family Medicine, that same student was assigned Version 1 for all of the additional examinations, also.

With three unique versions of the End of Rotation examinations, descriptive statistics were calculated. End of Rotation Version 1 ( $n=134$ ) had comparable descriptive statistics to the average of all seven EOR scores, with the exception of the psychiatry examination ( $M=75.63$ ). The students who took Version 2 of the EOR examinations had higher average scores on the PANCE ( $M=519.13$ ). However, this is still within one standard deviation of the PANCE mean when all seven EOR scores are accounts without separating the data. With all of the end of rotation examinations, regardless of version, all results were within one standard deviation of the mean when all seven EOR exams are accounted across all versions. There was insignificant variation in the descriptive statistics for by each end of rotation version.

**Table 2*****Descriptive Statistics by EOR Examination Topic and Version***

EOR Version		Family Medicine	Internal Medicine	Pediatrics	Psychiatry	General Surgery	Womens Health	Emergency Medicine	PANCE
1	N	134	134	134	134	134	134	134	134
	Mean	77.64	76.1	76.46	75.63	75.64	74.28	72.6	496.22
	Median	77	76	77	76	76	74.5	73	497.5
	Std. Error of Mean	0.58	0.73	0.54	0.48	0.52	0.49	0.63	7.10
	Minimum	54	54	58	58	60	54	51	322
	Maximum	92	93	91	89	91	87	87	766
	Std. Deviation	6.75	8.45	6.23	5.59	6.00	5.67	7.24	82.15
2	N	24	24	24	24	24	24	24	24
	Mean	71.88	75.25	78.33	84.96	76.58	76.54	76.42	519.13
	Median	71	75	79.5	86.5	77.5	77	76.5	511.5
	Std. Error of Mean	1.59	1.53	1.30	1.00	1.41	1.61	1.38	16.17
	Minimum	54	59	62	74	63	57	66	378
	Maximum	89	96	91	92	94	94	90	779
	Std. Deviation	7.77	7.50	6.37	4.92	6.89	7.87	6.78	79.21
3	N	165	165	165	165	165	165	164	165
	Mean	75.79	75.11	76.04	80.41	74.93	75.5	73.98	488.42
	Median	76	74	77	81	74	76	74	477
	Std. Error of Mean	0.49	0.64	0.47	0.47	0.49	0.49	0.57	5.66
	Minimum	60	59	55	62	55	57	53	316
	Maximum	90	91	88	93	89	91	91	800
	Std. Deviation	6.33	8.20	6.05	6.03	6.34	6.25	7.34	72.66
Total	N	323	323	323	323	323	323	322	323
	Mean	76.27	75.53	76.39	78.76	75.35	75.07	73.59	493.94
	Median	77	75	77	79	75	75	74	490
	Std. Error of Mean	0.38	0.46	0.34	0.36	0.35	0.34	0.41	4.31
	Minimum	54	54	55	58	55	54	51	316
	Maximum	92	96	91	93	94	94	91	800
	Std. Deviation	6.78	8.24	6.16	6.44	6.24	6.17	7.31	77.39

### **Cronbach's alpha are consistent by instrument version**

**Assumptions.** As stated in Chapter 4, For the Pearson  $r$  correlation, both variables should be normally distributed. Other assumptions included linearity and homoscedasticity. Linear regression assumed a linear relationship among the data, multivariate normality, little or no multicollinearity, no auto-correlation, and homoscedasticity. The alpha level was set to 0.05.

Cronbach's alpha was determined for the seven unique content examinations - Family Medicine, Internal Medicine, Pediatrics, General Surgery, Psychiatry, Women's Health, and Emergency Medicine. Cronbach's alpha when all seven examinations were included was  $\alpha = 0.85$ . The examinations were then separated by version, and Cronbach's alpha was  $\alpha = 0.84$ .

Inter-item correlation was performed to determine internal consistency reliability, see Table 3: Reliability Inter-Item Correlation Matrix. As the core exams measure different content, the correlations between the examinations were not expected to have strong correlations. The range of correlation are moderately positive correlation between internal medicine and family medicine at 0.54 and the psychiatry examination had a weakly positive inter-item correlation at 0.26 with both the family medicine and general surgery examinations.

