Fall 2023

PREDICTING PHYSICAL THERAPY STUDENT CLINICAL REASONING IN PRACTICE USING ALIGNED ASSESSMENTS IN A CURRICULAR DOMAIN

Joy L. Moulton
Southeastern University - Lakeland

PREDICTING PHYSICAL THERAPY STUDENT CLINICAL REASONING IN PRACTICE USING ALIGNED ASSESSMENTS IN A CURRICULAR DOMAIN

By

JOY L. MOULTON

A doctoral dissertation submitted to the
College of Education
in partial fulfillment of the requirements
for the degree Doctor of Education
in Curriculum and Instruction

Southeastern University
August, 2023

PREDICTING PHYSICAL THERAPY STUDENT CLINICAL REASONING IN PRACTICE USING ALIGNED ASSESSMENTS IN A CURRICULAR DOMAIN

by

JOY L. MOULTON

Dissertation Approved:

Sarah Yates

Sarah J. Yates, EdD, Dissertation Chair

Thomas Gollery

Thomas J. Gollery, EdD, Committee Member

Lance Cherry

Lance Cherry, PT, EdD, Committee Member

Lisa A. Coscia

Lisa A. Coscia, EdD, Dean, College of Education

ACKNOWLEDGMENTS

I would like to acknowledge and thank the professors in the College of Education at Southeastern University who guided me through the courses of the Doctor of Education degree program. Specifically, I would like to thank my chair, Dr. Sarah Yates, for her unwavering support and enthusiasm during my journey. Dr. Yates was always available to meet in person to discuss ideas for my dissertation and was instrumental in guiding me through the process. I would also like to thank Dr. Thomas Gollery, the second member of my dissertation committee and methodologist. I could not have completed the quantitative statistical analyses for this project without Dr. Gollery and his expertise.

A second group of people deserving special acknowledgement are the faculty in the Doctor of Physical Therapy department at Nova Southeastern University. First, Dr. Lance Cherry, who is the third member of my dissertation committee and content expert for the subject matter in this study. Dr. Cherry not only served as content expert but provided me with data from one of his courses to satisfy the third tier of the conceptual framework I used. Next, Dr. Robin Galley, the Director of Clinical Education, supplied data for the top tier of the pyramid from archived scores of the Clinical Internship Evaluation Tool. Finally, Dr. Keiba Shaw, who served as a reference for me when I applied to Southeastern University and supplied data for the competence tier of the conceptual framework for this study. Thank you all for your advice and support throughout this journey.

Abstract

Clinical reasoning (CR) is a difficult concept to assess and is a leading cause of failure for outpatient clinical experiences for physical therapy (PT) students (Silberman et al., 2018). The purpose of this study was to determine to what degree didactic assessments of CR could predict CR in clinical practice utilizing Miller's (1990) pyramid as a conceptual framework. Using a retrospective quantitative exploratory observational design, archived data were collected for graduates in the classes of 2022 and 2023 (*N* = 84) from a hybrid Doctor of Physical Therapy program. Each assessment pertained to the musculoskeletal domain of PT practice and aligned with corresponding tiers of Miller's (1990) pyramid. Assessments included biomechanics exams, clinical skills practical exams, musculoskeletal *Triple Jump* scores (tiers 1-3) and Clinical Internship Evaluation Tool scores (tier four). Regression analyses were completed for didactic assessments (tiers 1-3) to determine whether each individually and as a group could predict CR in clinical practice (tier four). No statistically significant results were found for any individual assessment or the group of assessments for prediction of CR in practice; however, follow-up ancillary analysis revealed that clinical experience cohort rotation was a significant predictor of CR in practice. Further research is needed to explore variables not included in this study; however, results of this study indicate that Miller's (1990) pyramid may not be the best conceptual framework for assessment of CR for PT students. Educational research is warranted for alternative models or frameworks to assess CR during a Doctor of Physical Therapy program.

Keywords: clinical reasoning, physical therapy students, assessment framework, prediction, education

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Clinical reasoning (CR) is one of the most difficult skills to learn for students in healthcare professions education programs. CR involves higher order cognitive processes to evaluate and treat patients in real time. Students in healthcare professions often do not have clinical experience as a reference point from which to draw conclusions or base decisions (Huhn et al., 2018). Furthermore, students initially learn foundational information required for CR in didactic courses during a healthcare education program and have limited opportunities to apply what they have learned. The volume of information necessary in foundational courses of healthcare programs causes many students to struggle to keep up with the pace of learning required as a base for CR (Schmidt & Rikers, 2007). A lack of foundational knowledge, application of knowledge, and clinical experience all contribute to poor clinical reasoning skills in healthcare professions students.

Physical therapy students are no exception to this trend. Clinical reasoning is one of the most common reasons for Clinical Experience (CE) failures in physical therapist students, especially in the outpatient setting (Silberman et al., 2018). Outpatient clinical settings usually involve patients with conditions in the musculoskeletal or orthopedic realm of physical therapy. In the outpatient setting, a physical therapist examines a patient, evaluates the findings of that examination to hypothesize a differential diagnosis, and then develops a treatment plan of care.

These tasks are taught during the course of a Doctor of Physical Therapy (DPT) program and require CR to perform competently.

To date, no standardized way of teaching and assessing CR for healthcare students has been agreed upon by faculty. The quest continues for the best ways to teach and assess clinical reasoning for physical therapy students. Previous studies have focused on assessing CR in one course during a DPT program or in the clinical setting (Huhn et al., 2018). However, a lack of research exists for assessment of CR as a whole during a DPT program. Faculty who teach in DPT education programs need a way to analyze student physical therapists' (SPTs) clinical reasoning before they embark upon full-time CEs. Ensuring sufficient CR in SPTs may help to prevent failures at this final level of the education sequence. A framework is needed to predict CE success based on CR during didactic coursework.

Background of the Study

Agreeing upon a standard definition of clinical reasoning is difficult for health professions educators. For the purposes of this study, CR will be defined according to how physical therapists use CR in clinical practice. Current assessments of CR vary greatly depending on the health profession and the level or year of the student within the education program. The variability of CR assessments has created a conundrum regarding when and how to assess CR for physical therapy students. The problem of SPT outpatient CE failures due to CR must be addressed; however, no framework has been established to do so. Physical therapy educators need a way to assess CR in SPTs before outpatient CEs so educators can predict potential failures and prevent them from occurring.

What is Clinical Reasoning?

Clinical reasoning is a complex and difficult concept to define. In the seminal work by

Edwards et al. (2004), eight reasoning strategies are described within two broad categories of CR in PT practice: diagnosis and management. Under the broad category of diagnosis are diagnostic reasoning and narrative reasoning. Diagnostic reasoning is a combination of forward reasoning and hypothetico-deductive reasoning. Forward reasoning involves organizing knowledge into “illness scripts” or recognizable patterns leading to a conclusion. Forward reasoning is fast and efficient, and usually employed by expert or experienced clinicians. Hypothetico-deductive reasoning, a process of using cues from the patient to generate a hypothesis, is a slower process employed by novice clinicians, or by experienced clinicians who encounter a novel problem (Huhn et al., 2019; Wainwright et al., 2011; Wainwright, 2020). Narrative reasoning constructs meaning by seeking to understand lived experiences of patients during the history-taking portion of the examination process.

Under the broad category of management are the remaining six reasoning strategies: reasoning about procedure, interactive reasoning, collaborative reasoning, reasoning about teaching, predictive reasoning, and ethical reasoning. Reasoning about procedure involves making decisions about and delivering treatment procedures. Interactive reasoning is the establishment and management of the patient-therapist relationship. Collaborative reasoning seeks consent of the patient and/or caregivers in the approach of the examination, goal setting, and progression of interventions. Reasoning about teaching relates to teaching in clinical practice and involves the content, method, and amount of teaching as well as assessment of understanding. Predictive reasoning is prognostic in nature and involves consideration of future scenarios with patients, including patient choices and implications thereof. Ethical reasoning is management of dilemmas affecting conduct of interventions and resultant goals, such as resource allocation, access issues, and billing for treatment times.

Edwards et al. (2004) asserted that because of the fluidity of clinical practice, the eight reasoning strategies work in combination based on cues at any given time during a patient interaction. The main category of diagnosis, however, directs the main category of management. For example, hypothetico-deductive reasoning within the category of diagnosis may lead to identification of exercises needed for a patient, but the reasoning strategies under the management category direct the instruction, performance, feedback, and patient learning of the exercises. Ultimately, Edwards et al. (2004) concluded with a dialectical model of CR moving between the diagnosis and management categories.

Clinical reasoning has been described as the defining characteristic of the health professions; however, the varied health professions view CR in different ways (Furze et al., 2022; Gruppen, 2017; Huhn et al., 2019). For the profession of pharmacy, CR includes critical thinking and problem-solving. For the profession of psychology, CR includes the attributes of critical thinking, reflection, weighing information, and flexibility in thinking. For nursing, CR focuses on competence, establishing a plan of care, and recognizing changes in signs and symptoms regarding patient status (Huhn et al., 2019). For the medical profession, CR requires integration of the physician's biomedical and clinical knowledge while gathering patient information to form a case representation of the problem. The physician repeats the reasoning process until a diagnosis and management plan can be determined (Gruppen, 2017). For the profession of physical therapy, CR includes a similar process of integrating foundational knowledge with a patient history and then formulating a differential diagnosis and treatment plan of care. Much of the literature defining CR has focused on the cognitive and metacognitive processes involved; however, the physical therapy profession also includes narrative and contextual factors as well as movement (Edwards et al., 2004; Huhn et al., 2019; Øberg et al., 2015; Sebelski et al., 2020). Huhn et al. (2019) argued that CR is complex and evolving, and as such cannot be saddled with a formal or concrete definition. Rather CR must be understood as a concept. Conceptualization of CR is broader in scope and includes the essence of CR—a combination of all its characteristics. CR in physical therapy includes attributes of intuition, patient and therapist perspectives, flexibility in thinking, reflection, negotiation, and a dialectical approach (Huhn et al., 2019). Huhn et al. (2019) conducted a concept analysis of CR by researching multiple healthcare disciplines and created a working description by connecting key elements of the concept from all disciplines. The Clinical Reasoning Curricula and Assessment Consortium of the American Council of Academic Physical Therapy (ACAPT) has adopted the description by Huhn et al. (2019). Therefore, for the purposes of this study, the description by Huhn et al. (2019) was adopted as well. CR was operationally defined as the adaptive, iterative, and collaborative process of integrating cognitive, psychomotor, and affective skills in clinical context, involving both therapist and patient perspectives, to result in a biopsychosocial approach to patient/client management (Huhn et al., 2019).

Literature on Clinical Reasoning in DPT Education

Clinical reasoning has been studied extensively in the medical field and, more recently, has been studied in allied health professions, such as physical therapy. However, the majority of research on clinical reasoning has been conducted within clinical practice and residencies rather than the DPT classroom (Cunningham et al., 2019; Edwards et al., 2004; Furze et al., 2022; Gilliland & Wainwright, 2017; Huhn et al., 2018; Wainwright et al., 2011).

Clinical Reasoning in DPT Education

Clinical reasoning has been an elusive part of DPT education. Because CR is difficult to define, it is difficult to teach and assess. Most faculty use multiple frameworks, tools, or assessments when teaching CR (Christensen et al., 2017; Huhn et al., 2018). Recent studies have called for improved CR education in DPT programs (Furze et al., 2022; Gilliland & Wainwright, 2017, 2020; Jensen et al., 2017; Jette et al., 2020), and the ACAPT includes the development of students' clinical reasoning skills in its Excellence Framework for Physical Therapist Education recommendations (American Council of Academic Physical Therapy [ACAPT], 2022). However, the best and most effective ways of teaching CR have not been discovered or agreed upon by academicians (Christensen et al., 2017; Furze et al., 2022; Huhn et al., 2018; Kobal et al., 2021; Reilly et al., 2022). Recent research has focused on learning science and teaching and learning activities to promote CR (Jensen et al., 2017; Reilly et al., 2022). The majority of educators (93.8%) use the Guide to Physical Therapist Practice Patient-Client Management Model and/or the International Classification of Functioning, Disability and Health (ICF) model as a foundation for teaching CR (American Physical Therapy Association [APTA], n.d.; Christensen et al., 2017; World Health Organization [WHO], n.d.). Some researchers have explored the idea of creating CR courses or modules within DPT curricula (Huhn, 2017; Stickley, 2018). Other studies have explored the use of case studies, simulation, authentic patient experiences, reflection, and theater or "thinking out loud" for teaching CR (Borleffs et al., 2003; Christensen et al., 2017; Delaney & Golding, 2014; Kobal et al., 2021). Several studies have examined the use of isolated frameworks to develop CR during clinical education or residencies (Baker et al., 2017; Horwitz et al., 2021). Most CR is learned by SPTs during upper-level clinical courses in the DPT curriculum. However, if foundational knowledge is lacking, CR does not develop in these courses or during full-time CEs (Furze et al., 2022). A lack of CR in clinical courses may lead to course failure or CE failure after didactic coursework is completed.

