

THE RELATIONSHIP BETWEEN FORMATIVE AND SUMMATIVE EXAMS AND LICENSURE EXAM PASS RATES

Can the past predict the future

Scott J. Massey PhD PA-C
Central Michigan University

WHY IS THIS IMPORTANT

- The relevance of this presentation is to inform participants how assessment instruments can be used to measure students acquisition of knowledge . This can be used to determine if students have met graduation requirements in the cognitive domain through summative examinations .
- This assessment methodology has been developed over 10 years of ongoing research. The presenter believes that any program that has a graduate licensure exam can develop a similar method of assessment



GENESIS ASSESSMENT SYSTEM

- Tutored students who failed high-stakes exams for 25 years
- Qualitatively determined that student learning deficits existed early in the education program
- Piloted summative exams over six years including 10 external PA programs
- .Developed recommendations for implementation based upon predicted scores used longitudinally
- Testing instruments refined psychometrically based upon correlational strength
- .Multiple test writers participated in project.
- Utilize nationally standardized blueprint for national certification exam



CONCLUSIONS OF 6 YEAR LONGITUDINAL STUDY

- this validates that program based formative and summative exams can provide statistically valid indicators of future performance on nationally standardized licensure exams
- The use of multiple instruments has demonstrated to be the most accurate in identifying students potentially at risk for failure and providing prompt intervention

Overview of paper

- The scores of the participants from five PA programs were gathered between 2011 and 2017 to determine correlational strength between PACKRAT and SUMM I and II and the PANCE.
- The SUMM I, SUMM II, and PACKRAT scores (N = 641) were combined in a regression model to generate a predicted PANCE score.
- The strength of correlation between the predicted PANCE score and actual PANCE score demonstrated an R value of 0.72 and an R² of 0.51. In 2014 -2017, Seven 120 question End of Rotation™ exams (EORE) created by PAEA were administered to students from seven PA programs (N= 276) Pearson coefficient between PANCE and EORE is 0.642 and in R² of 0.412 were obtained

Selected publications

- Peer-Reviewed Articles

2015 Massey, S., Holmurd, D., Hammond, J., Tipton, L., Rizzolo, D., Preston, T. "Correlation of Physician Assistant Education Association End of Rotation Examinations with the Physician Assistant National Certification Examination." *Journal of Physician Assistant Education* 26(3): 1-3.

2013 Massey, S., Lee, L., Young, S., Holmurd, D. "The Relationship Between Formative and Summative Examinations and PANCE Scores: A Multiprogram Study." *Journal of Physician Assistant Education* 25(1): 27-34.

2011 Massey, S., Stallman, J., Lee, L. "The Relationship Between Formative and Summative Examinations and PANCE Scores: Can the Past Predict the Future?" *Journal of Physician Assistant Education* 22(2): 23-30.



BACKGROUND

- In the field of PA education, there is a lack of validated formative and summative instruments that can accurately predict future PANCE performance. There is also a need for PA programs throughout the world to begin building formative and summative assessment tools to measure student progress and student achievement of program specific graduation requirements



INTRODUCTION

The identification of students at risk of failing the Physician Assistant National Certifying Exam (PANCE) or an analogous country specific Licensure examination is one of the largest challenges facing physician assistant (PA) educators.

The tasks of learning pedagogical methods, along with the necessity of assessing student performance, challenges experienced educators and overwhelms young, developing faculty. This reinforces the need for a standardized practice to analyze student performance and identify students at risk who still have adequate time for effective intervention.

PURPOSE

- The purpose of the study was to determine if program specific formative and summative assessment tools and the PAEA EORE and PACKRAT could be administered to PA students in multiple programs and achieve a consistent strength of correlation .
- This research study thereby could demonstrate that a system of program designed examinations could achieve external validity.
- This could provide a model for PA programs to develop multiple tools that can be triangulated to predict future PANCE (Country Specific Licensure Exam)performance combined with the system of early remediation and mentoring during the clinical year.