**Table 3*****Reliability Inter-Item Correlation Matrix between EOR examinations***

	Family Medicine	Internal Medicine	Pediatrics	Psychiatry	General Surgery	Womens Health	Emergency Medicine
Family Medicine	1.00	0.54	0.48	0.26	0.48	0.42	0.42
Internal Medicine		1.00	0.52	0.34	0.55	0.48	0.55
Pediatrics			1.00	0.38	0.52	0.41	0.50
Psychiatry				1.00	0.26	0.40	0.37
General Surgery					1.00	0.45	0.52
Womens Health						1.00	0.48
Emergency Medicine							1.00

**The regression analyses are replicable to determine if the National End-of-Rotation Examinations are predictive of the National PANCE exam based on the version of the instrument.**

**Assumptions.** Ordinary least squares regression linear regression assumes homoscedasticity.

For the regression modeling, the dependent variable is PANCE score and the variables are the examination topic and version. Again, all of the students who took Version 1 of one examination were also assigned Version 1 of every examination thereafter, so the regression analyses were performed utilizing all seven end of rotation examinations but were separated by version. The regression equation in the pilot study was:

$$y' = -517 + 13.46x$$

Parametric Correlations were performed to determine the Pearson  $r$  between each End of Rotation examination and the PANCE board examination (Table 4: Parametric Correlations). The examinations with the strongest positive correlations to the PANCE exam were family medicine ( $r=0.60$ ), internal medicine ( $r=0.69$ ), general surgery ( $r=0.64$ ), and emergency medicine ( $r=0.64$ ). All of these correlations were significant at the 0.01 level.

**Table 4*****Parametric Correlation***

		Family Medicine	Internal Medicine	Pediatrics	Psychiatry	General Surgery	Womens Health	Emergency Medicine	PANCE
Family Medicine	Pearson Correlation	1.00	.54**	.48**	.25**	.48**	.42**	.42**	.60**
	Sig. (2-tailed)		0	0	0	0	0	0	0
	<i>n</i>		323	323	323	323	323	322	323
Internal Medicine	Pearson Correlation		1.00	.52**	.34**	.55**	.48**	.55**	.69**
	Sig. (2-tailed)			0	0	0	0	0	0
	<i>n</i>			323	323	323	323	322	323
Pediatrics	Pearson Correlation			1	.38**	.52**	.41**	.50**	.60**
	Sig. (2-tailed)				0	0	0	0	0
	<i>n</i>				323	323	323	322	323
Psychiatry	Pearson Correlation				1	.26**	.40**	.37**	.39**
	Sig. (2-tailed)					0	0	0	0
	<i>n</i>					323	323	322	323
General Surgery	Pearson Correlation					1	.45**	.52**	.64**
	Sig. (2-tailed)						0	0	0
	<i>n</i>						323	322	323
Womens Health	Pearson Correlation						1	.48**	.50**
	Sig. (2-tailed)							0	0
	<i>n</i>							322	323
Emergency Medicine	Pearson Correlation							1	.64**
	Sig. (2-tailed)								0
	<i>n</i>								322
PANCE	Pearson Correlation								1
	Sig. (2-tailed)								
	<i>n</i>								

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

Parametric Correlations were performed to determine the Pearson  $r$  between the four most predictive End of Rotation examinations and the PANCE board examination (Table 5: Correlations by EOR Version Utilizing 4 EOR Examinations). There was a strong

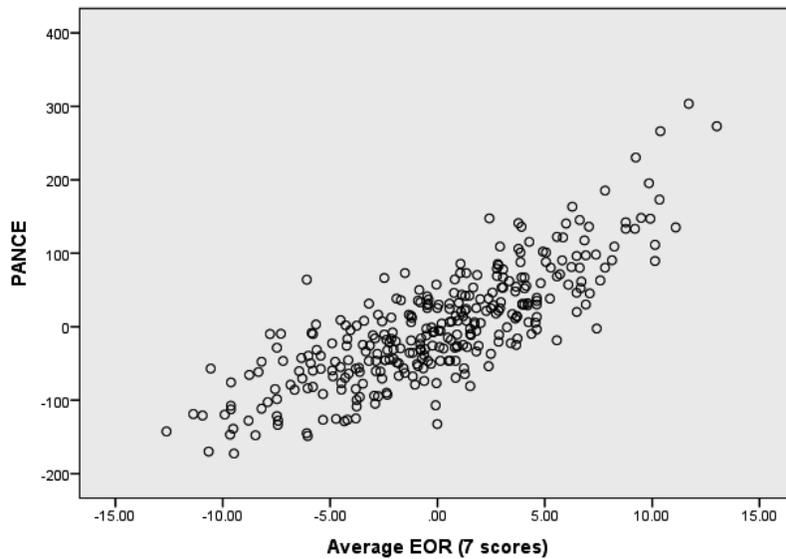
positive correlation between the Average score on the 4 EOR exams and PANCE ( $r=0.811$ ). There was a weak negative correlation between EOR version and PANCE scores ( $r=-0.052$ ). Listwise deletion was the method for accounting for the missing Emergency Medicine score,  $n = 322$ .

