UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PHYSICIAN ASSISTANT (PA) PRODUCTIVE FAILURE: THE EFFECTS OF PRODUCTIVE FAILURE IN PA EDUCATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By
BOBBY BOSSE
Norman, Oklahoma
2025

PHYSICIAN ASSISTANT (PA) PRODUCTIVE FAILURE: THE EFFECTS OF PRODUCTIVE FAILURE IN PA EDUCATION

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF INSTRUCTIONAL LEADERSHIP AND ACADEMIC

CURRICULUM

BY THE COMMITTEE CONSISTING OF

Dr. Crag Hill, Chair

Dr. Robert Jackson, Co-chair

Dr. Maeghan Hennessey

Dr. Diane Horm

Table of Contents

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my family. To my wife, Crystal, your unwavering love, encouragement, and patience have carried me through every stage of this dissertation. You stood by me during the most challenging moments, and your support has been my anchor. To my children, Sydney, Elijah, and Dylan, you are the rhythm of my heartbeat and vital to my success. I know this journey came with sacrifices of time together and shared moments, but your belief in me, even when I struggled to believe in myself, gave me strength. Thank you for your understanding, love, and constant motivation.

To my parents, you ingrained the work ethic, dedication, and love required to complete a project of this magnitude. You are and were the initial foundation needed for me to find success. To my siblings, you have always supported me and provided additional strength to complete my endeavors. Thank you for your continued support and love.

To my dissertation committee, thank you for your invaluable guidance, insight, and support throughout this process. Your thoughtful feedback and academic mentorship shaped the quality of this work and helped me grow as a scholar. Dr. Hill, your steady support and intuitive guidance often arrived just when I needed it most. Dr. Jackson, your detailed reviews of my work challenged me to reach a higher standard of academic rigor. Your expertise helped sharpen both the content and structure of my work. Dr. Hennessey, your kindness, thoughtful but meticulous critiques, and professional integrity were deeply appreciated throughout this journey. Your teaching and character shine through your instruction and allow students to know who you are. Dr. Horm, from the very beginning, your questions pushed me to think deeply, and your invitation to define education and teaching early in my coursework set the tone for the reflection and growth that followed. Your encouragement never wavered, and I am sincerely grateful.

I am also thankful to the Oklahoma City University PA Program faculty for their support and flexibility. Your willingness to absorb additional responsibilities made it possible for me to dedicate time and energy to this research. Joey, your mentorship, wisdom, and steady presence were instrumental. You offered guidance and friendship at every step. Sally, you instinctively knew when to step in and offer help, your support made a lasting impact on my career. Stacey, your support and direction in study design and guidance in data analysis were critical to the success of this project. Your continual encouragement brought me the strength to complete this project. To Becky and Niki, your roles in coordinating and facilitating the OSCEs were vital to its academic rigor. This project could not have come to life without your expertise and dedication. Your commitment to our students and high-quality simulation enriched every part of this study. To Sara and Nancy, thank you for the support you provide me and the talents you display by instructing our students.

To the OSCE facilitators: Shawn Altebaumer, Alvin Short, Taylor Airola, Caleb Bullard, and Cate Joyner, thank you for donating your time and expertise to help create authentic, meaningful learning environments. Your contributions made this study not only possible but impactful.

This dissertation would not have been possible without each of you. Thank you for walking beside me on this journey and for shaping my growth as both a scholar and a person.

In closing, I am profoundly grateful to everyone who played a part in this journey. Your encouragement, guidance, and belief in me helped me reach this important milestone and prepared me for the road ahead. Thank you for helping shape the person I've become.

.

Abstract

Introduction

Physician assistants (PAs) need highly developed critical thinking and problem-solving skills, along with the ability to adapt to diverse clinical situations. The current highly structured nature of curricular activities in medical and PA training often fails to effectively transfer knowledge and skills to the clinical practice environment. This study proposes the implementation of Productive Failure (PF) in PA education through simulation-based clinical patient experiences. Productive failure is an instructional approach where learners attempt to solve complex problems before receiving formal instruction. Productive Failure aims to enhance critical thinking, problem-solving skills, learner motivation, and resilience among PA students. However, there is a lack of research evaluating the impact of PF in PA education.

Methodology

This study uses high-fidelity simulation scenarios to create realistic and challenging clinical situations, allowing learners to engage in problem-solving and experience initial failure in a controlled setting. All clinical phase students enrolled in the Oklahoma City University Physician Assistant program were invited to participate in this study. Voluntary participants were stratified based on an emergency medicine rotation completion, then randomly assigned to either a control or intervention group in a randomized control trial (RCT). Both groups participated in an observed structured clinical experience (OSCE) simulation activity. Immediately before and after the OSCE simulation, both groups completed pre-simulation and post-simulation surveys to assess self-efficacy (confidence) and emotional stress responses. Participants' clinical competency in decision-making was also evaluated during the simulation. The intervention group experienced productive failure during their OSCE simulation as the sole intervention. This

involved a progressive decline in the patient's clinical status despite appropriate medical interventions initiated by participants. Immediately following the OSCE simulation, each group participated in a debriefing session. Following this, only the intervention group took part in focus group discussions to reflect on learning from OSCE activities that incorporated PF in the patient scenario. Pre-simulation and post-simulation self-efficacy, pre-simulation and post-simulation emotional stress responses, and competency scores were analyzed to assess the effects on participants and the effectiveness of PF in PA education.

Conclusion

This study highlights the impacts and potential benefits of integrating productive failure into PA education through OSCE simulation, suggesting that this approach can better prepare students for the complexities of clinical practice. Future research is needed to explore the long-term effects and benefits of PF and its applicability across different PA education contexts.

Chapter 1: Introduction

Background and Significance

The field of medicine is a dynamic, ever-evolving industry requiring practitioners to possess extensive medical knowledge, advanced critical thinking, problem-solving skills, and a high degree of adaptability to various settings and situations. Physician Assistants (PA) have transformed the healthcare system since the profession's inception in the 1960s. They play a critical role as members of the medical team and serve as integral components within the healthcare system. As clinical providers, PAs take on various responsibilities, which include patient evaluation, physical examination, diagnostic evaluation and interpretation, and the clinical and therapeutic management of patients.

Education programs designed to prepare PAs for their roles emphasize the practical application of theoretical knowledge and skills with the goal of graduating learners who are well-equipped to meet the demands of diverse clinical environments. The educational training of physician assistants is evolving to meet the changing demands of the healthcare industry and to mirror the educational changes occurring within medical school education. The healthcare industry is requiring providers to be more adaptable and better equipped to apply knowledge and practical skills to the fast-paced and progressively changing clinical environment. However, an enduring challenge in PA and medical education is the ongoing efforts to optimize teaching methods and learning strategies to enhance the clinical competence, confidence, and adaptability of future providers.

Progression of Medical Education to Application-based Training

Medical education has long utilized application-based learning curricula as a method of instruction. The apprentice model of instruction was the primary form of instruction before the early 1900s, when standardization of the American medical curriculum was implemented (Flexner, 2002). This standardized model of medical education included a two-phase curriculum, the didactic phase and the clinical phase. The didactic phase of medical instruction followed the behavioristic views of B.F. Skinner and John Locke, who believed that students are blank slates to be filled by the instructor and that the shaping of knowledge was through fixed repetitive conditioning (Palmer, 2001; Skinner, 1976). In this phase, content was traditionally delivered in the form of direct instruction, before application, shaping the learner's knowledge. The clinical phase of instruction followed an application-based instruction model similar to the previous apprenticeship form of instruction. This initial transformation of medical education prioritized teacher-centered knowledge transmission early in training over student-centered active learning, problem-solving, and learner adaptability.

Approximately fifty years after the standardization of medical education into two distinct phases, knowledge development (theoretical) and practical skill application, medical education underwent another transformation. This shift was prompted by the recognition that the existing model did not fully address the educational needs required for modern patient care. The focus transitioned to evidence-based knowledge development and competency-based outcomes, ensuring that learning was directly applicable to patient care.

This change was driven by advancements in cognitive psychology, which emphasized the importance of active engagement in learning, and by the increasing complexity of medical practice, which required learners to be adaptable to evolving patient needs. These pedagogical changes prioritized the individual learner, designing instruction around their specific needs and desired outcomes.

A learner-centered approach emerged, incorporating hands-on, practical educational experiences that actively involved learners in the process of knowledge acquisition. These methods, known as application-based and experiential-based learning, are designed to foster deeper understanding and skill development. Common instructional strategies in this approach include problem-based learning (PBL), Objective Structured Clinical Examinations (OSCE), and simulation-based learning (SBL), all of which are widely utilized in medical and PA education today.

Problem Development and Evolution of Productive Failure

Medical and PA education is designed to teach learners to critically think, analyze data, and make decisions based on the current data. Despite the current models of instruction implemented in medical education, learners continue to seek and desire a single right answer (SRA) for target concepts. This is further reinforced in training through high-stakes examinations and application-based curricula that employ a single right answer to problems (Witt, 2021). It is the SRA concept that may encourage the learner to devalue the process of education and seek only to find the concept outcome. These concept outcomes refer only to the measurable results that reflect a learner's understanding of the educational principle (Wiggins, 2005). Seeking this SRA can be an adaptive form of student survival and a coping mechanism employed during rigorous medical education (Witt, 2021). Additionally, modern medical learners, conditioned by technological advancements and unlimited access to resources, often find immediate, easily accessible correct answers to problems rather than a deeper and more complete solution (Judd, 2017).

Medical education's transformation in the mid-twentieth century to a learner-focused curriculum follows the theoretical underpinnings of constructivists. This educational philosophy posits learning as an active, constructive process where learners build new knowledge based on prior experiences (Piaget, 1953). However, in medical education, the process of creating knowledge through the constructive process of learning is not fully captured, despite the model of instruction currently utilized, direct instruction or application-based learning. Many medical target concepts and learning activities are highly structured and standardized, preventing true learner construction of knowledge as the learner focuses on the SRA for the target concept (Schmidt, 2015). Problem-based learning (PBL) is a constructivist instructional approach used in medical education that results in a single right answer of guided inquiry and is considered a productive successful learning activity (Kapur, 2016). Although there is substantial research on the benefits of PBL, OSCE, and other application-based problem-solving pedagogies, these approaches predominantly focus on highly structured learning experiences to minimize errors and arrive at an SRA. However, there is a lack of significant research in medical education involving ill-structured problem-solving activities that do not focus on an SRA

Although PBL, OSCE, and other application-based instructional activities allow learners to make the correct diagnosis and treatment, that is within their existing knowledge, they do not adequately create new knowledge. These traditional educational activities do not adequately prepare learners for the clinical practice environment, which involves ambiguity, concepts beyond current knowledge, and unsuccessful outcomes in clinical-patient interactions (Kapur, 2016). These educational approaches have merits and have served the industry well, but recent educational research suggests that experiences involving failure and learning from failure may be a more potent catalyst for a deeper understanding of theoretical knowledge and skill acquisition (Hattie, 2007; Kapur, 2008; Kapur, 2014; Steenhof, 2020). Activities that are beyond the learner’s current knowledge and ill-structured or ambiguous allow the learner to arrive at a cognitive impasse during critical thinking and problem-solving (Kapur, 2008). Research highlights that cognitive impasses in complex problem-solving create opportunities for generating new knowledge (Darabi, 2018; Kapur, 2008). This concept is referred to as productive failure (PF) and emphasizes the long-term learning benefits that arise through short-term failed performances when learners engage with challenges beyond their current expertise (Kapur, 2008). These initial failures encourage learners to explore and acquire new knowledge that extends beyond their existing understanding. Palominos (2021) further examines how ill-structured, ambiguous, and complex problems create meaningful learning experiences by providing opportunities for discovery in a non-threatening environment. These ill-structured, ambiguous, and complex problems utilized in PF require further investigation in medical and PA education.

Purpose and Research Aims

Experiential-based learning activities require the learner to actively participate and experiment within the learning environment to problem solve. However, medical education has failed to design effective curricula and educational activities that allow the learner to construct new knowledge through self-discovery, failure, and a reflective process on the learning experience. Productive failure is an instructional pedagogy that allows the learner to actively experiment within the learning experience to problem-solve solutions beyond their current knowledge and create new knowledge (Kapur, 2008). However, there is insufficient research to evaluate the use of productive failure in PA education.

The primary aim of this study is to evaluate and understand the impact of productive failure when implemented in PA education. A secondary aim of the study is to discover if participants identify any potential benefits of PF when utilized in PA education. These aims will be evaluated by implementing simulated productive failure into experiential-based learning activities in PA education. By accomplishing these study aims, the researchers hope to provide valuable insights that may help shape the future of PA education. These insights may improve the instructional methods utilized in PA education, impact learners' ability to adapt to various complex clinical problems, enhance clinical decision-making skills, improve confidence, and improve reflection on the clinical experience. Ultimately, improving learners' adaptability and reflection may lead to improved patient care and decreased provider burnout.

Chapter 2: Literature Review

Introduction

The shift in medical education to a learner-centered approach has been driven by multiple key factors, reflecting significant changes in educational theory, the requirements of modern healthcare, and the evolving changes in the knowledge and skills of professional healthcare providers. Traditional instructor-dominated approaches to medical education have not been sufficient to assist learners in developing skills and attitudes for the lifelong learning requirements of the modern medical professional (Towle, 1996). The transition to a learner-centered approach represents a paradigm shift from traditional instructor-centered methods to pedagogical strategies that prioritize the needs, experiences, and active participation of the learner (Knowles, 1980). Knowles advocated for adult learning experiences that promote self-inquiry supported by adequate resources, making the learning process an active one where learners discover meaning and knowledge through their experiences.

These changes in educational theory and subsequent shifts in medical education were influenced by constructivist and cognitivist theories. Cognitivists believe that learners acquire knowledge through active mental problem-solving, organizing information to remember and apply it to the learning environment (Yilmaz, 2011). Similarly, constructivists view knowledge as a process of experiences where learners build symbolic relationships to previous knowledge, representing reality through active discovery (Cooper, 1993). The principles of active mental problem-solving, structured learning environments, knowledge construction through experiences, and applying knowledge in real-world contexts have transitioned medical education to a learner-centered method over the past half-century. The goal of a learner-centered approach to medical education focuses on the need for enhancing critical-thinking and problem-solving skills, knowledge retention, motivation and engagement of the learner, and fostering lifelong learning skills.

The goal of a learner-centered approach in medical education is to enhance critical thinking and problem-solving skills, improve knowledge retention, increase learner motivation and engagement, and foster lifelong learning skills. These goals were largely driven by accrediting bodies in medical education to standardize education into a competency-based education (CBE) model. The Accreditation Council for Graduate Medical Education (ACGME) and the Accreditation Review Commission on Education for the Physician Assistant (ARC-PA) have promoted CBE as the training standard. Competency-based education focuses on learning outcomes rather than the process, emphasizing the demonstration of specific competencies required for effective medical practice. This approach aligns with learner-centered methods that prioritize developing practical skills and real-world knowledge application over rote memorization.

Research evidence has shown that the learner-centered model has significant benefits across multiple educational domains, including medical education. Studies indicate that active learning strategies, such as problem-based learning (PBL), simulation-based learning (SBL), and team-based learning (TBL), enhance learner retention of knowledge, clinical reasoning skills, lifelong learning development, and student satisfaction (Albanese, 2000; Candy, 1991; Deci, 1991; Freeman, 2014; Hmelo-Silver, 2004; Schmidt, 2015). These instructional methods encourage deeper student engagement with knowledge concepts and facilitate the application of knowledge to the clinical practice environment. Medical education has implemented various learner-centered pedagogies to impact learning outcomes and realize the long-term benefits. Observed structured clinical experiences (OSCE) and simulation-based learning (SBL) are learner-centered instructional methods that allow learners to apply theoretical knowledge to a clinical problem-solving scenario (Albanese, 1993; Joshi, 2017). These methods are practical examples of the experiential learning theory of education, emphasizing the learning process and bridging the gap between theory and practice.

Experiential Learning

What causes change in the learner? If instruction aims to move the learner from one knowledge point to another, then change must occur. If a change in the learner occurs, does it persist, or is it retained? What motivates the learner toward this change, and what is the learner's perspective on that change? These are some of the questions the experiential learning theory aims to answer. Kolb defined learning as "a process whereby knowledge is created through the transformation of experience" (Kolb 1984, p. 41). He believed that learning through experience integrates a more comprehensive view of learning by combining behavior, cognition, and perception. This process of learning through experiences builds on Piaget's theories of learning as an internalized reflection of adaptation to the environment. Experience provides the environment for the learner to process knowledge through accommodation and assimilation into abstract knowledge through active reflection (Piaget, 1972). Interaction between the learner and the environment creates experiential knowledge. "Experience does not go on simply inside a person. It does go on there, but it influences the formation of attitudes and purpose" (Dewey 1938, p. 38). This transformation of attitudes and purpose is crucial to experiential learning, as it creates change in the learner. Yardley (2012) identifies a few elements needed to construct learning through an experience. The first is pre-existing knowledge, which was created by previous experiences. In medical education, this would be the theoretical knowledge developed during the initial phase of training, as basic scientific knowledge or knowledge from previous undergraduate courses and employment experiences. The second element is the learner's ability to identify and differentiate between previous experiences and knowledge gained from the new experiences. This is done in a discursive manner, allowing the learner to reflect on the new experience and evaluate why this experience differs from previous ones. These elements are necessary for the learner to process and transform the learning experience into newly constructed knowledge.

Theoretical Foundations of Experiential Learning

Experiential learning is rooted in the theories of educational philosophers and psychologists such as John Dewey, Kurt Lewin, and David Kolb. This approach contrasts with traditional didactic learning, focusing instead on the direct application of knowledge in real-world contexts. The theoretical framework for experiential learning integrates several key theories and concepts that collectively explain how learning occurs through active involvement and reflection on the experience.

Knowledge Acquisition: Cognitive and Social Constructivists

"Knowledge emerges only through invention and reinvention" (Freire, 1970, p. 58). The epistemology of experiential learning begins with a pre-existing or base knowledge that the learner brings to any situation or experience. Dewey (1938) believed knowledge to be constructed socially through experiences in life. He argued that life is filled with uncertain and unstable experiences, where learners use base knowledge as a foundation to build new knowledge. Using this foundational knowledge, learners reconstruct life experiences through problem-solving strategies to progressively construct increasingly complex information.

Piaget (1953) argued that learners construct and reconstruct knowledge using organized representations of base knowledge he calls schemas. These blocks of base knowledge intersect with the social environment, causing conflict between the base knowledge and the perspectives of the social interaction. The conflict facilitates learners to adapt their thinking and expand their knowledge (Piaget, 1977)

For Dewey, discursive problem-solving follows a general pattern: initial equilibrium, discourse or the presenting problem, refinement or evaluation of the problem, formulation of a hypothesis in problem-solving, experimentation or application of the solution, and reflective review of the process (Kolb, 1984). This social interaction between the environment and the individual's experience allows learners to develop problem-solving skills and acquire new knowledge for application. This model of knowledge acquisition uses an active process of adaptation, centering the acquisition on the learner's experience rather than the instructor's transmission of knowledge. Additionally, the process involves layers of complexity in mental processing, as learners build new knowledge upon existing knowledge, using theories of Piaget's schemas and concepts of Vygotsky's *zones of proximal development (*ZPD).

Vygotsky rejected the prevailing theories of his time regarding the relationship between learning and development. He believed that learners come to school with pre-existing concepts, such as size, geometry, and mathematical skills, based on their prior experiences (Vygotsky, 1978). It is the introduction of new material or problems in school that allows learners to build on these pre-existing concepts to construct new knowledge in solving school-based problems. This progressive process of knowledge acquisition builds upon the base knowledge but is dependent on the development of the learner. Vygotsky defined "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" as *ZPDs* (Vygotsky 1978, p. 86). This establishes a structure of obtainable mobile targets in the process of knowledge acquisition and the environmental experiences in which the learner should be placed for progressive development.

In medical education, learners might be able to discuss basic science theories in the initial stages of training without assistance but would need further assistance from more advanced peers or instructors in abstract problem-solving. The learning environment is designed in this hierarchical model to progress learners through these obtainable distances in a progressively complex and independent manner to acquire the necessary knowledge. With the assistance of “capable peers”, knowledge acquisition occurs through socially co-constructed experiences where learners build on their base knowledge (Vygotsky, 1978). This assistance acts to support learners in the process of development until they can perform activities independently.

Bruner furthered Vygotsky's theories of social construction of knowledge and ZPDs by defining these supportive acts of more knowledgeable others (MKOs) as educational scaffolding (Bruner, 1973; Wood, 1976). He also believed that learning is an active process where learners rely on past knowledge to formulate hypotheses and make decisions on cognitive structures. Bruner (1966, p. 72) stated, “Knowing is a process, not a product”, describing the process of knowledge acquisition as one that stimulates inquiry and skill development, rather than mere memorization. This process of knowledge acquisition encourages active inquiry from learners, which is significant in developing critical thinking and problem-solving skills. Thus, it provides the potential for lifelong learning and meets the goal of the learner-centered model in medical education.

Experiential Learning Process

Kurt Lewin’s field theory and his work on action research provide a foundational understanding of the experiential learning process. Lewin argued that learning is best achieved through a cyclical process of planning, acting, observing, and reflecting (Lewin, 1946). This cycle is central to experiential learning, where learners engage in iterative cycles of experience and reflection. Lewin's theory also emphasizes the importance of experiencing change. He believed that real learning occurs when individuals are involved in activities that challenge their existing beliefs and practices, leading to both personal and professional growth. His model for the process was based on field studies of human behaviors and their interaction or tensions between individuals and the environment. There are four stages of this model that continuously evaluate the observed consequences within the interactions between the individual and the environment for directed actions (Kolb, 1984), as shown in Figure 1.

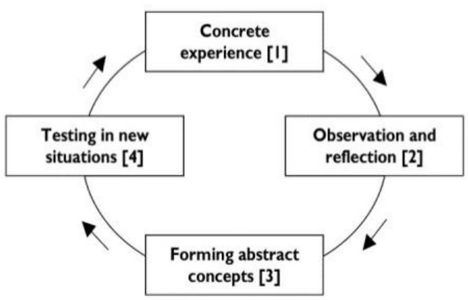


Figure 1 (Kolb, 1984).

The experience leads the individual to observe and reflect on the experience, followed by the formation of abstract concepts regarding the experience, and finally, experimentation with the actions formulated for a new situation. A significant similarity can be seen between the scientific method and this model of experimentation. Another model of the experiential learning process shares this similarity, as both Dewey and Lewin describe models of feedback association for final actions. Dewey (1938) described the process as differentiating the differences in continuity of the experiences that ultimately result in a change in action or goals. Like Lewin, this represents a type of feedback system in which the environment creates discourse, and adaptation or modification in the learner occurs to prepare for future experiences. This adaptation is additionally realized by Piaget in his "intelligent adaptation" regarding individuals' intelligence in response to various environmental stimulations (Piaget, 1953). Although his stages of cognitive development are more fixed, or non-variant, and do not involve this adaptive process, his theories on intelligence are more adaptive. Adaptation is a process of integration or substitution, where higher thought or deductive reasoning related to the environmental stimuli either integrates a new situation into a pre-existing schema or creates/substitutes a new schema to manage the situation. "Intelligence does not by any means appear at once derived from mental development, like a higher mechanism, and radically distinct from those which have preceded it. Intelligence presents, on the contrary, a remarkable continuity with the acquired or even inborn processes on which it depends and at the same time makes use of." (Piaget, 1953, p. 21).

Each of these models involves similar characteristics of the experiential learning process. These characteristics include 1) foundational or theoretical knowledge that is brought to the experience, 2) observational reflection of the experience, 3) abstract reasoning in problem-solving theories, and 4) practical application and experimentation for a solution. For the creative process of knowledge formation in an experience to be successful, each of these characteristics is utilized by the learner to balance the tension in problem-solving (Dewey, 1938; Kolb, 1984; Schön, 1983; Wallas, 2014). This process has the potential for enhanced problem-solving skills to meet the goal of the learner-centered model in medical education. By systematically engaging learners in learning experiences involving reflective observation, abstract reasoning, and practical experimentation, this process can cultivate adaptive expertise and critical thinking skills crucial for clinical decision-making.

Learner Motivation

Multiple components are involved in making each learning experience successful. The first component is the motivation of the learner within the learning environment. Rogers (1994) believes one of the key distinctions of experiential learning is that it meets the desires and needs of the learner in the learning process. The motivation of the learner is a balance of internal and external rewards (Bruner, 1966; Kusurkar, 2013; Stegers-Jager, 2012; Sobral, 2004; Wilson, 2009). Within each learning experience, the individual balances their degree of involvement in the learning process based on these perceived rewards (Abercrombie, 2022; Pelaccia, 2017). External rewards may include expected gains such as grades, success in completion, promotion, recognition, or other possibilities. Internalized rewards may consist of emotional empowerment, psychological enlightenment, autonomy, or personal goals or success. The external components focus on performance, whereas the internal ones focus on the process of learning. The motivation of the learner is based on the expected outcomes from the experiences. Tailoring the experience to the needs of the learner and focusing on the process of learning by placing the experience within obtainable distances for the learner improves the learner's motivation (Deci, 1991; Knowles, 1984; Vygotsky, 1978). The value of success or pertinence the learner places on the learning experience may be key to success in the process of learning (Spence, 1983). Holland (2016) showed that motivation in medical education focuses on successful graduation versus the internal motivation to learn. Additionally, Witt (2022) postulated that current medical learners focus on outcomes and single right answers (SRA) as a survival mechanism reinforced by current forms of instructional assessment. Thus, significant emphasis on the experience of the medical learner is needed to establish an environment that can positively impact the learner's motivation toward the process of learning.

The Learning Environment

The structure of the learning environment plays a significant role in the process of experiential learning. Dewey (1938) believed the learning environment should be a representation of real-life situations. The environment should involve social settings and interactions that would promote the creation of new knowledge. It should be simple and balanced to facilitate progression through the learning process. The goal of the environment is to set the stage for a successful educational experience and knowledge construction. Consequently, the structure of the environment should allow the interactions of the learner within the experience to foster a process of progressive knowledge acquisition.

For Vygotsky, this would create an environment where the goals of each activity would be carefully configured to reach expanding zones of knowledge development. According to Wood (1976), this concept creates educational "scaffolding" for the progressive learning process. This involves simplifying tasks and providing consistent, relevant direction. As the learning process expands, the environment can become more specialized or generalized.

Not only should the learning environment be structured progressively, but it should also be arranged to cultivate the learner’s natural propensity to learn. Rogers (1994) believed this environment should include setting a positive atmosphere for learning, clearly defining the stated purpose, providing structure and learning resources in the process, and balancing cognitive and emotional learning. If any of these elements are lacking, the learning outcome may suffer or be unproductive. In Rogers’ environment, the learner must be free to actively participate and self-direct their learning focus to meet their needs, enjoying the learning experience facilitated by the environment.

Bruner believed that learning occurs through social interactions for the construction of knowledge. Therefore, the goal of education is to facilitate the learner's constructive process through interactive problem-solving. He posited that learners use modes of representation in the learning process, including enactive, iconic, and symbolic modes (Bruner, 1966). Enactive representation involves hands-on direct experiences, iconic involves visual representations and reflections, and symbolic representation is the final mode, which progresses from the first two. In the symbolic mode, learners use language and symbols as a representation of what was learned. Using these concepts, the learning environment involves practical application, conceptualization, reflection, and analytical explanation of experiences.

The Experience

"The belief that all genuine education comes about through experience does not mean that all experiences are genuinely or equally educative" (Dewey, 1938, p. 25). This ultimately suggests that some experiences are counterproductive in the process of forming knowledge. The experience of learning involves exposure to new concepts and the modification of old ones, making learning a process of re-learning based on existing concepts. Poor learning experiences can result from a lack of motivation by the learner, insecurities of the learner in their pre-existing knowledge, rigidity of the experience, and poor continuity of the experience to existing knowledge (Bandura, 1997; Knowles, 1984). The goal of the instructor is to create an environment and experience that enables learners to build continuity from previous experiences to the new experience and acquire a progression in knowledge construction. For Dewey (1938), these progressive experiences in knowledge were to prepare individuals for the responsibilities and tasks in life through hands-on experiences.

As previously noted, characteristics of the experience should be built upon a progressive hierarchy of knowledge formation for the desired outcomes. This includes a strong foundation of knowledge to help the experience modify existing information and progress through observation and reflection, expanding the learner's understanding. It should also include abstract conceptualization for problem-solving and practical application of the solution in an experimental form. Dewey believed the experience should involve an interaction between the learner and the environment of learning. A "transaction" between the learner and this environment in which the outcome is a modification of knowledge and behavior (Dewey, 1938, p. 43).

Bruner and Vygotsky both believed that social interactions are involved in the constructive process of knowledge formation through interactive practical experiences. There exists a dialogue in the transaction between the individual and the environment. That dialogue is one that can be internalized by the learner, external, or occur in social groups of learners. This dialogue allows the learner to form abstract concepts of problem-solving within the experience. It is a conversation in the modification of old knowledge to form new knowledge. This new knowledge can now be utilized during the hands-on learning experience in the process of further experimentation. This allows the learner to continually re-evaluate the solution to the given problem and make adjustments or refinements in the constructive process of the new concept. Thus, potentially enhancing the lifelong problem-solving ability of the medical learner.

Learner Reflection

Reflective actions, whether internal or external, are involved in the experiential learning process. These occur as the learner is processing the pre-existing information and modifying or removing it to apply the new information. Jordi (2011, p. 181) defines reflection as "the rational analytical process through which human beings extract knowledge from their experience”. For Mezirow (1991), the reflective process involves a critique of the learner’s hypothesis to determine if their previously held views related to the experience are still valid. This discursive concept of reflection in learning is similar to those presented by Piaget, Dewey, Bruner, Vygotsky, and many of the other constructivist theorists. However, for Mezirow, this process of discursive reflection results in a transformation in the learner, an adaptation of the cognitive process and/or emotional behavior. The discursive process of reflection is an internal dialogue and one that balances or replaces the old with the new. Freire (1970) used the term “praxis” to describe the active process in this dialogue of reflection. This dialogue creates links between the theoretical and applied knowledge of the learner. Reflection is a continual cycle that allows the learner to invent and re-invent knowledge through the process.

Schön further expands the concepts of the practical application of the reflective process. He defines the reflective process as one in which professionals become cognizant of their knowledge base and learn from their personal experiences (Schön, 1983). He differentiates between *reflection-on-action* and *reflection-in-action*. In medical education, a significant amount of *reflection-on-action* occurs after a learning experience, often in collaboration with peers or instructors. This process involves analyzing and discussing the experience that has taken place identifying strengths, weaknesses, and areas for improvement. By externalizing their thoughts, learners gain new insights and perspectives, which enhances their understanding and helps them refine their approach to future situations.

In contrast, *reflection-in-action* is an internal, real-time process that takes place while professionals are actively engaged in practice. It involves drawing upon their mental frameworks, accumulated through prior experience, to assess and adapt to the situation at hand.

This form of reflection allows professionals to be aware of their current knowledge, recognize patterns, and adjust in the moment. By leveraging their experiences and existing knowledge base, they continuously reconstruct and refine their understanding, progressing toward greater competence and expertise. As such, *reflection-in-action* is *learning-in-action*. This dynamic process is essential for navigating the complexities of medical practice and fostering ongoing professional development. It is this developed skill that is most sought in medical providers. The external dialogue in the reflective process may be inadequate to describe what is learned or how it was learned for the individual (Boud, 1985). However, the internalized process of reflection in tacit knowledge, or knowing more than you can verbalize or describe, may be more useful in clinical practice (Schön, 1983). This implies that not only is the instrument of self-conceptualization a balance in the discursive process regarding the adaptation of theory and practical application important, but moreover, is the understanding of the internalized self-reflective process of *knowing-in-action* through the experience of learning. Schön places importance on the adaptation from known schemas to preferred ones and the setting of the problem, or environment, in this reflective process.