Student Physical Therapists' CR Abilities

Student physical therapists have difficulty grasping the concept of CR and putting it into practice during clinical courses and CEs (Huhn, 2017). Clinical reasoning ability is related to how well physical therapists (PTs) acquire and organize knowledge (Furze et al., 2022; Gilliland & Wainwright, 2017). Creating schemas or scripts in long-term memory is important for CR and is recognized in the reasoning processes of expert physical therapists (Huhn et al., 2019). Most SPTs do not have prior experience in physical therapy clinical settings and have no frame of reference for applying new knowledge. Studies have shown that SPTs exhibit similar CR strategies as novice PTs. Novice PTs and SPTs tend to focus on psychomotor skills and following protocols, while experienced PTs focus on engaging with patients allowing skills to fade into the background of clinical work. SPTs also prioritize physical therapy interventions and pathology-focused education (Gilliland & Wainwright, 2020). SPTs exhibit four patterns of reasoning: following protocol, the hypothetico-deductive process, reasoning about pain, and analyzing patient behavioral patterns (Gilliland & Wainwright, 2017). Following protocol includes using an examination template of some kind to guide the physical examination process. As part of diagnostic reasoning, the hypothetico-deductive process involves testing hypotheses for acceptance or rejection. Reasoning about pain includes a biomedical approach, which uses location and description of pain to develop hypotheses, and/or a behavioral approach, which assesses a patient's behavioral response and perspective regarding pain. Behavioral pattern analysis incorporates a patient's behavioral history or responses to movement or pain to predict prognosis and success with physical therapy treatment. Novice PTs and SPTs lack the skill of reflection-in-action, or reflection during the examination process, which results in a rule-based linear approach to reasoning. Instead, SPTs reflect-on-action, after the fact, to learn from a recent experience, compare that experience to others, and build a repertoire of prior experiences to draw from in future encounters (Gilliland & Wainwright, 2017).

The Commission on Accreditation in Physical Therapy Education (CAPTE) requires SPTs to pass CEs in a variety of settings (CAPTE, 2017). Clinical experiences are usually completed in three main settings: acute care, rehabilitation, and outpatient (CAPTE, 2016; Silberman et al., 2018). Silberman et al. (2018) found that the greatest number of incidences of CE difficulty occurred for the acute care setting (34.2%) followed by the outpatient setting (31.6%). Difficulties involving the cognitive learning domain resulted in the highest rate (40.9%) of CE failures. The cognitive domain of learning includes CR skills, such as foundational science knowledge, differential diagnosis, and using current literature for evidence-based practice. Foundational knowledge, application of evidence-based practice strategies, and CR skills are rated among the top elements necessary for SPTs to be ready to enter their first full-time CE (Timmerberg et al., 2019). Although the outpatient setting had the second highest degree of difficulty, it had the highest rate of cognitive issues leading to CE failure. Therefore, the outpatient setting was used as part of the conceptual framework for this study. Silberman et al. (2018) called for future research to investigate predictive factors of clinical difficulty to identify students at risk for failure.

Current Clinical Reasoning Assessments

The CAPTE includes teaching and assessment of CR in its Standards and Required Elements for Accreditation of DPT education programs (CAPTE, 2017). The ACAPT has a consortium for Clinical Reasoning Curricula Assessment and Research. The consortium is calling for the development of best practice guidelines for teaching and assessing CR in DPT education (Christensen et al., 2017). As the definition of CR varies between education programs, so does assessment of CR (Christensen et al., 2017; Furze et al., 2022). The majority of faculty use practical examinations, clinical fieldwork, written examinations, and written assignments to assess CR (Christensen et al., 2017). Generally, assessments range from self-assessment surveys to rubrics scored by faculty for an isolated course or event (Christensen et al., 2017; Kobal et al., 2021). Even CR assessment within clinical education varies with some sites using standardized tools from the American Physical Therapy Association and some sites using facility-specific tools (Huhn et al., 2018). Jensen et al. (2017) called for a longitudinal comprehensive approach to standardize learning outcomes for CR. Benchmarks have been recommended to assess CR performance; however, a need exists for standardization of CR teaching and assessment in DPT education programs (Christensen et al., 2017; Huhn et al., 2019; Reilly et al., 2022).

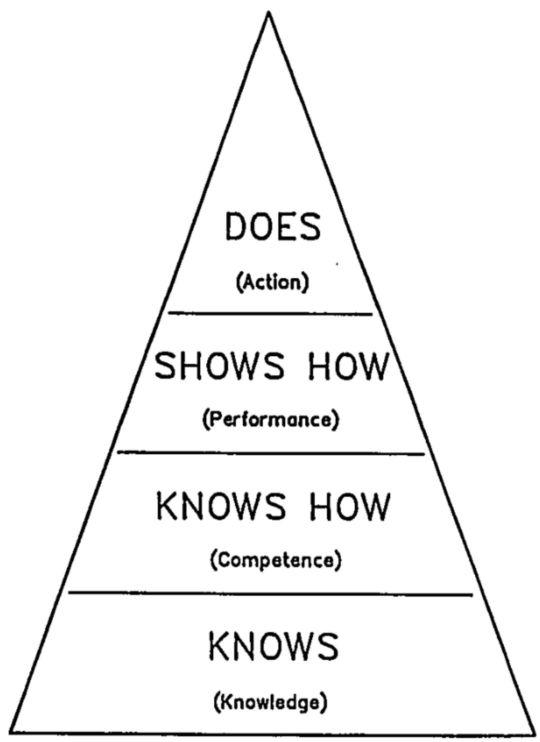
No single standardized assessment exists in DPT education for CR (Reilly et al., 2022). Therefore, the use and triangulation of multiple assessments have been suggested in recent studies (Fu et al., 2015; Furze et al., 2022). Furze et al. (2022) advocated for a shift to a theory-informed model for teaching, learning, and assessing CR. Comparing post-graduate residences to entry-level education, Furze et al. (2022) argued that a curricular paradigm is needed to establish proficiency in CR. This paradigm may include a domain of competence in CR, such as is currently used in residency education, that would benefit all physical therapy graduates, not only those pursuing residency. A proposed theory-informed paradigm for assessing CR in DPT education programs is discussed in the next section and is the focus of this study.

Conceptual Framework

The conceptual framework for this study comes from a seminal work by Miller (1990). Miller (1990) argued that no single assessment can provide enough data to judge any complex concept and that a framework of multiple assessments was instead required. Miller (1990) developed a pyramid (Figure 1) to model levels of assessment for complex concepts in medical education programs. The bottom tier of the pyramid represents a student's knowledge of what is required to perform the duties of the profession. Objective cognitive examinations are usually utilized to assess the bottom tier. However, knowledge by itself is insufficient for clinical practice. The second tier of the pyramid assures that the student knows how to apply the knowledge they have learned. The idea of competence in the second tier involves the ability to acquire information from a variety of resources, analyze and interpret data, and translate findings into a diagnosis or plan of care (Miller, 1990). Assessments, such as modified essay questions or patient management problems that use a clinical vignette, have been used for the second tier of the pyramid. Academic and psychomotor assessments may probe the first two tiers of the pyramid; however, these assessments fail to examine how a student reacts with a patient. Therefore, the third tier of the pyramid assesses whether the student can show how they integrate what was learned into a patient encounter. Assessments for the third tier involve some kind of patient substitute, such as a standardized patient who is either a paid actor or a licensed clinician or faculty member playing the role of a patient. These role-playing simulations also allow for assessment of students' oral reasoning, critical thinking processes, and judgment or decision-making. For medical programs, the Objective Structured Clinical Examination (OSCE) fits this type of assessment. The OSCE involves a multi-station examination of a standardized patient, data analysis and interpretation, and creation of a patient management plan. The top tier of the pyramid represents what the student does in the clinical practice setting. Miller (1990) offered no assessments of clinical or residency education in his article; however, instruments have been developed for these settings in recent years. Miller (1990) posed the question of whether the bottom three tiers of the pyramid, which he compared to an artificial examination setting, could predict the top tier in an authentic or real practice setting.

Figure 1

Framework for Clinical Competence Assessment



Note. From "The Assessment of Clinical Skills/Competence/Performance," by G. E. Miller, 1990, *Academic Medicine*, *65*(9), p. S63.

Problem Statement

Clinical reasoning is a difficult concept for physical therapy students to grasp and for DPT education faculty to teach and assess. CR deficits are responsible for CE failures, especially in the outpatient clinical setting. Currently, no standardized assessment or framework exists to evaluate CR in SPTs to predict CE success.

Purpose Statement

The purpose of this study was to determine to what degree CR performance in didactic coursework can predict CR performance in CEs for DPT students.

Overview of Methodology

This study used a quantitative exploratory observational design to predict one dependent variable from three independent variables. Archived assessments from a single DPT education program were utilized for data. The DPT program in this study is structured with a hybrid format of online and face-to-face learning and takes four years to complete. The curricular structure follows a traditional model with foundational science courses occurring during the first year, application courses occurring during the second year, clinical courses occurring during the third year, and full-time CEs occurring during the fourth year. The four-year curricular structure fits well with the four-tiered pyramid used for the conceptual framework in this study. For the bottom tier of the pyramid, *Knows*, scores from the first-year biomechanics course multiple-choice tests were utilized as a representation of foundational knowledge necessary for clinical reasoning and a factual recall type of assessment. For the second tier of the pyramid, *Knows How*, scores from second-year practical exams were utilized as a representation of interpretation and application of knowledge. For the third tier of the pyramid, *Shows How*, a practical examination called a *Triple Jump* was utilized as a representation of demonstration of knowledge in a supervised setting similar to outpatient clinics. For the top tier of the pyramid, *Does*, scores from a standardized assessment called the Clinical Internship Evaluation Tool (CIET) were utilized for clinical reasoning performance in actual clinical settings under the supervision of a clinical instructor who is a non-faculty licensed physical therapist. All assessments used for this study belong to the musculoskeletal or orthopedic domain of physical therapy practice. CIET scores were only used from outpatient clinical experiences because the outpatient setting mostly involves the musculoskeletal domain of practice.

Assessments utilized in this study belong to DPT students from the hybrid program's classes of 2022 and 2023. The classes of 2022 and 2023 were selected because prior to 2022, the assessment used for CE scoring was the Clinical Performance Instrument, which is a different tool with different scoring methods than the CIET. For simplicity and matching of data, only cohorts utilizing the CIET were included. Assessments were collected from each of the four years for the student population and were accessed through password-protected electronic sources, such as closed courses on the school's Learning Management System or archived documents in faculty files. All assessments and scores used for this study were de-identified for confidentiality and stored in the researcher's password-protected electronic files for security of information. The researcher used student identification numbers for aligning scores between independent variables and the dependent variable during data analysis.

Research Questions

This study addresses the following research questions:

1. To what degree will the level of knowing (factual recall of multiple-choice question assessments) predict study participant level of clinical reasoning performance in practice?
2. To what degree will the level of interpretation and application of knowledge be predictive of study participant level of clinical reasoning performance in practice?
3. To what degree will the level of demonstration predict study participant level of clinical reasoning performance in practice?

4. Considering study participant level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, which is most predictive of study participant level of clinical reasoning performance in practice?

Research Hypotheses

1. To what degree will the level of knowing (factual recall of multiple-choice question assessments) predict study participant level of clinical reasoning performance in practice?

H_0: The level of knowing will not significantly predict the level of clinical reasoning performance in practice.