**Table 5**

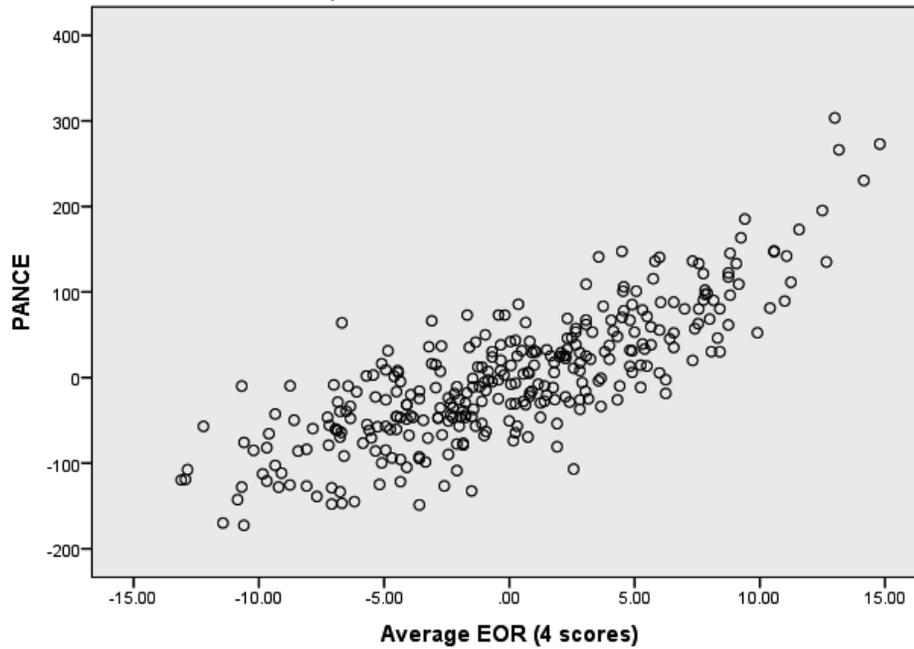
***Pearson Correlation of 4 EOR Examinations***

		PANCE	Average EOR (4 scores)	EOR Version	Program Identifier
Pearson Correlation	PANCE	1	0.81	-0.05	-0.10
	Average EOR (4 scores)		1	-0.04	-0.11
	EOR Version			1	0.35
	Program Identifier				1

The correlation of all seven End of Rotation examinations with the PANCE examination has a strong positive correlation of  $r=0.65$  as illustrated in the scatterplot in Figure 2: Correlation of PANCE with Average EOR Score (7 scores).

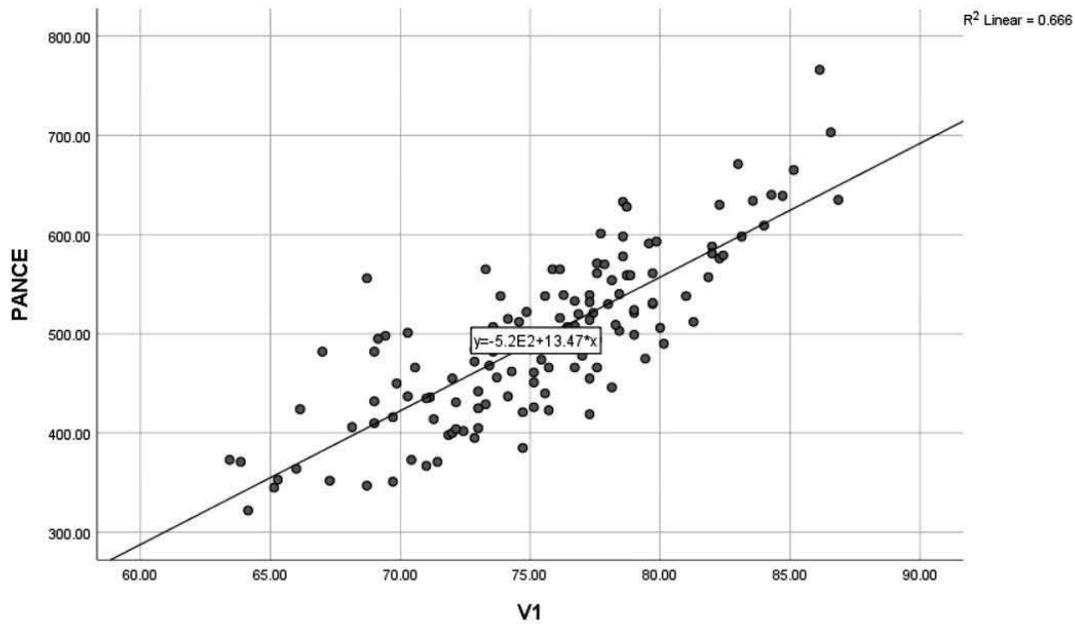
**Figure 2*****Correlation of PANCE with Average EOR Score (7 scores)***