Results of Experiential Learning

The results of experiential learning are highly dependent on the components involved in the process. These include, but are not limited to, the motivation of the learner, the environment, the experience, and the reflective process. The goal of experiential learning is not to follow a traditional transmission of knowledge from instructor to learner but to instill self-directed qualities, problem-solving skills, and create lifelong learners. This active learning process involves the invention and reinvention of knowledge and adaptation within the process, transforming the learner not only cognitively but also emotionally. For the adult learner, this process of teaching is more andragogy than pedagogy in that it moves the focus to learner-centered and progressive accumulation of knowledge through experiences (Knowles, 1980). This approach allows the learner to self-construct knowledge through active experimentation and reflection in practical, real-world applications.

This process of experimentation is commonly utilized in the second phase of medical training; however, introducing it earlier in training may further expand and deepen the learner's ability to "knowing-in-action". According to Schön (1983, p. 280), "doing and thinking are complementary. Doing extends thinking in the tests, moves, and probes of experimentation action, and reflection feeds on doing and its results". Reflecting on earlier knowledge and progressively deepening medical learners' understanding of the concepts aids in their transformation and provides more comprehensive training for medical practice. These cognitive transformations are accentuated by the emotional, behavioral changes, and improvements in learners' confidence within their professional practice, better meeting the needs of the medical provider.

The recognition of learning through direct experience is increasingly seen as essential in medical education. This approach offers numerous research-supported benefits, enhancing the knowledge, competence, and confidence of the medical learner (Yardley, 2012). These benefits include enhanced clinical skills and competence, improved patient safety, better retention and application knowledge, development of critical thinking and problem-solving skills, increased student engagement and motivation, and improved communication and teamwork skills.

Enhanced clinical skills and competence are notably achieved through simulation-based learning (Issenberg, 2005; McGaghie, 2011). A systematic review by Cook (2011) found that simulation-based learning improves knowledge and procedural skills in medical students and residents more effectively than traditional didactic methods. Zendejas (2013) also highlighted the competency benefits of simulation-based instruction for specific procedural skills such as airway management, gastrointestinal endoscopy, and central venous catheter insertion. Simulation-based learning has also been shown to improve patient safety by enhancing clinical skills, reducing error, and promoting better decision-making (McGaghie, 2010; Seymour, 2002; Wayne, 2008; Zendejas, 2013).

Enhanced retention and application of knowledge over time have been evidenced in both simulation-based (SBL) and team-based learning (TBL) instructional methods when compared to traditional instruction (Haidet, 2014; McGaghie, 2011). "Team-based learning helps bring up students at the bottom of the curve" by using Bruner's concept of the more knowledgeable other to inform lower-performing students and improve their retention and application of knowledge (Haidet, 2014, p. 6).

Two significant benefits of experiential learning are increased student engagement and motivation and improved critical thinking and problem-solving skills. Improved motivation and engagement are crucial, as they enhance the learner's value placed on the learning process, not just the outcome. Yardley (2012) reported higher levels of engagement and motivation in learners participating in experiential learning activities such as simulations and case studies compared to lecture-based learning. Critical thinking and problem-solving skills are improved when learners are placed in patient-based scenarios where they must apply their knowledge to make clinical decisions. Instructional methods such as PBL, SBL, and interactive case-based learning (CBL) are commonly used in medical education to foster these skills (Dolmans, 2005; Hmelo-Silver, 2004; Scalese, 2008; Shinnick, 2011). These approaches provide learners with opportunities to apply theoretical knowledge to practical situations, fostering the development of essential skills required for effective clinical practice.

Emergence of Productive Failure

As medical education sought instructional methods to better prepare learners for the ill-structured and ambiguous problems, to discover and construct knowledge beyond their current knowledge, and the unsuccessful patient outcomes that present in clinical practice, productive failure (PF) emerged as a potential solution. This experiential learning approach emphasizes the importance of struggling with complex problems before receiving formal instruction. Kapur (2008) described PF as a learning design where students are initially given complex problems to solve without any guidance. The initial struggle and failure in solving these problems set the stage for subsequent instruction, which helps students consolidate and make sense of the concepts. He defined "failure" to mean that learners "typically fail to generate or discover the correct solution(s) by themselves" (Kapur, 2014, p.1009). This method contrasts with traditional direct instruction by encouraging initial exploration and failure, thereby promoting deeper understanding and long-term retention of knowledge.

Learning from failure is not a new concept, as PF has been evaluated in many different settings and shown to have a moderately positive effect on learning (Darabi, 2018). The successes from PF have been realized in learning experiences that have challenged learners, but do not frustrate them within the complexity of the experience. Productive failure learning assists learners in developing conceptual understanding and the transfer of knowledge compared to more traditional learning methods (Kapur, 2015). Cognitive benefits, when comparing PF in a non-threatening simulated environment to direct instruction, have shown significantly improved outcomes in the areas of explanatory and application-based knowledge (Palominos, 2021).

Components of Productive Failure

Manu Kapur's model of productive failure involves two phases: phase 1- generation and exploration, and phase 2- consolidation (Kapur, 2012). During phase 1, learners are presented with a problem without prior instruction, encouraging them to draw on their prior knowledge to explore potential solutions. Problems presented to learners in this phase are ill-structured, open-ended, and complex, requiring learners to explore multiple solution paths. The complex and ill-structured nature encourages the learner to create new solutions without scaffolding or direct instruction. However, these problems should be challenging enough to allow for learner exploration without frustrating the learner and causing them to give up (Kapur, 2016). During exploration, learners individually or collaboratively work on the problem, generating multiple solutions, and are likely to struggle and encounter failure. The purpose of exploration and failure is to help students recognize the limitations of their current understanding and to create cognitive dissonance that motivates learning (DeCaro, 2012; Kapur, 2016; Loibl, 2014; Ohlsson, 1996). Generating multiple solutions to the problem promotes divergent thinking and the development of a deeper understanding of the problem space (Kapur, 2010).

Finally, in the consolidation phase (phase 2), learners receive direct instruction that helps them make sense of their experiences and integrate new knowledge. This targeted instruction provides explicit explanations that address the concepts and principles related to the problem. Additionally, it is informed by the common errors and misconceptions observed during the initial problem-solving phase. The purpose of this phase is to consolidate students' understanding and assist them in connecting their experiences during the initial phase with the correct conceptual framework. The scaffolded support in this phase helps gradually build students' confidence and competence in problem-solving while ensuring that they do not become overly frustrated or disengaged in the learning process (Kapur, 2008).

Theoretical Foundations of Productive Failure

Productive failure is an experience in which learning is accomplished in the face of failure despite what is attempted by the learner to achieve success. This concept emphasizes the cognitive and learning benefits derived from struggling with complex, ambiguous, or ill-structured problems that initially result in failure. The goal is not necessarily about the task being possible to complete successfully within the experience, but rather about how the struggle fosters deeper understanding and prepares learners for future success.

According to Kapur (2008, 2016), productive failure occurs when the learner attempts to solve challenging problems that are beyond their current knowledge. Despite failing to achieve a correct solution initially, the process of engaging with the task and exploring, hypothesizing, and making mistakes enables the learners to construct a deeper understanding. This struggle provides a foundation for improved learning outcomes when they subsequently receive instruction or feedback.

Productive failure is a form of experiential learning. Similar to experiential learning, the theoretical underpinnings of productive failure (PF) in PA education integrate principles from constructivist learning theories, cognitive psychology, and educational research. By examining the cognitive and affective processes involved in PF, the theoretical foundation aims to elucidate the mechanisms through which struggle and error correction can facilitate deeper understanding and skill acquisition in PA learners.

Piaget's Theory of Cognitive Development. Piaget's theory of cognitive development focuses on how learners go through stages of development, actively constructing knowledge as they interact with their environment. This interactive process of learning emphasizes the importance of disequilibrium in cognitive development. Disequilibrium occurs in situations where current knowledge is insufficient to solve a problem and acts as a substantial driver for cognitive development (Piaget, 1972). He believed learners build schemas, which are basic building blocks of knowledge and mental structures that help individuals organize and interpret information within the environment (Piaget, 1972). When learners encounter challenges that do not fit their existing schemas, they experience cognitive conflict or disequilibrium, which motivates conflict resolution. Piaget (1972) called this resolution of conflict a process of balancing assimilation and accommodation. He defined assimilation as the integration of new information into existing schemas and accommodation as the modification, or adaptation, of existing schemas or the creation of new schemas in response to new information (Piaget, 1953). In the context of PF, disequilibrium is deliberately introduced by presenting learners with complex problems without guidance. The ill-structured nature of the presented problem exceeds their current schemas, or knowledge, creating cognitive conflict that necessitates constructing new knowledge for a resolution. This resolution is the process of accommodation or adaptation to the learner's current schemas.

Vygotsky's Sociocultural Theory. Vygotsky's theory builds upon Piaget's concepts of the learning environment by emphasizing the role of social interactions and cultural tools in the learning process. While Piaget focused on individual cognitive development stages through interaction within the learning environment, Vygotsky highlighted the importance of community and interaction with others for cognitive growth. He introduced the concept of Zones of Proximal Development (ZPD), which represents the range of tasks, or problems, that a learner can perform with the assistance of a more knowledgeable other (MKO) (Bruner, 1973; Vygotsky, 1978).

These tasks are beyond the learner's independent capability but close enough to their current knowledge that the MKO can advance their cognitive development. This theory underscores the importance of advancing learners from one ZPD to another, similar to Piaget's assimilation or accommodation of schemas. Vygotsky (1962) considered language central to cognitive development and the primary means through which the MKO transmits information. He believed that language is a cultural tool that shapes the learner's thought, and learning leads to development, whereas Piaget believed development leads to learning. In the context of PF, ill-structured problems are situated within the ZPD, where learners initially struggle with the complex problem but later receive guidance and feedback from the MKO to master it. The struggle creates cognitive disequilibrium, and guidance provides an opportunity to achieve understanding through discovery and collaboration. The concept of ZPDs and collaborative support offers a structural framework for knowledge building. Bruner termed this structural framework "scaffolding", characterized by the temporary nature of support provided to assist the learner just enough to help them progress in completing the task, without completing the task for them (Wood, 1976). This temporary support after initial failure or struggle assists learners in building their understanding and skills, enabling them to eventually complete the tasks independently.

Sweller's Cognitive Load Theory (CLT). Sweller's Cognitive Load theory provides a framework for understanding the mental effort required for individuals to learn new information (Sweller, 1988). He described the cognitive architecture of memory and how to optimize instructional design to enhance learning. Cognitive load theory asserts that learning is most effective when instructional methods are designed to align with the cognitive architecture of working memory and long-term memory (Sweller, 1988). Working memory is described as one that has limited capacity and duration, and it is the site of conscious processing. Long-term memory has a vast capacity, is considered relatively permanent in nature, and is accessed by working memory. Sweller (1988) described multiple types of cognitive load impacting learners and instructional design. The first is intrinsic load, which is the inherent complexity of the material being learned. Second is the extraneous load, which is the cognitive load imposed by the manner the material is presented and is influenced by instructional design. Lastly, the germane load is the cognitive load dedicated to the processing, construction, and formation of schemas. Implementing CLT concepts includes strategies to provide worked examples, gradually transition to independent problem-solving, and provide timely, constructive feedback to help learners consolidate understanding and correct misconceptions. Productive failure can be viewed as an instructional method utilized to manage the intrinsic cognitive load by engaging learners in complex problem-solving activities that make the underlying structure of the problem more apparent when direct instruction is eventually provided. Recent studies have shown that when learners struggle with complex tasks before receiving explicit instruction, they often develop a deeper understanding and enhanced problem-solving skills (Kapur, 2010; Sweller, 2011).

Bjork's Desirable Difficulties and Metacognition. Desirable difficulties are a concept introduced by Bjork and refer to learning tasks that are challenging yet manageable, promoting deeper processing and long-term retention (Bjork, 2020). Although these tasks may initially slow down the learning process and be perceived as challenging, they lead to improved cognitive encoding and long-term retention through task experimentation. Principles of desirable difficulties include the spacing effect, interleaving, variability, testing effect, and generation effect (Bjork, 2020).

Interleaving involves mixing different topics or types of problems within a single study session, helping learners develop the ability to distinguish between the different types of problems and apply appropriate problem-solving strategies for each. Variability uses different conditions under which the learning occurs, allowing learners flexibility in finding solutions to problems. The generation effect refers to a deeper understanding of the concept and the construction of solutions to the problem. Productive failure aligns with these concepts by challenging learners with tasks that induce struggle, thereby enhancing their cognitive engagement in the task and facilitating long-term learning. The complex problems introduced in productive failure may include the interleaving of multiple topics and varying conditions that cause learners to generate answers in an application-based learning environment.

Bjork (1994) discussed the significance of the learner's awareness of their abilities and competence, termed metacognition. A goal of the learning experience includes the opportunity to produce a valid assessment of the individual's competence related to the target concept (Bjork, 1994; Schmidt, 1992). This assessment serves as feedback to correct misconceptions about their competence. Metacognitive strategies involve learners' awareness and control of their cognitive processes. "Training is frequently nonoptimal because it fails to incorporate variability, delays, uncertainties, and other challenges the learner can be expected to face in a real-world job setting" (Bjork, 1994, p. 201). Productive failure presents variability and challenges that encourage metacognitive reflection, requiring learners to evaluate their problem-solving strategies, identify errors, and adjust their approaches to find solutions. This self-regulatory process enhances the learner's ability to monitor and apply knowledge that is adaptable to various contexts.

Error Correction and Feedback. Studies on error correction emphasize the importance of feedback and learning from mistakes. Research in this area explores how different types of feedback and correction methods impact learning outcomes, retention, and student motivation. Shute (2008) indicates that immediate feedback is often more effective for simple tasks as correction occurs before ingraining the concept. This differs from delayed feedback, which may be more beneficial for complex tasks and promoting long-term retention because it encourages the learner to retrieve information from memory, reinforcing the learning process (Shute, 2008).

One of the primary goals of PF is to promote long-term knowledge retention by leveraging the benefits of struggle and reflection. Providing immediate feedback during the exploratory phase could disrupt the productive struggle that is essential to the learning process. When learners are quickly corrected by an MKO, they may miss the opportunity to fully engage with cognitive dissonance and critical reasoning, both of which are crucial for developing deeper understanding. This premature feedback can diminish the richness of the learning experience and reduce the benefits of PF, which depend on the learner's active construction of knowledge.

To support the learning process, PF involves two key phases. The first phase, generation and exploration, allows learners to grapple with complex problems, reflect on their approaches, identify gaps in their reasoning, and test hypotheses. Delayed feedback plays a vital role here, as it encourages learners to engage deeply with the problem and retrieve prior knowledge, reinforcing long-term retention.

The second phase, consolidation, follows shortly after the exploration phase and provides learners with feedback and targeted instruction. This phase helps them make sense of their experiences, integrate new knowledge, and refine their understanding. Together, these phases and the timing between them enable learners to connect their initial struggles with structured insights, fostering deeper learning and better preparing them for future challenges.

Error correction and feedback are critical components of the learning process, particularly in educational settings. Feedback allows the learner to identify errors in the process of learning and make meaningful changes. When learners receive timely and specific feedback on errors, they are better able to comprehend the nature of their misconceptions and correct them (Roediger, 2011). Studies on self-correction with guidance show that error correction promotes deeper understanding and enhances learning, especially when learners are encouraged to actively identify and engage with their mistakes to understand why they occurred (Narciss, 2008; Steenhof, 2020).

Guided error correction encourages learners to reflect on their learning experiences, providing a bridge to transfer gained knowledge to different contexts. Additionally, supportive feedback can improve learner motivation and self-efficacy in the process of learning (Deci, 1991). Guided feedback must balance identifying both learner strengths and areas for improvement to maintain learner motivation and to promote persistent effort in tasks and the learning process.

Productive failure incorporates feedback as a critical component, ensuring learners reflect on their errors and refine their conceptual understanding. As PF utilizes complex tasks, feedback is often delayed until after the exploration phase, allowing learners to internally reflect on the task and explore self-corrective paths. The delayed feedback can more effectively consolidate concepts for the learner and encode them for long-term retention. Using a balanced feedback approach in PF encourages learner motivation and self-efficacy while avoiding negative impacts on their perception of the experience (Kapur, 2014; Loibl, 2014).

Schön's Reflective Practice. Schön's works on reflection emphasize the importance of reflection in professional practice to improve skills and adapt to new situations. He viewed reflective practice as a cornerstone of professional development and education, especially in fields requiring complex decision-making and adaptive expertise, such as PA practice. Through Schön’s concepts of reflection-in-action, reflection-on-action, and reflection-for-action, professionals can continuously improve their skills, adapt to new situations, and enhance their decision-making abilities (Schön, 1983).

He described reflection-in-action as a form of reflection that occurs during a task. This process involves active thinking and decision-making that occurs immediately during the task. This reflection helps professionals be flexible to new and unpredictable situations, allowing them to continuously modify their actions to the unfolding scenario. Reflection-on-action occurs after the action, it involves analyzing the experience to better understand what occurred and why it occurred during the situation. It is a reflective process allowing deeper learning and understanding to refine practice and improve future actions (Schön, 1983). Key components of his theories involve experiencing, observing, analyzing, evaluating, and planning.

Experiencing involves active engagement in the complex tasks of professional practice. Observing involves noticing and attending to various aspects as part of the experience. Analyzing involves examining the experience to understand underlying principles and tasks. Evaluating involves assessing the outcomes and effectiveness of actions taken within the experience. Planning involves using insights gained from reflection to plan future actions and improve practice. Developing these skills helps professionals improve their competency through active refinement of their abilities, enhance problem-solving skills by reflecting and implementing changes to avoid past mistakes, increase flexibility enabling them to adjust to unpredictable situations, improve decision-making through a comprehensive understanding of their collective experiences, and promote personal growth through continuous self-improvement and a life-long learning process (Osterman, 2004; Schön, 1983; Schön, 2000).

The process of learning from PF utilizes many of Schön's concepts, as it is an active learning experience in a practical setting and requires the learner to *reflect-in-action* and *reflect-on-action* through guided feedback to enhance problem-solving skills and adaptability. For PA learners, these are fundamental skills required for the ever-changing clinical practice environment.

Benefits of Productive Failure

Research across multiple educational disciplines and contexts supports the efficacy of productive failure (Kapur, 2008; Kapur, 2012; Loibl, 2014; Schwartz, 1998). Key benefits identified include enhanced problem-solving skills, improved conceptual understanding, increased learner motivation and engagement, development of resilience and perseverance, better preparation for future learning, promotion of metacognitive skills, and fostering of collaborative learning (Deci, 1991; Kapur, 2008; Kapur, 2012; Loibl, 2014; Mylopoulos, 2009; Schwartz, 1998). Kapur (2010) found that students who struggle with complex problems explore various solutions and become more adept at identifying and understanding underlying principles, leading to better performance on subsequent problem-solving tasks. Learners forced to engage with material at a conceptual level develop a more nuanced understanding of the subject matter and improved conceptual grasp (Kapur, 2012). This enhanced conceptual understanding allows them to transfer their knowledge to new contexts more effectively (Loibl, 2014).

Loibl and Rummel demonstrated that PF can increase learners' intrinsic motivation and engagement in the learning process, as initial struggles encouraged them to understand the content better and appreciate subsequent instruction more holistically (Loibl, 2014). Dweck

(2016) identified the development of resilience and perseverance, viewing the challenges and failures as an opportunity for growth and learning. Learners experiencing productive failure learn how to learn through a deeper conceptual understanding and advanced problem-solving skills, better preparing them for future learning and problems (Kapur, 2008). This preparation is due to the promotion of learner metacognitive skills through reflection, as individuals become more aware of their learning strategies and how to adjust them to new situations (Kapur, 2007). Additionally, PF enhances collaborative learning skills as students learn to articulate their thoughts, listen to others, and build on each other's ideas for improved learning outcomes (Kapur, 2014).

Implementation of PF in Medical and Physician Assistant Education

In medical education, PF has shown promise in enhancing clinical reasoning and diagnostic skills. Schmidt (2015) highlighted that encountering diagnostic errors and reflecting on them can lead to improved diagnostic accuracy and clinical judgment. Altering the conditions of the acquisition phase of learning impacts the process of learning in practice and allows learners the opportunity to more effectively transfer knowledge for improved retention of the concept (Schmidt, 2015; Steenhof, 2019). This process effectively teaches learners to think critically during clinical reasoning. Clinical reasoning is an essential skill for medical and PA learners to develop, and PF offers opportunities to enhance this development. The development of critical thinking skills allows the healthcare provider to make informed clinical decisions using evidence-based theory with improved diagnostic accuracy to improve patient outcomes (Facione, 1996; Greenhalgh, 2019; Norman, 2010). These concepts align with PF principles, where initial struggle, analysis of the problem, and subsequent reflection facilitate deeper learning and improved clinical decision-making. These struggles are desirable as they push learners to their cognitive limits, creating new knowledge from prior understanding (Bjork, 1994; Steenhof, 2020; Woods, 2007).

Integrating PF into medical and PA education requires a careful balance of these desirable difficulties with the necessity of achieving proficiency in critical competencies. Various approaches to achieve these practical struggles have been incorporated into the experiential learning pedagogies currently used in medical and PA education. These approaches include problem-based learning, observed structured clinical experiences, and simulation-based learning, as these methods encourage active engagement of the learner in practical clinical settings.

Problem-based learning allows students to collectively work in groups to solve practical problems through active engagement and collaboration. This instructional method has been shown to provide benefits of improved critical thinking skills, deeper understanding, and improved learner motivation (Hmelo-Silver, 2004; Smits, 2002). Despite the benefits, PBL's open-ended nature can lead to cognitive overload for learners, inconsistent coverage of the curriculum, and challenges for accurate assessment due to variability in the facilitation of the experience (Colliver, 2000; Dolmans, 2005; Sweller, 1988; Van Der Vleuten, 2005). In addition, the implementation of PBL into medical and PA education encourages the concept of SRA, further prioritizing outcomes over the learning process and impacting the true construction of knowledge (Witt, 2022).

The OSCE has been widely used for assessing clinical competencies in medical education (Harden, 1975). It involves a circuit of stations or patient interactions where students perform clinical tasks while being observed and assessed by facilitators. It systematically assesses clinical skills, including history taking, physical examination, communication, and procedural skills (Harden, 1975). This instructional method provides a standardized and objective assessment of learner competencies and an opportunity for feedback that is critical for learning through real-life scenarios (Bartfay, 2004; Boursicot, 2005; Newble, 2004). Multiple studies show the general benefit, feasibility, acceptability, and positive student attitudes toward the OSCE in medical education (Furmedge, 2016; Hammad, 2013; Khairy, 2004; Lazarus, 1983; Siddiqui, 2013). The OSCE has also been implicated in the development of professional identity and how reflection in learning is incorporated into the process of this identity formation (Goldie, 2012). Despite the cognitive and non-cognitive benefits, the OSCE's structured design can introduce a degree of artificiality in the clinical scenario, lacking the complexity and unpredictability of real-world clinical encounters, thus limiting its ability to assess clinical reasoning skills (Gormley, 2011; Wass, 2001). Additionally, it may not reflect the true capabilities of the learner due to the increased stress and anxiety associated with the high-stakes nature of the assessment (Sloan, 1995).

Simulation-based learning is an instructional approach that has become increasingly utilized in medical and PA education due to its potential to represent real-world clinical encounters more accurately, enhancing clinical skills without risking patient safety (Gaba, 2004; Okuda, 2009). It allows learners to achieve a high level of clinical proficiency needed through repetition, feedback, and reflection (Issenberg, 2005; Rudolph, 2006). Incorporating PF through SBL offers advantages over PBL and OSCE alone by encouraging experiential learning without focusing on outcomes and more accurately simulating real-world encounters. Additionally, simulated experiences include active feedback and reflection on the experience required for effective learning in PF.

Identification of Research Questions

Over the last few decades, medical education has increasingly aimed to prepare learners for the ill-structured, ambiguous problems and unsuccessful patient outcomes encountered in clinical practice. Current pedagogical approaches have shifted from direct instruction to experiential learning, which includes more active learner engagement. Instructional methods such as SBL and the OSCE provide approaches that engage learners in the process, improving competencies and clinical decision-making.

Since its inception, PF has been considered a method to enhance complex problem-solving skills and improve the transferability of these skills to real-world applications. Implementing PF into experiential learning activities may create a more robust learning environment with better long-term skill retention. While the benefits of SBL and PF have been studied independently in medical education, there remains a dearth of research examining their intersection within the context of PA education.

Lee (2014) evaluated the potential effect of failure through a no-win clinical scenario on a medical learner’s psyche and the ability to learn from the experience. Lee concluded that the simulated death of a patient during the learning experience did not significantly affect the graduate medical learner's psychological or cognitive state. A one-month follow-up interview with participants showed that learners felt more prepared to manage similar patient scenarios in the future. This improved self-efficacy may have been influenced by the complexity of the clinical scenario. Further research is needed to evaluate the effects of PF on the medical and PA learners to identify possible benefits and disadvantages of this instructional method.

The effects of PF and failure on PA learners have not been thoroughly investigated despite the limited volume of medical research showing its benefits. The primary aim of the proposed study is to evaluate and understand the impact of productive failure when implemented in PA education. A secondary aim of the study is to discover if participants identify any potential benefits of PF when utilized in PA education. To accomplish these aims, the following three questions were developed: 1) What are the effects of simulated productive failure on physician assistant (PA) students' self-efficacy and emotional stress response when performing OSCE activities? 2) What are the effects of simulated productive failure on student competency when utilized in physician assistant (PA) OSCE activities? 3) How do physician assistant (PA) students reflect on learning from OSCE activities that utilize productive failure in the patient scenario?

Conclusion

Identifying the potential effects PF has on PA learners may provide evidence to support the utilization of this instructional method in PA education. As experiential learning pedagogies have enhanced the learner-centered nature of instruction, implementing PF may advance this instructional method to both meet the desired educational competencies and develop learners who are more adaptable to complex clinical problems. By developing this adaptability, the learner may possess a stronger sense of self-efficacy in conjunction with a high level of competency and resilience to endure the various struggles and setbacks of the clinical practice environment. Improving these components may enhance problem-solving skills and patient outcomes, and provide a more realistic expectation of provider satisfaction, potentially leading to decreased provider burnout.

Schön (1983) discussed the necessity for the professional to learn from experiences, highlighting how this learning impacts both cognitive and non-cognitive skills in professional practice. Understanding participants' perceptions related to learning from this instructional method may provide evidence to expand its utilization in PA curricula. Learning to reflect on experiences involving failure may provide learners with the critical skills required of professional healthcare providers.

Chapter 3: Methodology

Participants and Setting

This study took place during the clinical phase of training for physician assistant (PA) students at Oklahoma City University PA program (OCU PA program). It involved thirty-five OCU PA students who were asked to voluntarily participate in the study. These students were asked to participate in the study because they are in the clinical phase of their training and have engaged in numerous clinical simulation activities within the OCU PA program. There were no exclusionary factors for this study based on gender, ethnicity, or culture. Inclusion criteria for participation in this study included students enrolled in the OCU PA program within the clinical phase of training and between 20-59 years of age.

The study was conducted at the Oklahoma City University Meinders Simulation Center within the College of Health Professions. This simulation center is equipped with eight simulation suites utilizing high-fidelity mannequins for interactive simulation activities. The simulation suites allowed for complex medical simulations where participants could monitor, evaluate, and manage the clinical status, vital signs, and evaluation of a simulated clinical patient. The ability to simulate this interactive patient activity enabled participants to analyze and respond to the various changes encountered during the simulation, providing a complex problem-solving activity with changing variables within a safe learning space. Additionally, the simulation center contains multiple secured spaces for debriefing and feedback sessions, which allow for focus group participation.

Ethical Considerations

To ensure participant well-being and research integrity, several ethical considerations were addressed. Institutional Review Board (IRB) approval was obtained from both Oklahoma

City University (OCU) and the University of Oklahoma (OU), ensuring compliance with ethical guidelines for research involving human participants. The primary investigator (PI) conducted the study as a graduate student at OU. While the PI serves as the director of the OCU PA program, their role is indirect and does not involve direct supervision of didactic courses, clinical instruction, or course grading that could impact participants' academic progression.

Research facilitators, consisting of OCU PA program faculty, assisted in the study by indirectly assessing OSCE simulation performance in a manner consistent with standard curriculum practices. Participation in the study had no effect on students' academic grades or progression within the OCU PA program. All participants provided informed consent, ensuring they fully understood the study's purpose, procedures, potential risks, and benefits before participation (Beauchamp, 2019).

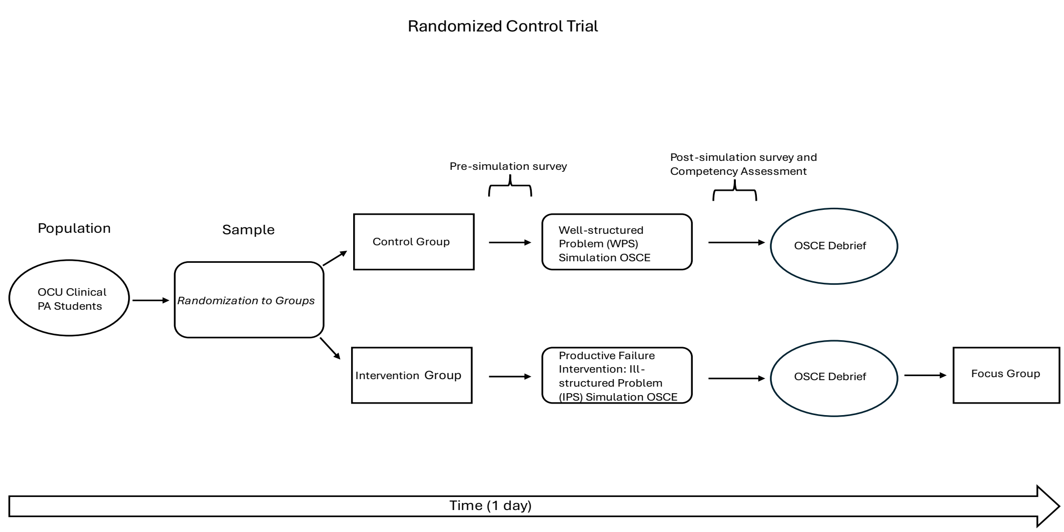
Research Design

The research employed a concurrent triangulation mixed-methods design incorporating both quantitative and qualitative approaches. The combination allowed for a comprehensive exploration of the impact of simulated productive failure on clinical PA students’ self-efficacy (confidence), emotional stress response, clinical competency in decision-making, and reflection on learning from this instructional method. In this mixed-methods design, the primary quantitative focus was to evaluate self-efficacy (confidence), emotional stress response, and clinical competency. The primary qualitative focus of the study was to evaluate the participants’ reflections on the learning process.

The study utilized a randomized control trial (RCT) design where participants were randomized into control and intervention groups using stratified randomization, with the completion of an emergency medicine rotation serving as the stratification factor. In the RCT, the productive failure experience was the sole intervention. Stratified randomization was employed to ensure that both groups had a balanced distribution of students with and without prior ER experience, minimizing potential bias and controlling for variations in clinical exposure (Kernan, 1999). Students were first categorized into two strata based on whether they had completed an ER rotation. Within each stratum, participants were then randomly assigned to either the control or intervention group using a computer-generated randomization process. This approach helped maintain comparability between groups, ensuring that differences in learning outcomes could be attributed to the intervention rather than disparities in prior clinical experience.