METHODS

- The six participating PA programs(N=641) administered the SUMM I, the SUMM II, and the PACKRAT examination to all second-year students .The Exams were administered at the same timeline . Participating programs were provided cut scores for each of the test instruments that identified students at critical risk for failing the PANCE
- The seven PA programs administering the PAEA generated examinations over 3 years .(N= 276) Correlational analysis was conducted to determine the relationship between performance on these examinations and the certification exam

INSTRUMENTATION

- ▶ 1. **SUMM I** is a 360 question PANCE simulation exam based upon the NCCPA blueprint – administered about one third of the way through the clinical year. This provides a first glimpse at student performance on PANCE simulation examinations
- ▶ 2. **PACKRAT** –a 225 question nationally standardized examination – administered midway through the clinical year. A lack of improvement can be a serious sign
- ▶ 3. **SUMM II** -a 700 question PANCE simulation based upon the NCCPA blueprint –administered about three months prior to graduation – this provides a final data source. Triangulation of all three instruments provides a final predicted score.

INSTRUMENTATION

- ▶ 4. Predicted PANCE Score is a numerical figure generated through logistic regression derived from retrospective data generated from all three instruments(SUMM I/SUMM II/PACKRAT)
- ▶ 5. EORE Seven 120-question End of Rotation examinations administered after specific clerkships measuring the following disciplines :family medicine, internal medicine, behavioral medicine, emergency medicine, surgery, OB/GYN, and pediatrics. This data was used to generate a Predicted PANCE Score based upon the average of the 7 EORE

HYPOTHESES AND STATISTICAL ANALYSES

In examining the relationship between formative and summative examinations and PANCE scores, three key questions were explored:

1. Is there a statistically significant correlation between formative and summative examinations and PANCE performance? (SUMMI/II/PACKRAT)
2. Is there a statistically significant relationship between predicted PANCE scores and actual PANCE scores generated by formative and summative assessment tools? Is there a statistically significant correlation between the predicted scores generated and the actual scores earned on PANCE
3. Is there a statistically significant correlation between the PAEA EORE scores and PANCE scores.

Results Question 1

When the cumulative database of **SUMM I** scores gathered between 2011-2017 (N= 641) were combined in a regression model to determine strength of relationship between the raw score in the actual PANCE Score, an R value of 0.51 and R2 of 0.284 were obtained

When the cumulative database of **PACKRAT** scores gathered between 2011-2017 (N= 641) were combined in a regression model to determine strength of relationship between the raw score in the actual PANCE Score, an R value of 0.68 and R2 of 0.477 were obtained

When the cumulative database of **SUMM II** scores gathered between 2011-2017 (N= 641) were combined in a regression model to determine strength of relationship between the raw score in the actual PANCE Score, an R value of 0.60 and R2 of 0.37 were obtained

Results Question 2

- When the SUMM I, PACKRAT, and SUMM II raw scores were combined in a regression model, a predicted score, based upon retrospective data gathered between 2011-2017, was demonstrated (N=641).
- The strength of correlation between the predicted score and actual PANCE score was generated using logistical linear regression. An R value of 0.72 and R² of 0.51 were obtained.

Results Question 3 2015 Data

- Forty-two students achieved a score of 400 or above on the PANCE, and thirty students scored below 400. An independent t test demonstrated a significant difference between the averages of these 2 groups of students.
- Students who scored 400 or above on the PANCE had an End of Rotation examinations average of 74% (SD = 7.4%) compared with those that scored below 400 who had an End of Rotation examinations average of 61% (SD = 6.1%) (P < .05)
- The mean PANCE score for students who achieved a 400 or above was 489 (SD = 49) compared with those who scored below a 400 who had a mean score of 336 (SD = 47) (P < .05).
- The Pearson coefficient between the PANCE and End of Rotation examinations scores is 0.856 (P < .05).



Results Question 3 2017 Data

- Data was obtained from 7 PA programs over a period of three years. N= 276 The average score was calculated from the seven exams.
- When this data was combined into a regression model a predicted score based upon the retrospective data was demonstrated.
- The strength of correlation between the predicted score and the actual PANCE score was generated using logistical Linear regression an R-value of 0.642 and in R2 of 0.412 were obtained