There is an even stronger correlation ( $r=0.66$ ) of the PANCE examination with the four EOR examinations that had the strongest individual correlations with the PANCE (Emergency Medicine, Family Medicine, Internal Medicine, and General Surgery), as is illustrated in the scatterplot in Figure 3: Correlation of PANCE with Average EOR Score (4 scores).

**Figure 3*****Correlation of PANCE with Average EOR Score (4 scores)***

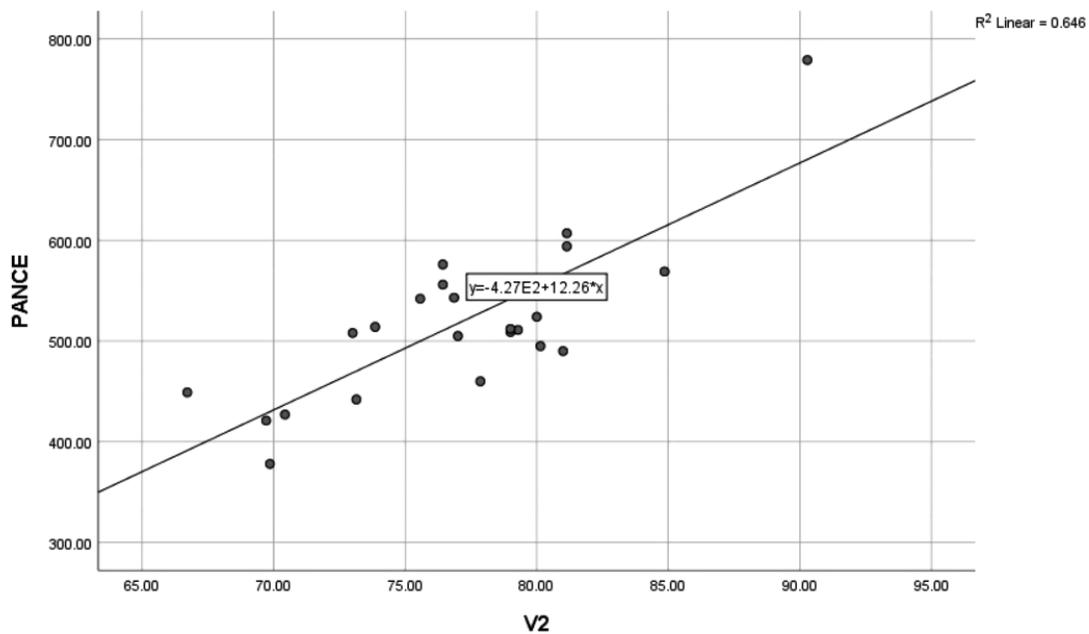
Regression analysis was performed with all versions separated. The regression equation for Version 1 EOR Exams is  $y = 13.47x + 520$