The use of RCTs in the experimental design served to reduce potential bias and minimize errors within the study design (Kendall, 2003). The primary goal of the study was to evaluate the potential effects of productive failure on clinical PA students when used as a method of instruction for OSCE simulation activities. This goal would be accomplished by examining participants' self-efficacy (confidence), emotional responses, competency, and perceptions of learning from PF. The research design is shown in Figure 2 below:

Figure 2: Randomized Control Trial Design



The qualitative portion of the study utilized phenomenological methods, including emergent focus group interviews and inductive thematic analysis, to answer the research questions. Emergent focus groups allowed for development and variability in the process, uncovering more effective meanings related to the study participants. As a theory already exists related to the productive failure phenomenon, this design was well-suited for the qualitative component of the study. Focus groups were conducted after the intervention and the completion of each of the other components in the study as a reflective process on the learning experience to discover meaning.

Rationale for Methodology

The goal of any empirical research is to provide new knowledge to the field of study. There is a lack of medical research related to productive failure (PF), especially in PA education. The concurrent triangulation mixed-methods design provided both quantitative and qualitative data, offering a holistic view of PF use in PA education. This mixed methods design is a research approach in which both qualitative and quantitative data are collected simultaneously but analyzed separately. The results are then compared and integrated to provide a more comprehensive understanding of the research problem. This design allows for validation of findings through cross-verification from different methods, enhancing the credibility and depth of the results. It is particularly useful when the researcher aims to balance the strengths and weaknesses of each method and gain complementary insights within a single phase of data collection. As the main research question in this study sought to evaluate the effects of PF on PA students during educational activities, this mixed-methods design allowed for the identification of these potential effects both quantitatively and qualitatively. This method provided a comprehensive understanding of the quantitative outcomes of PF, and the qualitative phase explored how and why participants benefited from PF by gathering insight into their experiences, thought processes, and engagement during failure and subsequent reflection. The different types of data provided the potential to reveal bidirectional effects, including potential positive and negative impacts of PF in PA education. Additionally, this mixed-methods design allowed for more contextualization of data, as participants' perceptions were open-ended in the qualitative component of the data. By combining these two forms of data, the qualitative data could better illustrate or illuminate the quantitative data, or vice versa, through the analytical process.

The quantitative component of the design served to emphasize the quantification of data using well-known statistical analysis. Through measurable data evaluation, trends and relationships between variables were discovered. These trends and relationships provided information related to the effect or impact of the intervention. This design method benefits from larger volumes of data and sample sizes, improving statistical power and allowing for generalizability to the greater population. The use of RCT in experimental design within this methodology allowed for the comparison of two groups and the evaluation of the potential causal relationship of the intervention.

The qualitative component of the design served to emphasize naturally occurring phenomena and interactive interpretation of the observed environment. This design analyzes data through interpretive processes to describe, decode, translate, and discover meaning. The basic features of qualitative design hold that knowledge is constructed through engagement with the realities of the world, and there is no absolute truth (Crotty, 2020). This study employed an instructional method based on interactive engagement with the realities of the experience, and the qualitative component of the design allowed for meaning and understanding from various points of view. Constructed meanings may differ significantly for each individual or group of individuals who participate. This method of design allowed for a more complete understanding of perspectives from multiple viewpoints without limiting the context of the phenomenon. This qualitative design typically involves smaller sample sizes and does not attempt to analyze relationships between variables or groups of data statistically. The outcome of this design was to discover meaning from data rather than generalizing outcomes to a larger population. It does provide potential transferability to similar groups in similar settings and environments. This design provided data through the participants' perceptions, constructed within the study.

The mixed-methods design offered significant benefits as it provides quantifiable objective data showing potential causation due to the intervention. Using RCT improved the data's reliability and generalizability of the intervention effect. Moreover, the qualitative data provided participant perspectives on the intervention and implications created from their reality. This mixed-methods design also allowed for the discovery of unanticipated influences throughout the process. The qualitative data may augment the quantitative data by providing meaning to the interactive social and environmental exposures in which the intervention occurred. As educational pedagogical learning research aims to improve the learning environment and process for practical use, combining these components through a mixed-methods design allowed for an improved evaluation of the intervention's effects compared to either design alone.

Productive failure was being considered in PA education to improve the learning process, environment, and outcomes. With this instructional method, potential reflection and transformation may occur for the participants within the learning process. Qualitative analysis provided a better opportunity to discover this data and to evaluate the potential impact on participants. Additionally, quantitative data may reveal statistical changes in learning related to the intervention compared to the control group. Identifying the potential effects of PF on PA students can assist in discovering the benefits or consequences of using this pedagogical method of instruction in PA and medical education.

Research Structure and Procedural Flow

Oklahoma City University clinical phase PA students meet on campus monthly for clinical call-back days (CBD) to participate in high-stakes assessments and curriculum delivery. After the conclusion of the academic curriculum on the CBD, the OCU PA clinical class was approached as a group to voluntarily participate in the educational research study. The duration, involvement, and voluntary nature of the study were discussed during the recruitment process. Before the beginning of the research, the OCU PA students were asked to voluntarily consent to participate in the study. After consent, a demographic paper survey was administered allowing for stratified randomization, with completion of an emergency medicine rotation as the stratification factor, to assign participants to 1) a control group that does not receive the PF experience in the OSCE simulation activity and 2) an intervention group that does receive the PF experience in the OSCE simulation activity. The study utilized single blinding, as participants in the study were blinded to group assignment. However, OSCE facilitators were aware of group assignments as knowledge of the OSCE simulation outcome related to productive failure was required for the simulation activity.

The intervention, PF, involves two design phases: Phase 1- generates and explores the constraints of the problem, and Phase 2- the consolidation phase, compares and contrasts failed solutions of the problem (Kapur, 2012). The sole intervention in the study consisted of an ill-structured problem and the outcome of the clinical patient scenario, representing productive failure. The control group's problem structure had limited, or constrained, degrees of freedom in the problem-solving activity, representing a well-structured problem (WSP). In the control group, medically appropriate treatments initiated by the participants resulted in clinical improvement of the patient's status and vital signs in the simulation. Spaced time intervals were utilized to progressively adjust the patient's vital signs dependent on the appropriate treatments initiated, or lack thereof. The lack of appropriate treatments resulted in the deterioration of the patient's clinical status, which progressed until the patient's demise.

The intervention group's problem structure utilized higher degrees of freedom and complexity, representing an ill-structured problem (ISP). For the intervention, medically appropriate treatments initiated by participants in this group resulted in the continued clinical decline of the patient's status and vital signs in the simulation. Identical to the control group, spaced time intervals were utilized to progressively deteriorate the patient's vital signs despite any appropriate treatments provided. This progression resulted in the patient's clinical status decline until the patient's demise. The OSCE simulation flow charts for the control and intervention groups are represented in Appendix 2 and Appendix 3, respectively.

Both the control and intervention groups participated in a pre-simulation survey to evaluate participants' self-efficacy (confidence) in performing OSCE simulation skills. This assessment was repeated for each group after the simulation activity had taken place to assess post-simulation participants' self-efficacy (confidence) in performing the OSCE simulation skills. For this study, the terms self-efficacy and confidence will be used interchangeably. However, self-efficacy refers to an individual's perception that they have the ability to successfully complete tasks (Bandura, 1977). This differs slightly from confidence, which refers to an individual's belief that they can complete tasks successfully (Merkel, 2006). Both concepts are subject to past experiences and the expectations related to the experience (Sanders, 2006).

Additionally, both the control and intervention groups participated in pre-simulation and post-simulation surveys to evaluate participant perceptions related to emotional stress responses.

Participant competency, or knowledge, during the OSCE activity was assessed by trained OSCE facilitators following standard procedures used in all previous OSCE simulation experiences at the OCU PA program. To ensure consistency and accuracy in competency assessment, OSCE facilitators underwent additional training specific to this study. The training consisted of two separate sessions, each lasting two hours, for a total of four hours. The first training session took place approximately eight weeks before the OSCE activities. This session provided an overview of the study's aims, research questions, procedural flow, and measurement tools, while also addressing any general questions from the OSCE facilitators. Additionally, an expert evaluator was assigned to ensure the accuracy and real-world applicability of the OSCE competency measure. This expert, a researcher with over 15 years of emergency medicine experience and a medical educator, reviewed the OSCE simulation case scenario to confirm its alignment with real-world complex clinical situations.

The second training session occurred approximately three weeks before the OSCE activities. During this session, the expert evaluator trained the OSCE facilitators on the proper use of the OSCE competency assessment tool and provided standardized guidelines to ensure consistent application during the simulation. The expert evaluator also served as a reference point for inter-rater reliability, both before and during the OSCE activity, ensuring that competency assessments remained objective and consistent across facilitators.

After completing the OSCE simulation activity, both the control and intervention groups participated in identical debriefing sessions designed to help them organize and understand their approaches to problem-solving during the simulation. Simulation-based education (SBE) debriefing is essential for enhancing learning outcomes, as it allows participants to reflect on their decisions and deepen their understanding (Abulebda, 2025). These sessions followed the Gather, Analyze, and Summarize (GAS) method, which uses a structured three-phase approach to the debriefing. The GAS method guides learners through describing their experiences, critically analyzing their actions, and summarizing key takeaways for future application (Sawyer, 2016).

For the intervention group, this debriefing session represented Phase 2 (Consolidation) of the Productive Failure (PF) process, where learners reflected on their problem-solving approaches and integrated new knowledge to reinforce learning. For the control group, the debriefing session served as a standard review process aimed at consolidating key target concepts introduced during the activity.

After the completion of all OSCE activities, only the intervention group participated in a focus group discussion, facilitated by the study investigators. The purpose of this session was to explore how participants reflected on their learning experience in OSCE activities that incorporated Productive Failure (PF) within the patient simulation.

Quantitative Data

Demographic Survey

Before the OSCE simulation activity, each of the study participants was asked to complete a pre-simulation survey that included a 5-item paper survey component about participants' demographics. This survey contained the following questions: 1) Which of the following categories best describes your age? 2) Which of the following best describes your race/ethnicity? 3)Which of the following gender categories best describes you? 4) Which of the following best describes your marital status? 5) Which of the following clinical rotations have you completed? (Circle all that apply.) This survey data was utilized to further describe participants within the study.

Self-efficacy Measure

Measurement of pre-simulation and post-simulation self-efficacy was used for both the control and intervention groups. The pre-simulation survey was conducted immediately before the OSCE simulation activity. This questionnaire was repeated a few minutes after the patient simulation interaction, before the debriefing session, and was represented as post-simulation self-efficacy. The Generalized Self-Efficacy Scale (GSES) was administered to assess the self-efficacy of the participants (Schwarzer, 2010). This is a 10-item survey that evaluates self-efficacy by evaluating the degree or strength of the participant's belief in their ability to respond to novel or difficult situations and to manage obstacles and setbacks. This survey asked participants to answer the degree to which they: 1) Can always manage to solve difficult problems, 2) When opposed, can find means and ways to get what they want, 3) Can stick to aims and accomplish goals, 4) Could deal efficiently with unexpected events, 5) Knows how to handle unforeseen situations, 6) Can solve most problems if invest the necessary effort, 7) Can remain calm when facing difficulties and can rely on coping abilities, 8) Can usually find several solutions to a problem, 9) Can think of something to do when in a bind, 10) Can handle what comes their way. The survey utilizes a 4-point Likert scale to evaluate participant responses for each of the questions, where responses for each item range from 1= Not at all true, and 4= Exactly true. The range for this instrument is 10 to 40, where higher scores represent higher levels of self-efficacy. The pre-simulation self-efficacy survey is represented in Appendix 4.

The GSES has demonstrated strong internal consistency, with Cronbach's alpha values ranging from 0.76 to 0.90 across different populations, indicating high internal consistency

(Schwarzer, 2010). The scale's construct validity correlates positively with measures of optimism, positive coping strategies, and well-being and negatively correlates with measures of anxiety, depression, and stress (Luszczynska, 2005). Criterion validity shows that higher self-efficacy scores are associated with better academic performance, health behaviors, and job-related success (Scholz, 2002).

Emotional Stress Response Measure

Measurement of pre-simulation and post-simulation emotional stress state was used for both the control and intervention groups. The pre-simulation survey was administered immediately before OSCE activities. The Short Stress State Questionnaire (SSSQ) was utilized to evaluate perceived stress and emotions experienced before the OSCE simulation activity (Helton, 2004). This questionnaire was repeated a few minutes after the patient simulation interaction, before the debriefing session, and was represented as the post-simulation emotional stress response. This 24-item assessment evaluates stress state and emotional responses to tasks in the domain of task engagement, distress, and worry at the present moment. The questionnaire utilizes a 5-point Likert scale where responses to each item range from 1= Not at all, to 5= Extremely. The range for this instrument is 24 to 120, where higher scores represent higher levels of emotional stress state. The pre-simulation self-efficacy survey is represented in Appendix 5.

This measurement tool was developed from the Dundee Stress State Questionnaire (DSSQ) used to evaluate subjective state response in stressful environments (Mathews, 1999, 2002). The SSSQ provides a transactional approach to the evaluation of stress in relation to the participants and the environment in which the experience takes place. This multidimensional approach to the measurement of stress may yield more robust data and conceptually aligns with the theories of experiential learning and the learning environment.

The SSSQ measures 3 dimensions of stress, including task engagement, distress, and worry (Helton, 2004). The scale has demonstrated satisfactory internal consistency across its subscales with Cronbach's alpha typically exceeding 0.70, indicating acceptable reliability for each dimension (Helton, 2004). Confirmatory factor analyses have supported the three-factor structure of the SSSQ, showing its construct validity mirroring the original Dundee Short Stress State Questionnaire (Helton, 2004).

Competency Measure

The measure of participant competency, or knowledge, was assessed during the OSCE simulation activity. Trained OSCE facilitators observed and evaluated the participants' abilities to evaluate and problem-solve during the patient simulation activity. The Global Rating Scale (GRS), complemented by the Task Specific Checklist (TSC), was used to evaluate participants' competency within the OSCE simulation activity.

The GRS is a 10-item assessment tool scored by OSCE facilitators using a 5-point scale. Participant scores for each item range from 1= Poor, to 5= Excellent. The GRS measures the participant's structured diagnostic approach, guideline-conforming therapies, application of therapies, overall patient safety, and overall participant performance in an emergency care patient setting. Scores range from 10 to 50 points, with the higher score representing higher competency in the skill.

The TSC complements the GRS and is used in non-traumatic emergencies. The TSC is a 25-item assessment tool that was scored by OSCE facilitators with scores that range from 0= Not done/Incorrect, to 1= Done correctly. This task checklist measures the participant's evaluation of patient safety, airway, breathing, circulation, disability, exposure, working diagnosis, temperature management, patient position optimization, treatment, and hospital referral. Scores for this measure range from 0 to 25, with the higher score representing higher competency in the skill. The post-simulation OSCE competency survey is represented in Appendix 6.

The construct validity of the novel combined GRS/TSC from Zoller (2021) has shown that the GRS significantly discriminates between the different levels of training when comparing novice to expert simulation participants, p-value of 0.005. The TSC component neared but did not meet significance with a p-value of 0.052. The GRS inter-rater reliability was shown to have a Kendall's coefficient of concordance of 0.88 and the TSC a coefficient of 0.80, both indicating a strong level of agreement (Zoller, 2021).

Individually, the GRS has shown high inter-rater reliability and improves with the use of structured rating criteria among evaluators (Regehr, 1998). It demonstrates strong construct validity as it correlates significantly with other measures of clinical competency, including standardized patient assessment and faculty evaluations (Regehr, 1998; Hodges, 1999). Compared to checklist-based assessments alone, the GRS is considered more sensitive to different levels of expertise as it allows evaluators to assess overall clinical reasoning (Hodges, 1999).

TSCs have shown high inter-rater reliability because they involve objective criteria that are either met or not met, providing higher agreement among raters (Regehr, 1998). Additionally, TSCs exhibit strong content validity as they are developed based on expert consensus and procedural guidelines (Ma, 2012).

Qualitative Data

Focus Groups

Focus groups are a qualitative research method that brings together a small group of participants to discuss a specific topic under the guidance of a skilled facilitator. This interactive format encourages participants to share personal experiences, beliefs, and opinions often yielding deeper insights than individual interviews or surveys (Krueger, 2015). In health professions education, focus groups are frequently used to explore learners' perceptions, uncover underlying motivations, and evaluate interventions in educational settings.

In the context of medical simulations and OSCEs, focus groups have been increasingly employed to assess learners' responses to simulation-based training, evaluate the realism and instructional effectiveness of scenarios, and explore emotional or cognitive reactions to high-stakes environments (Nestel, 2007; Steinert, 2014). These discussions can reveal nuanced perspectives on self-efficacy, stress, or perceptions of failure, dimensions that are often difficult to capture through quantitative methods alone.

The benefits of using focus groups in medical education include the generation of rich, context-specific data that can guide curriculum design and inform improvements in teaching strategies. Focus groups also enable triangulation with quantitative data, enhancing the overall validity and depth of findings (Barbour, 2005). Additionally, the group dynamic can prompt participants to articulate insights they might not have shared individually, fostering more reflective and layered responses.

However, focus groups are not without limitations. The presence of peers and facilitators can introduce social desirability bias, where participants may alter their responses to align with group norms or perceived expectations (Sim, 2019). Dominant voices may influence the discussion, potentially silencing quieter participants and skewing the results. Additionally, the logistical challenge of scheduling sessions with busy medical trainees may result in sampling bias, as only certain students, often those with fewer academic pressures, may opt to participate (Wilkinson, 1998).

Despite these challenges, focus groups remain a powerful tool in medical education research. When implemented thoughtfully, with attention to group composition and moderation techniques, they can provide valuable insights into learner experience and guide improvements in simulation-based learning and assessment practices.

In this study, after the conclusion of all OSCE activities and debriefing sessions, only the intervention group who participated in the ill-structured problem-solving of productive failure engaged in focus group discussions. Study investigators, who had previous training in qualitative research and focus group discussions, facilitated the in-person focus group discussion using semi-structured interviews to discover how participants reflect on learning from OSCE activities that utilize productive failure in the patient simulation. These focus group sessions lasted for a maximum duration of 60 minutes, allowing the participants to openly reflect on and discuss the OSCE simulation activity. The emergent nature of the focus group provided variability in the structure of the interviews to best facilitate the discovery of meaning through the reflective action of the participants. This group session was audio-recorded for transcribing purposes. The interview session focused on the following questions: 1) What challenges did you experience in the simulation activity? 2) How did today's simulation activity compare to other teaching methods previously used in the PA OSCE activities? 3) Reflecting on your participation in the productive failure simulation, how do you feel it influenced your learning and development as a PA student? By answering these questions, researchers hoped to better understand how PA students respond to simulation learning activities that utilize productive failure and the potential effect that these learning experiences have on their learning process.

Plan for Data Analysis

Quantitative Analysis

The primary aim of this study was to evaluate and understand the impact of productive failure when utilized in PA education. The quantitative data collected served to identify the effects of productive failure on PA students when utilized in OSCE simulation activities.

Analysis Strategies for Research Question One. To answer research question one, the analysis of data pertaining to participants' self-efficacy and emotional stress responses was conducted. The reliability, or internal consistency, of pre- and post-simulation self-efficacy and pre- and post-simulation emotional stress responses were evaluated by performing the Cronbach alpha coefficient. Descriptive analysis of the central tendency and distribution of scores for both pre- and post-simulation self-efficacy and emotional stress responses were performed, and results were displayed with pictographic representations and the evaluation of normality. The Shapiro-Wilk test evaluates if the continuous variable follows a normal distribution, as this test is best utilized for samples that have fewer than 50 observations. This measure of normality evaluates the skewness as the degree of asymmetry in the distribution, and kurtosis as the degree of "peakedness" of the distribution (DeCarlo, 1997).

An independent samples t-test analyzed pre- and post-simulation self-efficacy and emotional stress response scores for each of the groups (control and intervention) to determine if there was a significant difference in the means of the independent groups (Lomax, 2012). In this analysis, the independent variable (IV) was the group assignment, and the dependent variable (DV) was the study variable, self-efficacy, and emotional stress response. Additionally, this evaluation included Levene's test for homogeneity of variances, which assessed whether the variance of a DV is equal across different groups.

A paired t-test was also conducted to compare pre- and post-simulation self-efficacy scores within both the intervention and control groups to determine whether there was a statistically significant change in participants' self-efficacy over time. This test examined whether the mean difference between the pre- and post-simulation scores were significantly different from zero, indicating a change due to the simulation. Similarly, a paired t-test analysis was performed on participants' emotional stress response scores to assess whether their emotional responses significantly changed from pre- to post-simulation within each group. Cohen's d-test was performed as part of the paired t-tests to evaluate the magnitude of difference between the pre-simulation and the post-simulation scores within the same group for self-efficacy and emotional stress response.

Pearson's correlation coefficient test was used to analyze the correlation between learner OSCE competency with both pre-simulation and post-simulation self-efficacy and emotional stress response scores. This analysis was performed on all study variables to include self-efficacy, emotional stress response, and competency scores. Although Pearson's correlation does not provide causation, it does indicate the degree (or strength) and direction of the relationship between the variables (Pallant, 2020). By evaluating the correlation between participant competency, self-efficacy, and emotional stress response scores in both the control and intervention groups, this analysis provided data to identify the potential effects of PF on PA participants when performing OSCE simulation activities.

Analysis Strategies for Research Question Two. To answer research question two, data pertaining to participants' competency were analyzed. The score distributions and descriptive statistics were evaluated for normality using the Shapiro-Wilks test, and the results were displayed with pictographic representations to evaluate normality in the distribution of scores.

The comparison of means between the control and intervention groups was evaluated using an independent samples t-test. This test analyzed participant OSCE competency scores for each of the groups, where the independent variable (IV) was the group assignment, and the dependent variable (DV) was the OSCE competency score. The Levene's test was also performed for homogeneity of variances in this analysis.

Linear regression was used to evaluate both the control and intervention groups' predictive relationship between one or more independent variables and the single continuous variable (Lomax, 2012). In this analysis, OSCE competency scores were considered the DV, and pre- and post-simulation self-efficacy and pre- and post-simulation emotional stress responses were considered the IV. By evaluating this predictive relationship between variables, the analysis provided data to identify the potential effects of PF on PA participants' OSCE competency scores during OSCE simulation activities.

Finally, multiple regression was used to analyze potential demographic predictors for participant OSCE scores. In this evaluation, OSCE competency scores were considered the DV, and demographic characteristics were considered the IV.

Pre-participation competency was not assessed to prevent informing participants of the specific nature of the OSCE activities, which would diminish the potential effects of the intervention. If pre-participation competency were to be assessed, the most appropriate measure would likely consist of one or more of the Physician Assistant Education Association (PAEA) End of Rotation examinations (EOR). At the time of this study, not all students within the cohort or the study participants had completed these examinations.

Qualitative Analysis

Analysis Strategies for Research Question Three. To answer research question three, data was collected, recorded, and transcribed from the semi-structured focus group interview sessions. Thematic analysis was performed on the collected data. Initially, line-by-line coding of the transcript was performed to reduce the data to more manageable codes, which were then organized into categories and themes (Merriam, 2016). An inductive process of coding was used to identify both the manifested and latent meanings in this axial coding. Themes were derived from these meanings based on the comparison of multiple codes to the interpreted meanings.

The trustworthiness of the data and interpretations was enhanced through triangulation techniques in the development of overarching themes. Three research investigators participated in the triangulation of the data to develop overarching themes. In the triangulation process, data and themes were discussed and agreed upon for verification of the thematic development. The interpretive meaning derived from the data and themes served to answer how PA students reflect on learning from OSCE activities that utilize PF in the patient simulation. By answering this question, the study's secondary aim of understanding if PA students identify the benefits of PF in PA education may be realized. The insights gained through this process may be transferable to other similar PA training programs and provide evidence for pedagogical change in PA education.

Chapter 4: Results and Analysis

Introduction

This chapter presents the results and analysis of the mixed-methods study investigating the impact of a PF intervention introduced through OSCE simulation on PA students. The study's primary aim is to evaluate and understand the impact of productive failure when implemented in PA education. The secondary aim of the study is to discover if participants identify any potential benefits of PF when utilized in PA education. Quantitative data gathered from pre-intervention and post-intervention assessments include the General Self-Efficacy Scale (GES), Short Stress State Questionnaire (SSSQ), and OSCE evaluation scores. These assessments are presented alongside qualitative insights derived from participant reflections on learning from OSCE activities that utilize productive failure. By leveraging the OSCE format, the gold standard for assessing clinical skills, the study sought to replicate a high-stakes environment where productive failure could foster resilience, adaptability to various complex clinical problems, clinical decision-making skills, confidence, and deeper understanding through reflection on the experience (Harden, 1975; Kapur, 2008).

This chapter is structured as follows: First, the demographic profile of the participants is described to contextualize the findings, and the comparison of group participants are analyzed. Followed by the evaluation of the reliability and internal consistency of the measures utilized in the study. Next, the descriptive results of the pre-simulation and post-simulation assessments are presented, beginning with self-efficacy, followed by emotional stress response, and the OSCE competency scores. This chapter is then subdivided into sections that present and analyze data corresponding to each research question: 1) What are the effects of simulated productive failure on physician assistant (PA) students' self-efficacy and emotional stress response when performing OSCE activities? and 2) What are the effects of simulated productive failure on student competency when utilized in physician assistant OSCE activities? Statistical analyses to answer these questions included paired t-tests, independent samples t-tests, Pearson's correlation, linear regression, and multiple regression to evaluate the effects of the intervention. Research question three corresponds to the qualitative data of the study: 3) How do physician assistant students reflect on learning from OSCE activities that utilize productive failure in the patient scenario? This final section analyzed the qualitative themes derived from reflective narratives and integrated them with the quantitative findings to provide a comprehensive understanding of the intervention's impact on PA students.

Through the data analysis, this chapter aims to highlight how the intervention of productive failure utilized in OSCE simulation activities impacts study participants and can bridge the gap between theoretical knowledge and practical application, ultimately shaping competent, confident, and resilient healthcare professionals.

Demographics

The demographic information of this study is a fundamental component of the study design and provides insight into the characteristics of the study population, enables subgroup analysis, and assists in supporting the generalizability of findings to the broader population. These demographic characteristics are crucial for determining whether the study sample represents the target population and identifies potential biases that could arise from unbalanced representation (Creswell, 2018; Polit, 2021). This demographics section includes sample size, age range, gender distribution, marital status, number of completed core clinical rotations, completion of emergency medicine rotation, location, and setting.

Sample size indicates the number of participants included in the study, providing crucial information about the reliability and validity of the study findings. A sufficiently large sample size enhances reliability by reducing random error and increasing consistency across repeated measures, while also supporting validity by ensuring that the results more accurately reflect the true effects within the target population, thereby improving generalizability. Sample size impacts the statistical power of the study to detect meaningful differences or relationships. Insufficient sample size may increase the likelihood of type II errors, while excessive sample size may lead to wasteful resources (Creswell, 2018).

Information on age range, gender distribution, and marital status helps to contextualize participants' cognitive development, prior experiences, communication styles, responses to stress, and self-efficacy. As an example, younger participants may exhibit different stress responses or learning styles than older individuals (Field, 2018). In addition, married individuals might face additional life responsibilities compared to single participants, impacting their engagement with the intervention (Kapur, 2008). These characteristics provide crucial information about the sample group of the study and its potential representation of the larger population.

Accreditation standards for PA education require all students to complete seven core Supervised Clinical Practice Experiences (SCPEs), commonly referred to as clinical rotations, in the following specialties: family medicine, emergency medicine, internal medicine, surgery, pediatrics, women's health, and behavioral health. In this study, documenting the number of completed core clinical rotations, particularly whether participants have completed an emergency medicine rotation, provided critical context for assessing their prior medical experience and clinical knowledge. This information helps to ensure comparability between sample groups, reducing potential confounding variables.

Prior clinical experience, especially a completed emergency medicine rotation, may influence performance in OSCE simulation activities, as emergency settings often involve high-pressure, complex decision-making scenarios. Additionally, exposure to emergency medicine may impact participants' self-efficacy, influencing their confidence and approach to problem-solving during simulations.

By contextualizing participants' clinical backgrounds, the study can more accurately determine whether differences in outcomes result from the PF intervention rather than pre-existing variations in clinical experience. This strengthens the validity and reliability of the study's findings.

The location and setting of the study describe where and under what conditions the research takes place. This can include the geographic location, country, and the specific institution and type of institution where the study was conducted. The setting provides environmental conditions that can include in-person, physical location, and characteristics of the environment that may impact learning and performance. Each of these provides essential context for understanding, interpreting, and replicating research findings. Describing these elements helps ensure transparency, reliability, and applicability of the study results to other educational or clinical environments, improving generalizability or transferability.

Table 1: Study Demographics

Study Demographics	
Demographic	**Values**
Age Range	21-29
Gender Distribution	Female: 5 (62.5%), Male: 3 (37.5%)
Race/Ethnicity	White: 6 (75%), African American: 1 (12.5%), Asian: 1 (12.5%)
Marital Status	Single: 6 (75%), Married: 2 (25%)

Geographic Location	Oklahoma City, Oklahoma
Institution	Oklahoma City University
College	College of Health Professions
Program	Physician Assistant Program
Environmental Setting	Meinder's Simulation Center
Academic Year	2024
Participant Year in Cohort	2nd year
Academic Phase	Clinical
Core Clinical Rotations Completed (of 7 total)	Four Rotations: 5 (62.5%), Five Rotations: 3 (37.5%)
Completed Emergency Medicine Rotation	Yes: 6 (75%), No: 2 (25%)
Cohort Size	35
Sample Size	8
Non-Adherent Participant Withdrawals	2
Control Group	4
Intervention Group	4

Setting and Participants

This study was conducted at the Oklahoma City University (OCU) campus in Oklahoma City, Oklahoma, at Meinders Simulation Center within the College of Health Professions, during the academic year of 2024. It took place during the clinical phase and within the second year of the OCU PA class of 2025's training. The initial participant pool consisted of 35 second-year PA students. From this, 10 students provided informed consent to voluntarily participate in the study. The ten participants were randomly assigned to study groups using stratified randomization. Initially, the participants were categorized into two strata based on their prior completion of an emergency medicine (ER) rotation. Each stratum was randomized to either the control or intervention group. Eight participants completed all components of the study protocol, yielding a final sample size of 8 participants. Two participants voluntarily withdrew from the study after initial randomization but before any additional study activities were completed.

The final study sample contained 4 participants in the control group and 4 participants in the intervention group. Both the control group and the intervention group each contained 3 participants who had previously completed an ER rotation. Participants who completed the study present a demographic profile with an age range of 21-29 years old and a gender distribution of 62.5% female and 37.5% male.

Demographic Comparison of Groups

Ensuring demographic equivalence between the control and intervention groups is crucial for reducing bias and strengthening the study's internal validity. By verifying that both groups have similar characteristics at the start of the study, we can ensure that the observed differences in outcomes are due to the intervention rather than pre-existing factors. Additionally, understanding the demographic composition improves the generalizability of the study's findings. The Fisher Exact test was utilized to compare the control and intervention group participants. This statistical method was utilized to determine if there are significant associations between two categorical variables, particularly when sample sizes are small (Kim, 2017). It is an alternative to the chi-squared test of Independence when group sample sizes are less than five. The group comparison using the Fisher Exact test is shown in Table 2 below.

Table 2: Group Comparison

Demographic Group Comparison			
Demographic	**Control Group**	**Intervention Group**	**p-value**
Age Distribution	21-24: 1 (25%), 25-29: 3 (75%)	21-24: 3 (75%), 25-29: 1 (25%)	0.486
Gender Distribution	Female: 3 (75%), Male: 1 (25%)	Female: 2 (50%), Male: 2 (50%)	1.00
Race/Ethnicity	White: 4 (75%), Non-white 1 (25%)	White: 4 (75%), Non-white 1 (25%)	1.00
Marital Status	Single: 3 (75%), Married: 1 (25%)	Single: 3 (75%), Married: 1 (25%)	1.00
Core Clinical Rotations Completed (of 7 total)	Four Rotations: 3 (75%), Five Rotations: 1 (25%)	Four Rotations: 2 (50%), Five Rotations: 2 (50%)	1.00
Completion of Emergency Medicine Rotation	Yes: 3 (75%), No: 1 (25%)	Yes: 3 (75%), No: 1 (25%)	1.00

The p-values from the group comparison are all greater than 0.05, indicating there is no significant difference between the control and intervention groups across all demographic characteristics. This ensures that both groups are comparable at baseline, suggesting any observed differences in outcomes can be more confidently attributed to the intervention rather than pre-existing demographic differences.