2. To what degree will the level of interpretation and application of knowledge be predictive of study participant level of clinical reasoning performance in practice?

H_0: The level of interpretation and application of knowledge will not significantly predict the level of clinical reasoning performance in practice.

3. To what degree will the level of demonstration predict study participant level of clinical reasoning performance in practice?

H_0: The level of demonstration will not significantly predict the level of clinical reasoning performance in practice.

4. Considering study participant level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, which is most predictive of study participant level of clinical reasoning performance in practice?

H_0: Of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, none will significantly predict level of clinical reasoning performance in practice.

H_1: The combination of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, will significantly predict clinical reasoning performance in practice.

Overview of Analyses

Using a quantitative exploratory observational design, data were analyzed using statistical tests of prediction. Simple linear regression analysis was utilized for research questions 1-3. Multiple linear regression analysis was utilized for research question 4.

Preliminary Analysis

Preliminary analysis involved descriptive statistics for the student population included in the study as well as regression analyses for research questions 1-4. The class of 2022 had 43 students and the class of 2023 had 40 students who participated in outpatient CEs. A priori power analysis for multiple linear regression at a medium effect size (f^2=0.15, α=0.05, 1-β=.80) with three predictors required a sample size of 77 participants. This study had a sample size of 83, which is enough for statistically significant power.

Data Analysis by Research Questions

IBM SPSS 29.0 was used for data analysis. For research question 1, simple linear regression analysis was used to determine whether the bottom tier of the pyramid could predict the top tier. Scores for the first-year biomechanics course exams were entered into SPSS as the independent variable and the CIET scores were entered as the dependent variable. The same pattern was used for research questions 2 and 3. For research question 2, second-year practical exam scores were entered as the independent variable, and for research question 3, third-year *Triple Jump* scores were entered as the independent variable. For research question 4, a stepwise multiple linear regression analysis was conducted to determine which of the three independent variables best predicted the dependent variable of CIET scores. The multiple linear regression also determined the degree to which the three independent variables as a group predicted the dependent variable of CIET scores.

Delimitations

The limitations of this study are related to the sources of data. The researcher only has access to assessments from the DPT program used in this study; therefore, the participant population was limited to one location. If the researcher had access to assessments from other DPT programs, the participant sample size would increase and improve generalizability of results. However, many other DPT programs use the Clinical Performance Instrument for assessment of CE outcomes rather than the CIET. A benefit of using assessments from one DPT program is consistency in tools used during CEs.

Another limitation related to data sources is specific to the second tier of the pyramid, *Knows How*. Assessments fitting this category include those such as script concordance tests, key features tests, and virtual patients. The DPT program included in this study does not utilize any of these specific assessments. However, the main idea of the *Knows How* category is to use assessments that test application of knowledge rather than just recall. Usually, assessments in this category involve some kind of case scenario and a simulated patient. The practical exams used for this study to satisfy the *Knows How* category include a peer acting as a patient, a short case scenario, and application of knowledge specifically in the selection and performance of psychomotor skills.

Summary

Clinical reasoning is a complex concept with no current consensus amongst DPT educators regarding the best way to teach and assess it. CR is also a major factor contributing to

DPT students' CE failure in the outpatient setting. Many isolated assessments have been developed to measure CR in clinical settings or individual courses of a DPT program. However, no framework exists to comprehensively assess CR in DPT students across the educational curriculum. Furthermore, no framework exists to predict CR in the clinical setting based on CR measured during the DPT curriculum. This study aims to examine a potential framework for assessing CR in DPT students at multiple levels of the DPT education with the ultimate goal of predicting CR in the CE setting. If CR can be measured during the course of DPT education and can be predicted at the CE level of education, then faculty can identify CR deficits before students reach the level of CEs, remediate such deficits, and potentially prevent CE failures.

II. REVIEW OF LITERATURE

The purpose of this study was to determine to what degree CR performance in didactic coursework can predict CR performance in CEs for DPT students. CR is a major reason for failure in outpatient CEs (Silberman et al., 2018). Faculty need a system to assess DPT students' CR prior to CEs to determine clinical readiness and to try to prevent CE failures. Researchers have studied isolated assessments of CR during DPT education programs (Christensen et al., 2017; Furze et al., 2022; Huhn et al., 2018); however, no comprehensive framework has been developed to assess CR throughout a DPT program or for the purpose of predicting CE success. Miller (1990) created a framework for assessing clinical skills, competence, and performance for medical students. Miller's (1990) framework consisted of a pyramid with four tiers of assessment. From the bottom up, the four tiers are *Knows*, meaning acquisition of foundational knowledge; *Knows How*, meaning application of foundational knowledge; *Shows How*, meaning performance of a comprehensive task; and *Does*, meaning independent practice with little to no supervision. Miller's pyramidal framework provides a foundation for application to DPT education. In chapter two, literature is reviewed on the types of assessments included in the predictive analysis of the comprehensive framework.

Assessment for Each Level of the Pyramid

Knowledge-Level Assessment

The bottom level of Miller's (1990) pyramid is titled Knows and reflects the foundational knowledge required as a basis for clinical reasoning. Knowledge of domain-specific content is integral to the CR process. Proficiency or emerging ability is expected prior to the first CE in foundational knowledge subjects like anatomy, kinesiology, physiology, and tissue mechanics (Timmerberg et al., 2019). At the Knows level of the pyramid, any assessments for student's ability to recall knowledge within a domain of practice are applicable. Knowledge-level assessments can include multiple-choice question (MCQ) tests, free-response tests, and oral tests focusing on verbal repetition of facts (Miller, 1990; Pugh et al., 2019). Christensen et al. (2017) found that DPT programs use written exams (87.5%) to assess CR; however, written exams in isolation are insufficient to predict clinical performance.

Literature is scant for MCQ exams or written exams in DPT education and the connection to CR. However, literature exists for MCQ exams in other healthcare professions. Daniel et al. (2019) conducted an exploratory systematic review to create a practical collection of CR assessment methods to guide medical educators. Using a constructivist research paradigm and a scoping methodology, Daniel et al. (2019) used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis protocol to search multiple databases from the time of each database's inception through the year 2016 for articles about CR assessments. Inclusion criteria were all study design types, any article that explicitly studied a method or tool of CR assessment, and any health profession at any stage of training or practice. After discovering the volume of articles fitting inclusion criteria, the authors decided to narrow the scope to only the medical health profession. Exclusion criteria were non-English articles, essays or commentaries, and if the article focused on decision-making applied only to a specific clinical problem instead of broader cognitive processes. Pairs of authors analyzed articles and reached consensus using a survey regarding each method's ability to assess different components of CR. Interrater agreement was assessed using Cohen's kappa with the following ranges: 0.0-0.5 poor, 0.6-1.0 average, 1.1-1.5 good, and 1.6-2.0 very good.

Daniel et al. (2019) examined 377 articles spanning 50 years of publication. Participants were medical students, residents, and practicing physicians. Articles included 20 individual assessment methods which the authors placed into three categories: non-workplace-based assessments, simulated clinical environment assessments, and workplace-based assessments. The MCQ examination method was categorized under non-workplace-based assessments. MCQs in the articles studied were comprised of clinical vignette stems followed by up to five alternatives including: single best alternative, matching, true/false, and a combination of alternatives. Psychometric analysis of MCQ exams demonstrated the capacity to test a wide range of knowledge in a short time period. MCQs in recent years have been used for medium-high stakes exams, especially in summative formats like licensure exams. MCQs are one of the most frequently used non-workplace-based assessments and provide broad sampling that helps minimize context specificity. Advantages of MCQs are that they have the best chance of high internal consistency and, therefore, the best utility for high-stakes assessments. The high consistency, meaning there is a "right" answer, allows a measurement of accuracy. MCQs have strong content validity due to expert consensus and blueprinting, and they assess a standardized problem set.

Daniel et al. (2019) had the best consensus about MCQs assessments for the CR components of Leading Diagnosis (1.9) and Management and Treatment (1.8). MCQs had less consensus for the following CR components: Information Gathering (0.9), Hypothesis Generation (0.3), Problem Representation (0.0), Differential Diagnosis (0.6), and Diagnostic Judgment (0.0). Daniel et al. (2019) acknowledged limitations of their study, including time constraints for analyzing all references, which meant the possibility of excluding some new developments. A late decision was made to exclude other health professions from the search, which was pragmatic but limited the scope of the review. Because MCQs are considered part-task assessments of CR, Daniel et al. (2019) advocated programmatic assessment including multiple methods or sources to comprehensively examine CR competence. Future research should gather information longitudinally from multiple CR assessments and triangulate findings to guide decisions regarding clinical readiness or achievement of competence.

Surry et al. (2017) explored MCQ exams in relation to CR processes and behaviors observed clinically. Using a qualitative design sensitized by dual-process theory, Surry et al. (2017) challenged assumptions about the relationship of MCQ exams to CR in practice. MCQ exams are assumed to provide meaningful information about CR; however, MCQs have been criticized as testing factual recall and how well an individual can take a test rather than assessing higher level cognitive processing. Two major assumptions exist regarding MCQ exams and CR. The first assumption is that "real-world" CR should be elicited by the MCQ assessment tool. The second assumption is that MCQ exam scores give useful information about the quality of CR. Evidence from the first assumption should support the second assumption. Surry et al. (2017) challenged the first assumption with three expectations: non-analytical and analytical reasoning should co-occur at the item level, some evidence should exist of emotional reactions to MCQs, and items should elicit some uncertainty, which is associated with analytical reasoning because clinical patient encounters involve much uncertainty. If findings aligned with the three expectations, then MCQ exams would significantly reflect CR processes in practice.

Fourteen participants were recruited: three medical students, five residents, and six faculty physicians affiliated with Surry et al.'s (2017) institution. Only three participants were female. Participants completed three sets of five timed MCQs (two minutes per item) from the American College of Physicians Medical Knowledge Self-Assessment Program in the field of internal medicine. Participants then completed retrospective think aloud sessions to describe their thoughts when solving the MCQ items. Surry et al. (2017) used the dual-processing theory as the conceptual framework for the study, and a constant comparative analysis for thematic analysis of the data. Dual processing theory presents two categories of reasoning. Type 1, or System 1, non-analytical reasoning is fast, subconscious, requires low effort, and involves pattern recognition, heuristic use, and intuition. Type 2, or System 2, analytical reasoning is slow and effortful and includes hypothetico-deductive reasoning, as well as what Surry et al. (2017) called pathophysiology-based reasoning, meaning use of foundational science knowledge in reasoning processes.

Surry et al. (2017) found three categories of participant behaviors: CR behaviors, test-taking behaviors, and reactions to MCQs. CR behaviors were defined as strategies to make judgement, generate solutions, and reach conclusions during the MCQ exam. Data analysis revealed two categories of CR behaviors consistent with dual-processing theory. Non-analytical (Type 1) reasoning was described in 100% of MCQ items. The most common strategy used by participants was pattern recognition (88.1%), followed by heuristics (44.3%), and then intuition (37.6%). Analytical (Type 2) reasoning was described in 97.1% of MCQ items and all co-occurred with non-analytical (Type 1) reasoning processes. Identifying key features was described for 95.2% of MCQ items, which involved extracting important pieces of information from the question stem to decide on an answer. Pathophysiology reasoning occurred in 19% of MCQ items, and hypothetico-deductive reasoning only occurred in 3.8% of items. The co-occurrence of analytical and non-analytical processes supported Surry et al.'s (2017) first expectation. Co-occurrence was nearly universal (97.1% of items), indicating that cognitive processes utilized during the MCQ exam were similar to processes used for CR in clinical practice.

The other two categories Surry et al. (2017) found, test-taking behaviors and reactions to MCQs, addressed the other two expectations of the study. Test-taking behaviors were defined as activities intended to leverage aspects of the exam items to improve scores. Process of elimination was the most common strategy (67.6%), followed by clue seeking (32.4%). Process of elimination involved removing answer choices to narrow the range of possibilities and improve chances of answering correctly. Clue seeking used clues from previous items on the exam to answer a current question. The remaining test-taking strategies were not as common: inattention (10%), guessing (6.2%), changing answers (3.3%), and time management (2.8%). Reactions to MCQs were defined as self-reflections or emotional responses prompted by the exam items. Uncertainty was the most common reaction (70%), followed by emotional responses (16.7%) of laughter or frustration, and then confidence (7%). The expectation of emotional reactions with reasoning as indicators of authentic engagement with MCQ items was only evident in a few participants, and thus did not align with Surry et al.'s (2017) second expectation. However, the third expectation that items should elicit uncertainty was supported by the findings. Uncertainty, associated with analytical reasoning, was found in participant descriptions of 70%

of MCQ items. Content domains with more uncertainty also had more items with pathophysiology reasoning, which supports analytical (Type 2) reasoning processes.