Figure 4

**Regression for Version 1 of EOR Exams**

The regression equation for Version 2 EOR exams is  $y = 12.26x + 427$

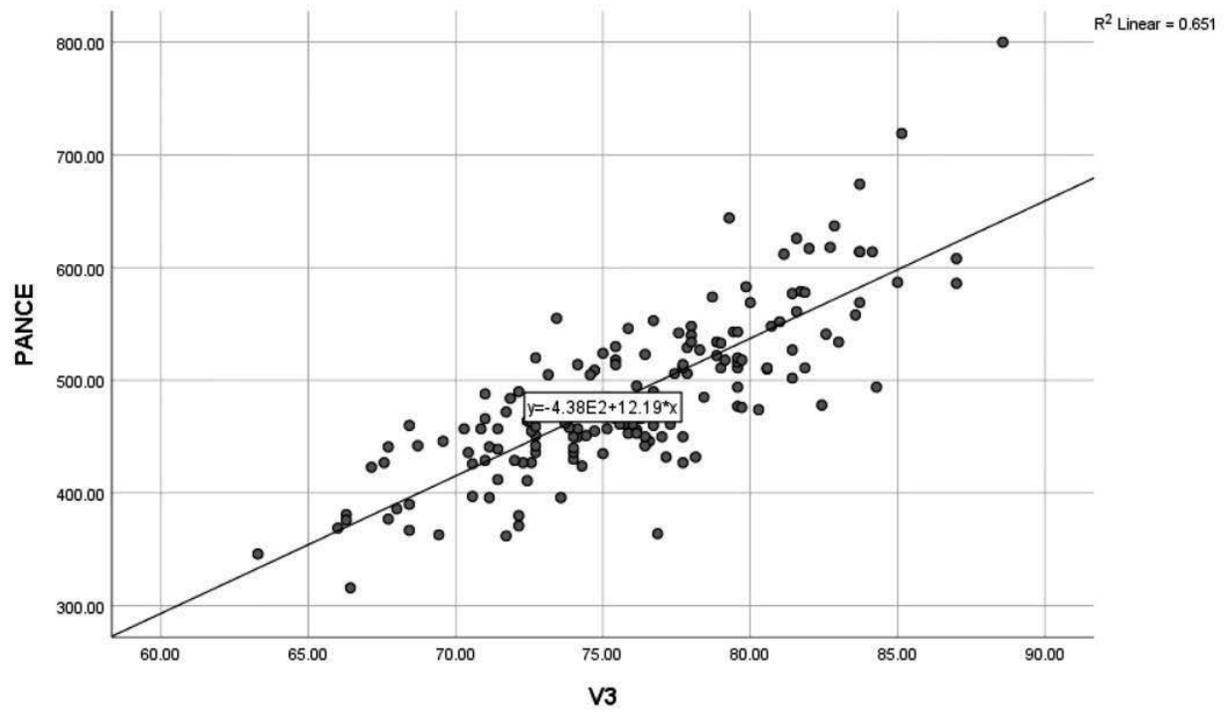
Figure 5

**Regression for Version 2 of EOR Exams**

The regression equation for Version 3 EOR exams is  $y = 12.19x + 438$

**Figure 6**

***Regression for Version 3 of EOR Exams***



## CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

### Conclusions

Restatement of Hypothesis A: descriptive statistics are homogeneous by version. This is important because if the mean and standard deviations for each version of the National End-of-Rotation Examinations are markedly different, then the versions of the instrument cannot be considered equivalent. Substantiating this, however, will provide support for the total score analyses (i. e., combining of versions) in Gietzen et al. (2018), and will generate promise for successful use of future versions of the National End-of-Rotation Examinations in the form of a total score. An important limitation, of course, is a more sophisticated equivalency would require establishing similar item difficulty index and item discrimination index values, which may be a topic for another study.

The mean and standard deviations for each version tested of the instrument were not significantly different. This further supports the use of total score analyses to predict student outcomes on the national board examination. This significantly impacts PA education because there is equivalency among the versions of the End of Rotation examinations, so test version is not a confounding variable of the student performance on the instrument.

Restatement of Hypothesis B: the magnitude of Cronbach's alpha are consistent by version. This is important, because the statistical engine of the Cronbach alpha is the Pearson product-moment correlation, which is attenuated by the number of items. The implication might be the total score reliabilities used in Gietzen et al. (2018) were inflated due to the attenuation. Therefore, reconducting a reliability analysis by version will rule out that possibility.

Cronbach's alpha was consistent by version of each End of Rotation examination. The total score reliabilities in the pilot study were validated by separating the examinations by version, therefore reducing the number of items. The concern for inflation due to attenuation in the pilot study was disproven by separating the examinations and having consistent results of Cronbach's alpha.

The Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine examinations correlate most closely with PANCE scores ( $r = 0.637, 0.599, 0.644, \text{ and } 0.689$ , respectively). The lowest correlations were Psychiatry at  $r = 0.393$  and Women's Health at  $r = 0.501$ . All correlations were statistically significant. Cronbach's alpha was calculated at 0.847 overall. Psychiatry was the only examination that increased this measure when removed, to 0.849.

Restatement of Hypothesis C: replicate the regression analyses to determine if the National End-of-Rotation Examinations are predictive of the National PANCE exam based on the version of the instrument. The results of Gietzen et al. (2018) indicated there was predictive validity, but it was based on concatenating the versions of the National End-of-Rotation Examinations.

The National End-of-Rotation Examinations are predictive of the National PANCE exam regardless of the version of the instrument. The regression analyses were replicated and the examination version did not impact the predictive validity of the End of Rotation examinations.

As noted in the pilot, "the  $R^2$  was 0.661 for the model incorporating all 7 EOR exams, and 0.657 for the 4 exams with the greatest correlation to certification exam score (Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine)."

(Gietzen et al., 2018). Therefore, the models explain 66.1% of the variance in PANCE scores when including all seven examinations and 65.7% of the variance in PANCE scores when including only the 4 examinations with the strongest correlation (Gietzen et al., 2018). The version of the end-of-rotation exam did not significantly influence the 4-exam regression model. For the 7-exam composite score model, while exam version had a statistically significant ( $p = 0.006$ ) but very small effect (increase in  $R^2$  by .008). This may be influenced by student remediation efforts and taking multiple versions of the same examination, which could not be discerned from in the data set.

## **Discussion**

As noted in the pilot study, “the good reliability of the seven EOR exam scores and the significant and generally strong correlations of the exams with an outside “gold standard” (PANCE) provide support for use of these instruments as medium to high-stakes assessment tools” (Gietzen et al., 2018). The Emergency Medicine, Family Medicine, General Surgery, and Internal Medicine EOR examinations correlate most closely with PANCE scores (Gietzen et al., 2018). The literature review shows studies performed over the years to predict student outcomes on the national board examination – from undergraduate behaviors to PA program performance, but there were no strong correlations between any predictive factors and student board examination performance.

The creation and implementation of the End of Rotation examinations was the first national high-stakes examination method utilized in PA education. As PA education is lock-step, a failing grade on one of these examinations could result in student dismissal from a program. Therefore, the reliability testing was crucial in the event that a student may challenge these examinations if dismissed from a program due to performance. The

good reliability of these examinations supports the future use of these examinations in PA programs nationally.

While the End of Rotation examinations are reliable and predictive of student success on the national board examination, not every program is able to administer all seven unique exams. For programs with financial constraints on testing or a logistical need to limit testing, the 4-exam model of internal medicine, family medicine, emergency medicine, and general surgery may be ideal. There was strong correlation between examination scores and PANCE scores when only these four examinations were utilized in the model.

As the frequency of the tests were administered, multiple versions of each examination were created to protect test integrity. The multiple versions of the examination having little variance in the correlations of the examination version with performance on the national PANCE board examination supports the use of any random assignment of examination version to the students. It also supports the use of a second version of the examination to be administered as a remediation effort if a student is unsuccessful on a first attempt of the examination.

The pilot and current analyses have several limitations. First, there was no ability to analyze the demographics of students to determine how representative the sample is of the population of PA students across the country. Including demographics could identify confounding variables in the regression analyses. The data was collected from programs that varied in terms of class size and setting, but a further demographic analysis would be a possibility for a future study.

Another limitation was the individual information reported to the PA programs regarding student performance on each examination. The data only included the examination topic, version, and final student score. There was no information available to compare student performance by examination question or to identify the topic of study on the examination question.

### **Significance of Research**

The analyses from both the pilot and current study support the validity of using national EOR exam scores as one tool to help identify students who may be at risk of failing the national certification exam. Importantly, as noted in the pilot study, “an average score from the 4 exams with the greatest correlation to certification exam score predicts 65.7% of variance on board exams, whereas the inclusion of all 7 exams predicts 66.1%” (Gietzen et al., 2018). The current study shows that student performance on the national board examinations does not vary dramatically based on exam versions, suggesting that it could reliably identify students at risk for certification exam failure regardless of examination administered.