Reliability and Internal Consistency of Measures

Self-Efficacy and Emotional Stress Response Reliability

In research, ensuring the reliability and internal consistency of measures is critical to producing valid and trustworthy results. Reliability refers to the extent to which a measurement instrument produces consistent results over time and across different conditions. If a measure is not reliable, the conclusions drawn from a study may be unstable, leading to weak or non-replicable outcomes (Tavakol, 2011). Internal consistency, a specific form of reliability, assesses whether multiple items within a test or survey are measuring the same construct and yielding similar results.

One of the most common ways to assess internal consistency is through Cronbach's alpha, which measures the average correlation between items within a scale. A high Cronbach's alpha, typically above 0.70, suggests that the items work well together in measuring a single construct, while a low alpha may indicate that some items are unrelated or redundant (Cronbach, 1951). Ensuring internal consistency is especially important in surveys, psychological scales, and educational assessments where multiple questions contribute to an overall score.

Ensuring reliability and internal consistency in research maintains the accuracy and trustworthiness of data collection instruments. Without reliable measures, research findings may be inconsistent and difficult to generalize across populations or contexts. By conducting reliability analyses, the researcher can refine measurement tools, eliminate weak items, and improve overall study quality, leading to more replicable and meaningful conclusions (DeVellis, 2022). The Reliability and Internal Consistency of Pre-simulation and Post-simulation Self-Efficacy and Emotional Stress Response are shown in Table 3 below.

Table 3: Cronbach's Alpha for Pre-simulation and Post-simulation Measures

Reliability			
Variable	**Cronbach Alpha**	**Mean of Scale**	**N of Items**
Self-Efficacy	0.719	27.313	20
Emotional Stress Response	0.909	66.875	48

Assessing the reliability of self-efficacy and emotional stress response measures before and after the simulation is a crucial step in the process of ensuring the consistency and validity of the data collected. Reliability and internal consistency for the pre-simulation and post-simulation self-efficacy measure demonstrated a Cronbach's alpha of 0.719, with a mean scale score of 27.313 across twenty items. This suggests a moderate to strong level of internal consistency, meaning that the items used to assess self-efficacy reliably measure the same underlying construct (Cronbach, 1951; Tavakol, 2011). Similarly, the pre-simulation and post-simulation emotional stress response measure yielded a Cronbach's alpha of 0.909, with a mean score of 66.875 across forty-eight items. While significantly higher than self-efficacy, this value falls within a strong level of internal consistency, indicating the measure is reliable for evaluating emotional stress responses.

These findings suggest that both self-efficacy and emotional stress response measures are consistent and reproducible before and after the simulation activities. The reliability scores support the validity of these instruments in capturing participants' perceived self-efficacy and emotional reactions in simulated environments.

OSCE Evaluation Reliability

Evaluating the reliability of the OSCE measure is similarly important in maintaining internal consistency and accuracy of the data collected. The OSCE evaluation revealed a mean scale score of 47.875 across thirty-five items. The number of items (N = 35) suggests a comprehensive assessment covering multiple clinical competencies, which can contribute to strong internal consistency. The reliability and internal consistency of the OSCE evaluation measure demonstrated a Cronbach's alpha of 0.900, suggesting a strong level of internal consistency, meaning that the items within the scale measure the same underlying construct The Reliability and Internal Consistency of Observed Structured Clinical Experience (OSCE) is shown in Table 4 below.

Table 4: Cronbach's Alpha for OSCE Measure

Reliability			
Variable	**Cronbach Alpha**	**Mean of Scale**	**N of Items**
OSCE Evaluation	0.900	47.875	35

Quantitative Data and Analysis of Study Variables

The descriptive analysis of study variables is an essential step in evaluating the distribution and characteristics of each variable within a given sample. This step provides fundamental insight into the data, ensuring that appropriate statistical methods are applied in further analyses. Key descriptive statistics such as measures of central tendency, dispersion, and normality assessments help determine the structure and reliability of the scale in a given population.

Central tendency metrics summarize the dataset by identifying a representative value. The mean indicates the average score, while the median provides the middle value, which is particularly useful in skewed distributions. The standard deviation and variance offer insight into score variability, helping assess whether score levels are homogeneously or widely distributed among respondents.

Examining normality is crucial, as many statistical tests assume a normal distribution of the dependent variable. A pictographic representation of normality, such as histograms, Q-Q plots, and box plots, visually depicts the data's distribution pattern, helping to identify skewness, kurtosis, or outliers that may affect statistical conclusions. These visual tools provide an intuitive way to determine whether a data set follows a normal bell-shaped curve or deviates from normality.

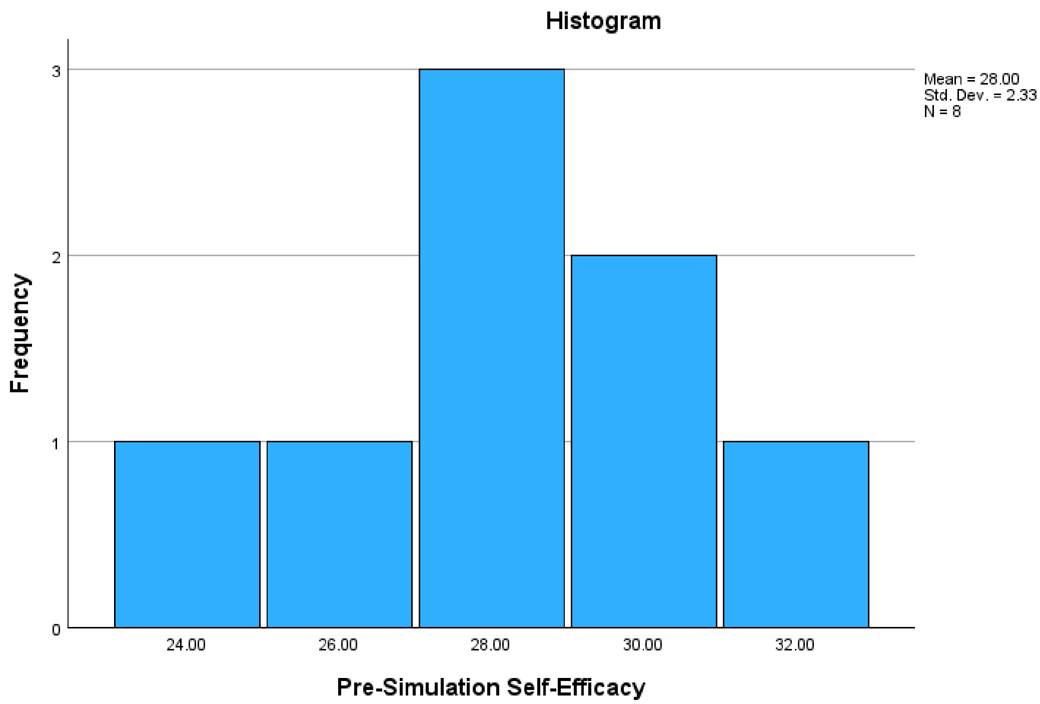
To statistically assess normality, the Shapiro-Wilk test is commonly employed, as it evaluates whether the sample distribution significantly deviates from a normal distribution. If the p-value of the Shapiro-Wilk test is below 0.05, it suggests a significant departure from normality, requiring the use of non-parametric tests for further analysis (Shapiro, 1965). Combining both the visual and statistical techniques to evaluate normality ensures a robust understanding of the data, guiding appropriate methodological choices in subsequent inferential analyses.

Descriptive Analysis of Generalized Self-Efficacy Scale (GES): Self-Efficacy

Pre-Simulation Self-Efficacy. A descriptive analysis of self-efficacy was performed to evaluate the distribution of pre-simulation and post-simulation scores. The mean value for pre-simulation self-efficacy scores was 28.0, with a median value of 28.0, and a 95% confidence interval for the mean of 26.052 to 29.948 representing the central tendency for these results. The variability for these results included a standard deviation of 2.330, a range of 8.0, with a minimum score of 24.0, a maximum score of 32.0, and a variance of 5.429. Distribution of the scores revealed skewness of -0.09 with a kurtosis value of 1.266. The normality of the distribution was evaluated using Shapiro-Wilk and Kolmogorov-Smirnov testing, with a graphical representation using a histogram. The p-value for the Shapiro-Wilk test was greater than 0.05, suggesting the data was normally distributed and is best used for less than 50 observations. The Distribution of Pre-Simulation Self-Efficacy Data is shown in Table 5 below.

Table 5: Pre-Simulation Self-Efficacy Data

Distribution of Pre-Simulation Self-Efficacy Data			
Test	**Statistic**	**Degrees of Freedom**	**p-value**
Kolmogorov-Smirnov	0.250	8	0.150
Shapiro-Wilk	0.936	8	0.574

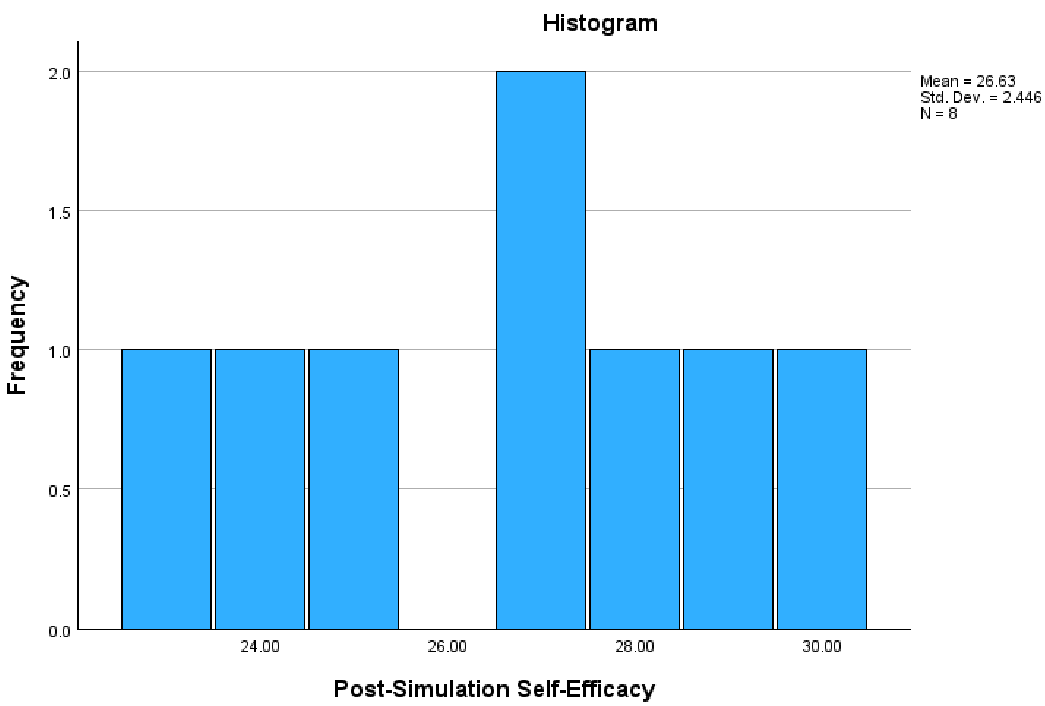


Post-Simulation Self-Efficacy. The Post-simulation self-efficacy scores revealed a mean value of 26.625, a median value of 27.0, and a 95% confidence interval for the mean of 24.580 to 28.670, representing the central tendency for these results. Variability for these results included a standard deviation of 2.446, a range of 7.0 with a minimum score of 23.0, a maximum score of 30.0, and a variance of 5.982. Distribution of the scores revealed skewness of -0.201 with a kurtosis value of -1.141. The normality of the distribution was evaluated using Shapiro-Wilk and

Kolmogorov-Smirnov testing, with a graphical representation using a histogram. The p-value for the Shapiro-Wilk test was greater than 0.05, suggesting the data was normally distributed and is best used for less than 50 observations. The Distribution of Post-Simulation Self-Efficacy Data is shown in Table 6 below.

Table 6: Post-Simulation Self-Efficacy Data

Distribution of Pre-Simulation Self-Efficacy Data			
Test	**Statistic**	**Degrees of Freedom**	**p-value**
Kolmogorov-Smirnov	0.186	8	0.816
Shapiro-Wilk	0.188	8	0.506

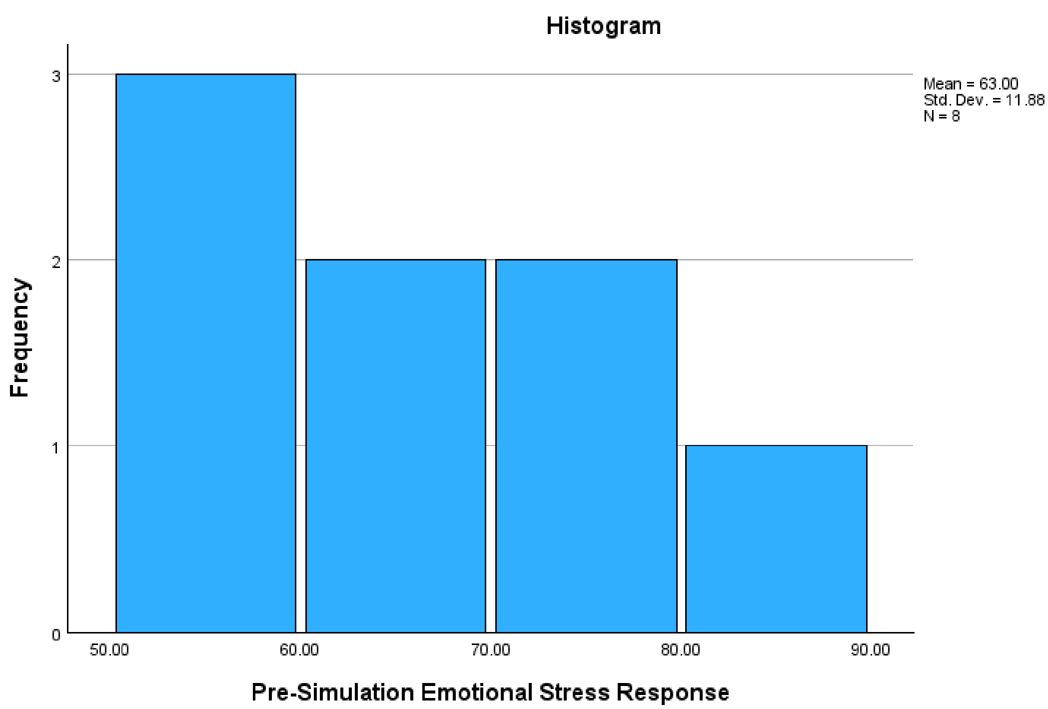


Descriptive Analysis of Short Stress State Questionnaire (SSSQ): Emotional Stress Response

Pre-Simulation Emotional Stress Response. The mean value for pre-simulation emotional stress response scores was 63.0, with a median value of 61.0, and a 95% confidence interval for the mean of 53.068 to 72.932 representing the central tendency for these results. Variability for these results included a standard deviation of 11.880, a range of 32.0, with a minimum score of 50.0, a maximum score of 82.0, and a variance of 141.143. Distribution of the scores revealed skewness of 0.423 with a kurtosis value of -1.216. The normality of the distribution was evaluated using Shapiro-Wilk and Kolmogorov-Smirnov testing, with a graphical representation using a histogram. The p-value for the Shapiro-Wilk test was greater than 0.05, suggesting the data was normally distributed and is best used for less than 50 observations. The Distribution of Pre-Simulation Emotional Stress Response Data is shown in Table 7 below.

Table 7: Pre-Simulation Emotional Stress Response Data

Distribution of Pre-Simulation Emotional Stress Response Data			
Test	**Statistic**	**Degrees of Freedom**	**p-value**
Kolmogorov-Smirnov	0.192	8	0.200
Shapiro-Wilk	0.909	8	0.374

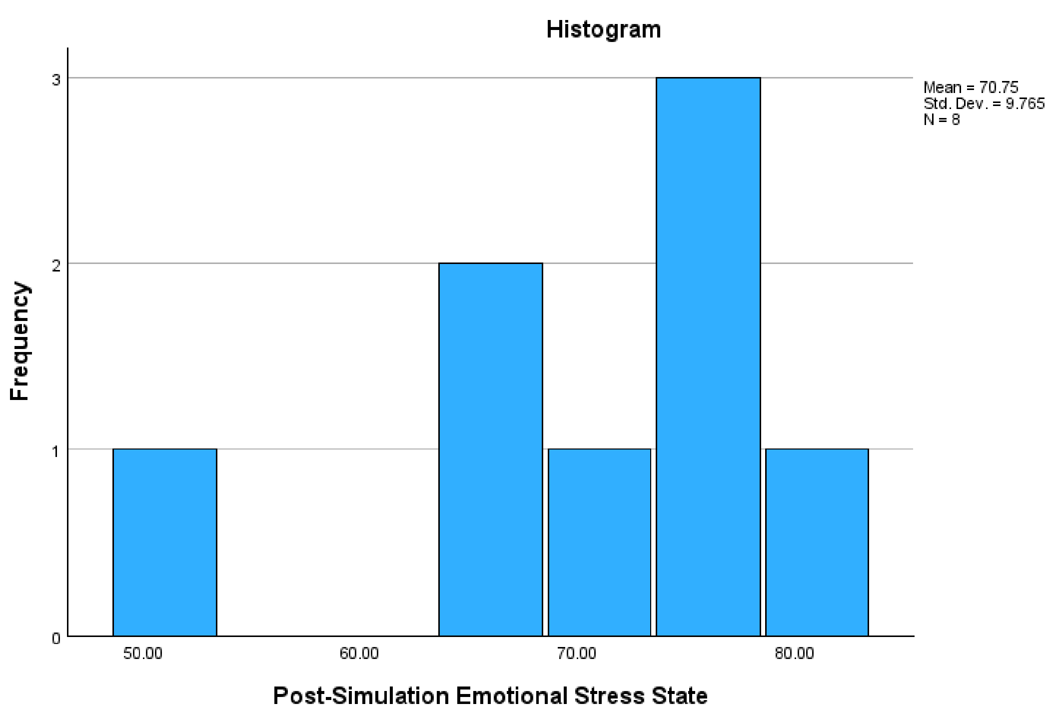


Post-Simulation Emotional Stress Response. The mean value for post-simulation emotional stress response scores was 70.750, with a median value of 72.50, and a 95%

confidence interval for the mean of 62.586 to 78.913 representing the central tendency for these results. Variability in these results included a standard deviation of 9.765, a range of 32.0, with a minimum score of 51.0, a maximum score of 83.0, and a variance of 95.357. Distribution of the scores revealed skewness of -1.118 with a kurtosis value of 1.850. The normality of the distribution was evaluated using Shapiro-Wilk and Kolmogorov-Smirnov testing, with graphical representation using a histogram. The p-value for the Shapiro-Wilk test was greater than 0.05, suggesting the data was normally distributed and is best used for less than 50 observations. The Distribution of Pre-Simulation Emotional Stress Response Data is shown in Table 7 below.

Table 7: Pre-Simulation Emotional Stress Response Data

Distribution of Pre-Simulation Emotional Stress Response Data			
Test	**Statistic**	**Degrees of Freedom**	**p-value**
Kolmogorov-Smirnov	0.188	8	0.200
Shapiro-Wilk	0.929	8	0.508

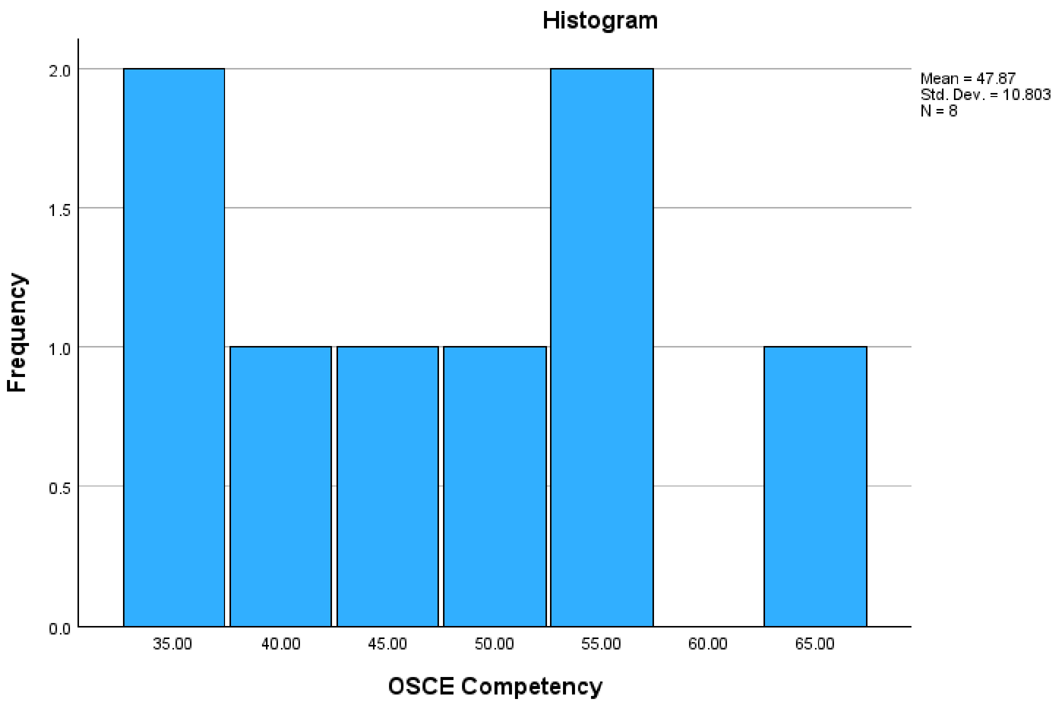


Descriptive Analysis of Global Rating Scale and Task Specific Checklist (GRS-TSC): OSCE Competency

A descriptive analysis of OSCE competency was performed to evaluate the distribution of scores. The mean value for OSCE competency scores was 47.875, with a median value of 47.50, and a 95% confidence interval for the mean of 38.844 to 56.906, representing the central tendency for these results. Variability for these results included a standard deviation of 10.803, a range of 30.0, with a minimum score of 35.0, a maximum score of 65.0, and a variance of 116.696. Distribution of the scores revealed skewness of 0.212 with a kurtosis value of -1.125. The normality of the distribution was evaluated using Shapiro-Wilk and Kolmogorov-Smirnov testing, with a graphical representation using a histogram. The p-value for the Shapiro-Wilk test was greater than 0.05, suggesting the data was normally distributed and is best used for less than 50 observations. The Distribution of OSCE Competency Data is shown in Table 9 below.

Table 9: OSCE Competency Data

Distribution of OSCE Competency Data			
Test	**Statistic**	**Degrees of Freedom**	**p-value**
Kolmogorov-Smirnov	0.174	8	0.200
Shapiro-Wilk	0.913	8	0.526



Descriptive Analysis of Study Variables for All Groups

The dataset presents descriptive statistics for five key variables measured in both the control and intervention groups across all eight participants. These variables include pre-simulation and post-simulation self-efficacy, pre-simulation and post-simulation emotional stress response, and OSCE competency evaluation. The descriptive analysis includes measures of central tendency (mean), variability (standard deviation, range), and distribution characteristics (skewness and kurtosis) for each variable. The summary statistics for all study variables are displayed in Table 10 below.

Table 10: Descriptive Statistics of All Groups

Descriptive Statistics of All Groups									
Group	Variable	N	Range	Minimum	Maximum	Mean	Std Deviation	Skewness	Kurtosis
Control	Pre-simulation Self-Efficacy	4	4.0	28.0	32.0	29.250	1.893	1.659	2.615
	Post-simulation Self-Efficacy	4	6.0	24.0	30.0	27.000	2.449	0.000	1.500
Intervention	Pre-simulation Self-Efficacy	4	6.0	23.0	29.0	27.000	2.646	-0.864	-0.286
	Post-simulation Self-Efficacy	4	6.0	23.0	29.0	26.250	2.754	-0.323	-3.033
Control	Pre-simulation Emotional Stress Response	4	21.0	61.0	82.0	69.250	10.210	0.628	-2.492
	Post-simulation Emotional Stress Response	4	17.0	66.0	83.0	74.000	7.257	0.345	-0.577
Intervention	Pre-simulation Emotional Stress Response	4	23.0	50.0	73.0	56.750	10.996	1.833	3.357
	Post-simulation Emotional Stress Response	4	27.0	51.0	78.0	67.500	11.902	-1.210	1.126
Control	OSCE Competency	4	30.0	35.0	65.0	47.500	15.000	0.370	-3.901
Intervention	OSCE Competency	4	14.0	42.0	56.0	48.250	6.850	0.268	-4.276

As shown in Table 10, the descriptive statistics indicate that the control group had a higher pre-simulation self-efficacy mean of 29.25 compared to the intervention group, 27.00. After the simulation, self-efficacy decreased in both groups, with the intervention group scoring slightly lower at 26.25 than the control group, 27.00. For the emotional stress response, the control group initially reported higher stress levels at 69.25 compared to the intervention group, 56.75. Although post-simulation stress levels increased in both groups, a more prominent rise was seen in the intervention group's score of 67.50, with an increase of 10.75 compared to the control group change of 4.75.

Regarding the OSCE competency evaluation, the control group had a wider score range of 35 to 65 than the intervention group, 42 to 56. Despite this, both groups had similar mean scores of 47.50 for control and 48.25 for the intervention. The control group also showed greater variability, with a standard deviation of 15.00 compared to 6.85 in the intervention group.

Overall, both groups experienced changes in self-efficacy and stress levels before and after the simulation, while the control group exhibited more variability in OSCE competency scores.

Inferential Analysis of Study Variables

Inferential statistics are essential for drawing conclusions about a population based on sample data. Unlike descriptive statistics, which summarize and describe data, inferential statistics allow researchers to test hypotheses, determine relationships, and make predictions based on the sample data. By using techniques such as t-tests, correlation analysis, and regression models, researchers can evaluate differences between groups, assess the strength and direction of relationships between variables, and measure the effectiveness of interventions. These methods help ensure that findings are not due to chance and provide generalizable insights, making them crucial for evidence-based decision-making in research.

Research Question 1: Effects of PF on SE and Emotional Stress Response. Question one asks what are the effects of simulated productive failure on PA students' self-efficacy and emotional stress response when performing OSCE activities?

Paired t-test. As this question investigates changes in self-efficacy and emotional stress response, it involves comparing within groups, both control and intervention groups, for pre-simulation and post-simulation scores. The paired t-test is a statistical method used to compare the mean values of two related groups to determine if there is a significant difference between them. It is applied when measuring the same research participants under two different conditions, such as before and after an intervention. The test assumes the differences between paired observations are normally distributed (Fields, 2018). It also accounts for the dependency between paired observations, reducing variability and increasing statistical power. The purpose of using a paired t-test is to assess whether an intervention leads to a meaningful change in the measured variable over time, making it particularly useful for pre- and post-test study designs. To answer RQ1, this study employed the paired t-test to evaluate the difference in means within the same groups, control and intervention, in the pre-simulation and post-simulation scores for self-efficacy and emotional stress response. The paired t-test results are displayed in Table 11 below.

Table 11: Paired t-test for Self-Efficacy and Emotional Stress Response

Paired t-test Self-Efficacy								
Variable	**Group**	**N**	**Mean Score**	**Correlation**	**Mean Difference**	**Std Deviation**	**t-statistic**	**p-value (two-tailed)**
Pre-simulation Self-Efficacy	Control	4	29.250	0.216	2.250	2.754	1.634	0.201
Post-simulation Self-Efficacy			27.000					
Pre-simulation Self-Efficacy	Intervention	4	26.750	0.778	0.500	1.732	0.577	0.604
Post-simulation Self-Efficacy			26.250					
Paired t-test Emotional Stress Response								
Pre-simulation Emotional Stress Response	Control	4	69.250	0.958	-4.750	3.862	-2.460	0.091
Post-simulation Emotional Stress Response			74.000					
Pre-simulation Emotional Stress Response	Intervention	4	56.750	0.376	-10.750	12.812	-1.678	0.192
Post-simulation Emotional Stress Response			67.500					

The results of the paired t-test evaluating self-efficacy showed that the control group's mean pre-simulation score was 29.25, which declined to 27.00 post-simulation. Although the mean difference was 2.25, the two-tailed p-value of 0.201 indicated that the decline was not statistically significant. Similarly, the intervention group's self-efficacy scores decreased slightly from 26.75 pre-simulation to 26.25 post-simulation. The mean difference was 0.50 with a two-tailed p-value of 0.604. These results suggest that the observed changes were not statistically significant.

The paired t-test results for emotional stress response showed an increase in the control group from 69.25 pre-simulation to 74.00 post-simulation, with a mean difference of -4.75 and a two-tailed p-value of 0.091, indicating a non-significant increase in emotional stress response within the group. The intervention group also demonstrated an increase in emotional stress response from 56.75 to 67.50, with a larger mean difference of -10.75; however, the two-tailed p-value of 0.192 again did not reach statistical significance. It is important to note that the small sample size and corresponding low statistical power may limit the ability to detect meaningful differences. These non-significant results do not necessarily imply the absence of an effect but rather suggest that the study may have been underpowered to detect potentially meaningful changes in self-efficacy and emotional stress response.

Cohen's d. Additionally, a Cohen's d-test was performed with the paired t-test to further evaluate self-efficacy and emotional stress response. This test measures the effect size, indicating the magnitude of the difference between the pre-simulation and the post-simulation scores within the same group. It quantifies the practical significance of the change by standardizing the mean difference in relation to the pooled standard deviation of the paired differences (Cohen, 2013b). The higher Cohen d-value suggests a stronger effect of the intervention or treatment, while a lower value indicates a smaller impact. Cohen's d-values of 0.2 indicate a small effect, values of 0.5 indicate a medium effect, and values of 0.8 or higher indicate a large effect (Cohen, 2013b). This test assists in the analysis of the paired t-test by understanding whether a significant statistical difference of means in the paired t-test also has practical importance due to the larger Cohen d-value. In this study, each of the Cohen d-values shows a large effect size as they are greater than 0.8. Hedges' *g* is a bias-corrected version of Cohen's d-value used especially when sample sizes are small, typically with $n < 20$. It applies a correction factor to adjust Cohen's d-value downward to account for small sample bias. The paired t-test, Cohen's d-values, and Hedges' g-values are displayed in Table 12 below.

Table 12: Effect Size for Self-Efficacy and Emotional Stress Response

Effect Size Self-Efficacy					
Variable	**Group**	**Cohen's d**	**Point Estimate**	**Hedges' Correction**	**Hedges' g**
Pre-simulation Self-efficacy Post-simulation Self-Efficacy	Control	2.754	0.817	3.806	2.600
Pre-simulation Self-efficacy Post-simulation Self-Efficacy	Intervention	1.732	0.209	2.394	1.640
Effect Size Emotional Stress Response					
Variable	**Group**	**Cohen's d**	**Point Estimate**	**Hedges' Correction**	**Hedges' g**
Pre-simulation Emotional Stress Response Post-simulation Emotional Stress Response	Control	3.862	-1.230	5.337	3.650
Pre-simulation Emotional Stress Response Post-simulation Emotional Stress Response	Intervention	12.816	-0.839	17.711	12.120

Independent Samples t-test and Levene's test. An independent t-test is a statistical method used to compare the means of two independent groups to determine whether there is a significant difference between them (Field, 2018). This test assumes that the two groups are unrelated and that their variances are approximately equal. It is commonly used in RCTs to assess the effect of an intervention by comparing the performance of a control group versus an experimental group (Gravetter, 2013). The purpose of using an independent t-test is to evaluate whether observed differences in mean values are due to chance or if they reflect a true effect in the sample (Field, 2018).