Surry et al. (2017) addressed challenges regarding test-taking behaviors of clue seeking, guessing, and process of elimination as inauthentic to real-world practice. Clue seeking behaviors can be seen clinically in using clinical support tools, algorithms, protocols, or medical references. Guessing and the process of elimination can be seen in differential diagnosis processes. Therefore, seemingly inauthentic aspects of test-taking during MCQ exams can relate to authentic CR processes during practice. In summary, Surry et al. (2017) found evidence confirming the presence of CR processes occurring during MCQ exams that reflect CR processes in practice, thus validating MCQ exams as a tool for assessing aspects of CR. Limitations of Surry et al.'s (2017) study include small sample size and poor representation of the study population (only three medical students and only three female participants). Surry et al. (2017) suggested future research to replicate the study with larger sample sizes and different specialties in the medical field. Surry et al. (2017) also suggested future research to explore how MCQs designed to assess rote memorization relate to CR processes, how CR behaviors observed relate to MCQ exam performance, and how CR behaviors vary based on experience and overall performance, which would be evidence for the second assumption of MCQ exams providing information about the quality of CR.

Application-Level Assessment

The second tier of Miller's (1990) pyramid is the Knows How level, which includes application of foundational knowledge to context specific situations or scenarios. Timmerberg et al. (2019) found that an emerging level of competence is expected prior to the first CE for students' understanding and skill to perform examination and intervention procedures.

Assessments at the Knows How level include written exams or assignments, case-based tests, Script Concordance Tests (where students respond to a series of items following brief case scenarios), Key Features Tests (where students answer questions focusing on key features of a case scenario), virtual patients, patient management problems, and clinical reasoning problems (Fu, 2020). Application-level assessment of CR typically involves basic procedural thinking and applying foundational knowledge to short clinical scenarios to begin situational awareness.

Reilly et al. (2020) used application-level assessment as part of a model implemented into the curriculum of a DPT program to teach clinical skills. Reilly et al. (2020) created a model of integrated labs as part of an entire curriculum revision for a single DPT program. The model was designed to reinforce motor learning strategies in clinical skills development, increase experiential learning opportunities, and promote CR (Reilly et al., 2020). Using a retrospective design, data were collected for 62 SPTs who participated in the integrated labs. Four outcomes were measured: Skills Competency Tests, Clinical Competence Performance Exams (CCPEs), the Clinical Performance Instrument (CPI), and the National Physical Therapy Exam (NPTE).

The model was designed to progress the labs in complexity and difficulty over a period of six semesters. Two to three weeks after initial instruction of skills, students were examined through application-level assessments called Skills Competency Tests (SCTs). The SCTs progressed from basic skill performance to performing skills as they relate to a patient case to performing skills on authentic patients. The SCTs allowed faculty to evaluate whether students had acquired the psychomotor abilities required for clinical practice and to provide feedback related to skill performance and a level of assurance that students were ready to progress to the next phase of the lab. Each integrated lab course culminated in a CCPE, a case-based final exam that integrated skills from all clinical sciences taught up to that point. CCPEs were assessed using the Clinical Reasoning Grading Rubric and determined students' readiness for CEs. CEs included short interspersed clinical rotations during the didactic curriculum as well as full-time terminal CEs at the end of the didactic curriculum. CEs were assessed with the CPI, specifically the CPI components of CR, screening, examination, evaluation, diagnosis, prognosis, procedural interventions, and educational interventions.

All students passed the SCTs during semesters 1-6, given a maximum of three trials each. All students, except one, passed the CCPEs at the end of each semester in one of three trials. All students met expectations on the CPI for four CEs. Finally, all students successfully passed the NPTE. Given the positive outcomes, Reilly et al. (2020) concluded that the model for the integrated labs was successful. However, Reilly et al. (2020) acknowledged that similar outcomes existed prior to the implementation of the model into the curriculum. Therefore, more research is needed to determine the impact of the model on student outcomes versus outcomes without the model. Because the model was implemented as part of curricular revision for the entire DPT program, confounding variables prevented isolation of outcomes related to the model alone. Finally, Reilly et al. (2020) used application-level assessment, SCTs, as part of the overall model, and CR was assessed as part of the SCTs. However, future research should examine CR skills resulting from the model implementation as a whole.

Furze et al. (2015) conducted a qualitative longitudinal study to explore CR development in PT students. Application-level assessment was one of the tools in Furze et al.'s (2015) study, contributing to a broader understanding of the progression of CR development over time. Furze et al. (2015) used the Dreyfus model as a conceptual framework. The Dreyfus model included three phases for moving from novice to expert. Phase one involved moving from a system of absolute rules to incorporating past experiences to guide decisions. Phase two involved moving away from gathering unrelated pieces of information to seeing a cohesive picture for a situation. Phase three involved viewing oneself as an integral part of the situation rather than as an outsider. Furze et al. (2015) recruited 98 students from two consecutive cohorts (Class A and Class B) at a Midwest DPT program. Most of the students were female and Caucasian, and the mean age was 23. The 3-year, 8-semester curriculum for the DPT program began with foundational science and clinical skills courses, followed by clinical integrative and contextual science courses, and ended with full-time CEs. Shorter CEs were interspersed between semesters and case applications were also interspersed throughout the curriculum, beginning with simple cases and progressing to complex cases. At the end of each semester, students participated in a Clinical Competence Performance Examination (CCPE), a practice-based case simulation assessment where faculty played the role of patient and evaluator.

Two instruments were utilized to collect data for Furze et al.'s (2015) study. The Clinical Reasoning Reflection Questionnaire (CRRQ) was a six-question online survey developed by Furze et al. (2015). The questions addressed concepts of metacognition, struggling with uncertainty, critical self-reflection and growth, students' ability to apply the ICF model to a patient case, and describing students' CR processes. Students completed the CRRQ 20 minutes before the CCPE and 20 minutes after the CCPE. The second instrument utilized was the Clinical Performance Instrument (CPI), specifically narrative responses from the Clinical Instructor (CI) for item 7 on Clinical Reasoning. Data were collected at four time points during the curriculum. At the end of semester four, Class A completed the CRRQ, and data were collected from the CPI for a 3-week CE. At the end of semester six, Class A and Class B completed the CRRQ. At the end of semester seven, CPI midterm data was collected during a 20-week CE for Class A and Class B. At the end of semester eight, Class A and Class B completed the CRRQ, and CPI data were collected for both classes at the end of a 16-week CE. Using the constant comparative method, data were triangulated and coded for key words, concepts, or phrases. Preliminary codes were analyzed until saturation was reached and concepts were categorized. Then, during a course of six meetings, Furze et al. (2015) conducted axial coding for themes. Finally, an external audit was conducted by an experienced qualitative researcher.

Eight themes from three semesters emerged that aligned with the Dreyfus model. Themes from semester four aligned with adherence to absolute rules. The theme of "focus on self" meant that students were focused on their own skill rather than being patient-centered during the encounter. The theme of "compartmentalized thinking" involved checklists, the use of routine tests or measures without regard for patient response, difficulty organizing data gathered, and difficulty understanding how to interpret data. The theme of "limited acceptance of responsibility" included students blaming external sources, like instructors, for their own mistakes. Themes for semester six aligned with Dreyfus' phase two of starting to see the bigger picture. "Procedural performance" was a theme of focusing on procedural aspects of patient management without recognizing context, and difficulty incorporating factors beyond the textbook. "Initial stages of recognition using context" was a theme showing more incorporation of context, like patient age and environmental factors, into simple cases. Students were able to envision a picture of the patient based on initial case information, however, they could not adapt to information during the encounter that differed from their expectations. "Improved reflection on performance" was a theme of reflection-on-action. Students started directing attention to patient needs rather than to their own performance. Themes for semester eight reflected Dreyfus' third phase of integrating self into the situation. The theme of "dynamic patient interaction" showed students being more flexible and able to change direction during the patient exam, actively listening, anticipating the unknown, and reflecting-in-action. The theme of "integrating situational awareness" involved students applying contextual information to patient management, relating to the patient's story, and narrative thinking.

Overall, Furze et al. (2015) found data to support the Dreyfus model as a possible framework to categorize CR development over time. Themes supported the Dreyfus phases shifting focus from internal to external and shifting reasoning from rigid to flexible. Students began with more scripted, procedural reasoning that was self-focused and only able to manage simple cases. The early phase of this level of reasoning is common at the application-level of assessment, or the second tier of Miller's (1990) pyramid. Over time, students shifted to fluid reasoning that was patient focused. Furze et al. (2015) asserted that curricular design may heavily influence students' reasoning development. Early curricular courses and content focused on skill development and testing for a right answer, which could have contributed to shallow reflection by students. Early attention to psychomotor skill development may have been a reason for the internal or self-focus of students during these points in the curriculum. Furze et al. (2015) advocated for early learning experiences and assessments that foster consideration of broader situational context. Limitations of Furze et al.'s (2015) study included using only one DPT program, as well as having an incomplete cohort at the end of semester eight due to graduation. Curricular design was also a limitation because it focused on procedural outcomes early, which may have contributed to student responses at the early time points. Also, the interspersed curriculum cases progressed from simple to complex over time, which may have impacted student responses on surveys. Future research, suggested by Furze et al. (2015), should develop standards for each level of development of CR to provide benchmarks for competence. Future research should also determine the best tools to assess CR as well as strategies to facilitate the

CR learning process in both didactic and clinical curricula.

McDevitt et al. (2019) conducted a study using the Clinical Reasoning Assessment Tool (CRAT) to determine whether the tool reflected student progress in developing CR over time as well as to determine whether case context impacted student CR performance. The CRAT was created to assess CR abilities of SPTs over time. The CRAT was developed based on theoretical frameworks including structure and application of foundational knowledge, skill acquisition based on the Dreyfus model, and consensus-based descriptions of CR from experts to set the evaluation scale from beginner to proficient. The CRAT represents CR as having three domains: content knowledge, procedural knowledge and psychomotor skills, and conceptual reasoning. Using a cross-sectional design, McDevitt et al. (2019) recruited 55 SPTs from two consecutive cohorts at a single DPT program. Eleven faculty and clinical assessors scored 172 CRATs using a visual analog scale of 0-16 points representing a learner continuum of beginner, intermediate, competent, and proficient.

The four time points assessed chronologically were

- Year 1, Semester 2 using the application-level assessment of a clinically based practical exam;
- Year 1, Semester 3 using a standardized patient (SP) formative assessment simulation;
- Year 2, Semester 4 during a full-time CE; and
- Year 2, Semester 5 using a SP summative practical simulation.

For the first time point, two assessors examined students in two skills-based stations (examination and intervention). Students performed an examination and selected an intervention based on a patient vignette. The patient was played by a second-year student, and assessors completed the CRAT after the student performance. For the second time point, three assessors examined students across three content areas (musculoskeletal, neurological, and medical conditions). Students had three consecutive encounters with three different SPs. Assessors watched video recordings of student performance and met with students two weeks later for 30-minute feedback sessions. Assessors completed the CRAT after the feedback sessions. For the third time point, CIs completed the CRAT while observing students during a single patient encounter. Outpatient, inpatient, home health, and pediatric clinical settings were included in the third time point measurements. For the fourth time point, the SP simulations were conducted exactly the same as the second time point. McDevitt et al. (2019) analyzed data using analysis of variance (ANOVA) to determine whether time and case context were predictors of performance in each of the three CR domains. A second ANOVA was used to determine if content area was a predictor of student CR performance during the SP simulations in each of the three CR domains.