### **Suggestions for Future Research**

A topic for future research is analyzing student demographics in this study. The literature review suggests that demographics such as age, gender, and race play a role in student performance in PA programs and on the national PANCE examination. Obtaining a larger sample of student End of Rotation examination scores with demographic information would allow for a more detailed analysis of demographic factors influencing performance on End of Rotation examinations, and the ability to predict

performance on the national PANCE examination.

Additionally, the reliability portion of this study could be enhanced if the specific performance by question was available for analysis. A more sophisticated equivalency for the End of Rotation examinations would require establishing similar item difficulty index and item discrimination index values. The available data is limited at this time due to the national platform host of the examinations, but this may be a topic for another study.

Another topic to explore is whether remediation efforts on End of Rotation examinations is predictive of student performance on the National PANCE examination. With the costs of the examinations being a limiting factor for many programs, examining the correlations between initial testing performance and remediation performance with National PANCE examination performance could lead to early intervention for students at risk of not passing the national board examination.

## **Closing**

Physician Assistant education is a new subset of medical education that has rapidly evolved to meet the medical shortage in the United States. As regulations from Accrediting agencies continue to impose stronger sanctions on student performance markers, it is imperative that there are reliable and valid examination instruments to evaluate students. The PAEA National End of Rotation examinations have been identified as reliable and valid predictors of student performance on the National PANCE certifying board examination. These examinations can be used to identify students at risk, leading to early academic intervention, which is crucial for success of the student and the future patients they will care for when practicing medicine.

## REFERENCES

- AAPA. (2015). What Is a PA? Retrieved March 01, 2016, from <https://www.aapa.org/what-is-a-pa/>
- Andreeff, R. (2014). Predictors of Student Success on the Physician Assistant National Certifying Examination. *The Journal of Physician Assistant Education, 25*(3), 36-39.
- ARC-PA. (2015). Retrieved February 08, 2016, from [http://www.arc-pa.org/acc\\_programs/](http://www.arc-pa.org/acc_programs/)
- Asprey, D., Dehn, R., & Kreiter, C. (2004). The Impact of Age and Gender on the Physician Assistant National Certifying Examination Scores and Pass Rates. *The Journal of Physician Assistant Education, 15*(1), 38-41.
- Brown, G., Imel, B., Nelson, A., Hale, L. S., & Jansen, N. (2013). Correlations Between PANCE Performance, Physician Assistant Program Grade Point Average, and Selection Criteria. *The Journal of Physician Assistant Education, 24*(1), 42-44.
- Cawley, J. (2002). Physician Assistant Education and PANCE Performance: A Passing Controversy? *The Journal of Physician Assistant Education, 79-80*.
- Downing, S. M. (2004). Reliability: on the reproducibility of assessment data. *Medical Education, 38*(9), 1006–1012. <https://doi.org/10.1111/j.1365-2929.2004.01932.x>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education*. New York: McGraw Hill.
- Gietzen L, Roman C, Hegmann T. (2018). Reliability and Validity of National End of Rotation Examinations: An Update. *The Journal of Physician Assistant Education, 29*(2), 86-88.

- Hall, F. R., & Bailey, B. A. (1992). Correlating students' undergraduate science GPAs, their MCAT scores, and the academic caliber of their undergraduate colleges with their first-year academic performances across five classes at Dartmouth Medical School. *Academic Medicine*, *67*(2), 121-3.
- Hegmann, T., Roscoe, M., & Statler, M. (2015). Reliability and Validity of PAEA End of Rotation™ Examination Scores for Predicting Performance on the Physician Assistant National Certification Examination. *Journal of Physician Assistant Education*, *26*(4), 187–192. <https://doi.org/10.1097/jpa.0000000000000044>
- Higgins, R., Moser, S., Dereczyk, A., Canales, R., Stewart, G., Schierholtz, C., . . . Arbuckle, S. (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.
- Hocking, J. A., & Piepenbrock, K. (2010). Predictive Ability of the Graduate Record Examination and Its Usage Across Physician Assistant Programs. *The Journal of Physician Assistant Education*, *21*(4), 18–22. <https://doi.org/10.1097/01367895-201021040-00002>
- Lane, D. M. (n.d.). Homogeneity of Variance. Retrieved April 18, 2016, from <http://davidmlane.com/hyperstat/A45619.html>
- Massey, S., Holmerud, D., Hammond, J., Preston, T., Tipton, L., & Rizzolo, D. (2015). Correlation of the Physician Assistant Education Association End of Rotation Examinations with the Physician Assistant National Certifying Exam. *Journal of Physician Assistant Education*, *26*(3), 144–146. <https://doi.org/10.1097/jpa.0000000000000039>

Merriam-Webster. (2015). Retrieved February 08, 2016, from <http://www.merriam-webster.com/dictionary/attrition>

McDaniel, J. (2012). *CASPA Cycle 11 Report* (Rep.).