This study employed an independent t-test to evaluate the differences between the control and intervention groups' mean scores on pre-simulation self-efficacy, post-simulation self-efficacy, pre-simulation emotional stress response, and post-simulation emotional stress response. The IV is the group assignment, and the DV is either pre-simulation self-efficacy, post-simulation self-efficacy, pre-simulation emotional stress response, or post-simulation emotional stress response. The Independent Samples t-test results for self-efficacy and emotional stress response are displayed in Table 13 below.

Table 13: Independent Samples t-test for Pre- and Post-simulation Self-Efficacy and Emotional Stress Response

Independent t-test for Pre- and Post-simulation Self-Efficacy						
Variable	**Levene p-value**	**t-statistic**	**Degrees of Freedom**	**p-value (two-tailed)**	**Cohen's d**	**Hedges's g**
Pre-simulation Self-efficacy Intervention vs Control	0.606	-1.715	6.0	0.137	2.0616	1.793
Post-simulation Self-Efficacy Intervention vs Control	0.473	-0.407	6.0	0.698	1.893	1.646
Independent t-test for Pre- and Post-simulation Emotional Stress Response						
Variable	**Levene p-value**	**t-statistic**	**Degrees of Freedom**	**p-value (two-tailed)**	**Cohen's d**	**Hedges's g**
Pre-simulation Emotional Stress Response Intervention vs Control	0.972	-1.666	6.0	0.147	10.611	9.227
Post-simulation Emotional Stress Response Intervention vs Control	0.460	-0.933	6.0	0.394	9.857	8.571

The results indicate that Levene's test for equality of variances confirmed the assumption of equal variances for all variables, as the p-values were greater than 0.05. For pre-simulation self-efficacy, the t-statistic and p-values showed no statistically significant difference between the control and intervention groups, although the intervention group scored slightly lower than the control group. Similarly, for post-simulation self-efficacy, pre-simulation emotional stress response, and post-simulation emotional stress response, the intervention group consistently scored lower than the control group. However, these differences were not statistically significant. Overall, the analysis shows no statistically significant differences between the control and intervention groups in terms of self-efficacy or emotional stress response, both before and after the simulation. It is important to note that the small sample size and resulting low statistical power may limit the ability to detect true effects. Thus, while the current data does not demonstrate a measurable impact of the intervention, the possibility of meaningful differences cannot be entirely ruled out under these study conditions.

Pearson's Correlation Coefficient. Pearson's correlation coefficient (Pearson's r) is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. The coefficient ranges from -1 to 1, where a value close to 1 indicates a strong positive correlation, a value close to -1 represents a strong negative correlation, and a value around 0 suggests little to no correlation (Field, 2018). Pearson's r is widely used to determine whether changes in one variable are associated with changes in another, making it an essential tool for exploratory data analysis and hypothesis testing (Cohen, 2013a).

The primary purpose of Pearson's correlation is to evaluate relationships between variables without implying causation. For instance, Pearson's r can be used to assess whether there is a significant relationship between students' self-efficacy and their OSCE competency evaluation. A positive correlation would suggest that higher self-efficacy is associated with better OSCE competency, while a negative correlation would indicate an inverse relationship (Pallant, 2020). However, it is crucial to interpret these findings cautiously, as correlation does not establish a cause-and-effect relationship between variables.

In this study, Pearson's r is used to evaluate the relationship between all study variables. Understanding the correlation between variables helps to determine whether further statistical modeling is appropriate. For example, if self-efficacy and OSCE competency scores show a significant correlation, researchers may proceed to test whether self-efficacy predicts competency using regression analysis (Tabachnick, 2019). Thus, Pearson's r serves as a foundational statistical test in many quantitative research designs, providing valuable insights into variable relationships and guiding further analytical approaches. Pearson's r-correlation test was performed on all continuous variables within the study and is displayed in Table 14 below.

Table 14: Pearson's r-correlation for All Study Variables

Pearson's Correlation Statistics						
Variable		**Pre-sim Self-Efficacy**	**Post-sim Self-Efficacy**	**Pre-sim Emotional Stress Response**	**Post-sim Emotional Stress Response**	**OSCE Competency Evaluation**
Pre-simulation Self-Efficacy	Pearson's Correlation	1.0	0.526	0.738	0.320	-0.153
	Sig (2-tailed)		0.180	0.037	0.439	0.717
	N	8.0	8.0	8.0	8.0	8.0
Post-simulation Self-Efficacy	Pearson's Correlation	0.526	1.0	0.167	0.181	0.376
	Sig (2-tailed)	0.180		0.692	0.668	0.099
	N	8.0	8.0	8.0	8.0	8.0
Pre-simulation Emotional Stress Response	Pearson's Correlation	0.738	0.167	1.0	0.644	-0.623
	Sig (2-tailed)	0.037	0.692		0.085	0.990
	N	8.0	8.0	8.0	8.0	8.0
Post-simulation Emotional Stress Response	Pearson's Correlation	0.320	0.181	0.644	1.0	-0.725
	Sig (2-tailed)	0.439	0.668	0.085		0.042
	N	8.0	8.0	8.0	8.0	8.0
OSCE Competency Evaluation	Pearson's Correlation	-0.153	0.376	-0.623	-0.725	1.0
	Sig (2-tailed)	0.717	0.099	0.990	0.042	
	N	8.0	8.0	8.0	8.0	8.0

This initial analysis showed a strong positive correlation between pre-simulation self-efficacy and pre-simulation emotional stress response, indicated by a correlation coefficient of 0.738 and a significance level of 0.037, revealing statistical significance. Additionally, there is a strong negative correlation between post-simulation emotional stress response and OSCE competency evaluation, indicated by a correlation coefficient of -0.725 and a significance level of 0.042, revealing statistical significance. Several other correlations did not meet the statistical significance of a p-value less than 0.05 but showed a moderate-to-strong correlation. Pre-simulation self-efficacy and post-simulation self-efficacy with a strong positive correlation of 0.526, and a p-value of 0.180. Post-simulation self-efficacy and OSCE competency evaluation revealed a moderate positive correlation of 0.376 and a p-value of 0.358. Lastly, pre-simulation emotional stress response and OSCE evaluation scores showed a strong negative correlation of -0.623 and a p-value of 0.099.

Further evaluation was performed by repeating Pearson's correlation that included the group assignment for each of these variables. In the control group, pre-simulation self-efficacy and pre-simulation emotional stress response showed a strong positive correlation, indicated by a correlation coefficient of 0.737 and a significance level of 0.263, revealing non-statistical significance. The pre-simulation emotional stress response and post-simulation emotional stress response showed a very strong positive correlation, indicated by a correlation coefficient of 0.958 and a p-value of 0.042, revealing statistical significance. Pre-simulation self-efficacy and post-simulation emotional stress response showed a non-statistically significant positive correlation with a correlation coefficient of 0.679 and a p-value of 0.321. Post-simulation self-efficacy showed a moderate negative correlation with post-simulation emotional stress response, indicated by a correlation coefficient of -0.563 with a non-significant p-value of 0.437. Lastly, post-simulation emotional stress response and post-simulation self-efficacy showed a moderately negative correlation indicated by a correlation coefficient of -0.480 with a p-value of 0.520, revealing non-statistical significance.

In the intervention group, pre-simulation self-efficacy showed a strong positive correlation with post-simulation self-efficacy, indicated by a correlation coefficient t of 0.709 and a non-significant p-value of 0.291. Pre-simulation self-efficacy and pre-simulation emotional stress response showed a strong positive correlation indicated by a correlation coefficient of 0.510 and a p-value of 0.490, revealing non-statistical significance. Post-simulation self-efficacy and post-simulation emotional stress response showed a strong positive correlation with a correlation coefficient of 0.514 and a non-statistically significant p-value of 0.486. Pre-simulation emotional stress response and post-simulation emotional stress response showed a moderate positive correlation with a coefficient of 0.376 and a non-significant p-value of 0.624. Post-simulation self-efficacy and pre-simulation emotional stress response reveal a strong positive correlation with a correlation coefficient of 0.564 and a p-value of 0.436, revealing non-statistical significance. Lastly, post-simulation emotional stress response and pre-simulation self-efficacy reveal a weakly negative correlation with a coefficient of -0.212 and a p-value of 0.788, indicating non-statistical significance.

Research Question 2: Effects of PF on Competency. Question two asks what are the effects of simulated productive failure on student competency when utilized in physician assistant (PA) OSCE activities.

Independent Samples t-test and Levene's test. As this question investigates whether simulated PF influences participants' OSCE competency scores, it involves comparing between groups, both the control and intervention groups for OSCE competency scores. To answer RQ2, this study employed a second independent samples t-test to evaluate the difference in means between control and intervention groups for OSCE competency evaluation scores. The Independent Samples t-test was performed on OSCE competency scores and is displayed in Table 15 below.

Table 15: Independent Samples t-test for OSCE Competency Evaluation

Independent t-test for OSCE Competency Evaluation						
Variable	**Levene p-value**	**t-statistic**	**Degrees of Freedom**	**p-value (two-tailed)**	**Cohen's d**	**Hedge's g**
OSCE Competency Evaluation Intervention vs Control	0.022	0.091	4.199	0.932	11.660	10.140

The results indicate that Levene's test for equality of variances showed the p-value was 0.022. Since the p-value is less than 0.05, the null hypothesis of equality of variances was rejected and indicated the variances of the two groups are not equal. Consequently, the Welch independent samples t-test, which does not assume equal variances, was used to analyze the data. The t-statistic was 0.091 and represents the difference in mean scores between the two groups relative to the variability in the data. The small t-value indicates minimal differences in means, and the 4.199 degrees of freedom reflect the adjustment for unequal variances. Additionally, the two-tailed p-values for the t-statistics indicated no statistically significant differences in mean scores between the intervention and control groups. Overall, the data show no significant difference in OSCE competency evaluation scores between the groups. However, the small sample size and corresponding low statistical power may limit the ability to detect true differences, suggesting that potential effects of the intervention may not be fully captured under the current study conditions.

Linear Regression. Linear regression is a statistical method used to examine the relationship between one or more independent variables, predictors, and a dependent variable, outcome. This test helps determine how changes in the predictors are associated with changes in the outcome, allowing for predictive relationships between variables. This evaluation is particularly useful in determining the strength and nature of relationships, identifying significant predictors, and making predictions about future outcomes based on existing data (Field, 2018; Tabachnick, 2019).

As previously stated in this chapter, Pearson's correlation (Table 14) revealed a strong negative relationship that was statistically significant between post-simulation emotional stress response (IV) and OSCE competency evaluation (DV). Linear regression was performed to further explain the nature of that relationship. Additionally, linear regression was performed to evaluate the relationship between self-efficacy scores and OSCE competency scores and emotional stress response with OSCE competency scores. For each of these regression analyses, OSCE competency evaluation scores were considered the DV, and the other variables were considered the IV. Linear Regression of OSCE Competency scores is displayed in Table 16 below.

Table 16: OSCE Competency Score Linear Regression

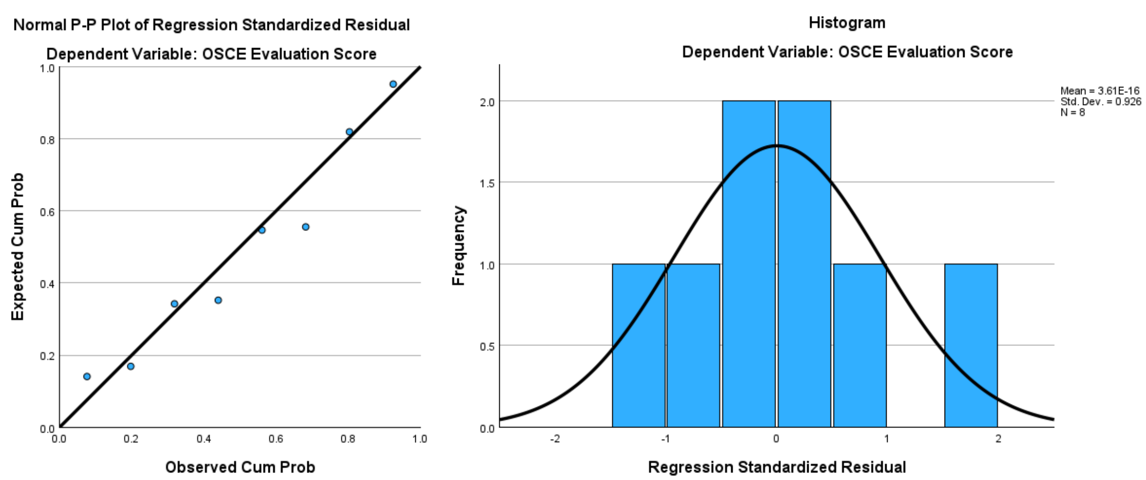
Linear Regression of Self-Efficacy with OSCE Competency Evaluation Scores							
Variable	**R-squared**	**Adjusted R-squared**	**F-value**	**Regression p-value**	**Slope (B)**	**Beta coefficient**	**Coefficient p-values**
Pre-simulation Self-efficacy	0.230	-0.139	0.144	0.717	-0.711	-0.153	0.717
Post-simulation Self-efficacy	0.142	-0.001	0.991	0.358	1.663	0.376	0.358
Linear Regression of Emotional Stress Response with OSCE Competency Evaluation Scores							
Pre-simulation Emotional Stress Response	0.389	0.287	3.813	0.099	-0.567	-0.623	0.099

Post-simulation Emotional Stress Response	0.525	0.446	6.643	0.420	-0.802	-0.725	0.042

Overall, the regression analysis of pre-simulation and post-simulation self-efficacy scores did not significantly predict OSCE competency evaluation scores. The regression analysis revealed that 23% of the variance in OSCE competency scores was explained by the pre-simulation self-efficacy, with an F-value of 0.144 and a regression p-value of 0.717, indicating the model is not statistically significant. The slope of -0.711 and the Beta coefficient of -0.153 suggest a weak negative relationship between pre-simulation self-efficacy and OSCE competency, but the coefficient p-values of 0.717 confirm this relationship is not statistically significant. Similarly, the post-simulation self-efficacy regression revealed an F-value of 0.991 and a regression p-value of 0.358, indicating the model was not statistically significant. The slope was 1.663, and the Beta coefficient was 0.376, suggesting a positive relationship between post-simulation self-efficacy and OSCE competency. However, the coefficient p-value of 0.358 showed the relationship is not statistically significant.

Pre-simulation emotional stress response regression showed that 38.9% of the variance in OSCE competency score was explained by pre-simulation emotional stress response. The F-value was 3.813, and the regression p-value was 0.099, suggesting a trend toward significance, but it does not meet statistical significance. Slope and Beta coefficients of -0.567 and -0.623, respectively, indicate a moderate negative relationship. The coefficient p-value of 0.099 indicates the relationship is not statistically significant. Lastly, regression analysis of post-simulation emotional stress response revealed it was a statistically significant predictor of OSCE competency evaluation scores, with higher stress associated with lower competency scores. This identified that 52.5% of the variance in OSCE competency scores is explained by post-simulation emotional stress response, with an adjusted R-squared of 0.446 indicating a strong model fit. F-values of 6.643 and a regression p-value of 0.042 indicate the model is statistically significant. Slope and Beta coefficients of -0.802 and -0.725 reveal a strong negative relationship between post-simulation emotional stress and OSCE competency evaluation scores, with a coefficient p-value of 0.042 confirming this relationship. It is important to note that the small sample size may have limited the statistical power of these analyses, reducing the ability to detect potentially meaningful effects in these models. Linear Regression of OSCE competency evaluation scores and post-simulation stress response scores are displayed in Figure 2 below.

Figure 2: Linear Regression of OSCE Competency Score and Post-Simulation Emotional Stress Response



Analysis of Demographic Predictors of OSCE Competency Scores. To evaluate the potential impact of demographic characteristics on OSCE competency scores within the study, a multiple regression analysis was conducted. The analysis examined the effects that age, race, gender, marital status, number of core rotations, and completion of an ER rotation had on OSCE competency scores.

Multiple regression is a statistical technique used to examine the relationship between a single dependent variable and multiple independent variables (Field, 2018). In this test, the OSCE competency scores were considered the DV, and demographic characteristics were considered the IVs. This analysis examined the extent to which each predictor contributes to the outcome while controlling for other variables (Hair, 2014). The primary purpose of multiple regression is to identify significant predictors, assess the strength and direction of relationships, and determine how much variance in the dependent variable can be explained by the independent variables (Tabachnick, 2019). Multiple Regression of OSCE competency evaluation scores and Demographic Characteristics are displayed in Table 17 below.

Table 17: Multiple Regression of OSCE Competency Score and Demographic Characteristics

Multiple Regression of OSCE Competency Scores and Demographic Predictors							
Demographic Characteristic	**R-squared**	**Adjusted R-squared**	**F-value**	**Regression p-value**	**Slope (B)**	**Beta coefficient**	**Coefficient p-values**
Age	0.957	0.700	3.721	0.377	-2.979	-0.147	0.700
Race					-4.063	-0.28	0.586
Gender					-7.604	-0.364	0.483
Marital Status					-17.854	-0.765	0.278
Number of Core Rotations					-1.125	-0.054	0.893
Completion of ER Rotation					16.25	0.696	0.222

The multiple regression analysis of demographic predictors of OSCE competency scores revealed there were no statistically significant relationships between these variables. The completion of an ER rotation had a positive B-value of 16.250, indicating that completing this rotation is associated with increased OSCE scores. Additionally, the highest Beta coefficient was for completion of an ER rotation, 0.696, suggesting this variable has the strongest effect on

OSCE scores. However, the coefficient p-value of 0.222 indicated the relationship was not statistically significant.

Qualitative Data and Analysis- Research Question 3

Research question three (RQ3) asks how PA students reflect on learning from OSCE activities that utilize productive failure in the patient simulation. To answer this question, a qualitative approach using focus group discussions was employed in this study.

Introduction to Focus Groups

The primary purpose of focus groups in qualitative research is to obtain a deeper understanding of participants' thoughts, emotions, and behaviors in a social context. Unlike surveys or individual interviews, focus groups encourage discussion and debate, providing richer data that can uncover shared meanings and diverse viewpoints (Kitzinger, 1995). This method is particularly useful for exploring complex issues, generating new ideas, and identifying potential areas for further investigation.

In this study, focus groups were used as a qualitative research method to gather in-depth insight into participants' attitudes, perceptions, experiences, and opinions on the intervention of PF. This method involved semi-structured discussions facilitated by study investigators. These focus groups are composed of small groups who share common characteristics relevant to the research question (Krueger, 2015). The interactive nature of focus groups allowed for the exploration of individual responses, in addition to the dynamics of the group discussion, including how participants influenced each other's perspectives and how opinions evolved through the group conversation (Morgan, 1996). This process allows for the co-construction of knowledge, as participants build on each other's responses, clarify their views, and reveal underlying motivations or assumptions that may not have emerged in one-on-one interviews (Barbour, 2007).

To answer RQ3, focus group discussions were centered around three primary qualitative focus group questions: 1) What challenges did you experience in the simulation activity? 2) How did today's simulation activity compare to other teaching methods previously used in the PA OSCE activities? 3) Reflecting on your participation in the productive failure simulation, how do you feel it influenced your learning and development as a PA student? Participants were prompted to respond to questions with open-ended responses and to have an open discussion within the group. The discussion data was audio-recorded and then transcribed for analysis.

Thematic Analysis of Qualitative Data

Thematic analysis of the transcribed data was performed initially utilizing line-by-line coding for each segment of text. These codes included short labels that described the meaning of the specific words, phrases, or sentences. The initial codes were descriptive in nature and closely tied to the raw data or participant quotes within the transcript. This approach to the thematic analysis ensures all aspects of the data are considered and prevents researchers from overlooking details (Charmaz, 2006). Line-by-line coding assisted by reducing the data into manageable segments and allowed for a more detailed analysis of patterns in participants' responses.

The second step in the analysis process was organizing codes into categories. After completing the line-by-line coding, related code groups were placed into categories based on their shared meaning or patterns of observation in the data (Saldaña, 2021). The final step in the process was to identify overarching themes that emerged from the categories. The themes represent the broader patterns of meaning across the dataset and provide insight into the research question. Unlike categories, the themes are conceptual and interpretative concepts (Braun, 2006).

The following section of this chapter presents a thematic analysis of the focus group questions and responses. For each focus group question, an initial set of codes was developed based on participants' responses. These codes were then grouped into broader categories that captured the key ideas and patterns within each question. From these categories, distinct themes were developed for each focus group question, reflecting the participants' shared experiences and insights.

Finally, the themes generated from each focus group question were synthesized to create overarching themes that addressed the broader research inquiry. These overarching themes provided a comprehensive response to research question three, ensuring that the analysis was grounded in participants' perspectives while offering a cohesive understanding of the PF OSCE simulation experience. This structured approach allowed for a clear and systematic development of themes, enhancing the depth and clarity of the research findings.

Thematic Analysis Triangulation

Triangulation was integrated into the thematic analysis process through a collaborative and systematic approach involving three researchers. These three individuals have prior experience in qualitative research and thematic analysis, and each holds a graduate-level educational degree. Each researcher independently coded and analyzed the focus group data by reviewing transcripts and developing initial codes, categories, and themes in response to each focus group question. This independent analysis ensured that a diverse range of perspectives was considered, minimizing individual bias. Each researcher brings a unique blend of academic training and practical experience in qualitative methodologies, including focus group facilitation, coding, and interpretive analysis. Their combined expertise strengthens the credibility of the thematic findings, as they are well-versed in ensuring methodological rigor, reflexivity, and consistency throughout the qualitative process.

Following this independent coding, the researchers convened to discuss and compare their identified codes, categories, and themes. Any discrepancies or differences in interpretation were thoroughly discussed and resolved through consensus, leading to the development of final themes that accurately represented the participants' experiences. This collaborative process ensured that the themes were not influenced by any single researcher's perspective but were instead a balanced and collective interpretation.

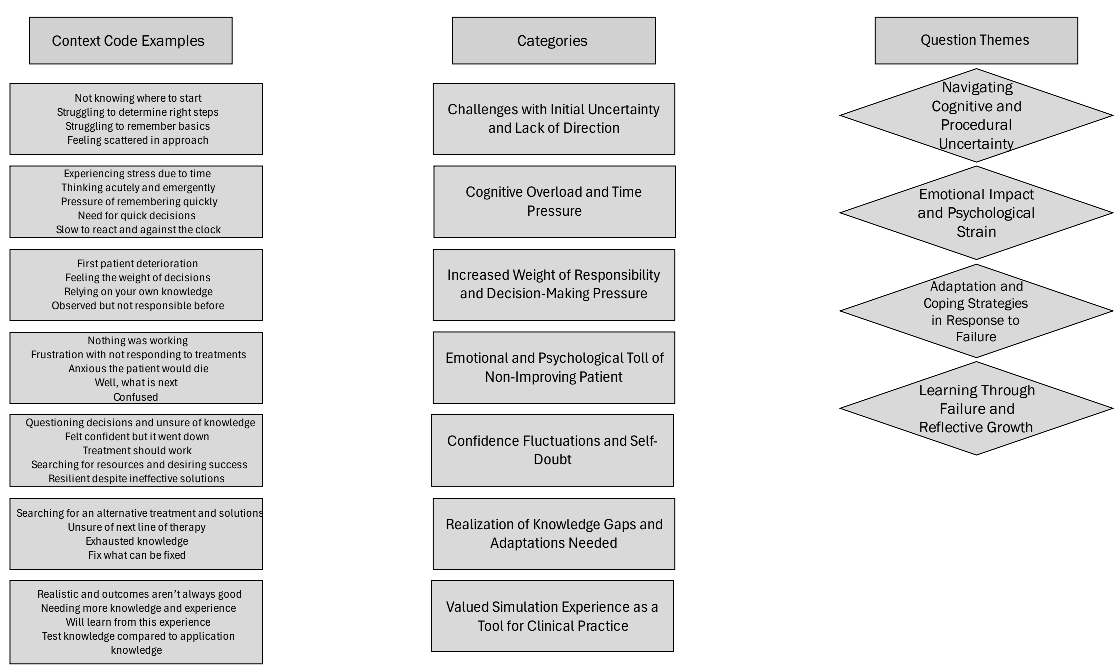
The triangulation process provided a robust foundation for the thematic analysis, enhancing the reliability and credibility of the findings. The final, agreed-upon themes were used to address the third research question, offering a well-rounded and in-depth understanding of the PA students' experiences in the PF OSCE simulation. By incorporating multiple viewpoints and reaching consensus, the thematic analysis was strengthened, ensuring that the research findings were both reliable and comprehensive.

Qualitative Focus Group Question One

The initial focus group question, "What challenges did you experience in the simulation activity?" was designed to prompt participants to reflect on the challenges encountered during the PF OSCE experience. This question encouraged both individual and collective reflection and discussion, allowing participants to express deep insights into the cognitive, emotional, and procedural difficulties they faced. By exploring these challenges, researchers gained a better understanding of the internal thought processes, uncertainties, and emotional responses that influenced participants' clinical decision-making and overall learning experience.

Through this thematic analysis process, several themes emerged pertaining to this focus group discussion*: Navigating Cognitive and Procedural Uncertainty, Emotional Impact and Psychological Strain, Adaptations and Coping Strategies in Response to Failure, and Learning Through Failure and Reflective Growth*. These themes emerged from seven categories: *Challenges with Initial Uncertainty and Lack of Direction, Cognitive Overload and Time Pressure, Increased Weight of Responsibility and Decision-Making Pressure, Emotional and Psychological Toll of Non-Improving Patient, Confidence Fluctuations and Self-Doubt, Realization of Knowledge Gaps and Adaptations Needed, Valued Simulation Experience as a Tool for Clinical Practice*. Multiple codes were utilized to develop these categories, and examples are provided in Figure 3. A summary of the thematic analysis of focus group question one is displayed in Figure 3 below.

Figure 3: Thematic Analysis Summary of Focus Group Question One



A primary goal of this focus group question was to identify potential barriers that hinder effective clinical performance during the simulation. Thematic analysis revealed several key challenges experienced by participants, including uncertainty, cognitive overload, lack of confidence, time pressure, increased responsibility, knowledge gaps, and emotional strain. For instance, one participant highlighted the difficulty of initiating care, stating, "Simply not knowing where to start," while another expressed, "I struggled with determining what are the right steps to make and what appropriate order." Emotional frustration was also evident, particularly regarding time constraints and treatment decisions, as participants noted, "I feel like I was against the clock the whole time" and "The stress of like, nothing I'm doing is working." These challenges contributed to the development of categories such as *Challenges with Initial Uncertainty and Lack of Direction* and *Cognitive Overload and Time Pressure*.

The category *Increased Weight of Responsibility and Decision-Making Pressure* emerged from numerous codes related to participants experiencing patient deterioration, feeling the burden of clinical decisions, and transitioning from observers to active decision-makers. One participant reflected, "I've seen it happen, but I haven't like, felt the weight of the situation," while another added, "Watching people who have done it day in, day out, but not putting myself in their shoes and actually trying to understand how they're thinking." These comments highlighted the participants' struggles with the new responsibilities encountered during the simulation.

Emotional and Psychological Toll of Non-Improving Patient and *Confidence Fluctuations and Self-Doubt* were categories developed from participant responses to unsuccessful treatment outcomes. Participants shared how failed treatments affected their confidence, with one stating, "My confidence went down," and another adding, "Confidence definitely went down…and I learned from it.” Emotional responses such as frustration, anxiety, and shock were also expressed, as indicated by statements like, “I’m getting frustrated,” “I was stunned,” and “I was getting anxious the patient was gonna die.” These reflections provided insight into how participants managed setbacks and confidence challenges during the simulation.

Participants also recognized gaps in their knowledge and the need for adaptation, leading to the categories *Realization of Knowledge Gaps and Adaptations Needed* and *Valued Simulation Experience as a Tool for Clinical Practice.* One participant described the experience as “Like real, like clinical practice, like, not everything you do is a positive outcome,” while another stated, “I’m not studying for a test, but I’m studying like for like medical practice.” Another participant noted, “I think it helps me see why you need to know that stuff like the back of your hand too.” These comments underscored the importance of identifying knowledge gaps and applying clinical knowledge effectively through simulation.

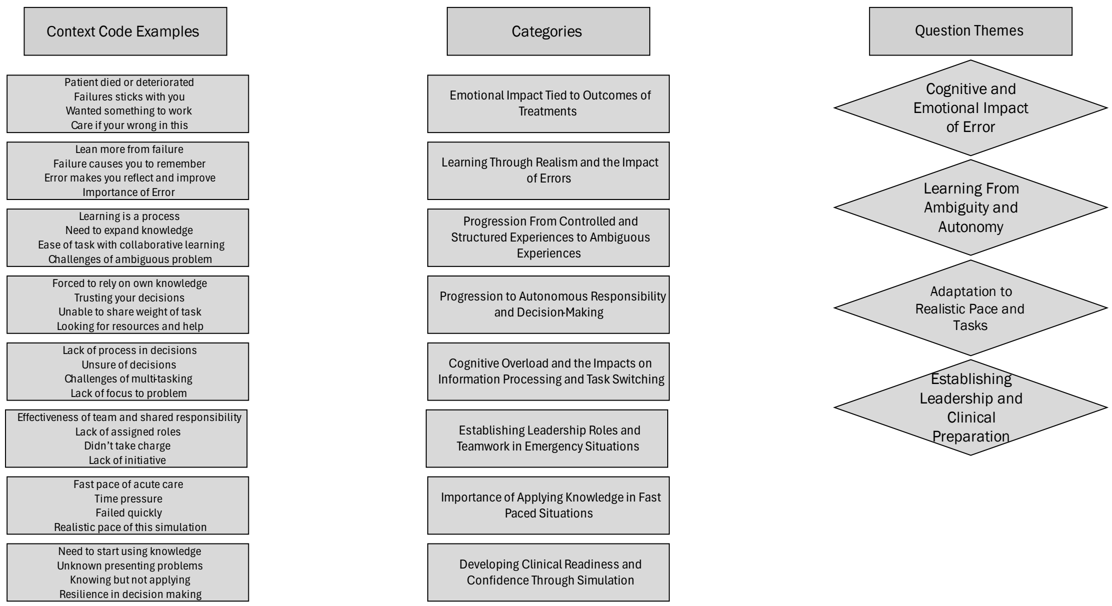
From these categories emerged the focus group question themes of *Navigating Cognitive and Procedural Uncertainty, Emotional Impact and Psychological Strain, Adaptations and Coping Strategies in Response to Failure, and Learning Through Failure and Reflective Growth*, providing meaning across the data set. Collectively, these insights helped researchers understand the root causes of participants' challenges, such as knowledge gaps, stress and time management difficulties, and the pressures of making rapid decisions in high-stakes scenarios. Furthermore, the focus group discussions highlighted the value of simulation as both an educational and reflective tool, emphasizing the need for PA students to build confidence through practice, reflection, and feedback to prepare for real-world clinical environments where quick and accurate decision-making is essential.

Qualitative Focus Group Question Two

The second focus group question, "How did today's simulation activity compare to other teaching methods previously used in the PA OSCE activities?" aimed to explore participants' perceptions of the PF OSCE simulation experience in relation to prior traditional OSCE activities. This question encouraged participants to reflect on and discuss their different learning experiences, to identify how the PF simulation influenced their learning process, whether it enhanced their understanding or presented new challenges.