McDevitt et al. (2019) found that mean scores in each of the three CR domains increased at each time point assessed. Each time point was significantly predictive of student CR performance in each of the three domains ($p < .0001$ for each). For the first time point, sample behaviors on the CRAT observed by assessors included moderate evidence of foundational knowledge and application of the ICF model; moderate accuracy in performing tests, measures, and interventions; evidence of ability to justify most tests, measures, and interventions; and ability to identify relevant patient problems while creating a working hypothesis and patient problem list. For the fourth time point, sample behaviors on the CRAT observed by assessors included more refined skills, stronger evidence of foundational knowledge and patient-related ICF components, stronger accuracy and efficiency in performing examination and intervention skills, appropriate communication, improved ability to confirm or disprove working hypotheses, and stronger justification for decisions. Thus, a steady improvement in CR abilities was observed over time. However, for students lacking foundational knowledge or psychomotor skills, CR did not develop independently. Therefore, foundational and clinical sciences appear to be essential ingredients for the development of CR. Results of the second ANOVA showed that case context is predictive of procedural knowledge (p = .007) and conceptual reasoning (p = .0297), but not content knowledge (p = .326). Also, students performed better in musculoskeletal case contexts than in neurological or medical conditions contexts.

McDevitt et al. (2019) accomplished the purpose of their study by demonstrating that the CRAT is a useful tool for measuring CR performance in SPTs over time. McDevitt et al. (2019) showed that the CRAT is a generalizable tool for didactic and clinical portions of PT education programs, across time points and clinical settings, and across assessors. However, the intra- and inter-rater reliability for the CRAT has not been established in the literature, and McDevitt et al. (2019) suggested exploration of the tool's reliability for future research. Limitations of McDevitt et al.'s (2019) study include only using a single DPT program, which limits generalizability of the results. The curricular content being taught at each time point examined may have influenced CRAT scores and student performance. McDevitt et al. (2019) suggested future research explore case context and curricular influence on CR performance. Each time point assessed also varied in structure, environment, and case content. Using the CRAT across multiple clinical sites created inconsistency with how the tool was used or scored by multiple CIs. The cross-sectional design limited the sample in that more cohorts were not followed over time. Finally, McDevitt et al. (2019) only used students in a DPT program, and suggested future research explore CRAT scores for graduates, early career PTs, residents, and fellows. Another suggestion for future research was to gather qualitative data to examine how the CRAT influences teaching, learning, and assessment in PT education.

Competence-Level Assessment

The third tier of Miller's (1990) pyramid is titled Shows How and indicates a complex performance of knowledge and skill from the student. Demonstration of CR skills for non-complex patients is expected prior to beginning the first full-time CE. An emerging level of competence is expected for CR skills like applying evidence-based strategies to guide decision-making, articulating rationale during a patient evaluation, developing and linking patient goals to activity limitations or participation restrictions, determining whether a patient is appropriate for the physical therapy scope of practice, interpreting examination findings, and screening to rule in or rule out conditions (Timmerberg et al., 2019). The Shows How level requires a more realistic situation with a simulated patient portrayed either by a faculty member or a SP. The component of a SP introduces uncertainty and unpredictability as well as the personal aspect of therapist-patient interaction. The task is more comprehensive in nature than the second tier of Miller's (1990) pyramid and usually involves a complete patient evaluation from start to finish and a hypothesis or diagnosis for the source of the pathology or problem. Christensen et al. (2017) found that DPT education programs use practical exams (99%) to assess CR at the performance level. A practical exam for this level may include stages or phases of the exam in a certain sequence to simulate the process of patient evaluation. In medical education programs, the Objective Structured Clinical Examination (OSCE) has been used to assess performance level CR by using sequential stations for each part of the medical examination process. In DPT education programs, practical exams similar to OSCEs have been developed, such as the TASPE (Fu, 2015) and the *Triple Jump*.

Pelland et al. (2022) conducted a systematic review to appraise the scientific rigor of the OSCE in health sciences education programs. OSCEs have been used to determine clinical readiness in medical students; however, different health sciences programs have adapted the original structure of the OSCE from that used for medical students to a design fit for individual program needs. Therefore, research was warranted to assess the OSCE's validity and reliability in other health sciences professions. Pelland et al. (2022) searched six databases using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol. A review panel including a health sciences librarian, two PTs, and a health education scholar searched for articles in which the OSCE was used in PT education. Due to the lack of literature on PT education, the search was expanded to include related healthcare disciplines such as occupational therapy and speech-language pathology. Studies on medical or nursing education programs were excluded as well as non-English language studies. Data were extracted using the utility model, which evaluated validity, educational impact, reliability, acceptability, and costs of the OSCEs. Although 106 articles were identified, only 13 studies met eligibility criteria. The 13 studies were published between 2003-2020 and included global locations, four PT-specific articles, five studies from graduate programs, and eight studies from undergraduate programs. OSCEs in the 13 studies covered practice areas of cardiorespiratory, neurologic, musculoskeletal, exercise prescription, and professional behaviors. All OSCEs in the 13 studies were utilized for summative assessment of students.

Pelland et al. (2022) found that OSCEs used in health sciences programs had low to fair concurrent validity, low to acceptable reliability, and poor to low predictive validity for clinical readiness. Concurrent validity exhibited a large range of correlation values between the OSCE scores and traditional clinical assessments and was affected by the number of practice areas sampled within the OSCE. Concurrent validity was lower when multiple areas of practice were combined within one OSCE compared to when only a single practice area was included in the OSCE. Internal consistency and reliability depended on several factors. Interstation consistency was low to good (Cronbach's alpha, 0.48-0.77). Interdomain consistency was good to high (Cronbach's alpha, 0.70-0.83). Professional and cognitive domains had higher agreement in scores (8% variance of between-student scores and ICC 0.39-1.0) than psychomotor domain scores (30% variance of between-student scores and ICC 0.25-0.63). Predictive validity for OSCEs predicting CE scores ranged from low (-0.01 to -.16) to moderate-strong (0.36-0.62). Predictive validity was higher for psychomotor skills than for professional behaviors. Overall, OSCEs predicted less than 50% of the variance in CE scores and were influenced by the timing of the OSCE during the education program (early $r = 0.32$, late $r = 0.63$).

Pelland et al. (2022) concluded that the scientific rigor of the OSCE is insufficient for its use in making academic decisions regarding a student's status in an educational program or a student's clinical readiness. Limitations of the study by Pelland et al. (2022) include selection bias due to the retrospective study design, inadequate control of confounding variables in the articles included, and a gap in the literature (only 13 articles eligible) for the rigor of OSCEs used in health sciences education programs. Pelland et al. (2022) suggested that OSCEs may be more appropriate for formative assessment and education rather than summative, and that DPT programs should develop a culture of assessment that aligns better with the emerging competency-based education framework sweeping DPT education nationwide. Although the OSCE may be insufficient to predict clinical readiness alone according to Pelland et al. (2022), an earlier study examined the OSCE in combination with other assessments to predict clinical readiness.

The OSCE was examined in a study by Terry et al. (2020), along with two other discrete course assessments in relation to SPTs' clinical performance. Terry et al. (2020) designed a retrospective observational cohort study to determine relationships between SPT assessment scores in pre-clinical coursework and future performance in CEs. In semesters two and three of an Australian DPT program, three assessments' mean scores were collected: OSCEs, written exams, and seminar presentations. The written exams included a combination of short answer questions and MCQs. The seminar presentations were only done in the cardiopulmonary class, whereas the other two assessments were part of all classes examined: cardiopulmonary, orthopedics, musculoskeletal, and neurological. OSCEs assessed professional conduct, communication, clinical reasoning, technique performance, and safety. The dependent variable of CE performance was measured by the valid and reliable Assessment of Physiotherapy Practice (APP) instrument. Each CE was full-time and five weeks in duration immediately following the subject or course in which students had just completed the OSCE, written exams, and/or seminar presentations. Final APP mean scores were included for analysis.

Terry et al. (2020) recruited 118 SPTs (62 male, 56 female) from four consecutive cohorts (2010-2013) at a single Australian DPT education program. Participants were excluded from the study if they did not successfully complete the first semester of the program. Data analysis included Pearson's correlations of mean scores for each independent variable ($p < .05$) and the mean final APP scores using a Bonferroni adjusted p-value of 0.0125 (0.05/4). OSCE mean scores resulted in the strongest correlation to APP mean scores ($r = 0.57$, $p < .001$), followed by a moderate correlation for written exams ($r = 0.39$, $p < .001$) and a weak correlation for seminar presentations ($r = 0.29$, $p = .012$). Terry et al. (2020) highlighted the moderate correlation of written exams to APP mean scores as supporting the idea that students' foundational knowledge revealed in written assessments is related to their application of that knowledge in CE settings. The written assessment findings reinforce literature supporting written exams as a means of measuring the Knows tier of the pyramidal framework.

Terry et al. (2020) also performed a multiple regression analysis to determine the effect of each pre-clinical assessment on APP scores. The model including mean scores of the OSCE, written exams, and seminar presentations resulted in the best fit and was a moderate predictor of APP scores ($F_{[3,114]} = 20.26$, $p < .001$, $R^2_{adj} = .003$, SEE = 6.88). However, the OSCE mean scores resulted in the best individual predictor of APP scores (adjusted $R^2 = 0.33$, $p < .001$, ß = 0.49, 95% CI 0.46-0.98). OSCEs explained 32% of the variance in SPTs' future CE performance, especially for practice categories of intervention, assessment, analysis, and planning. Therefore, OSCEs have potential for use as early indicators of students' future CE performance.

The study by Terry et al. (2020) had implications for future research into the design, implementation, and evaluation of early intervention protocols to improve clinical readiness in SPTs. However, the study had several limitations. Sampling was from a single DPT program, which limited generalizability. The structure of the stations in the OSCEs was variable between courses examined, and the timing of the OSCEs was immediately following written exams, which may have influenced students' performance during the OSCEs. The CEs included in the study were each only five weeks in duration and immediately followed the course matching the subject matter and setting for the CE, which may have resulted in higher APP scores than if CEs had occurred with a time lapse between courses matching the subject or setting of the CE. Finally, the assessments included in the study by Terry et al. (2020) only occurred in semesters two and three of the DPT program, which does not reflect factors that may have influenced student performance at varying points throughout the duration of the entire DPT program.

Fu (2015) developed an assessment similar to the OSCE for DPT education programs and for assessment of CR specifically. The Think Aloud Standardized Patient Exam (TASPE) is a practical exam wherein students provide verbal reasoning at three points sequentially. Students were given a patient history and then verbalized and defended what they thought the three most likely diagnoses were for the case. Then students conducted a physical examination on a SP, after which they again verbalized and defended what they thought were the three most likely diagnoses. Third, students explained and justified their decisions for selecting the three most appropriate interventions for their top choice of diagnosis. Because of its three-part design, the TASPE is very similar to a *Triple Jump* assessment.

Fu (2015) explored the reliability and usefulness of the TASPE with an experimental design involving 28 SPTs and four examiners. The examiners scored students' performance during the TASPE event. Modest interrater reliability was found using Spearman rho for total think aloud scores (0.63-0.98). High internal consistency was found for the ordinal alpha (0.93). To determine usefulness, both examiners and students completed surveys ranking their level of agreement with survey questions. Examiners rated the TASPE higher than students in terms of usefulness; however, no significant difference was found between the two groups except for one question on the survey. The results showed that the TASPE may be a reliable and useful tool for assessing SPTs' CR; however, Fu's (2015) study had several limitations. The TASPE was developed using only the hypothetico-deductive reasoning model to define CR. The TASPE also was only applied to the musculoskeletal domain of practice and needs further research to explore other practice domains. A convenience sample of students and examiners was utilized, limiting generalizability of results. Fu (2015) recommended future research to determine the validity and applicability of the TASPE. Although the TASPE is an assessment addressing the third tier of

Miller's (1990) pyramid, it is limited in its scope for a comprehensive assessment of CR in SPTs throughout the length of a DPT program. Fu (2015) recommended triangulating assessments at each level of Miller's pyramid for a more comprehensive approach to assessing CR in SPTs.