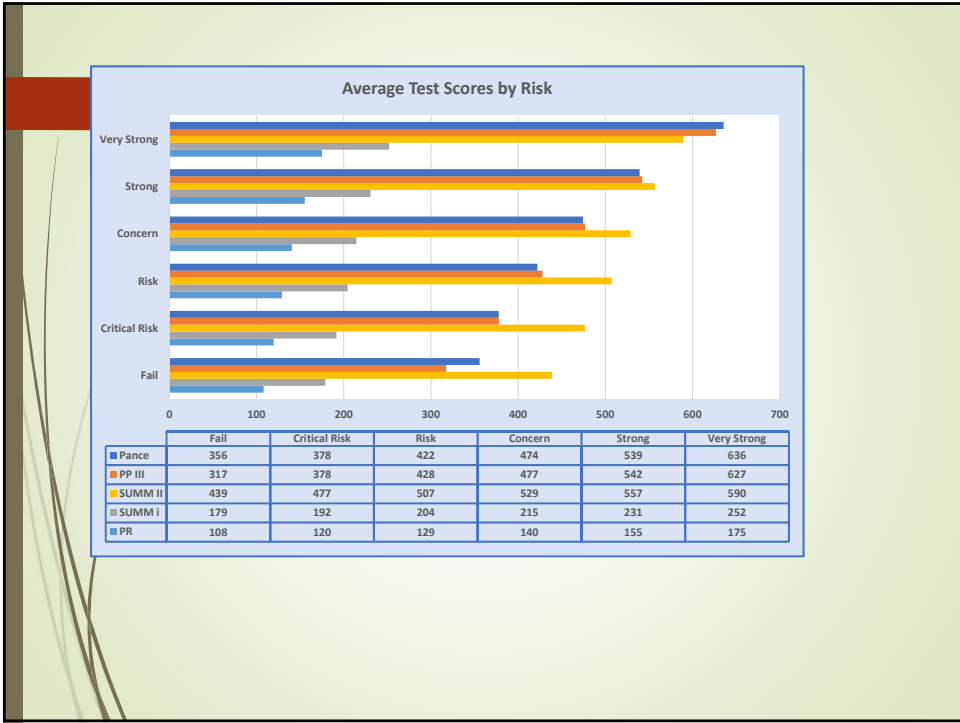
EXAMPLE OF PRACTICAL APPLICATION

CENTRAL MICHIGAN UNIVERSITY POLICIES

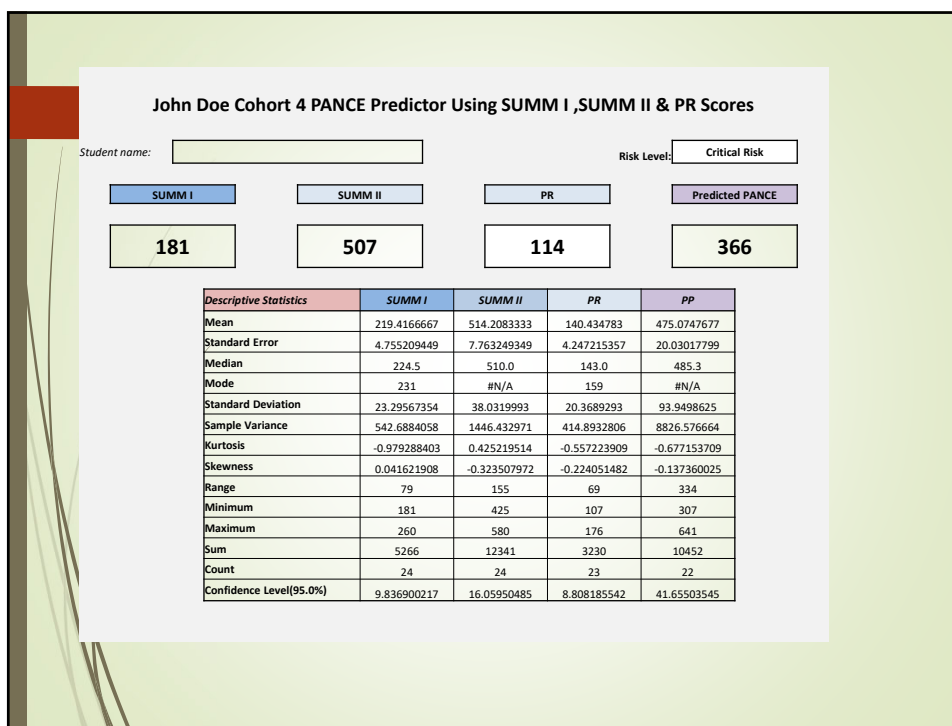
- To receive final signature for graduation students must complete all formative and summative assessments in the 2nd year. SUMM I, PACKRAT, and SUMM II.
- The grade in these courses is not based upon the numerical score achieved in these formative and summative examinations.
- The scores as defined in the graphic below which is entitled "average test scores by risk" is used to determine whether students require additional remediation and mentoring during the clinical year.
- Students with the highest level of risk will be required to complete more extensive remediation assignments defined in the study contract