Through this thematic analysis process, several themes emerged pertaining to this focus group discussion: *Cognitive and Emotional Impact of Error, Learning From Ambiguity and Autonomy, Adaptation to Realistic Pace and Tasks, Establishing Leadership and Clinical Preparation*. These themes emerged from eight categories: *Emotional Impact Tied to Outcomes of Treatments, Learning Through Realism and the Impact of Errors, Progression From Controlled and Structured Experiences to Ambiguous Experiences, Progression to Autonomous Responsibility and Decision-Making, Cognitive Overload and the Impacts on Information Processing and Task Switching, Establishing Leadership Roles and Teamwork in Emergency Situations, Importance of Applying Knowledge in Fast-Paced Situations, Developing Clinical Readiness and Confidence Through Simulation*. Multiple codes were utilized to develop these categories, and examples are provided in Figure 4. A summary of the thematic analysis of focus group question two is displayed in Figure 4 below.

Figure 4: Thematic Analysis Summary of Focus Group Question Two



By comparing the PF simulation to more conventional teaching methods, this question sought to determine if the simulation provided a more realistic and immersive learning environment that better prepared participants for clinical practice. The category *Emotional Impact Tied to Outcomes of Treatments* emerged as participants frequently expressed feelings of frustration, anxiety, and emotional strain when treatments did not yield successful results. One participant noted, “Well, no patients ever died on us, but they probably should have” and another stated, “The impact on the people, you take it home, I think”, highlighting the emotional toll that unsuccessful clinical interventions can have on participants, particularly when they feel responsible for patient outcomes.

The categories of *Learning Through Realism and the Impact of Errors* were developed from comments emphasizing the realism of the simulation and the lessons drawn from making mistakes. For instance, one participant shared, “It felt real, like in an actual clinical setting where every decision matters, and you have to learn quickly from your errors." Another reflected, "You learn more from failures than you will from successes sometimes". Such reflections underscore how realistic simulations provide valuable learning experiences, allowing participants to process and reflect on their clinical decisions.

The category *Progression From Controlled and Structured Experiences to Ambiguous Experiences* was informed by participants who discussed the shift from structured OSCE scenarios to more complex, unpredictable situations. One participant commented, "In previous OSCEs, I knew exactly what to expect, but this time, there was so much uncertainty, and I had to figure it out as I went," illustrating how increasing ambiguity in simulations challenged participants to adapt and build critical thinking skills. Another commented, "The stress level wasn't there for the first ones (prior OSCEs) cause you know everything going into it".

Progression to Autonomous Responsibility and Decision-Making was derived from participants' discussions about taking on more responsibility during the simulation. One participant stated, "In this case, like you didn't like, you couldn't have time to call your physician …you had to know it" and "it makes you like, rely more on yourself and your own knowledge", indicating the importance of fostering autonomy in clinical decision-making.

Participants also highlighted the cognitive challenges they faced, leading to the category *Cognitive Overload and the Impacts on Information Processing and Task Switching.* One participant remarked, "To go back to like, not being in one lane. You're in all four lanes on the highway and I couldn't stick with one. I mean, I was out of each one and with no blinkers. I'm in and out of everything". Another commented, "You had to juggle everything" demonstrating how the high-pressure simulation tested participants' cognitive capacity and ability to manage multiple tasks simultaneously.

Establishing Leadership Roles and Teamwork in Emergency Situations emerged from transcript examples where participants reflected on the importance of clear communication and leadership in team settings. One participant noted, "I should have probably gone over and, like, been like, I'll take compressions, will you need to start bagging (the patient)" and another stated, "I wasn't necessarily like taking charge of or physically doing it. Because I wasn't sure…of what you do versus what I do". These showcased the necessity of developing leadership skills and establishing roles under pressure.

The category *Importance of Applying Knowledge in Fast Paced Situations* was based on participants' experiences with time-sensitive clinical decisions. One participant shared, "The pace was so much faster and like really fast", and another stated, "I couldn't get to what I wanted to do fast enough" illustrating how the simulation reinforced the need for rapid, accurate clinical decision-making.

Finally, *Developing Clinical Readiness and Confidence Through Simulation* was derived from participants who acknowledged the value of the simulation in preparing them for real-world practice. One participant reflected, "So, this was a good experience of reality…I do need to know this stuff. I need to understand this and so I think that would help a student moving into clinicals", and another stated "So, you learn more from this type of situation, at least like for myself, than in the classroom. But then you get put into reality, and then it starts to click on what you need to remember, fix, and understand", emphasizing how productive failure simulations contribute to clinical competence and self-assurance.

The collective evaluation of these categories provided the foundation for the development of the focus group question themes *Cognitive and Emotional Impact of Error, Learning From Ambiguity and Autonomy, Adaptation to Realistic Pace and Tasks,* and *Establishing Leadership* *and Clinical Preparation*. These conceptual themes provided interpretive meaning and insight to answer the focus group question.

In summary, this question aimed to explore the extent to which the PF simulation fostered active learning, adaptability under pressure, and retention of clinical knowledge compared to traditional instructional methods. Additionally, asking participants to reflect on these comparisons prompted them to evaluate their learning journeys, highlighting how the PF OSCE simulation contributed to the development of their clinical reasoning, confidence, and overall readiness for real-world clinical practice. The insights gathered from these discussions can help to inform future curriculum development, ensuring that PA students receive comprehensive and effective preparation for the demands of clinical care.

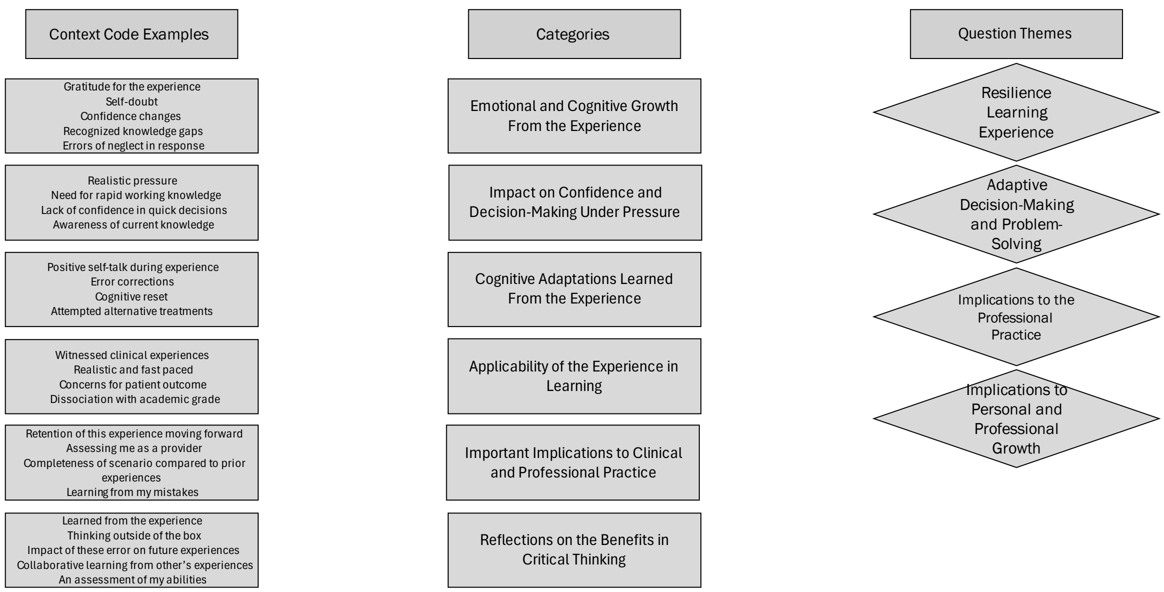
Qualitative Focus Group Question Three

The focus group question, “Reflecting on your participation in the productive failure simulation, how do you feel it influenced your learning and development as a PA student?” aimed to explore participants' insights into how the PF OSCE simulation impacted their educational journey and growth as future healthcare professionals. This question encouraged participants to reflect deeply on their experiences, allowing them to articulate how facing challenges, making mistakes, and navigating difficult clinical scenarios contributed to their learning process.

Through this thematic analysis process, several themes emerged pertaining to this focus group discussion: *Resilience Learning Experience, Adaptive Decision-Making and Problem-Solving, Implications to Professional Practice, and Implications to Personal and Professional Growth*. These themes emerged from six categories: *Emotional and Cognitive Growth From the Experience, Impact on Confidence and Decision-Making Under Pressure, Cognitive Adaptations*

Learned From the Experience, Applicability of the Experience in Learning, Important Implications to Clinical and Professional Practice, Reflections on the Benefits in Critical Thinking. Multiple codes were utilized to develop these categories, and examples are provided in Figure 5. A summary of the thematic analysis of focus group question three is displayed in Figure 5 below.

Figure 5: Thematic Analysis Summary of Focus Group Question Three



The primary goal of this question was to assess the perceived value of the productive failure approach in fostering critical thinking, clinical reasoning, and resilience. The question sought to uncover whether the experience helped them identify gaps in their understanding, improve their decision-making under pressure, and enhance their confidence in clinical settings. The category *Emotional and Cognitive Growth From the Experience* emerged from participants expressing how the challenges they faced contributed to both their emotional resilience and cognitive development. One participant noted, “It was frustrating, I felt pretty confident I knew what was going on, but I was like, I can’t stabilize this patient. Like what can I do? Nothing was working…I wish when I heard all the beeping I would just like stop and look at all the vital signs at one time", and another stated, "It was like a huge lesson for myself…there's like a lot that I need to know like the back of my hand now" highlighting how the experience facilitated growth through adversity.

Impact on Confidence and Decision-Making Under Pressure was drawn from participant reflections on how the simulation influenced their self-assurance and decision-making skills. Comments like, "I can't let my confidence go down right now…I have to do something, like, there is no feeling sorry for yourself right now…let's get this going" and "I've gotta keep going and try to save this guy" demonstrated the simulation's role in enhancing confidence and resilience in high-pressure scenarios.

The category *Cognitive Adaptations Learned From the Experience* stemmed from discussions on how participants adapted their cognitive processes during the simulation. One participant shared, "Nothing was working…well let's try something else" illustrating cognitive flexibility and adaptation.

Applicability of the Experience in Learning was developed from participants emphasizing how the simulation experience reinforced their learning. Statements such as, "I didn't feel like I knew the ACLS protocol as well as I should have" and "Oh yeah, I didn't even think about getting an EKG" underscored the importance of experiential learning in solidifying learning concepts.

Important Implications to Clinical and Professional Practice emerged from participants connecting the simulation to their future careers. One participant noted, "even though this was a mannequin, I'm not gonna forget this you know". Another commented, "This taught me that in real practice, things don't always go perfectly, and like, it's okay as long as you learn from it" emphasizing the simulation's relevance to clinical practice.

Lastly, *Reflections on the Benefits in Critical Thinking* were developed from participants' insights recognizing the simulation's role in enhancing their critical thinking abilities. Comments like "I think this will be beneficial going forward" and "Now I know what I know and what I don't know" highlight the value of PF simulations in fostering analytical thinking essential for clinical practice.

From these categories emerged the focus group question themes of *Resilience Learning Experience, Adaptive Decision-Making and Problem-Solving, and Implications to Personal and Professional Growth*. These themes provided collective meaning across the dataset to answer the focus group question.

In conclusion, this question aimed to gather insights into how the PF simulation contributed to the participants' ability to adapt to complex situations, manage cognitive load, and develop practical skills essential for clinical practice. It also provided an opportunity for participants to reflect on their emotional and psychological growth, highlighting how overcoming initial failures and challenges in the simulation helped build their resilience and readiness for real-world patient care. The responses to this question were instrumental in understanding how productive failure as a pedagogical approach may shape the learning trajectory of PA students. Insights gained can inform the refinement of simulation -based training, ensuring that PA education not only emphasizes clinical competence but also nurtures adaptability, problem-solving abilities, and continuous self-improvement in future healthcare providers.

Development of Overarching Themes

In concluding the thematic analysis, the focus group discussion's codes, categories, and themes were collectively reviewed to develop the overarching themes that answer research question three, "How do PA students reflect on learning from OSCE activities that utilize productive failure in the patient simulation?". Overarching themes emerged from the focus group question discussion to include *Emotional and Cognitive Growth Under Pressure, Development of Clinical Confidence and Decision-Making Abilities, Adaptability and Cognitive Flexibility in Complex Situations, Progression Towards Autonomous Clinical Practice, Leadership and Teamwork in High-Stakes Environments, Clinical Relevance and Lifelong Learning, Enhanced Critical Thinking and Reflective Practice*. These themes provided foundational meaning and insight into how PA students reflected on the learning experience of productive failure.

Chapter 5: Discussion and Conclusion

Introduction

This research study sought to explore the impact and potential benefits of the instructional method of productive failure when utilized in PA OSCE simulation activities. This investigation was initiated because the currently implemented methods of instruction in medical education fail to allow the learners to construct new knowledge through self-discovery, failure, and a reflective process within the learning experience. These instructional methods encourage and enable a learner attitude that devalues the process of learning and places an emphasis on problem-solving that finds a singular correct answer to complex problems. In doing so, this problem-solving method limits the creative critical thinking needed to find various potential solutions to complex clinical concerns that are present in medical practice. Conversely, the instructional method of PF encourages learners to seek out and discover a vast array of potential solutions to complex problems through short-term failed experiences. These complex problems are ill-structured and beyond the current knowledge of the learner, resulting in short-term failures that invoke a cognitive impasse in the learner and encourage the discovery of novel solutions to the concern (Kapur, 2008). Recent research suggests that learning from these failed experiences may provide a more potent catalyst for a deeper understanding of theoretical knowledge and skill acquisition compared to the current instructional methods (Steenhof, 2020).

To address the problem, the study was designed to pursue answers to three primary research questions that addressed the broader primary and secondary aims of the study. The study aimed to evaluate and understand the impact of PF when implemented in PA education, and to discover potential benefits of PF when utilized in PA education. To accomplish these aims, the study explored the potential effects of simulated PF on PA students during OSCE simulation activities. Specifically, the study evaluated the effect on participants' self-efficacy, emotional responses, and knowledge competency. In addition, the study sought to discover meaning from how participants reflected on learning from these PF OSCE activities.

These aims and questions evaluated the impact of productive failure in PA education, focusing on its influence on clinical reasoning, decision-making, confidence, and adaptability in complex clinical situations. By integrating PF into OSCE simulations, the research sought to provide insights into how this pedagogical approach could enhance PA education, foster critical thinking, and improve clinical competence. The findings discussed in this chapter highlight how PF can shape future PA education, enhance adaptability to clinical challenges, improve decision-making and confidence, and ultimately contribute to better patient care and reduced provider burnout.

Summary of Key Findings

The most significant finding of this study was the perceived benefits participants expressed regarding the use of PF in OSCE simulations. While the intervention group initially faced challenges that triggered self-doubt in medical decision-making and heightened stress, their reflective insights revealed that these emotional responses were ultimately outweighed by the significant benefits of the experience. Participants reported that, despite the initial frustration and uncertainty, the PF simulations fostered personal resilience in the face of challenges, enhanced self-confidence, and improved adaptability in managing complex clinical scenarios. Additionally, participants noted a marked improvement in their critical thinking skills, emphasizing the value of struggling through uncertainty as a means of deepening their clinical reasoning. These findings suggest that incorporating PF simulations into the PA curriculum could better prepare students for the unpredictability and demands of clinical practice.

Also important, while this study found no statistically significant negative impact of PF-based OSCE simulations on participants' self-efficacy, emotional stress responses, or competency evaluations, the small sample size and limited statistical power may have constrained the ability to detect subtle or meaningful effects. While participants in the intervention group experienced temporary declines in self-efficacy and heightened emotional stress during the simulation, comparisons between the control and intervention groups did not reveal any statistically significant differences in these areas. Most importantly, the PF simulation did not lead to any measurable decline in knowledge competency, nor did it negatively affect participants' overall performance in the OSCE assessments. This finding challenges the traditional reluctance in medical education to incorporate failure-based learning experiences due to concerns about their potential adverse effects on competency and confidence (Toufiq, 2024).

The results of this study provide compelling evidence that PF OSCE simulations are not inferior to traditional OSCE learning experiences. In fact, given the reported benefits in resilience, confidence, adaptability, and critical thinking, the findings suggest that the advantages of integrating PF into PA education may outweigh the perceived risks often cited by medical educators. These results support the notion that controlled failure within a structured learning environment can be a powerful tool for fostering the cognitive and emotional growth necessary for future healthcare providers. Allowing students to experience failure in OSCE simulations may ultimately enhance their ability to navigate the complexities and pressures of real-world clinical practice, leading to improved patient care and a more resilient workforce. However, it is imperative to acknowledge that more extensive research with a larger sample size is required to establish significant implications.

Interpretation of Results

This section of the chapter provides a comprehensive evaluation and interpretation of the study's findings in relation to the three research questions. The first two research questions investigate the potential effects of incorporating PF into OSCE simulations on PA students' self-efficacy, emotional stress responses, and knowledge competency. Specifically, these questions explore whether the introduction of PF impacts students' confidence in their clinical decision-making, influences their ability to manage stress during high-pressure scenarios, and affects their overall performance in competency-based assessments.

The third research question seeks to uncover participants' reflections on their learning experiences within the PF-based OSCE simulations. This inquiry aims to explore how students perceive the value of struggling through initial failure and uncertainty, how they process and adapt to challenges, and whether they identify PF as a beneficial instructional method for their clinical training. By examining participants' qualitative insights, this study aims to shed light on the broader implications of PF for cognitive adaptability, resilience, and critical thinking in PA education.

Through this analysis, the researchers sought to develop a deeper understanding of the pedagogical impact of PF within PA OSCE simulations. By evaluating both the measurable outcomes and subjective experiences of participants, this evaluation aims to determine whether PF serves as an effective instructional strategy for enhancing PA students' preparedness for real-world clinical practice.

Research Question 1 (RQ1)

This study examined the effects of implementing PF in OSCE simulation activities on participants' self-efficacy and emotional stress responses. Descriptive data from both the control and intervention groups indicated a general decline in self-efficacy and an increase in emotional stress responses following the simulation. This was reflected in lower post-simulation self-efficacy scores and higher post-simulation emotional stress scores in both groups. However, a paired t-test revealed that the differences between pre-simulation and post-simulation scores within each group were not statistically significant over time. The change closest to statistical significance was observed in the control group's emotional stress response, which decreased by -4.750 with a p-value of 0.091.

Further analysis using an Independent Samples t-test demonstrated that the mean differences between pre-simulation and post-simulation scores were not statistically significant between the control and intervention groups. In other words, self-efficacy and emotional stress response scores remained comparable before and after the simulation, regardless of whether participants were in the control or intervention group. This finding suggests that the PF intervention did not have a statistically significant negative effect on self-efficacy or emotional stress responses.

Interestingly, the intervention group consistently reported lower self-efficacy and emotional stress response scores both before and after the simulation. This suggests that, while the control group initially had higher self-efficacy, they also experienced greater emotional stress following the simulation compared to the intervention group. However, these differences did not reach statistical significance. Given the small sample size and limited statistical power, these trends may reflect meaningful effects that the study was underpowered to detect, despite moderate to large effect sizes observed in some comparisons.

To explore potential predictive relationships between study variables, a Pearson's r test was conducted. This analysis sought to determine whether any study variables were significantly correlated with one another and to assess the potential effects of the PF intervention. Two statistically significant relationships were identified. First, a strong positive predictive relationship was found between pre-simulation self-efficacy scores and post-simulation emotional stress responses. Participants who reported higher self-efficacy before the simulation tended to experience greater emotional stress afterward. This finding suggests that individuals who enter the simulation with higher confidence may be more affected by challenges and failures encountered during the experience. However, when this relationship was analyzed within each group separately, no statistically significant predictive relationship was found between pre-simulation self-efficacy and post-simulation emotional stress responses for either the control or intervention group.

Second, a very strong positive correlation was identified between the control group's pre-simulation and post-simulation emotional stress responses. This indicated that participants in the control group who had higher emotional stress before the simulation were more likely to experience increased emotional stress afterward. Notably, this relationship was statistically significant only for the control group, further reinforcing that the PF intervention did not have a significant impact, positive or negative, on emotional stress responses.

In summary, this study found no statistically significant impact of PF in OSCE simulation activities on participants' self-efficacy or emotional stress responses. Moreover, there was no evidence that PF had a detrimental effect on these variables, addressing a common concern in medical education regarding the potential risks of introducing failure-based learning experiences. These findings support the notion that PF can be integrated into PA education without negatively affecting students' confidence or emotional well-being. Despite the underpowered constraints of this study, the absence of statistically significant negative effects provides preliminary evidence about the safe use of PF in simulation-based learning

Research Question 2 (RQ2)

This study examined the effects of implementing PF in OSCE simulation activities on participants' OSCE competency scores. To assess whether the PF intervention influenced performance, an Independent Samples t-test was conducted to compare mean OSCE scores between the control and intervention groups. The results indicated no statistically significant difference between the two groups, suggesting that the PF intervention did not have a positive or negative impact on participants' knowledge competency or OSCE performance. Notably, despite the limitations of a small and underpowered sample, these findings provide early evidence that integrating PF into PA OSCE simulations does not compromise students’ clinical competency performance.

To further explore potential relationships between OSCE competency scores and other study variables, a linear regression analysis was conducted. While multiple variables were examined, only post-simulation emotional stress response demonstrated a statistically significant predictive relationship with OSCE competency scores. Prior Pearson’s r correlation analysis revealed a strong negative relationship between post-simulation stress and OSCE performance, which was further supported by these regression results. Specifically, for every one-unit increase in post-simulation emotional stress, OSCE scores decreased by approximately 0.802 points, with a p-value of 0.042. This suggests that emotional stress following the simulation had a more substantial impact on OSCE performance than self-efficacy or pre-existing stress levels. Consequently, managing post-simulation stress may be a crucial factor in optimizing competency outcomes in PF-based OSCE simulations.

Additionally, pre-simulation emotional stress response exhibited a moderate negative relationship with OSCE competency scores. However, while the regression model approached statistical significance, p-value of 0.099, it did not meet the threshold for significance. This finding suggests that both pre-simulation and post-simulation stress responses warrant further investigation, particularly with larger sample sizes, to better understand their influence on OSCE performance.

Overall, these results highlight the importance of considering emotional stress management when incorporating PF into PA OSCE simulations. While in this study, PF did not negatively impact clinical competency, higher post-simulation stress levels may hinder performance, emphasizing the need for strategies to help students regulate stress and build resilience in medical training.

To further evaluate RQ2, a multiple regression analysis was conducted to examine whether demographic characteristics had any significant impact on OSCE competency scores. Specifically, the analysis assessed whether age, race, gender, marital status, number of core clinical rotations, and completion of an ER rotation influenced participants' performance in OSCE simulations. The results revealed there was no statistically significant relationship between any of these demographic variables and OSCE competency scores. However, completion of an ER rotation showed a positive association with higher OSCE scores, suggesting that prior emergency medicine experience may be beneficial in clinical performance. Despite this observed trend, the effect did not reach statistical significance, with a p-value of 0.222, indicating that while the relationship appears strong, it could not be conclusively supported within the sample size of this study.

These findings suggest that demographic factors did not play a significant role in influencing OSCE performance, reinforcing that clinical competency in OSCE simulations is likely shaped more by experiential learning and training methodologies, rather than inherent demographic characteristics. Further research with larger sample sizes may be needed to explore whether specific prior experiences, such as ER rotations, provide measurable advantages in competency-based assessments.

Research Question 3 (RQ3)

The thematic analysis of the focus group discussions identified several overarching themes that provided valuable insight into how PA students reflected on their learning experience with PF. These themes emerged from frequent and recurring patterns in the qualitative data related to participants' experiences during the PF OSCE simulation. The first common theme that emerged through the analysis was *Emotional and Cognitive Growth Under Pressure*. This theme highlighted the overall positive impact of the PF OSCE simulation on participants, emphasizing emotional resilience and cognitive development in response to high-stress clinical scenarios. While participants acknowledged the heightened stress levels associated with facing failure in clinical decision-making and the demands of fast-paced environments, they consistently recognized that these challenges contributed to their cognitive growth and adaptability.

Many participants expressed that experiencing failure within the simulation strengthened their critical thinking, adaptability, and problem-solving skills, reinforcing their desire to improve their clinical performance. Despite the initial emotional strain, participants viewed the experience as a valuable opportunity to refine their clinical reasoning, self-confidence, and decision-making abilities.

These qualitative findings align with the study's quantitative data, which demonstrated that PF simulations did not result in statistically significant negative impacts on participants' emotional stress responses or competency scores. This supports the hypothesis that incorporating PF OSCE experiences into PA education provides meaningful learning benefits that outweigh potential emotional and cognitive risks.

The second prominent theme that emerged was the *Development of Clinical Confidence and Decision-Making Abilities.* The focus group analysis revealed that participants initially experienced a short-term decline in confidence following failed clinical decisions during the PF OSCE simulation. However, rather than discouraging them, these setbacks served as a catalyst for growth, encouraging participants to persevere, reassess their approach, and seek innovative solutions to their clinical challenges. Ultimately, this process led to an improvement in their overall confidence, problem-solving abilities, and critical thinking skills.

This theme aligns with the study's quantitative findings, which indicated that despite initial emotional stress and a temporary decline in self-efficacy, participants' confidence and knowledge competency did not significantly deteriorate as a result of the PF intervention. Instead, participants described a process of rebuilding their confidence and refining their critical thinking skills in response to failure.

This process directly relates to the third overarching theme*, Adaptability and Cognitive Flexibility in Complex Situations*. Participants reported that encountering failure required them to "reset" their approach, adjust their decision-making strategies, and adapt to the dynamic nature of the clinical environment. Through this experience, they became more cognitively flexible, recognizing knowledge gaps and modifying their clinical reasoning to bridge those gaps.

Furthermore, these findings are consistent with constructivist learning theories proposed by Bruner, Dewey, Kolb, Piaget, and Vygotsky. These theorists emphasized that learning is an active process that occurs through interaction with the environment, reflection, and adaptation to that environment (Bruner, 1966; Dewey, 1938; Kolb, 1984; Piaget, 1977; Wood, 1976). These theorists suggest individuals learn by engaging with challenges, constructing new knowledge, and developing adaptive responses, all of which were observed in participants' experiences with PF OSCE simulations.

The fourth overarching theme identified in the analysis was *Progression Toward Autonomous Clinical Practice*. This theme emerged from participants' shared experiences of increased responsibility and independent clinical decision-making, a significant shift from their previous learning experiences in both the didactic and clinical phases of PA training.

Many participants expressed that the PF OSCE simulation was their first exposure to fully autonomous decision-making in a clinical setting, an experience they had not encountered in prior training. They frequently described these moments as "real-world" experiences that closely mirrored the pressures and responsibilities of actual clinical practice. Participants recognized that the weight of making independent clinical judgments in the PF OSCE reflected a necessary progression toward real-world clinical responsibilities, marking an important step in their transition from learners to competent practitioners.

Notably, many participants emphasized that this was the first time they truly felt accountable for patient outcomes, reinforcing the importance of exposure to realistic, high-stakes decision-making scenarios in PA training. Their reflections highlighted the need for more real-world experiences in PA education and underscored the value of productive failure as a training tool. The perceived authenticity and relevance of the PF OSCE experience were highly valued, further supporting its integration into PA curricula to enhance clinical preparedness, independent reasoning, and decision-making confidence.

The fifth overarching theme identified was *Leadership and Teamwork in High-Stakes Environments.* This theme emerged as participants reflected on the challenges of functioning in acute clinical scenarios that required both collaborative teamwork and decisive leadership. Many participants recognized their knowledge gaps and expressed a strong need for support from a clinical team when faced with high-pressure decision-making. At the same time, they also identified the importance of stepping into leadership roles to direct the team effectively in critical situations.

This unexpected but significant finding reinforces the value of PF in exposing personal strengths, areas for growth, and essential team-based skills necessary for high-stakes healthcare environments. The experience highlighted the necessity of both adaptability and assertiveness in clinical teamwork, demonstrating that PF not only strengthens individual competence but also fosters collaborative dynamics crucial in acute care settings.

The sixth overarching theme was *Clinical Relevance and Lifelong Learning*. Participants overwhelmingly described the PF OSCE simulation as a realistic and immersive learning experience that bridged the gap between theoretical knowledge and practical application. They emphasized that the hands-on nature of PF simulations helped solidify critical clinical concepts, making the learning experience more tangible and meaningful.

Additionally, many participants expressed that the lessons learned from their mistakes during the PF experience would remain with them for years, shaping their clinical decision-making and reinforcing the importance of avoiding similar errors in future clinical practice. This theme supports the integration of PF into PA education by demonstrating its role in cultivating a mindset of continuous learning and self-improvement, a crucial trait for lifelong competency in medical practice.

The final overarching theme that emerged was *Enhanced Critical Thinking and Reflective Practice*. While similar to the second and third themes, which focused on developing confidence and adaptability, this theme specifically emphasized the role of self-reflection in cognitive adaptation. Participants recognized that their ability to analyze and refine their clinical decision-making was significantly influenced by their engagement in a reflective process during the PF OSCE experience.

Notably, participants described engaging in positive self-talk and internal dialogue in response to their failed clinical decisions. Rather than becoming discouraged, they actively reflected in real-time, considering alternative approaches and generating new potential solutions to the clinical problems presented. This reflective process enhanced their critical thinking abilities, ultimately improving their approach to patient care.

These findings align with Schön's (1983) theory of reflective practice, which differentiates between *reflection-in-action* (real-time critical thinking and adjustment) and *reflection-on-action* (post-experience learning). The PF OSCE experience facilitated the *reflection-in-action*, reinforcing the idea that real-time cognitive processing under pressure strengthens clinical reasoning and problem-solving skills.

The significance of these insights further supports the integration of PF into PA education, highlighting its potential to cultivate a deeper, more dynamic learning process. By encouraging students to think critically, adapt, and reflect during their clinical experiences, PF provides a valuable instructional framework that better prepares PA students for the complex and unpredictable nature of medical practice.

In conclusion, across multiple themes, participants consistently recognized the value of failure as a learning tool, acknowledging its role in strengthening clinical reasoning, adaptability, and decision-making. Participant reflections on their experiences described how the challenges of failure initially led to increased stress and temporary declines in confidence but ultimately enhanced their problem-solving skills, self-efficacy, autonomous decision-making, cognitive adaptability, and clinical preparation for practice. These insights demonstrate that PA students perceive PF OSCEs as a valuable educational tool that fosters resilience, deeper learning, and clinical readiness. Despite the failure initially increasing stress, they ultimately enhanced critical thinking, adaptability, and confidence, reinforcing the importance of integrating PF into PA training to better prepare students for real-world medical practice.

Limitations of the Study

The limitations section of this chapter plays a vital role in recognizing potential weaknesses in the study's methodology, external factors that may have influenced the results, and constraints in the research design. Its primary purpose is to ensure transparency by identifying key factors that could have impacted the study's outcomes, such as sample size, measurement tools, data collection methods, and external validity concerns (Creswell, 2018). By addressing these limitations, the researchers aim to help readers interpret the findings with appropriate caution and within the proper context.

Acknowledging study limitations is essential for enhancing credibility and maintaining academic integrity, as it prevents misleading interpretations of the results. This critical evaluation of potential biases and constraints strengthens the academic rigor of the research (Polit, 2021). Additionally, discussing these limitations helps assess the applicability and generalizability of the findings to a broader population, contributing to the advancement of knowledge in the field.

Moreover, by identifying areas for improvement, this section provides valuable guidance for future research on productive failure (PF), helping refine methodologies and improve study designs in subsequent investigations.

Sample Size

The study population included a maximum of 35 potential participants from a single institution, thus constraining the sample size. The result of a small sample size can reduce the statistical power, leading to an increased likelihood of type II errors, failing to detect a true effect when one exists (Cohen, 2013a). Additionally, a small sample size limits the ability to generalize findings beyond the study population, reducing external validity (Polit, 2021).