Performance-Level Assessment

The top tier of Miller's (1990) pyramid, Does, involves performance-level assessments. The top tier of the pyramid represents students' performance in authentic clinical settings. For SPTs, assessments at the performance level include those completed by CIs during full-time CEs. According to Christensen et al. (2017), DPT programs use CEs (94.8%) to assess CR. Until recent years, the main assessment tool used for full-time CEs was the CPI. The CPI has 18 criteria for CIs to assess at midterm and at the end of a full-time CE. The 7th criteria on the CPI is Clinical Reasoning. CPI line items are graded on a visual analog scale from Beginner to Entry-level. CIs may write comments for each line item as well. Students also assess themselves using the CPI and compare their self-assessment to that of the CI. The CPI is time consuming for students and CIs to complete, and for clinical education faculty to review. Many DPT programs still use the CPI; however, in recent years some DPT programs have switched to using the Clinical Internship Evaluation Tool (CIET). The CIET has 42 items divided into two categories: Professional Behaviors (18 items) and Patient Management (24 items). CIs score items using a scale of *Never* (0) to *Always* (4) for the Professional Behaviors section and *Well below* (1) to *Well above* (5) for the Patient Management section. At the end of the CIET is a Global Rating score for CIs to provide an overall impression of the student's competence for clinical practice. The developers of the CIET, Fitzgerald et al. (2007), found that all items on the CIET were representative of skills and behaviors necessary for competent PT practice. Internal consistency for the CIET was .98 for Patient Management items, and a moderate correlation (.76) was found for Patient Management section scores and the Global Rating scores at the end of the tool. CIs (96%) responded to surveys strongly agreeing that the CIET was short and easy to use.

One DPT program reported the process of implementing the CIET to replace the CPI. North and Sharp (2020) did a case report to disseminate the process and findings of exploration and implementation of the CIET. The process involved four steps: initial exploration and comparison of the CPI versus the CIET, a pilot trial of the CIET simultaneous with the CPI, processes involved in transitioning to the CIET, and evaluating the transition to the CIET. To explore and compare the two tools, key stakeholders were identified and consulted throughout the process. Stakeholders included the clinical education program within the DPT program, the academic program, clinical partners separated as Site Coordinators of Clinical Education (SCCEs) and CIs, and students in the DPT program. The clinical education program included the Director of Clinical Education (DCE) and Assistant Director of Clinical Education (ADCE). The clinical education program examined time, cost, and data management of the two tools for feasibility, and compared tool items to determine which items from the CPI were present or absent in the CIET. The clinical education program also consulted with other DPT programs in the region and with the CIET developers. The academic program included the DCE and ADCE, as well as the Program Director, and a Director of Educational Innovation and Evaluation. The academic program analyzed the tools' correlation to required accreditation standards as well as to the program's annual goal assessment and strategic plan. Clinical partners included clinical sites hosting more than 10% of the program's students per calendar year within the past five years, or those having active research collaborations with the program. SCCEs answered surveys regarding training needs and formats for the CIET as well as support for a new assessment tool. CIs answered surveys regarding their opinions of the CPI and their interest in adopting an alternate tool. Finally, third-year students in the program with previous experience using the CPI were consulted. Results of the initial exploration and comparison phase of the process were analyzed to inform the next phases of the process. The clinical education program found that the items of the CPI were also represented in the CIET. Of note was that the CIET embeds CR across multiple categories rather than as an isolated element like the CPI. The academic program found that accreditation standards for curricular content correlated with CIET items and that the program had three goals linked to CIET outcomes. SCCEs provided unanimous support for adopting the CIET. Fifty-six CIs completed surveys regarding openness to an alternate tool (25.7% response rate). CIs reported that the CPI had redundant categories, took too long to complete, and 90% of respondents recommended the program switch to the CIET. Student responses to surveys echoed those of the CIs.

The next phase of the process was a pilot trial of CIET simultaneous with CPI. Sixteen pairs of students and CIs volunteered for the pilot; however, they received minimal training on the CIET. Survey response rates were 68.8% for CIs and 81.3% for students. For the CIs, 63.6% preferred the CPI over the CIET; however, mixed qualitative feedback indicated uncertainty in CI understanding of performance expectations using the CIET. For the students, 84.6% preferred the CIET over the CPI. North and Sharp (2020) then described the processes involved in transitioning to the CIET. The CIET was implemented with the first CE for the cohort entering their final year of the DPT program.

Overall, it took 19 months for the total transition to the CIET. One cohort of 50 third-year students and their paired CIs completed the CIET five times as the sole assessment tool during terminal CEs. To evaluate the transition to the CIET, each stakeholder group played a role. The DCE and ADCE examined the tool's ability to provide a comprehensive picture of student performance and its ability to identify students needing additional support. The CIET showed a better level of detail for student performance and red flag markers that allowed for a proactive rather than reactive response for student support. The CIET provided aggregate data outcomes for the entire tool and for each tool section, which was not a feature of the CPI. The academic program and DCE evaluated outcomes of the transition. The CIET provided more detailed student outcomes in each performance category and allowed more refinement of performance assessment across the cohort. The SCCEs participated in open informal communication with the program and CI survey responses were tracked by the DCE and ADCE. The SCCEs reported no major barriers or concerns for using the CIET. CIs reported some challenges with the CIET online platform; however, 87% of CIs preferred the CIET over the CPI. Students were surveyed this time only on the CIET because this cohort of students did not have experience with the CPI. Students also reported challenges with the online platform; however, they liked the fact that they received substantial feedback from the CIET.

North and Sharp (2020) found the CIET to be a viable tool to measure student performance during clinical practice. CIs and students spent about half of the time completing the CIET compared to the CPI. The CIET enhances quality performance assessment elements and feedback capabilities. The CIET also allows for early identification of student performance concerns and proactive student support. Limitations of North and Sharp's (2020) study include a cut off in the data collected due to the timing of the report. The first cohort had not progressed through all terminal CEs; therefore, the ability to determine whether all students met final CIET performance expectations was limited. Another concern was that students who only used the CIET, and later would become CIs, would not have any experience with the CPI. North and Sharp (2020) suggested that future research examine use of the CIET across institutions for reliability, validity, and comparative research. Future research should also examine the CIET as a benchmark in the context of the competency-based education model.

Birkmeier et al. (2022) answered the call for examining the CIET across multiple institutions. Birkmeier et al. (2022) conducted an exploratory cross-sectional study involving CIs and students from five geographically diverse DPT programs across the United States. The purposes of the study were two-fold: to explore CI and student perceptions of the CIET as an accurate, user-friendly measurement of clinical performance; and to compare CI and student perceptions of the CIET versus the CPI. Participants of the study (56 students, 45 CIs) were recruited individually, not as matched pairs, and were trained on both the CIET and CPI. At midterm and the final week of intermediate and terminal CEs, the CIs and the students completed both tools. After final completion of the tools, the CIs and the students completed a survey developed by Birkmeier et al. (2022). The survey included Likert-scale responses as well as open-ended responses. Descriptive statistics were used to summarize demographics and perceptions of the tools to assess clinical performance. Wilcoxon sign rank tests with a selected medium effect size ($r = 0.30$) were used to determine significant differences in perceptions on utility of the tools. Content analysis was used to examine open-ended survey responses with frequency counts for words or phrases associated with perceptions of the tools.

Regarding perceptions of the tools as assessments of clinical performance, median survey scores of 4.0 for CIs and 4.5 for students indicated the CIET had better ease of use and little redundancy compared to the CPI. CPI completion time was reported as greater than 45 minutes by 66.1% of students and 73.3% of CIs. CIET completion time was reported as less than 45 minutes by 91.1% of both students and CIs. Median values on surveys were equal, however, for CPI and CIET pertaining to detection of change in performance over time, student performance and status, utility in guiding meaningful discussion, and effectiveness of training. Open-ended responses revealed perceptions that most items on both tools were relevant; however, the CPI was too long and redundant, and the CIET training needed further development. Comparison of the CIET as a clinical performance tool to the CPI revealed significant differences in perceptions of ease of use (z-score = -5.42, $p < .001$), appropriate completion time (z-score = -7.25, $p < .001$), and little redundancy of items (z-score = -7.17, $p < .001$).

Birkmeier et al. (2022) found that 71.4% of students and 68% of CIs preferred the CIET over the CPI for ease of use, time for completion, and the CIET's ability to accurately measure clinical performance. The CIET fosters meaningful discussions between CI and student, captures all necessary elements of clinical performance, and identifies learning priorities just as well as the CPI. Limitations of Birkmeier et al.'s (2022) study include low variance among student participants. All students passed their CEs, which limited the ability to investigate the CIET as an assessment of poor performers. Convenience sampling and low participation rates also decreased the external validity of the study. Participants were recruited individually instead of in matched CI-student pairs, which prevented the exploration of the CIET's responsiveness and validation as an effective clinical education assessment tool. Finally, scores between the CPI and CIET were not correlated in Birkmeier et al.'s (2022) study. Only survey questions on the CIET's ability to capture the same constructs as the CPI were analyzed. Therefore, future research should focus on comparing the constructs being assessed in the two tools.

Comprehensive Framework as a Predictor of CE Success

Reilly et al. (2022) developed and published a framework in PT education literature. Reilly et al. (2022) developed a learner blueprint to guide educators in teaching, learning, and assessing CR across the continuum of PT education. Initial validation of the blueprint was attempted through a four-phase process: literature review and content analysis, creation of a draft of the blueprint, expert analysis with initial modifications, and vetting of the blueprint with key stakeholders and final modifications. Reilly et al. (2022) analyzed 260 articles and categorized them according to teaching, learning, and assessment of CR. Then a draft blueprint was created incorporating four learner levels or performance descriptors (beginner, intermediate, competent, and proficient) as well as three CR domains (content knowledge, procedural knowledge and skills, and conceptual reasoning). The learner level and CR domains intersected with each other throughout the continuum of PT education. The competent learner level represented the entry-level graduate of a DPT education program, and the proficient learner level represented a licensed clinician with some amount of expertise. Phases three and four of the validation process included focus groups. Group members made recommendations for changes to the blueprint that were implemented by Reilly et al. (2022). However, the focus groups were small (n = 7) and consisted of faculty, students, and clinical instructors affiliated with only one university, which increased bias and limited geographic impact of the blueprint.

The learner blueprint created by Reilly et al. (2022) has great potential for future use to not only assess CR at multiple levels of education but also remediate CR at the level and time needed by the learner. Although the blueprint incorporates intersection of the level of the learner with CR domain, allowing for assessment and remediation at multiple levels of PT education, it goes beyond entry-level education to that of residency or fellowship. The blueprint allows for assessment of a complex CR domain at a beginner level as well as a simple CR domain at the proficient level, which is a strength of the framework. However, the blueprint does not stop on the continuum of PT education after the student graduates or contain assessment solely within the DPT education program. The blueprint also does not predict, or attempt to predict, CE outcomes.

Summary

Clinical reasoning in PT education is difficult to teach and assess. Lack of CR skills is a major cause of outpatient CE failures for PT students. Many isolated assessments have been created and utilized to measure CR in different ways; however, no comprehensive assessment of CR has been developed to examine CR throughout the duration of a DPT program and to predict CE success. Miller’s (1990) pyramidal framework provides a structure for assessing CR at different time points during a healthcare education program. The bottom tier of the pyramid, Knows, reflects foundational knowledge required for CR processes. Multiple-choice question exams have been used at the Knows level for part-task CR assessments (Daniel et al., 2019). Surry et al. (2017) found that CR cognitive processes occur while taking a MCQ exam and that MCQs can assess aspects of CR. The second tier of Miller’s (1990) pyramid, Knows How, involves application of foundational knowledge in a procedural way given some context. The intermediate phase of CR development has been examined through assessments like skills competency tests (Reilly et al., 2020) or lower-level practical exams (Furze et al., 2015). The third tier of Miller’s (1990) pyramid, Shows How, involves demonstration of CR skills in more complex case scenarios typically requiring a full patient examination. Medical education programs have been using a sequential practical exam called an OSCE to measure competence and CR skills at the Shows How level. The OSCE is equivalent to a *Triple Jump* or TASPE (Fu, 2015) for PT education programs. The OSCE in combination with other assessments moderately predicts CE success. The OSCE was the best predictor when examined with other assessments; however, the OSCE is insufficient alone to predict CE success (Terry et al., 2020; Pelland et al.

2022). For the top tier of Miller's (1990) pyramid, Does, assessment of CR occurs during full-time CEs. For PT education, the CPI has long been the standard for assessing competence during CEs. However, recent studies have shown the CIET as being more efficient for all stakeholders to use. The CIET accurately measures CE performance, includes all elements from the CPI, and incorporates aspects of CR throughout the tool (Birkmeier et al., 2022; North & Sharp, 2020).