McDaniel, J. M., Thrasher, A., & Hiatt, T. (2013). The Use of Noncognitive Factors in Physician Assistant Admissions. *The Journal of Physician Assistant Education*, 24(1), 15-23.

Mulitalo, K., & Straker, H. (2007). Diversity in Physician Assistant Education. *The Journal of Physician Assistant Education*, 46-51.

NCCPA. (2014). NCCPA. Retrieved March 01, 2016, from <http://www.nccpa.net/Public>

Oakes, D. L., MacLaren, L. M., Gorie, C. T., & Finstuen, K. (1999). Predicting Success on the Physician Assistant National Certifying Examination. *The Journal of Physician Assistant Education*, 10(2), 63-69.

Orfus, S. (2008). The Effect Test Anxiety and Time Pressure on Performance. *The Huron University College Journal of Learning and Motivation*, 46(1).

PACKRAT. (2015). PAEA PACKRAT Exam. Retrieved March 01, 2016, from <http://www.packratexam.org/>

Physician Assistant Education Association. *Twenty-Eighth Annual Report of Physician Assistant Educational Programs in the United States* (Rep.). (2014).

Simmons, S. W. (2003). Predictors of Academic Success for Underrepresented Minorities in Physician Assistant Programs. *The Journal of Physician Assistant Education*, 14(2), 110–113. <https://doi.org/10.1097/01367895-200314020-00010>

*Validity Evidence: Predicting Success in Graduate Education*. ETS. (2016).

[https://www.ets.org/gre/research/validity\\_evidence\\_predicting\\_grad\\_success](https://www.ets.org/gre/research/validity_evidence_predicting_grad_success).

**ABSTRACT****THE USE OF STANDARDIZED END OF ROTATION EXAMINATIONS IN GRADUATE  
PHYSICIAN ASSISTANT PROGRAMS**

by

**LINDSAY GIETZEN****May 2021****Advisor:** Dr. Shlomo Sawilowsky**Major:** Education Evaluation and Research**Degree:** Doctor of Philosophy

Graduate Physician Assistant Education programs are increasing in size and prevalence in the United States. PA Students are trained in the medical model, but not until only recently there have been standardized instruments to measure student progress versus national data before a student takes their certifying board examination. Many programs are now shifting focus to predicting student performance on national board examinations based upon student performance on national end of rotation examinations. These end of rotation examinations vary in topic and version to protect the test integrity. Examining the reliability and validity of the standardized end of rotation examinations, including the differences in examination version are paramount to creating reliable predictions for student success on the high stakes national board examinations.

## **AUTOBIOGRAPHICAL STATEMENT**

Lindsay is an Associate Professor and Program Director at Oakland University where she specializes in simulation and interprofessional education. As a physician's assistant, she practiced in emergency/trauma medicine and then specialized in neurological surgery, leading a team that pioneered many surgical procedures for the first time in the United States. She is also the Co-Creator and Director of Active Violence Emergency Response Training (AVERT) for the Health and Safety Institute.

A Wayne State University graduate of the Master of Science in Physician Assistant Studies, Lindsay also earned a Bachelor of Arts in Sociology and University Honors at the University. She began her academic career as a Clinical Coordinator and Assistant Professor at Wayne State in the Department of Physician Assistant Studies where she received the Dale Sillix Award for Teaching Excellence. Lindsay enjoyed working with Wayne State University PA students for almost a decade before getting the opportunity to design an interprofessional PA and DO program at Michigan State University. She is now the Program Director at Oakland University, designing their inaugural PA Science program which focuses on multidisciplinary integration of the medical sciences.

She is a Member of the Michigan Academy of Physician Assistants, American Academy of Physician Assistants, Physician Assistant Education Association, and Director at Large of the Association of Neurosurgical Physician Assistants. Lindsay is currently serving an appointment to the American Council on Education, as an expert faculty evaluator for military medical programs on behalf of the Department of Defense.

Her research interests include advancements in neurological surgery, special needs populations, student predictive factors, and interprofessional education. When she is not fulfilling her academic duties, she enjoys spending time with her children and traveling.