Participant attrition can further reduce the effective sample size. Some students may opt out of participation, experience schedule conflicts, fail to complete all study components, or experience other external factors that can impact their ability to complete the study components. The variability in student engagement can also contribute to increases within-group variance, reducing the ability to detect meaningful differences between experimental conditions. In this study, the recruitment and resulting small sample size may have been influenced by attrition and the timing of the study implementation. The study was conducted during the fall semester, approximately six months before graduation, a critical period for students in the PA program. At that point, most students were actively preparing for and needed to complete three or more high-stakes, nationally standardized clinical rotation exams, all required for program progression. This intense academic and cognitive demand may have led some students to triage their participation in additional learning activities like this study, thereby affecting recruitment. The added cognitive load and competing priorities could have contributed significantly to the reduced participation rate, further limiting the study's statistical power and generalizability.

To address these limitations, future studies should consider recruiting PA student participants from multiple institutions to contribute to a larger sample size. However, a multi-institute study does pose a challenge in finding a “common point” in PA student training that provides equality across participant experiences and knowledge. Additionally, future researchers should consider a power analysis before conducting the study to determine the minimum required sample size needed to detect statistically significant effects. A post-hoc power analysis conducted for this study indicated that a sample size of approximately 26 participants would be necessary to improve the likelihood of detecting true effects and enhance both effect size estimation and statistical power.

Selection Bias

Sampling bias can be a significant limitation in studies involving small, specialized populations such as PA students. This bias can arise if the sample does not reflect the diversity of the broader PA student population. This can be particularly noted in demographic characteristics of age, race, and prior experiences. Additional factors such as convenience sampling, voluntary participation, and institutional constraints may lead to an overrepresentation of certain student characteristics, such as higher motivation levels, previous clinical experience, or academic performance, which could influence study outcomes (Creswell, 2018). These factors can limit the external validity of the findings, making it difficult to generalize results to other PA programs or clinical training environments.

An additional concern is the demographic homogeneity of participants. A study that includes participants from a single PA program or similar educational backgrounds may not account for variations in training methods, faculty instruction, or student diversity seen across different programs. Additionally, self-selection bias can occur if students who are more confident in their abilities or have prior OSCE experience are more likely to participate, potentially skewing self-efficacy and competency results (Polit, 2021). Since this study was conducted before the completion of all core clinical curricular examinations, the added cognitive and emotional demands may have influenced students' decisions to decline participation in additional research activities. As a result, those who chose to participate may differ systematically in motivation, stress tolerance, or academic bandwidth compared to those who declined, potentially skewing the findings and limiting generalizability. This can result in underestimating or overestimating the actual impact of productive failure on self-efficacy, emotional stress responses, and OSCE competency.

To mitigate sampling bias, future studies should consider larger, more diverse samples across multiple PA programs and geographic regions to improve generalizability. Random sampling techniques and stratified sampling methods can help ensure that the participant pool reflects the broader PA student population (Etikan, 2017). Collecting demographic data on participants and comparing study results across different subgroups can also help researchers better understand the extent of bias and its impact on findings. Additionally, future research should consider implementing PF interventions closer to, or after, the completion of core curricular requirements to minimize the influence of excessive cognitive load. This timing may help reduce self-selection bias by encouraging broader participation from students who might otherwise opt out due to academic stress and competing high-stakes assessment demands.

Misrepresentation of Effect Size

Effect size estimates, such as Cohen’s d, are highly sensitive to sample size variations. Small sample sizes may yield inflated effect sizes or wide confidence intervals, reducing the precision and reliability of the results (Button, 2013). The observed effect may be overestimated, leading to misrepresentations of the outcome and the intervention's efficacy. In future studies, improving sample size and calculating power analysis can mitigate this potential limitation.

Timing of Implementation within Training

Many external factors can affect the implementation of research or new pedagogical approaches to instruction in the dense schedule of the PA curriculum. This study was confined to implementation in the clinical phase of PA student training and to the fall semester, approximately six months before program completion, due to the availability in student schedules. The timing of the implementation can have an impact on the homogeneity of participants within a study, as it pertains to prior training experiences and knowledge. PA student clinical experience schedules and context present a robust array of variation within and between programs. The timing of implementation, for PF research, can pose a significant challenge during the clinical phase, given these variations. Finding a "common point" of knowledge and experiences in PA training, both within or among PA programs, can impact the comparability of participants and the study outcomes.

To mitigate this limitation, a larger and more diverse sample should be sought and randomized. Additionally, implementing PF experience near or after the completion of the required clinical experience may improve the quantity and equality of participants within a study.

Short Study Duration

A short study duration can present significant limitations when investigating the effects of PF on PA student learning through OSCE simulation activities. The limited timeframe may restrict the ability to capture long-term learning outcomes, particularly when evaluating self-efficacy, emotional stress responses, and competency. Learning processes, especially those involving productive failure, often need extended time for reflection, adaptation, and mastery

(Kapur, 2016). With a shorter study period, participants may not have sufficient time to process and integrate the learning gained from the PF experience, potentially underestimating the intervention's true impact on OSCE performance and clinical skills development.

Furthermore, a short study duration may hinder the generalizability of findings by failing to account for variations in learning trajectories across different clinical rotations or curricular stages. PA students' self-efficacy and emotional stress responses might fluctuate depending on their clinical exposures and experiential learning opportunities.

Moreover, short-duration studies may limit the ability to establish causal relationships between PF interventions and long-term clinical performance. OSCE simulations, by design, provide immediate performance feedback. However, sustained improvements in clinical reasoning and decision-making may require repeated practice and ongoing feedback (Issenberg, 2005). To mitigate these limitations, future research should consider implementing longitudinal designs that track PA students' performance over multiple OSCEs across varied clinical settings. Incorporating follow-up assessments could provide valuable insights into the lasting effects of productive failure on self-efficacy, emotional stress, and competency development.

Lack of Pre-assessment of Knowledge Competency

The lack of a comprehensive pre-assessment of knowledge competency represents a significant limitation in studies exploring PF. Without a baseline understanding of participants' existing knowledge and skills, it becomes challenging to accurately interpret performance outcomes following the PF intervention. Pre-existing competency levels can influence self-efficacy, emotional stress responses, and OSCE performance. The absence of pre-assessment data may lead to misattributions of learning gains to the PF intervention when, in fact, they may stem from pre-existing knowledge differences.

Additionally, the heterogeneity of participants' prior experiences further complicates data interpretation in the absence of pre-assessment measures. PA students enter training programs with varying levels of clinical exposure, academic preparation, and practical skills. Failing to account for these initial differences can introduce confounding variables, ultimately weakening the study's internal validity (Creswell, 2018).

Future studies should consider employing robust pre-assessment competency protocols to enhance result validity and contribute to a more nuanced understanding of how PF influences PA student learning during OSCE simulations. Completion and scores on the Physician Assistant Education Association (PAEA) End of Rotation (EOR) examinations may serve as a potential pre-assessment tool, as these are nationally standardized assessments and provide national means and comparisons.

Measurement Constraints

Measurement constraints can represent a critical limitation in research. The Generalized Self-Efficacy Scale (GSES) is a validated tool used to assess self-perceived competence in coping with challenging situations. However, self-efficacy is inherently subjective and may not accurately reflect actual clinical performance in OSCE tasks. Self-reported biases, such as social desirability or the overestimation or underestimation of abilities, may distort scores, particularly in higher-stakes simulation-based settings (Schwarzer, 2010).

Similarly, the Short Stress State Questionnaire (SSSQ) was designed to measure cognitive, emotional, and physiological stress responses, which may face sensitivity limitations when applied to OSCE simulations. The OSCE environments are designed to simulate real-world clinical scenarios and may elicit a stress response distinct from those captured by the SSSQ's standardized framework (Helton, 2004). Certain situational factors, such as evaluator observation and time constraints, might introduce extraneous variables that the SSSQ cannot fully account for, thereby affecting the consistency of the stress response measurement across participants.

The Global Rating Scale (GRS) and Task-Specific Checklist (TSC) present their own measurement challenges. While the GRS provides a holistic performance rating, it may be susceptible to subjective rater bias. Conversely, the TSC offers objective, behavior-based assessment, but may overlook critical contextual factors influencing task execution. The inter-rater reliability of these tools also requires careful consideration to ensure consistency across different OSCE scenarios and evaluators. The combined use of these instruments, while comprehensive, may present measurement constraints that can affect the accuracy of findings in PF-based research.

Future research may consider either longitudinal designs or a repeated measures design with inter-rater reliability analysis within the design to improve validity. Specific facilitator training on the use of each of these measures, and an evaluation of inter-rater reliability, may improve the credibility of the design and utilization of these instruments. Additionally, larger sample sizes may allow for validation of the GSES and SSSQ instruments use in OSCE applications.

Intervention Complexity

The intervention of PF is inherently complex as they are designed to foster learning through struggle and error. This requires careful design and implementation to effectively replicate clinical challenges. The intricacy of these interventions can impact participant engagement, performance outcomes, and the overall validity of the study's findings.

Moreover, the design and execution of PF tasks in OSCE simulations demand high-fidelity environments that mimic clinical practice. The complexity of these simulations includes the use of clinical equipment, time-sensitive tasks, and roles played by facilitators, the ER technician. These components can introduce confounding factors such as technical issues, inconsistent role-play performance, or variability in the timing of tasks. Each of these might influence student responses to the simulation environment, obscuring the effects of the PF intervention.

Additionally, PA students' cognitive load during simulated tasks may be heightened due to the novelty of the PF experience, potentially impairing performance outcomes (Sweller, 1988). An additional layer of complexity arises from the multifaceted nature of PF learning. PF tasks are designed to engage higher-order thinking, requiring participants to analyze, problem-solve, and self-reflect on clinical decision-making (Kapur, 2012). Overly complex tasks may overwhelm participants, leading to stress and disengagement, while tasks that are too simple might fail to trigger meaningful productive failure experiences. This delicate balance makes standardizing the intervention across diverse learner profiles particularly challenging.

Furthermore, facilitator training and consistency present additional complexities. PF interventions require facilitators to maintain neutrality during student struggles while providing constructive feedback after the simulation (Loibl, 2017). Differences in facilitator experience, feedback delivery styles, and observational consistency can impact how students perceive and respond to the PF experience.

Addressing these intervention complexities necessitates meticulous planning, pilot testing, and continuous quality monitoring. Future research might benefit from standardized intervention protocols, facilitator calibration sessions, and participant feedback mechanisms to refine intervention design and improve study reliability.

Ethical and Emotional Considerations

Productive failure-based learning introduces deliberate challenges designed to promote deeper learning through initial struggle (Kapur, 2008). However, the emotional toll associated with failure experiences and the ethical obligations to protect participants' well-being must be carefully considered.

Ethical considerations in PF OSCE simulations primarily revolve around participant autonomy, well-being, and informed consent. Given the inherent nature of productive failure, students must be made aware that initial performance struggles are an intentional part of the learning process. Misunderstanding this instructional design may lead to unnecessary distress or feelings of incompetence (Kapur, 2016). Additionally, protecting student anonymity and confidentiality presents ethical challenges. OSCE assessments are often evaluated by faculty, who might also participate in the research process. Clear boundaries must be established to prevent conflicts of interest and ensure that performance data does not impact academic evaluations or progression. Finally, students' participation in the study must be voluntary, with mechanisms in place to withdraw without academic repercussions.

The emotional impact of productive failure constitutes another significant limitation. Productive failure interventions are designed to induce temporary struggle to promote long-term learning gains (Kapur, 2012). However, exposure to repeated failures during high-stakes OSCE simulations may evoke stress, frustration, and diminished self-efficacy, particularly in novice learners. Emotional distress induced by productive failure experiences might negatively impact performance outcomes or discourage future engagement in simulation-based learning.

Furthermore, simulated clinical environments require participants to manage both clinical tasks and emotional responses simultaneously. The emotional toll of repeated failure can overload cognitive resources, leading to increased anxiety and reduced learning efficacy (Sweller, 1988). Participants' psychological well-being must be monitored, with support mechanisms in place, such as debriefing sessions or psychological check-ins that are integrated into the study design (Kolbe, 2021).

To mitigate these limitations for future studies, clear explanations should be implemented in the informed consent process of PF research. Structured debriefing sessions to support emotional responses can be implemented by framing failure experiences as learning opportunities. Finally, the implementation of appropriate faculty training and understanding of the process and goals of PF research. Faculty training should include an understanding of impartiality in the failure process, but additionally include their role in providing supportive feedback during debriefing sessions to mitigate the emotional impact of failure.

Implications of Productive Failure

This section of the chapter focuses on the insights gained from the study and the potential implications of these findings to education, clinical practice, and future research. These implications are designed to demonstrate how the study findings can be used in practical settings both now and in the future. This section aims to align the study's findings with current research on PF and expand upon it, contributing knowledge to the discipline and concept.

Implications for Education

The primary objective of medical education is to prepare learners for the complexities of clinical practice. However, clinical environments are inherently unpredictable, requiring healthcare providers to navigate uncertainty, complex decision-making, and inevitable failures in patient care. Despite these realities, PA and medical curricula often do not systematically incorporate failure as a learning tool. This gap raises an essential question: How can medical education effectively prepare students for real-world clinical challenges if it does not expose them to failure in a structured and meaningful way?

The findings of this study provided early evidence supporting the integration of PF into PA education. Despite low statistical power and effect size, this study demonstrated that the perceived negative consequences of PF, such as increased emotional stress, decreased self-efficacy, or reduced competence, were not statistically significant. Instead, the benefits of PF in fostering resilience, adaptability, and critical thinking far outweigh these concerns. Existing research in non-medical disciplines has shown that students who initially struggle in a learning task yet persist and actively engage in finding solutions, develop deeper conceptual understanding compared to those who receive direct instruction without challenge (Kapur, 2008). These struggles present opportunities for cognitive and emotional growth, as the learners develop perseverance and intrinsic motivation in response to failure (Dweck, 2016; Loibl, 2014). This study confirmed these findings through qualitative focus group discussions, where the themes of *Emotional and Cognitive Growth Under Pressure* and *Adaptability and Cognitive Flexibility in Complex Situations* emerged. Participants acknowledged that initial failures provided a powerful learning experience, allowing them to reflect on their weaknesses, adapt their approach, and build confidence in their ability to problem-solve. The significance of these findings in medical education implies an effective preparation for real-world clinical experiences. Providing these experiences in training curricula allows opportunities for students to begin to develop and advance these cognitive and emotional skills, preparing them for clinical practice.

Medical practice also exposes providers to complex scenarios with complex decision-making requirements. This study showed that by allowing participants to initially struggle with complex problems, they recognized their own knowledge gaps and developed adaptive strategies to resolve the clinical problems. Additionally, participants recognized that the reflective thinking process was significantly enhanced during the PF simulation. Instead of simply recalling facts or following standard protocols, they actively engaged in critical self-assessment, considering what went wrong, why it happened, and how they could improve their approach in future situations. Many participants reported that these failed experiences were so impactful that they would likely never forget them and that the lessons learned would directly influence their future clinical practice as PAs. This deep level of self-reflection and adaptive learning aligns with Schön's (1983) theory of reflective practice, which emphasizes that professionals develop expertise not just by accumulating knowledge but by continuously reflecting on their experiences, learning from mistakes, and refining their approach in real-time.

The findings of this study confirm that PF can be safely and effectively integrated into PA education curricula to enhance student learning and clinical preparedness. The results suggest that incorporating PF into training offers significant advantages over traditional instructional methods, which often focus on direct instruction and minimizing failure. Importantly, the study found that the potential concerns associated with failures, such as increased emotional stress responses, decreased self-efficacy, or diminished competence, were not statistically significant. This suggests that while students initially experience challenges and stress, these difficulties do not have a lasting negative impact on their confidence or competence abilities. Instead, they fostered resilience, adaptability, and deeper learning, ultimately benefiting the students' long-term professional development.

Given these insights, medical educators should prioritize the creation of safe, structured, and realistic learning environments where failure is framed as a valuable part of the learning process. Future educational strategies should focus on intentionally designing PF-based activities that allow students to struggle with complex clinical problems while receiving appropriate guidance and support. This approach ensures that the lessons learned from failure are constructive rather than discouraging, ultimately preparing PA students more effectively for the unpredictable and demanding nature of clinical practice.

Implications for Medical Practice

Integrating PF into PA education and training carries significant implications for patient care, medical decision-making, and professional development. By exposing learners to failure in a controlled, low-stakes educational setting, PF better prepares healthcare providers for the uncertainties and challenges they will inevitably face in real-world clinical environments. PA providers frequently encounter unexpected complications and evolving patient conditions, requiring them to adapt their clinical decisions accordingly. PF-based simulations provide structured opportunities for learners to develop resilience, critical thinking, and problem-solving skills, which are essential for managing these challenges effectively in clinical practice.

This study demonstrated that PF fosters clinical confidence and autonomy in decision-making under pressure. Current research suggests that students who experience controlled failure learn to rebuild their confidence through reflection and adaptation, a process that is essential for handling real-world clinical responsibilities (Kapur, 2012). In demanding medical environments such as emergency medicine or intensive care, providers must be able to quickly recognize, assess, and correct errors to ensure the best possible patient outcomes. However, even when providers deliver high-quality care, optimal patient outcomes are not always achieved. In such cases, doubts in clinical decision-making and decreased confidence can contribute to provider burnout.

By incorporating failed experiences into PA training through high-stakes OSCEs, educators may help develop coping strategies and self-reflective processes that mitigate burnout. Burnout is a critical issue in healthcare, as it affects provider well-being, patient safety, and retention of skilled medical professionals. Implementing PF as part of professional training programs may equip providers with the resilience and reflective capacity needed to navigate clinical uncertainties, ultimately reducing burnout and improving provider longevity. Addressing provider burnout is essential for maintaining high-quality patient care, retaining competent healthcare professionals, and ensuring access to quality care.

Another key implication of PF in clinical practice is its impact on interprofessional teamwork. Research has shown that many healthcare errors stem from communication failures among medical teams (Salas, 2005). PF-based simulations encourage collaborative problem-solving, effective communication, and shared decision-making, all of which are critical for reducing preventable medical errors. This study revealed that participants recognized the need for clear communication, leadership, and teamwork in high-pressure clinical settings. These findings suggest that PF-based simulations more effectively prepare medical professionals for real-world interdisciplinary teamwork, leading to improved patient safety and overall healthcare outcomes.

The integration of PF into medical education and clinical practice has the potential to transform how healthcare professionals are trained. Productive failure enables learners to develop a deeper understanding of clinical reasoning, build resilience, and refine their decision-making skills, all of which are essential attributes for competent and adaptable medical providers. By further refining and integrating PF strategies into PA curricula and professional training, educators and clinicians can cultivate more confident, competent, and adaptable healthcare professionals, ultimately enhancing patient care, provider well-being, and the overall effectiveness of the healthcare system.

Implications for Future Research

One critical area for future research is conducting longitudinal studies to evaluate PF's impact on competency development over time. While this study found no statistically significant negative impact on competency, larger-scale quantitative studies with improved statistical power and robust pre-intervention and post-intervention assessments are needed. These studies should track whether students trained with PF demonstrate better competency scores, improved clinical reasoning, or greater retention of knowledge and skills compared to those trained using traditional methods. Establishing long-term learning benefits and real-world application of PF-based training will be crucial in determining its overall effectiveness in PA education and medical training.

Future research should also focus on quantifying the balance between emotional stress and the failed experience in PF-based simulations. The current study revealed variability in participants' emotional responses, with some individuals experiencing heightened stress both before and after the simulation, even in the control group that did not receive the PF intervention. This finding raises concerns about how individual emotional stress levels may influence learning outcomes, suggesting the need for effective methods to evaluate and account for emotional responses prior to the simulation.

One potential future approach is to stratify randomization based on emotional stress factors rather than clinical experience alone. By measuring participants' baseline emotional stress responses and incorporating this factor into the study design, researchers may gain a better understanding of how different learners react to PF and develop strategies to optimize the emotional experience for maximum learning benefit.

As PF in medical education remains relatively underexplored, future studies should seek to identify the optimal balance between failure and instructional guidance. This study provided minimal guidance during the OSCE simulation, with the only external aids being an ER technician directing participants to the automated external defibrillator (AED) and a contraindication sheet for thrombolytic therapies. While PF emphasizes learning through failure, excessive failure without adequate scaffolding can lead to frustration, discouragement, and disengagement from learning (Kapur, 2016).

Future research should investigate how much instructional support is necessary to maintain motivation without undermining the benefits of productive struggle. This could involve testing different levels of scaffolding, such as providing limited instructor prompts, offering structured debriefing sessions, or adjusting difficulty levels dynamically based on learner performance. Identifying an optimal balance between independence and guidance will help ensure that PF fosters resilience and problem-solving skills without overwhelming learners.

Lastly, this study focused on PF implementation in an emergency room-based OSCE simulation. However, future studies should evaluate PF across a variety of medical specialties and clinical environments to determine its broader applicability. Different specialties, such as surgery, primary care, psychiatry, and internal medicine, present unique challenges and decision-making processes, which may require tailored PF approaches.

By exploring how PF influences learning and competency development across diverse medical fields, researchers can determine its global utility in medical education. Additionally, studying PF in varied patient care settings, including outpatient clinics, operating rooms, and intensive care units, will provide valuable insight into how different clinical environments shape the learning experience and whether PF is universally beneficial or requires specialized adaptations based on the context.

While this study provides preliminary evidence supporting the integration of PF into PA education and medical training, further research is needed to optimize its implementation, assess its long-term impact on competency, and refine instructional strategies to balance emotional resilience with effective learning. Expanding PF research across varied clinical specialties and environments will help establish best practices for incorporating failure-based learning into medical curricula, ensuring that future healthcare providers are better equipped to navigate complex clinical challenges with confidence, adaptability, and critical problem-solving skills.

Recommendations and Conclusion

Based on the findings of this study, PF should be intentionally integrated into PA education to enhance clinical preparedness, decision-making, adaptability, reflective processes, and resilience in future healthcare providers. The evidence suggests that exposing learners to controlled failure in a structured, supportive environment fosters deeper conceptual understanding and the transferability of theoretical to practical knowledge. Given the need to effectively prepare students for clinical practice, PA programs should strategically implement PF-based learning experiences, ensuring that students engage in realistic, yet psychologically safe challenges that mirror clinical realities.

To maximize its effectiveness, PF should be incorporated into existing curricula through structured simulation-based learning, problem-based case studies, and competency-driven assessments. High-stakes OSCEs that simulate challenging, uncertain, and evolving patient cases would allow students to confront failure, reflect on their decision-making processes, and develop adaptive strategies for improvement. Additionally, supportive debriefing sessions should be standardized to help students process their experiences, recognize the value of failure in learning, and build constructive coping mechanisms that support long-term clinical confidence.

Furthermore, educators should carefully calibrate the level of failure introduced in PF-based exercises to prevent overwhelming students with frustration. Research should continue to refine optimal scaffolding techniques, such as incorporating guided reflection and instructor-led discussions to maintain engagement and motivation. Faculty development programs should also train instructors on how to effectively facilitate PF experiences, ensuring that failure is framed as an opportunity for learning rather than a punitive or discouraging event.

Finally, longitudinal studies should be conducted to evaluate the long-term impact of PF on competency development and patient care outcomes. Future research should explore how PF-trained students perform in clinical rotations and post-graduate practice compared to those trained under traditional instructional methods. Additionally, expanding PF studies across multiple specialties and patient care settings will help determine its broader applicability in medical education. By continuing to refine and implement PF-based training, PA programs can cultivate more competent, confident, and adaptable practitioners, ultimately improving patient care, clinical outcomes, and provider wellness.

References

Abercrombie, S., Carbonneau, K. J., & Hushman, C. J. (2022). (Re)Examining academic risk taking: Conceptual structure, antecedents, and relationship to productive failure. *Contemporary Educational Psychology*, *68*, 102029. https://doi.org/10.1016/j.cedpsych.2021.102029

Abulebda, K., Auerbach, M., & Limaiem, F. (2025). Debriefing Techniques Utilized in Medical Simulation. In *Stat Pearls*. Stat Pearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK546660/

Albanese, M. (2000). Problem-based learning: Why curricula are likely to show little effect on knowledge and clinical skills. *Medical Education*, *34*(9), 729–738. https://doi.org/10.1046/j.1365-2923.2000.00753.x

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues [published erratum appears in Acad Med 1993 Aug;68(8):615]. *Academic Medicine*, *68*(1), 52–81. https://doi.org/10.1097/00001888-199301000-00012

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Barbour, R. (2007). *Doing Focus Groups*. SAGE Publications Ltd. https://doi.org/10.4135/9781849208956

Barbour, R. S. (2005). Making sense of focus groups. *Medical Education*, *39*(7), 742–750. https://doi.org/10.1111/j.1365-2929.2005.02200.x

Bartfay, W. J., Rombough, R., Howse, E., & Leblanc, R. (2004). Evaluation. The OSCE approach in nursing education. *The Canadian Nurse*, *100*(3), 18–23.

Beauchamp, T., & Childress, J. (2019). *Principles of Biomedical Ethics*: Marking Its Fortieth Anniversary. *The American Journal of Bioethics*, *19*(11), 9–12. https://doi.org/10.1080/15265161.2019.1665402

Bjork, R. A. (1994). Memory and Metamemory Considerations in the Training of Human Beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition* (pp. 185–206). The MIT Press. https://doi.org/10.7551/mitpress/4561.003.0011

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning*. Kogan Page.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Boursicot, K., & Roberts, T. (2005). How to set up an OSCE. *The Clinical Teacher*, *2*(1), 16–20. https://doi.org/10.1111/j.1743-498X.2005.00053.x

Bruner, J. S. (1966). *Toward a theory of instruction*. Belknap Press of Harvard University Press.

Bruner, J. S. (1973). Organization of Early Skilled Action. *Child Development*, *44*(1), 1. https://doi.org/10.2307/1127671

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Candy, P. C. (1991). *Self-direction for lifelong learning: A comprehensive guide to theory and practice* (1st ed). Jossey-Bass.

Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.

Cohen. (2013a). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441

Cohen, J. (2013b). *Statistical Power Analysis for the Behavioral Sciences* (0 ed.). Routledge. https://doi.org/10.4324/9780203771587

Colliver, J. A. (2000). Effectiveness of Problem-based Learning Curricula: Research and Theory. *Academic Medicine*, *75*(3), 259–266. https://doi.org/10.1097/00001888-200003000-00017

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. J. (2011). Technology-Enhanced Simulation for Health Professions Education: A Systematic Review and Meta-analysis. *JAMA*, *306*(9). https://doi.org/10.1001/jama.2011.1234

Cooper, P. A. (1993). Paradigm shifts in designed instruction: From Behaviorism to Cognitivism to Constructivism. *Educational Technology*, *33*(5), 12–19.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Crotty, M. (2020). *The foundations of social research: Meaning and perspective in the research process* (1st ed.). Routledge. https://doi.org/10.4324/9781003115700

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of the empirical studies. *Educational Technology Research and Development*, *66*(5), 1101–1118. https://doi.org/10.1007/s11423-018-9579-9

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292–307. https://doi.org/10.1037/1082-989X.2.3.292

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*(4), 552–568. https://doi.org/10.1016/j.jecp.2012.06.009

Deci, E., Vallerand, R., Pelletier, L., & Ryan, R. (1991). Motivation and Education: The Self-Determination Perspective. *Educational Psychologist*, *26*(3), 325–346. https://doi.org/10.1207/s15326985ep2603&4_6

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (Fifth edition). SAGE.

Dewey, J. (1938). *Experience and education*. Macmillan.

Dolmans, D. H. J. M., De Grave, W., Wolfhagen, I. H. A. P., & Van Der Vleuten, C. P. M. (2005). Problem-based learning: Future challenges for educational practice and research. *Medical Education*, *39*(7), 732–741. https://doi.org/10.1111/j.1365-2929.2005.02205.x

Dweck, C. (2016). *Mindset: The new psychology of success* (Updated edition). Random House.

Etikan, I. (2017). Sampling and Sampling Methods. *Biometrics & Biostatistics International Journal*, *5*(6). https://doi.org/10.15406/bbij.2017.05.00149

Facione, N. C., & Facione, P. A. (1996). Externalizing the critical thinking in knowledge development and clinical judgment. *Nursing Outlook*, *44*(3), 129–136. https://doi.org/10.1016/S0029-6554(06)80005-9

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th edition, North American edition). Sage Publications Inc.

Flexner, A. (2002). Medical education in the United States and Canada. From the Carnegie Foundation for the Advancement of Teaching, Bulletin Number Four, 1910. *Bulletin of the World Health Organization*, *80*(7), 594–602.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Freire, P. (1970). *Pedagogy of the oppressed* (11. printing). The Seabury Press.

Furmedge, D. S., Smith, L.-J., & Sturrock, A. (2016). Developing doctors: What are the attitudes and perceptions of year 1 and 2 medical students towards a new integrated formative objective structured clinical examination? *BMC Medical Education*, *16*(1), 32. https://doi.org/10.1186/s12909-016-0542-3

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, *13 Suppl 1*(Suppl 1), i2-10. https://doi.org/10.1136/qhc.13.suppl_1.i2

Goldie, J. (2012). The formation of professional identity in medical students: Considerations for educators. *Medical Teacher*, *34*(9), e641–e648. https://doi.org/10.3109/0142159X.2012.687476

Gormley, G. (2011). Summative OSCEs in undergraduate medical education. *The Ulster Medical Journal*, *80*(3), 127–132.

Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the behavioral sciences* (9th ed). Wadsworth Cengage Learning.

Greenhalgh, T. (2019). *How to read a paper: The basics of evidence-based medicine and healthcare* (Sixth edition). Wiley Blackwell.

Haidet, P., Kubitz, K., & McCormack, W. T. (2014). Analysis of the Team-Based Learning Literature: TBL Comes of Age. *Journal on Excellence in College Teaching*, *25*(3–4), 303–333.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7. Auflage, Pearson new internat. ed). Pearson.

Hammad, M., Oweis, Y., Taha, S., Hattar, S., Madarati, A., & Kadim, F. (2013). Students' opinions and attitudes after performing a dental OSCE for the first time: A Jordanian experience. *Journal of Dental Education*, *77*(1), 99–104.

Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *BMJ*, *1*(5955), 447–451. https://doi.org/10.1136/bmj.1.5955.447

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242. https://doi.org/10.1177/154193120404801107

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hodges, B., Regehr, G., McNaughton, N., Tiberius, R., & Hanson, M. (1999). OSCE checklists do not capture increasing levels of expertise: *Academic Medicine*, *74*(10), 1129–1134. https://doi.org/10.1097/00001888-199910000-00017

Holland, C. (2016). Critical review: Medical students' motivation after failure. *Advances in Health Sciences Education*, *21*(3), 695–710. https://doi.org/10.1007/s10459-015-9643-8

Issenberg, S., Mcgaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, *27*(1), 10–28. https://doi.org/10.1080/01421590500046924

Jordi, R. (2011). Reframing the Concept of Reflection: Consciousness, Experiential Learning, and Reflective Learning Practices. *Adult Education Quarterly*, *61*(2), 181–197. https://doi.org/10.1177/0741713610380439

Joshi, M. K., Srivastava, A. K., Ranjan, P., Singhal, M., Dhar, A., Chumber, S., Parshad, R., & Seenu, V. (2017). OSCE as a Summative Assessment Tool for Undergraduate Students of Surgery—Our Experience. *Indian Journal of Surgery*, *79*(6), 534–538. https://doi.org/10.1007/s12262-016-1521-y

Judd, T., & Elliott, K. (2017). Selection and Use of Online Learning Resources by First-Year Medical Students: Cross-Sectional Study. *JMIR Medical Education*, *3*(2), e17. https://doi.org/10.2196/mededu.7382

Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, *38*(6), 523–550. https://doi.org/10.1007/s11251-009-9093-x

Kapur, M. (2014). Productive Failure in Learning Math. *Cognitive Science*, *38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2015). Learning from productive failure. *Learning: Research and Practice*, *1*(1), 51–65. https://doi.org/10.1080/23735082.2015.1002195

Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*, *51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kapur, M., & Kinzer, C. K. (2007). Examining the effect of problem type in a synchronous computer-supported collaborative learning (CSCL) environment. *Educational Technology Research and Development*, *55*(5), 439–459. https://doi.org/10.1007/s11423-007-9045-6

Kendall, J. M. (2003). Designing a research project: Randomized controlled trials and their principles. *Emergency Medicine Journal*, *20*(2), 164–168. https://doi.org/10.1136/emj.20.2.164

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified Randomization for Clinical Trials. *Journal of Clinical Epidemiology*, *52*(1), 19–26. https://doi.org/10.1016/S0895-4356(98)00138-3

Khairy, G. A. (2004). Feasibility and acceptability of objective structured clinical examination (osce) for a large number of candidates: Experience at a university hospital. *Journal of Family & Community Medicine*, *11*(2), 75–78.