Although evidence exists to support the use of isolated assessments at each tier of Miller's (1990) pyramid, each assessment is insufficient to measure CR in a way that paints a comprehensive picture or that can predict CE success. Researchers consistently called for future research to triangulate multiple assessments over time for comprehensive CR assessment. Reilly et al. (2022) created a blueprint or framework to answer the calls for comprehensive assessment. The blueprint provides a structure for future research to examine CR assessment over time; however, the blueprint does not stop at the level of DPT education. The blueprint continues along the continuum of PT education to include residencies and fellowships. Miller's (1990) pyramid fits the format of DPT education, stopping at the level of graduation from the program, and provides a solid structure for assessing CR during the DPT program as well as predicting CE success.

III. METHODOLOGY

The purpose of this study was to determine to what degree CR performance in didactic coursework can predict CR performance in CEs for DPT students. Researchers studying assessment of CR in DPT education have called for a comprehensive framework to triangulate data collected throughout the duration of DPT education programs. A comprehensive framework is necessary to identify deficits in CR prior to beginning full-time CEs so that remediation can occur earlier in the education process. A framework would also inform faculty of students' clinical readiness to begin full-time CEs, and possibly could predict CE success. Using Miller's (1990) pyramid as a conceptual framework, data were collected and analyzed at each level of the pyramid to determine whether each assessment could predict CR during a full-time outpatient CE.

Description of Methodology

A retrospective quantitative exploratory observational design was used for this study. Archived data from completed assessments were collected from a 4-year DPT program in Florida. The DPT program employed a hybrid format where students learned online for three weeks, and during the fourth week each month, students came to campus for a 4-day weekend institute to refine hands-on skills and complete cognitive and psychomotor assessments. Each year of the DPT program contained assessments that correlated to each level of Miller's (1990) pyramid. The first year of the DPT program consisted of foundational science courses. In the second year, courses focused on applying foundational knowledge to clinical skills. The third year of the program involved clinical courses focusing on major systems encountered in the majority of clinical practice settings such as musculoskeletal, cardiopulmonary, and neuromuscular systems. The fourth year of the program consisted of three full-time CEs. The first two CEs lasted 12 weeks, the third CE lasted 11 weeks for a total of 35 weeks of clinical education. One of the three CEs is required by accreditation standards to be in the outpatient clinical setting.

Participants

Archived assessment data for students from the DPT program in the classes 2022 and 2023 ($N = 84$) were utilized in this study. Assessment data were included if the student began a full-time outpatient CE during the final year of the DPT program. Data were excluded if the student never began a full-time outpatient CE. Because the dependent variable for this study was the score from the full-time outpatient CE in the final year of the DPT program, data could only be used for students who began the CE, even if they failed or did not finish the CE because a score would still be recorded in those cases.

Students in the DPT program lived all across America with the majority (70% of the class of 2022 and 88% of the class of 2023) living in Florida. The class of 2022 participants ($n = 44$) included 18 males and 26 females with a mean age of 27. Fifty percent of the class of 2022 were White, with the other 50% comprised of Asian, African American, and unknown or multiple ethnicities. The class of 2023 participants ($n = 40$) included 15 males and 25 females with a mean age of 27.5. Seventy-five percent of the class of 2023 were White, with the other 25% comprised of Asian, African American, Native American, or multiple ethnicities.

The 2022 and 2023 cohorts were a convenience sample utilized for this study because the CIET was first used in the DPT program with these two cohorts. Prior to the year 2022, the CPI was used to assess student performance during full-time CEs. For consistency of data points, no graduate data prior to the year 2022 were collected for this study. No recruitment or informed consent of participants was necessary because data were archived, and no identifiable information was used for participants other than student numbers for alignment of assessment data. This study was approved by the Internal Review Boards of Southeastern University and the participating DPT program.

Archival Data

Data were collected from the online learning management system, Canvas, used by the DPT program. Data from Canvas satisfied the lower three tiers of Miller's (1990) pyramid. Data for the top tier of the pyramid was collected from an online platform called Exxat, which was also used by the DPT program for clinical education information storage. Both Canvas and Exxat are password-protected platforms accessed by faculty within the DPT program. Retroactive access to closed Canvas courses was granted by the information technology department at the university of the DPT program for collection of data at the lower three tiers of the pyramid. The Director of Clinical Education granted access to Exxat for data collection at the top tier of the pyramid. All data were stored in the researcher's password-protected electronic cloud storage files. Data were only viewed by the researcher and methodologist for this study. Data will be destroyed three years after the publication of this study through deletion from the researcher's electronic cloud storage files.

Knowledge-Level Assessment

Foundational knowledge is imperative for CR (Furze et al., 2022; Gilliland & Wainwright, 2017), and MCQs can assess the beginning of reasoning processes (Daniel et al., 2019; Surry et al., 2017). Therefore, mean scores from MCQ exams were collected from a first-year biomechanics course for the Knows tier of Miller's (1990) pyramid. Biomechanics for physical therapy requires integration of prerequisite knowledge of anatomy, physiology, and physics for application to movement analysis and understanding of movement pathology. The biomechanics course is a foundational science course in the curriculum of the DPT program and is a building block in the sequence of courses in the musculoskeletal domain of practice. Four exams occur during the biomechanics course that test students' knowledge and integration of foundational concepts. Each exam had 45 multiple-choice questions addressing both factual recall of information as well as application to function, movement, or short scenarios. Exam one focused on biomechanical principles, posture, and balance. Exam two focused on the upper extremity. Exam three focused on the spine, temporomandibular joint, ribs and respiration, and the pelvis and hip. Exam four focused on the lower extremity and gait. Because each exam focused on different topics, mean percentage scores for the four exams as a whole were used to provide a more comprehensive picture of biomechanical knowledge as a foundation for CR.

Application-Level Assessment

Timmerberg et al. (2019) found that an emerging level of competence is expected prior to the first CE for students' understanding and skill to perform examination and intervention procedures. Application-level assessment of CR involves procedural knowledge and applying foundational knowledge to context. For the Knows How tier of the pyramid, mean scores for practical exams from a second-year clinical skills course were collected. The clinical skills course focused on physical therapy interventions in the musculoskeletal domain of practice, such as gait training with assistive devices, therapeutic exercise, and manual therapy techniques. Three practical exams were conducted during the clinical skills course. During each practical exam, students were paired with a classmate who played the role of the patient for them. Students were given a short clinical scenario from which they selected and applied an appropriate intervention. Selection of the intervention requires clinical decision-making skills related to CR, and the correct performance of the intervention involves procedural knowledge, which is another component of CR. Because each practical exam focused on a different type of intervention, the mean percentage of the three exams as a whole was used for data analysis to provide a more comprehensive picture of CR during intervention selection and performance overall during the second year of the DPT program.

Competence-Level Assessment

The third tier of Miller's (1990) pyramid, Shows How, requires assessment mimicking realistic clinical situations. For this study, percentage scores from a practical exam called a *Triple Jump* were utilized. The *Triple Jump* is similar to an OSCE or TASPE in that it has three stations or phases to the practical exam. Part one of the *Triple Jump* involves students reading a patient history and creating a plan for a physical examination. During part two of the *Triple Jump*, students conduct the physical examination on a simulated patient, who is portrayed by an instructor. For part three of the *Triple Jump*, students document their assessment or evaluation of physical examination, which includes a PT diagnosis. Part three documentation also includes a prognosis, patient goals, and a plan of care to guide future treatment for the patient. The *Triple Jump* closely mimics a PT patient evaluation done in an outpatient clinical setting. The DPT program used in this study had a series of three musculoskeletal (MSK) courses. Each MSK course had a *Triple Jump* as the culminating comprehensive exam for the lab portion of the course. Percentage scores of the *Triple Jump* for this study were taken from the third MSK course because the third course is the last in the MSK series and students completed *Triple Jumps* in the preceding two MSK courses. By the time students complete the *Triple Jump* in the third MSK course, they are well-versed in the format and procedure of this type of practical exam.

Part one of the *Triple Jump* is not included in the total score because it is not graded and students only read a case history and prepare for the remainder of the examination. Parts two and three of the *Triple Jump* have a maximum score of 72 points. Part two (52 points) includes items assessing oral reasoning, examination planning and organization, tests and measures selection and performance, diagnostic hypothesis, selection and performance of an intervention, therapeutic presence, and safety for therapist and patient. Part three (20 points) includes items assessing documentation of the PT diagnosis, prognosis, goals, and three planned interventions to be carried out during the plan of care. Parts two and three of the *Triple Jump* include items that all correlate to the conceptualization of CR by Huhn et al. (2019). Also, Timmerberg et al. (2019) found that an emerging level of competence is expected for CR skills prior to beginning CEs, such as applying evidence-based strategies to guide decision-making, articulating rationale during a patient evaluation, developing and linking patient goals to activity limitations or participation restrictions, determining whether a patient is appropriate for the physical therapy scope of practice, interpreting examination findings, and screening to rule in or rule out conditions. The *Triple Jump* contains all CR elements described by Huhn et al. (2019) and Timmerberg et al. (2019). Therefore, *Triple Jump* total percentage scores were used for data analysis.

Performance-Level Assessment

The top tier of Miller's (1990) pyramid, Does, represents students' performance in authentic clinical settings. During the fourth and final year of the DPT program, students complete three full-time CEs, one of which must be in the outpatient setting. CIs teach, coach, and mentor students in the clinical settings; however, the goal is for the student to become independent with patient management by the end of each CE. To measure student performance, which includes CR, CIs complete a midterm assessment halfway through the CE, and a final assessment at the end of the CE. The DPT program utilized in this study used an instrument called the Clinical Internship Evaluation Tool (CIET) to collect performance data during full-time CEs.

The CIET (Appendix) has 43 items divided into two categories: Professional Behaviors (19 items) and Patient Management (24 items). CIs score items using a scale of *Never* (1) to *Always* (5) for the Professional Behaviors section and *Well below* (1) to *Well above* (5) for the Patient Management section. For the DPT program utilized in this study, the Professional Behaviors category also allowed for a score of zero in the Communication subsection if the task was not observed. A score of zero was not part of the original CIET instrument. At the end of the CIET is a Global Rating score for CIs to provide an overall impression of the student's competence for clinical practice. The developers of the CIET, Fitzgerald et al. (2007), found that all items on the CIET were representative of skills and behaviors necessary for competent PT practice. Internal consistency for the CIET was .98 for Patient Management items and a moderate correlation (.76) was found for Patient Management section scores and the Global Rating scores at the end of the tool. CIs (96%) responded to surveys strongly agreeing that the CIET was short and easy to use. Final CIET scores from outpatient CEs were collected for this study. If a student began an outpatient CE but did not finish it due to failure or other reasons, the midterm CIET score was utilized because the midterm score still provides an assessment of performance in the outpatient setting that can be included as a dependent variable.

Both the Professional Behaviors and Patient Management categories contained items related to CR as conceptualized by Huhn et al. (2019). The Professional Behaviors category addressed cognitive and affective aspects of CR through subcategories of safety, professional ethics, initiative (regarding learning and problem solving), and communication. To receive 100% in the Professional Behaviors category, a raw score of 95 was necessary. The Patient Management category addressed cognitive and psychomotor aspects of CR through subcategories of examination, evaluation, diagnosis/prognosis, and intervention. To receive 100% in the Patient Management category, a raw score of 120 was necessary. A rating of *Well above* (5) for the Patient Management items was not required by the DPT program in this study. The DPT program required entry-level ratings of *At that level for all patients* (4) for the Patient Management items. However, in the interest of examining the highest possible levels of competence on the CIET as the dependent variable, the total score raw score of 215 for both Professional Behaviors and Patient Management categories was used to determine percentages for data analysis. The Global Rating scale at the end of the CIET form was excluded from data analysis because it is separate from the two major categories throughout the instrument and has a 10-point rating scale that does not match the 5-point rating scales other two major categories.

Procedures

The three lower tiers of Miller's (1990) pyramid were utilized as independent variables to predict the top tier of the pyramid. The biomechanics exam scores were used for the first tier of the pyramid, the clinical skills practical exam scores were used for the second tier, and the MSK three *Triple Jump* scores were used for the third tier. The CIET scores in the outpatient CE setting were used as the dependent variable - the top tier of the pyramid.