BIBLIOGRAPHY RELATED ARTICLES

- The program will use the risk analysis score stratification to determine whether a student requires remediation and mentoring during the clinical year.
- Any student who achieves scores within category (three-risk) stratification or below risk category will be required to participate in a learning contract during the clinical year.
- If the student shows sufficient improvement during subsequent formative or summative examinations they may be released from the supervised study contract.
- Any student who achieves scores in the critical risk or fail category will be required to complete all elements of a study contract supervised by a faculty advisor. The faculty advisor can determine if the student has not satisfied the required remediation which can result in a delay in graduation



PANCE Results by Risk Category							
	Fail	Critical Risk	Risk	Concern	Strong	Very Strong	Total
Taken	31	63	111	143	234	59	641
Passed	18	47	94	137	231	59	586
Failed	13	16	17	6	3	0	55
Percentage	58.1%	74.6%	84.7%	95.8%	98.7%	100.0%	91.4%



LIMITATIONS

- Inability to control for confounding variables
- The examination questions must be connected to program specific disease blueprints with appropriate numbers of topical questions
- The distribution of blooms taxonomy must be roughly equivalent in program generated testing instruments and the licensure/registry examination
- Program curriculum must be mapped against content blueprint for the required licensure/registry examination to achieve any validity for the formative and summative examinations
- The power numbers to achieve statistical validity may require several classes (>100)

DISCUSSION AND CONCLUSIONS

- ▶ This research needs to be replicated at a larger scale to determine if there are widespread implications for other professional level programs. The test instruments (SUMM I, SUMM II) developed by the principal investigator can be replicated by professional programs using a standardized approach.
- ▶ Constructing test instruments using the licensure exams blueprint and incorporating questions at higher levels of Bloom's taxonomy are essential to attempt to replicate a similar instrument.
- ▶ The ability to identify students who are not achieving required performance standards and intervene with remediation is the most valuable insights gained from this research

DISCUSSION AND CONCLUSIONS

The data derived from the study are encouraging. The study suggests that professional programs could develop a system to better predict future performance by using program-specific examinations and nationally normed examinations .

The data analysis suggest that there was external validity when these test instruments were administered to students from several PA programs as similar correlational data were demonstrated.

The guidelines developed to mentor students who were identified as at-risk have promise and could be replicated and implemented by any pa program to measure preparedness for national registry examinations and program graduation



Questions?

Scott L Massey PhD PA-C
Director and Professor
Physician Assistant Program
Health Professions 1223
Central Michigan University | Mount Pleasant, MI 48859
Contact: 989-774-2337 | masse2sl@cmich.edu
Fax: 989-774- 2433



Scott L Massey PhD PA-C
Director and Professor
Physician Assistant Program
Health Professions 1223
Central Michigan University |
Mount Pleasant, MI 48859
Contact: 989-774-2337 |
masse2sl@cmich.edu
Fax: 989-774- 2433

IMPLICATIONS FOR FURTHER RESEARCH

- Validate similar assessment instruments for IAPAE member programs through longitudinal analysis
- Connect summative examinations with program specific graduation competencies/Disease topics
- Develop predictive summative examinations in countries that require a Registry /Licensing examination to enter practice
- Replicate this study at other US based PA programs

REFERENCES

- Massey, S., Lee, S., Holmurd, D.; Hammond, J, Tipton, L.; Rizzolo, D; Preston (2015) Correlation of Physician Assistant Education Association End of Rotation Examinations with the Physician Assistant National Certification Examination : *Journal of Physician Assistant Education*. J Physician Assist Educ 2015;26(3):1-3
- Massey, S., Lee, L Young S., Holmurd D., (2013). The relationship between formative and summative examinations and PANCE scores: A multiprogram Study *Journal of Physician Assistant Education*. 25(1) 27-34
- Massey, S., Stallman, J Lee, L. (2011). The relationship between formative and summative examinations and PANCE scores: can the past predict the future? *Journal of Physician Assistant Education*. 22(2) 23-30.
- Massey, S., Sedrak, M., & Lee, L. (2008). The effects of an Interactive Learning System on PANCE Performance. *Journal of Physician Assistant Education*. 19(3):30-34.



Contact Me

Scott L Massey PhD PA-C
Director and Professor
Physician Assistant Program
Health Professions 1223
Central Michigan University | Mount Pleasant, MI 48859
Contact: 989-774-2337 | masse2sl@cmich.edu
Fax: 989-774- 2433