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152. https://doi.org/10.5395/rde.2017.42.2.152

Kitzinger, J. (1995). Qualitative Research: Introducing focus groups. *BMJ*, *311*(7000), 299–302. https://doi.org/10.1136/bmj.311.7000.299

Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy* (Rev. and updated, 1. print). Association Press.

Knowles, M. S. (Ed.). (1984). *Andragogy in action* (1st ed). Jossey-Bass.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Kolbe, M., Schmutz, S., Seelandt, J. C., Eppich, W. J., & Schmutz, J. B. (2021). Team debriefings in healthcare: Aligning intention and impact. *BMJ*, n2042. https://doi.org/10.1136/bmj.n2042

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th edition). SAGE.

Kusurkar, R. A., Ten Cate, Th. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: A structural equation modelling analysis. *Advances in Health Sciences Education*, *18*(1), 57–69. https://doi.org/10.1007/s10459-012-9354-3

Lazarus, J., & Kent, A. P. (1983). Student attitudes towards the objective structured clinical examination (OSCE) and conventional methods of assessment. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, *64*(11), 390–394.

Lee, S., & Skog, A. (2014). 20 Kobayashi Maru in Medical Simulation: The Educational Utility of the No Win Patient Scenario. *Annals of Emergency Medicine*, *64*.

Lewin, K. (1946). Action Research and Minority Problems. *Journal of Social Issues*, *2*(4), 34–46. https://doi.org/10.1111/j.1540-4560.1946.tb02295.x

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review*, *29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts* (3. ed). Routledge.

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The General Self-Efficacy Scale: Multicultural Validation Studies. *The Journal of Psychology*, *139*(5), 439–457. https://doi.org/10.3200/JRLP.139.5.439-457

Ma, I. W. Y., Zalunardo, N., Pachev, G., Beran, T., Brown, M., Hatala, R., & McLaughlin, K. (2012). Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Advances in Health Sciences Education*, *17*(4), 457–470. https://doi.org/10.1007/s10459-011-9322-3

Matthews, G. (1999). Validation of a Comprehensive Stress State Questionnaire: Towards a State "Big Three"? *Personality Psychology in Europe*, *7*, 335–350.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315–340. https://doi.org/10.1037/1528-3542.2.4.315

McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does Simulation-Based Medical Education With Deliberate Practice Yield Better Results Than Traditional Clinical Education? A Meta-Analytic Comparative Review of the Evidence: *Academic Medicine*, *86*(6), 706–711. https://doi.org/10.1097/ACM.0b013e318217e119

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009: Simulation-based medical education research 2003–2009. *Medical Education*, *44*(1), 50–63. https://doi.org/10.1111/j.1365-2923.2009.03547.x

Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, *135*(3), 391–408. https://doi.org/10.1037/0096-3445.135.3.391

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (Fourth edition). Jossey-Bass.

Mezirow, J. (1991). *Transformative dimensions of adult learning* (1st ed). Jossey-Bass.

Morgan, D. L. (1996). Focus Groups. *Annual Review of Sociology*, *22*(1), 129–152. https://doi.org/10.1146/annurev.soc.22.1.129

Mylopoulos, M., & Regehr, G. (2009). How student models of expertise and innovation impact the development of adaptive expertise in medicine. *Medical Education*, *43*(2), 127–132. https://doi.org/10.1111/j.1365-2923.2008.03254.x

Narciss, S. (2008). *Feedback Strategies for Interactive Learning Tasks. Handbook of Research on Educational Communications and Technology* (3rd ed.). Routledge.

Nestel, D., & Tierney, T. (2007). Role-play for medical students learning about communication: Guidelines for maximising benefits. *BMC Medical Education*, *7*(1), 3. https://doi.org/10.1186/1472-6920-7-3

Newble, D. (2004). Techniques for measuring clinical competence: Objective structured clinical examinations. *Medical Education*, *38*(2), 199–203. https://doi.org/10.1111/j.1365-2923.2004.01755.x

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning: Diagnostic error and reasoning. *Medical Education*, *44*(1), 94–100. https://doi.org/10.1111/j.1365-2923.2009.03507.x

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, *103*(2), 241–262. https://doi.org/10.1037/0033-295X.103.2.241

Okuda, Y., Bryson, E. O., DeMaria, S., Jacobson, L., Quinones, J., Shen, B., & Levine, A. I. (2009). The utility of simulation in medical education: What is the evidence? *The Mount Sinai Journal of Medicine, New York*, *76*(4), 330–343. https://doi.org/10.1002/msj.20127

Osterman, K. F., & Kottkamp, R. B. (2004). *Reflective practice for educators: Professional development to improve student learning* (Second edition). Corwin Press, a Sage Publications Company.

Pallant, J. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS* (7th ed.). Routledge. https://doi.org/10.4324/9781003117452

Palmer, J. (2001). *Fifty Major Thinkers on Education: From Confucius to Dewey*. Routledge.

Palominos, E., Levett-Jones, T., Power, T., Alcorn, N., & Martinez-Maldonado, R. (2021). Measuring the impact of productive failure on nursing students' learning in healthcare simulation: A quasi-experimental study. *Nurse Education Today*, *101*, 104871. https://doi.org/10.1016/j.nedt.2021.104871

Pelaccia, T., & Viau, R. (2017). Motivation in medical education. *Medical Teacher*, *39*(2), 136–140. https://doi.org/10.1080/0142159X.2016.1248924

Piaget, J. (1953). *The origin of intelligence in the child*. /z-wcorg/.

Piaget, J. (1972). *The psychology of the child*. Basic Books.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Viking Press.

Polit, D. F., & Beck, C. T. (2021). *Nursing research: Generating and assessing evidence for nursing practice* (Eleventh edition). Wolters Kluwer.

Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination: *Academic Medicine*, *73*(9), 993–997. https://doi.org/10.1097/00001888-199809000-00020

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Rogers, C. R., & Freiberg, H. J. (1994). *Freedom to learn* (3rd ed). Merrill; Maxwell Macmillan Canada; Maxwell Macmillan International.

Rudolph, J. W., Simon, R., Dufresne, R. L., & Raemer, D. B. (2006). There's No Such Thing as "Nonjudgmental" Debriefing: A Theory and Method for Debriefing with Good Judgment: *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *1*(1), 49–55. https://doi.org/10.1097/01266021-200600110-00006

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4e ed.). SAGE.

Sanders, P. (2006). Understanding Academic Confidence. *Psychology Teaching Review*, *12*, 29–39.

Sawyer, T., Eppich, W., Brett-Fleegler, M., Grant, V., & Cheng, A. (2016). More Than One Way to Debrief: A Critical Review of Healthcare Simulation Debriefing Methods. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *11*(3), 209–217. https://doi.org/10.1097/SIH.0000000000000148

Scalese, R. J., Obeso, V. T., & Issenberg, S. B. (2008). Simulation Technology for Skills Training and Competency Assessment in Medical Education. *Journal of General Internal Medicine*, *23*(S1), 46–49. https://doi.org/10.1007/s11606-007-0283-4

Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Medical Education*, *49*(10), 961–973. https://doi.org/10.1111/medu.12775

Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science*, *3*(4), 207–218. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Scholz, U., Gutiérrez Doña, B., Sud, S., & Schwarzer, R. (2002). Is General Self-Efficacy a Universal Construct? 1. *European Journal of Psychological Assessment*, *18*(3), 242–251. https://doi.org/10.1027//1015-5759.18.3.242

Schön, D. (1983). *The reflective practitioner; How professionals think in action*. Basic Books.

Schön, D. A. (2000). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions* (1 ed., [Nachdr.]). Jossey-Bass.

Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition and Instruction*, *16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Schwarzer, R. (2010). The general self-efficacy scale (GSE). *Anxiety, Stress, and Coping*, *12*, 329–345.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study. *Annals of Surgery*, *236*(4), 458–464. https://doi.org/10.1097/00000658-200210000-00008

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Shinnick, M. A., Woo, M. A., & Mentes, J. C. (2011). Human Patient Simulation: State of the Science in Prelicensure Nursing Education. *Journal of Nursing Education*, *50*(2), 65–72. https://doi.org/10.3928/01484834-20101230-01

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Siddiqui, F. G. (2013). Final year MBBS students' perception for observed structured clinical examination. *Journal of the College of Physicians and Surgeons--Pakistan: JCPSP*, *23*(1), 20–24. https://doi.org/01.2013/JCPSP.2024

Sim, J., & Waterfield, J. (2019). Focus group methodology: Some ethical challenges. *Quality & Quantity*, *53*(6), 3003–3022. https://doi.org/10.1007/s11135-019-00914-5

Skinner, B. F. (1976). *About behaviorism*. Vintage Books.

Sloan, D. A., Donnelly, M. B., Schwartz, R. W., & E, W. (1995). The Objective Structured Clinical Examination The New Gold Standard for Evaluating Postgraduate Clinical Performance: *Annals of Surgery*, *222*(6), 735–742. https://doi.org/10.1097/00000658-199512000-00007

Smits, P. B. A. (2002). Problem based learning in continuing medical education: A review of controlled evaluation studies. *BMJ*, *324*(7330), 153–156. https://doi.org/10.1136/bmj.324.7330.153

Sobral, D. T. (2004). What kind of motivation drives medical students' learning quests? *Medical Education*, *38*(9), 950–957. https://doi.org/10.1111/j.1365-2929.2004.01913.x

Spence, J. T. (Ed.). (1983). *Achievement and achievement motives: Psychological and sociological approaches*. W.H. Freeman.

Steenhof, N., Woods, N. N., & Mylopoulos, M. (2020). Exploring why we learn from productive failure: Insights from the cognitive and learning sciences. *Advances in Health Sciences Education*, *25*(5), 1099–1106. https://doi.org/10.1007/s10459-020-10013-y

Steenhof, N., Woods, N. N., Van Gerven, P. W. M., & Mylopoulos, M. (2019). Productive failure as an instructional approach to promote future learning. *Advances in Health Sciences Education*, *24*(4), 739–749. https://doi.org/10.1007/s10459-019-09895-4

Stegers-Jager, K. M., Cohen-Schotanus, J., & Themmen, A. P. N. (2012). Motivation, learning strategies, participation and medical school performance: Motivation, learning strategies and participation. *Medical Education*, *46*(7), 678–688. https://doi.org/10.1111/j.1365-2923.2012.04284.x

Steinert, Y. (Ed.). (2014). *Faculty Development in the Health Professions: A Focus on Research and Practice*. Springer Netherlands. https://doi.org/10.1007/978-94-007-7612-8

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (Seventh edition). Pearson.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Toufiq, M., Ahmed, A., Ahmed, K., Toufiq, N., Rajaratnam, U., Williams, S., Jamileh, E., & Ahmed, M. (2024). Understanding Medical Students' Perceptions of Failure in Medical School. *Cureus*. https://doi.org/10.7759/cureus.74024

Towle, A., & Cottrell, D. (1996). Self directed learning. *Archives of Disease in Childhood*, *74*(4), 357–359. https://doi.org/10.1136/adc.74.4.357

Van Der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programs. *Medical Education*, *39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

Vygotsky, L. (1962). *Thought and language.* (E. Hanfmann & G. Vakar, Eds.). MIT Press. https://doi.org/10.1037/11193-000

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wallas, G. (2014). *The art of thought*. Solis Press.

Wass, V., Van Der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet*, *357*(9260), 945–949. https://doi.org/10.1016/S0140-6736(00)04221-5

Wayne, D. B., Didwania, A., Feinglass, J., Fudala, M. J., Barsuk, J. H., & McGaghie, W. C. (2008). Simulation-Based Education Improves Quality of Care During Cardiac Arrest Team Responses at an Academic Teaching Hospital. *Chest*, *133*(1), 56–61. https://doi.org/10.1378/chest.07-0131

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed). Association for Supervision and Curriculum Development.

Wilkinson, S. (1998). Focus Groups in Health Research: Exploring the Meanings of Health and Illness. *Journal of Health Psychology*, *3*(3), 329–348. https://doi.org/10.1177/135910539800300304

Wilson, J. I. (2009). A two factor model of performance approach goals in student motivation for starting medical school. *Issues in Educational Research*, *19*(3), 271.

Witt, E. E., Onorato, S. E., & Schwartzstein, R. M. (2022). Medical Students and the Drive for a Single Right Answer: Teaching Complexity and Uncertainty. *ATS Scholar*, *3*(1), 27–37. https://doi.org/10.34197/ats-scholar.2021-0083PS

Wood, D., Bruner, J. S., & Ross, G. (1976). THE ROLE OF TUTORING IN PROBLEM SOLVING. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Woods, N. N., Brooks, L. R., & Norman, G. R. (2007). It all makes sense: Biomedical knowledge, causal connections and memory in the novice diagnostician. *Advances in Health Sciences Education*, *12*(4), 405–415. https://doi.org/10.1007/s10459-006-9055-x

Yardley, S., Teunissen, P. W., & Dornan, T. (2012). Experiential learning: Transforming theory into practice. *Medical Teacher*, *34*(2), 161–164. https://doi.org/10.3109/0142159X.2012.643264

Yilmaz, K. (2011). The Cognitive Perspective on Learning: Its Theoretical Underpinnings and Implications for Classroom Practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *84*(5), 204–212. https://doi.org/10.1080/00098655.2011.568989

Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient Outcomes in Simulation-Based Medical Education: A Systematic Review. *Journal of General Internal Medicine*, *28*(8), 1078–1089. https://doi.org/10.1007/s11606-012-2264-5

Zoller, A., Hölle, T., Wepler, M., Radermacher, P., & Nussbaum, B. L. (2021). Development of a novel global rating scale for objective structured assessment of technical skills in an emergency medical simulation training. *BMC Medical Education*, *21*(1), 184. https://doi.org/10.1186/s12909-021-02580-4

Appendices

Appendix 1: IRB Approval Letter

The UNIVERSITY of OKLAHOMA

Institutional Review Board for the Protection of Human Subjects

Approval of Initial Submission – Expedited Review – AP01

Date: May 15, 2024 **IRB#:** 17244
Principal Investigator: Bobby Bosse
Approval Date: 05/15/2024
Status Report Due: 04/30/2025
Study Title: Physician Assistant (PA) Productive Failure: The Kobayashi Maru Patient Scenario

Expedited Category: 7
Collection/Use of PHI: No

On behalf of the Institutional Review Board (IRB), I have reviewed and granted expedited approval of the above-referenced research study. To view the documents approved for this submission, open this study from the *My Studies* option, go to *Submission History*, go to *Completed Submissions* tab and then click the *Details* icon.

Requirements under the Common Rule have changed. The above-referenced research meets one or more of the circumstances for which continuing review is not required. However, as Principal Investigator of this research, you will be required to submit an annual status report to the IRB.

As principal investigator of this research study, you are responsible to:

- Conduct the research study in a manner consistent with the requirements of the IRB and federal regulations 45 CFR 46.
- Obtain informed consent and research privacy authorization using the currently approved, stamped forms and retain all original, signed forms, if applicable.
- Request approval from the IRB prior to implementing any/all modifications.
- Promptly report to the IRB any harm experienced by a participant that is both unanticipated and related per IRB policy.
- Maintain accurate and complete study records for evaluation by the HRPP Quality Improvement Program and, if applicable, inspection by regulatory agencies and/or the study sponsor.
- **Submit an annual status report to the IRB to provide the study/recruitment status and report all harms and deviations that may have occurred.**
- **Submit a final closure report at the completion of the project.**

If you have questions about this notification or using iRIS, contact the IRB @ 405-325-8110 or irb@ou.edu.

Cordially,

Lara Mayeux, Ph.D.

Lara Mayeux

Chair, Institutional Review Board

Appendix 2: OSCE Control Group Simulation Flow Chart

Start: Bp 85/50, pulse 130, resp rate 30, pulse ox- 82%

ER Tech communication with provider 1-2 min

Goal #1: Recognize **Unstable Vital signs**

If Unrecognized hypotension or hypoxia within **1-2min- Blood pressure drops to 70/40mmHg, HR increases to 140 bpm, pulse ox**

Treatment #1: Initiate **IVF** (1-2L bolus NS or LR) and initiate **Oxygen**

Initiated IVF- no change in blood pressure

Initiated Oxygen by Mask or ambu bag- pulse ox improves to 90%

If No IVF initiated within 1-2 minutes – **Blood pressure drops by 10mmHg and HR increases by 10 bpm**

If No Oxygen initiated by mask or ambu bag pulse ox decreases by 10%

Goal #2: Manage **Persistent Hypotension**

Treatment #2: Initiate **Vasopressor therapy**

Initiated Epinephrine or Norepinephrine therapy- improve blood pressure to 85/50mmHg, HR 140

Initiated Additional IVF or blood products- blood pressure NO CHANGE

If No Vasopressor therapy initiated within 2-3 minutes – **Blood pressure drops 60/30, HR increases 150 bpm**

Goal #3: Consider Potential Diagnoses- **Pulmonary Embolism**

Treatment #3: **Thrombolytic Therapy**

Initiated Thrombolytic therapy- stabilization of rhythm blood pressure increases to 110/70mmHg, pulse 110bpm, pulse ox to 90% = Return of spontaneous circulation, Pt asks "where am I and what is going on?"

Initiate Patient Transfer = END OF SIMULATION

If No Thrombolytic Therapy initiated within 2-3 minutes – **V-Fib Arrest with Blood pressure drops to 0/0mmHg**

***IF RETURNING FROM CARDIAC ARREST (Treatment Step #4) and no thrombolytics within 1-2 min = Patient Demise, Asystole, and END OF SIMULATION**

IF NEEDED: Ventricular Fibrillation Arrest

Goal #4: **Manage Cardiac Arrest- Ventricular Fibrillation**

Treatment #4: **ACLS and Defibrillation**

Initiated CPR and Defibrillate – No change after first Defibrillation

Initiate 2nd Defibrillation AND/OR Epinephrine 1mg IV (amiodarone or lidocaine) = stabilization of rhythm sinus tachycardia, blood pressure increases to 70/40mmHg, pulse 140bpm, pulse ox to 90% = Return of spontaneous circulation AND RETURN TO TREATMENT STEP #3 Thrombolytic Therapy within 1-2 min

If No 2nd Defibrillation and/or Epinephrine after 3-5 min of CPR = **Patient Demise, Asystole, and END OF SIMULATION**

END OF

SIMULATION

Appendix 3: OSCE Intervention Group Simulation Flow Chart

Start: Bp 85/50, pulse 130, resp rate 30, pulse ox- 82%

ER Tech communication with provider 1-2 min

Goal #1: Recognize **Unstable Vital signs**

If Unrecognized hypotension or hypoxia within **1-2min- Blood pressure drops to 70/40mmHg, HR increases to 140 bpm, pulse ox decreases to 75%**

Treatment #1: Initiate **IVF** (1-2L bolus NS or LR) and initiate **Oxygen**

Initiated IVF- no change in blood pressure with IVF

Initiated Oxygen by Mask or ambu bag- pulse ox improves to 90%

After 1-2 minutes – **Blood pressure drops by 60/30 mmHg, HR 140 bpm**

Goal #2: Manage **Persistent Hypotension**

Treatment #2: Initiate **Vasopressor therapy**

Initiated Epinephrine or Norepinephrine therapy- no change in blood pressure

Initiated Additional IVF or blood products- no change in blood pressure

After 2-3 minutes – **Blood pressure drops by 50/30 mmHg, HR increases to 150 bpm**

Goal #3: Consider Potential Diagnoses- **Pulmonary Embolism**

Treatment #3: **Thrombolytic Therapy**

Initiated Thrombolytic therapy- no change in vital signs

After 2-3 minutes – **V-Fib Arrest with Blood pressure drops to 0/0mmHg**

Ventricular Fibrillation Arrest

Goal #4: **Manage Cardiac Arrest- Ventricular Fibrillation**

Treatment #4: **ACLS and Defibrillation**

Initiated CPR and Defibrillate – No change in vital signs

After 3-5 min of CPR/ACLS (2 cycles) = **Patient Demise, Asystole and END OF SIMULATION**

END OF SIMULATION

Appendix 4: Pre-simulation Self-Efficacy Survey

Please indicate the extent to which each statement applies to you: (circle answer) (1- Not at all true, 2- Barely true, 3- Moderately true, 4- Exactly true)

		Not at all true	Barely true	Moderately true	Exactly true
1.	I can always manage to solve difficult problems if I try hard enough.	1	2	3	4
2.	If someone opposes me, I can find means and ways to get what I want.	1	2	3	4
3.	It is easy for me to stick to my aims and accomplish my goals.	1	2	3	4
4.	I am confident that I could deal efficiently with unexpected events.	1	2	3	4
5.	Thanks to my resourcefulness, I know how to handle unforeseen situations.	1	2	3	4
6.	I can solve most problems if I invest the necessary effort.	1	2	3	4
7.	I can remain calm when facing difficulties because I can rely on my coping abilities.	1	2	3	4
8.	When I am confronted with a problem, I can usually find several solutions.	1	2	3	4
9.	If I am in a bind, I can usually think of something to do.	1	2	3	4
10.	No matter what comes my way, I'm usually able to handle it.	1	2	3	4

Appendix 5: Pre-simulation Emotional Stress Response Survey

Please indicate how well each statement describes how you feel AT THE MOMENT:
(circle answer) (1- Not at all, 2- A little bit, 3- Somewhat, 4- Very much, 5- Extremely)

		Not at all	A little bit	Somewhat	Very much	Extremely
1.	I feel dissatisfied.	1	2	3	4	5
2.	I feel alert.	1	2	3	4	5
3.	I feel depressed.	1	2	3	4	5
4.	I feel sad.	1	2	3	4	5
5.	I feel active.	1	2	3	4	5
6.	I feel impatient.	1	2	3	4	5
7.	I feel annoyed.	1	2	3	4	5
8.	I feel angry.	1	2	3	4	5
9.	I feel irritated.	1	2	3	4	5
10.	I feel grouchy.	1	2	3	4	5
11.	I am committed to attaining my performance goals.	1	2	3	4	5
12.	I want to succeed on the task.	1	2	3	4	5
13.	I am motivated to do the task.	1	2	3	4	5
14.	I am trying to figure myself out.	1	2	3	4	5
15.	I am reflecting about myself.	1	2	3	4	5
16.	I am daydreaming about myself.	1	2	3	4	5
17.	I feel confident about my abilities.	1	2	3	4	5
18.	I feel self-conscious.	1	2	3	4	5
19.	I am worried about what other people think of me.	1	2	3	4	5
20.	I feel concerned about the impression I am making.	1	2	3	4	5
21.	I expect to perform proficiently on this task.	1	2	3	4	5
22.	Generally, I feel in control of things.	1	2	3	4	5
23.	I thought about how others have done on this task.	1	2	3	4	5
24.	I thought about how I would feel if I were told how I performed.	1	2	3	4	5

Appendix 6: Post-simulation OSCE Competency Survey

Simulation Scoring Rubric

Student name:

Systematic assessment approach	1 Incomplete assessment	2	3 Adequate assessment, minor flaws	4	5 Complete and structured assessment
Physical examination	1 Too little or unnecessary	2	3 Appropriate & suitable to situation	4	5 Very accurate & reasonable
Vital signs& Monitoring	1 Possibly unaware of critical condition	2	3 Attention to most important warning signs	4	5 Very adequate and anticipatory
Patient's position	1 Disadvantageous position during the scenario	2	3 Appropriate position after an acceptable time	4	5 Beneficial position during the scenario
Temperature management	1 no attention to patient's temperature	2	3 Appropriate temperature management most of the time	4	5 Beneficial temperature management the whole time
Handling of equipment	1 Many difficulties with the equipment	2	3 Appropriate use of the equipment	4	5 Advanced use of the equipment
Therapy and medication	1 Inadequate or harmful therapy	2	3 Adequate therapy, no harm	4	5 Sufficient therapy based on current guidelines
Drug application	1 Unsafe and insecure	2	3 Safe, minor insecurities	4	5 Safe and secure
Patient safety overall	1 Many risk factors for patient safety	2	3 Few risk factors for patient safety occurred	4	5 Safe handling, maximal elimination of risk factors
Overall performance	1 Poor	2	3 Average	4	5 Excellent

	Not done/Incorrect	Done correctly
Check for Safety	0	1
Airway		
Assessment	0	1
Secure (if necessary)	0	1
Breathing		
Respiratory rate	0	1
Auscultation	0	1
Monitoring (saturation)	0	1
Clinical symptoms (e.g. cyanosis)	0	1
Oxygen therapy	0	1
Circulation		
Check for pulse, CRT	0	1
Clinical symptoms (e.g. skin color)	0	1
Monitoring (blood pressure, ECG)	0	1
Establish venous access	0	1
Adequate volume therapy	0	1
Disability		
Orientation	0	1
Neurological assessment (e.g. FAST)	0	1
Blood glucose level	0	1
Exposure		
Physical examination	0	1
SAMPLE anamnesis	0	1
Working Diagnosis	0	1
Temperature Management	0	1
Optimize Patient's Position	0	1
Treatment		
Guideline-based therapy	0	1
Check before drug application	0	1
Hospital		
Choose appropriate hospital	0	1
Provide essential information	0	1

CRT: capillary refill time, ECG: electrocardiogram, FAST: acronym meaning **F**ace, **A**rms, **S**peech, **T**ime, SAMPLE: acronym meaning **S**ymptoms, **A**llergies, **M**edication, **P**ast medical history, **L**ast oral meal, **E**vents

Appendix 7: Focus Group Interview Protocol

Introduction

Thank you for your time and willingness to participate in this research project and the interviews.

As you know, I am interested in evaluating and understanding the potential effects of complex simulation activities for clinical PA students. This study specifically aims to evaluate how PA students reflect on learning from OSCE activities that utilize productive failure in the patient simulation.

Please feel free to volunteer any detail you wish pertaining to the questions. You also have the option of declining to answer or passing on any of the questions.

Do you have any questions before we begin?

Interview Questions

1. What challenges did you experience in this simulation activity?

2. How did today's simulation activity compare to other teaching methods previously used in the PA OSCE activities?

3. Reflecting on your participation in the productive failure simulation, how do you feel it influenced your learning and development as a PA student?

Closing

Again, I want to thank you for your willingness to participate in this research project.

Now that I have completed all my questions, do you have any questions you'd like to ask me about this research project?

If you want to contact me later regarding this project, here is my contact information (I will provide a contact information sheet).

Also, I may need to contact you later for additional questions or clarification. If you agree to doing so, would you provide me with your follow-up contact information?

Appendix 8: Focus Group Thematic Analysis

Figure 3 – Focus Group Question 1 Thematic Analysis

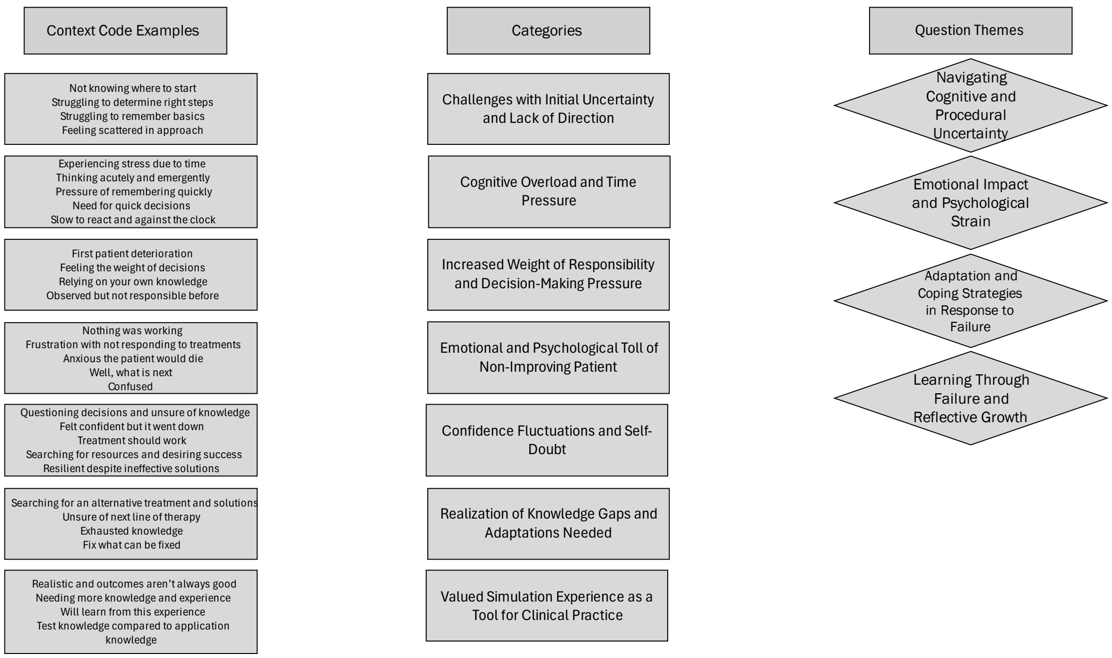


Figure 4 – Focus Group Question 2 Thematic Analysis

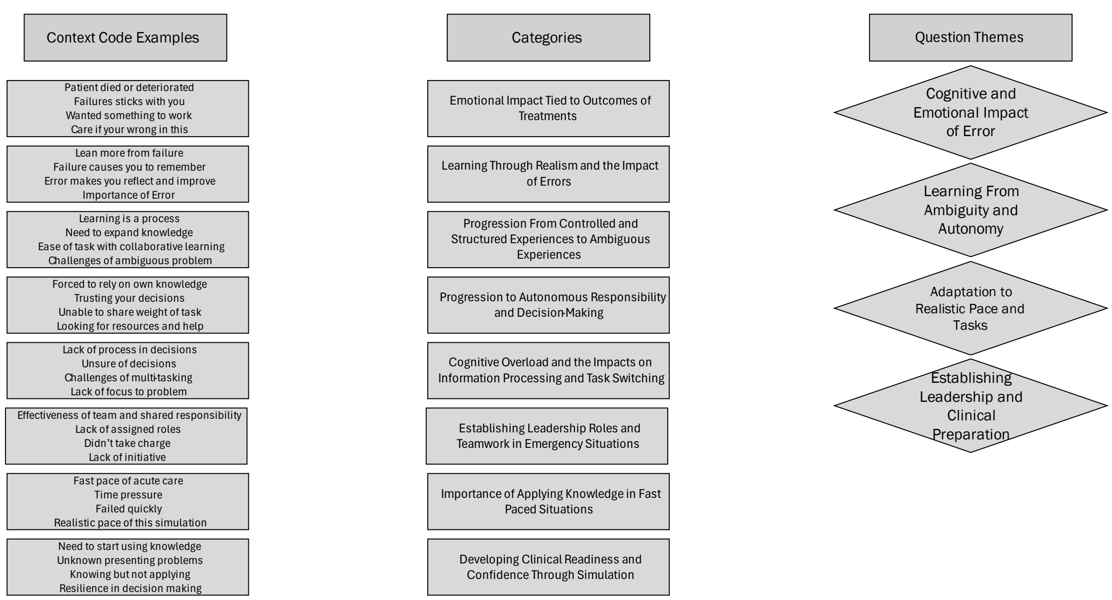


Figure 5 – Focus Group Question 3 Thematic Analysis