Mean biomechanics exam scores, mean clinical skills practical exam scores, and mean *Triple Jump* scores were collected from Canvas, the DPT program's learning management system. CIET scores were collected from Exxat, the clinical education platform utilized by the DPT program. Scores were aligned with corresponding student numbers for data analysis.

Data Analysis

Data were analyzed using statistical tests of prediction in IBM SPSS Version 29.0. Simple linear regression analysis was utilized for research questions 1-3. Multiple linear regression analysis was utilized for research question four.

Preliminary Analysis

Preliminary analysis included descriptive statistics for the participants in the study as well as regression analyses for research questions 1-4. The class of 2022 had 44 students who participated in outpatient CEs. The class of 2023 had 40 students who participated in outpatient CEs. A priori power analysis for multiple linear regression at a medium effect size ($f^2 = 0.15$, $\alpha = 0.05$, $1-\beta = .80$) with three predictors required a sample size of 77 participants. This study had a sample size of 84, which is enough for statistically significant power.

Research Question 1

1. To what degree will the level of knowing (factual recall of MCQ assessments) predict study participant level of CR performance in practice?

   H_0: The level of knowing will not significantly predict the level of CR performance in practice.

Simple linear regression analysis was used to determine whether the bottom tier of Miller's (1990) pyramid predicted the top tier. Mean biomechanics exam scores were examined to determine whether they were predictive of CIET scores.

Research Question 2

2. To what degree will the level of interpretation and application of knowledge be predictive of study participant level of CR performance in practice?

    H_0: The level of interpretation and application of knowledge will not significantly predict the level of CR performance in practice.

Simple linear regression analysis was used to determine whether the second tier of the pyramid predicted the top tier. Mean scores from clinical skills practical exams were tested for prediction of CIET scores.

Research Question 3

3. To what degree will the level of demonstration predict study participant level of CR performance in practice?

    H_0: The level of demonstration will not significantly predict the level of CR performance in practice.

Simple linear regression analysis was used to determine whether the third tier of the pyramid could predict the top tier. *Triple Jump* scores were examined to determine whether they predicted CIET scores.

Research Question 4

4. Considering study participant level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, which is most predictive of study participant level of CR performance in practice?

H_0: Of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, none will significantly predict level of CR performance in practice.

H_1: The combination of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, will significantly predict CR performance in practice.

Multiple linear regression analysis was used to determine which of the three independent variables best predicted the dependent variable, and to determine the degree to which the three independent variables together predicted the dependent variable. Biomechanics exam scores, clinical skills practical exam scores, and *Triple Jump* scores were examined to determine which of the three best predicted CIET scores, and to determine how well the grouping of the three sets of scores predicted CIET scores.

Summary

Using a retrospective quantitative exploratory observational design, this study collected archival data from a DPT program and correlated the data to Miller's (1990) pyramid. A convenience sample of participants was utilized including students in the classes of 2022 and 2023. Each level of Miller's (1990) pyramid corresponded to an assessment from each year of the DPT program. For the bottom three levels of the pyramid, biomechanics exam scores, clinical skills practical exam scores, and *Triple Jump* scores were utilized as independent variables. The top tier of the pyramid served as the dependent variable, and CIET scores from outpatient full-time CEs were collected for this level. Regression analyses were used for each of the independent variables to determine whether each predicted the dependent variable, which of the three was the best predictor of the dependent variable, and whether the independent variables as a group predicted the dependent variable. The design of this study answered the research questions and the overall purpose of predicting CR at the performance level by triangulating lower level assessments.

IV. RESULTS

The purpose of this study was to determine to what degree CR performance in didactic coursework can predict CR performance in CEs for DPT students. Chapter IV contains a presentation of the formal reporting of findings achieved in the study. Four research questions and hypotheses were stated to address the study's topic and research purpose. Descriptive and inferential statistical techniques were used to analyze study data at the foundational, preliminary level and for the four research questions. The following represents the formal reporting of study findings by preliminary, foundational descriptive analyses and for analyses associated with the study's research questions and hypotheses.

Descriptive Statistical Findings

Demographic Identifying Information

The study's demographic information was evaluated using descriptive statistical techniques. The study's demographic information was specifically addressed using the descriptive statistical techniques of frequencies (*n*) and percentages (*%*). Table 1 contains a summary of findings for the descriptive statistical analysis of the study's demographic identifying information of participant gender, age category, group (cohort year), and outpatient status (whether the student had the outpatient CE first, second, or third in sequence).

Table 1

Descriptive Statistics Summary Table: Demographic Information (Gender; Age Category; Group; and Outpatient Status)

Variable	*n*	%	Cumulative %
Gender			
Female	52	61.90	61.90
Male	32	38.10	100.00
Missing	0	0.00	100.00
Age Category			
28 and Younger	46	54.76	54.76
29 and Older	38	45.24	100.00
Missing	0	0.00	100.00
Group (Year)			
2022	44	52.38	52.38
2023	40	47.62	100.00
Missing	0	0.00	100.00
Outpatient Status			
Rotation #1	42	50.00	50.00
Rotation #2	36	42.86	92.86
Rotation #3	6	7.14	100.00
Missing	0	0.00	100.00

Descriptive Statistics: Constructs

Descriptive statistical techniques were utilized to assess the study's performance data for the constructs of knowing (Knows), application of knowledge (Knows How), demonstration of knowledge (Shows How), and CR performance (Does). The study's performance data within the constructs were specifically addressed using the descriptive statistical techniques of frequencies (*n*), measures of central tendency (mean scores), variability (minimum/maximum; standard deviations), standard errors of the mean (SE_M), and data normality (skew; kurtosis). Table 2 contains a summary of findings for the descriptive statistical analysis of the study's performance data for the constructs of knowing using mean scores from biomechanics exams, application of knowledge using mean scores from clinical skills practical exams, demonstration of knowledge using *Triple Jump* scores, and CR performance using CIET scores.

Table 2

Descriptive Statistics Summary Table: Knowing, Application of Knowledge, Demonstration of Knowledge, and CR Performance

Construct	*M*	*SD*	*n*	SE_M	Min	Max	Skew	Kurtosis
Knowing	83.14	7.88	84	0.86	65.00	98.33	-0.03	-0.69
Knowledge Application	93.77	3.94	84	0.43	84.25	100.00	-0.53	-0.49
Knowledge Demonstration	88.34	8.20	84	0.89	34.72	98.61	-3.80	21.11
CR Performance	188.70	10.04	84	1.10	124.00	210.00	-3.20	19.16

Table 3 contains a summary of findings for the descriptive statistical analysis of the study's performance data for the constructs of knowing, application of knowledge, demonstration of knowledge, and CR performance by group (cohort graduation year).

Table 3

Descriptive Statistics Summary Table: Knowing, Application of Knowledge, Demonstration of Knowledge, and CR Performance by Group

Year/Construct	*M*	*SD*	*n*	SE_M	Min	Max	Skew	Kurtosis
Class of 2022								
Knowing	82.19	8.67	44	1.31	65.00	98.33	-0.03	-0.84
Knowledge Application	92.16	4.06	44	0.61	84.25	100.00	-0.12	-0.82
Knowledge Demonstration	86.98	10.44	44	1.57	34.72	98.61	-3.22	13.00
CR Performance	188.59	6.99	44	1.05	166.00	208.00	-0.68	2.19
Class of 2023								
Knowing	84.19	6.86	40	1.08	72.78	97.22	0.24	-0.99
Knowledge Application	95.54	2.94	40	0.46	86.87	100.00	0.76	0.30
Knowledge Demonstration	89.84	4.29	40	0.68	80.56	97.22	0.26	-0.56
CR Performance	188.82	12.67	40	2.00	124.00	210.00	-3.31	15.42

Table 4 contains a summary of findings for the descriptive statistical analysis of the study's performance data for the constructs of knowing, application of knowledge, demonstration of knowledge, and CR performance by age category of study participant.

Table 4

Descriptive Statistics Summary Table: Knowing, Application of Knowledge, Demonstration of Knowledge, and CR Performance by Age Category of Study Participant

Age Category/Construct	*M*	*SD*	*n*	SE_M	Min	Max	Skew	Kurtosis
28 and Younger								
Knowing	82.28	8.33	46	1.23	65.00	98.33	0.17	-0.87
Knowledge Application	93.52	3.78	46	0.56	84.99	100.00	-0.26	-0.70
Knowledge Demonstration	89.19	5.69	46	0.84	61.80	97.22	-2.41	9.67
CR Performance	189.33	6.48	46	0.96	176.00	210.00	0.44	1.52
29 and Older								
Knowing	84.18	7.26	38	1.18	65.56	97.78	-0.28	-0.18
Knowledge Application	94.08	4.15	38	0.67	84.25	99.07	-0.83	-0.18
Knowledge Demonstration	87.32	10.45	38	1.70	34.72	98.61	-3.46	15.09
CR Performance	187.95	13.18	38	2.14	124.00	208.00	-3.13	12.86

Findings by Research Question & Hypothesis

Four research questions and hypotheses were stated to address the study's topic and research problem. The following represents the reporting of findings achieved in the study by research question and hypothesis stated.

Research Question 1

To what degree will the level of knowing (factual recall of multiple-choice question assessments) predict study participant level of clinical reasoning performance in practice? Simple linear regression analysis was used to evaluate the degree to which knowing, represented as factual recall in biomechanics exams, was statistically significant in predicting study participant CR performance (CIET) in research question one. The assumptions of linear regression analysis were addressed and satisfied through statistical means (independence of error; normality of residuals; influential outliers) and visual inspection of scatter plots (linearity; homoscedasticity).

The predictive model was non-statistically significant (F (1,82) = 0.63, p = .43, R^2 = .01), indicating that knowing, reflected as factual recall via biomechanics MCQ exams, did not explain a significant proportion of variation in study participant CR performance (CIET). As a result, the model's independent predictor variable of knowing (factual recall in biomechanics) was not further evaluated. Table 5 contains a summary of the predictive model used to evaluate the predictive ability of knowing (factual recall in biomechanics) for study participant CR performance (CIET) in research question one.

Table 5

Linear Regression Summary Table: Knowing (Factual Recall in Biomechanics) Predicting Study Participant CR Performance (CIET)

Model	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	179.44	11.71	[156.15, 202.73]	0.00	15.33	< .001
Knowing (Factual Recall)	0.11	0.14	[-0.17, 0.39]	0.09	0.79	.43

***H*₀ *1*:**

The level of knowing will not significantly predict the level of clinical reasoning performance in practice. Considering the non-statistically significant predictive relationship between levels of knowing expressed as biomechanics factual recall assessments and study participant CR performance, the null hypothesis in research question one was retained.

Research Question 2

To what degree will the level of interpretation and application of knowledge be predictive of study participant level of clinical reasoning performance in practice? Simple linear regression analysis was used to evaluate the degree to which knowledge application, reflected as clinical skills practical exams, was statistically significant in predicting study participant CR performance (CIET) in research question two. The assumptions of linear regression analysis were addressed and satisfied through statistical means (independence of error; normality of residuals; influential outliers) and visual inspection of scatter plots (linearity; homoscedasticity).

The predictive model was non-statistically significant (F (1,82) = 0.04, p = .85, R^2 = .00), indicating that application of knowledge did not explain a significant proportion of variation in study participant CR performance (CIET). As a result, the model's independent predictor variable of application of knowledge was not further evaluated. Table 6 contains a summary of the predictive model used to evaluate the predictive ability of the application of knowledge for study participant clinical reasoning (CIET) in research question two.

Table 6

Linear Regression Summary Table: Application of Knowledge (Clinical Skills Practical Exams) Predicting Study Participant CR Performance (CIET)

Model	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	193.69	26.43	[141.12, 246.26]	0.00	7.33	< .001
Application of Knowledge	-0.05	0.28	[-0.61, 0.51]	-0.02	-0.19	.85

***H_0 2*:**

The level of interpretation and application of knowledge will not significantly predict the level of clinical reasoning performance in practice. Considering the non-statistically significant predictive relationship between the application of knowledge expressed as clinical skills practical exams and study participant CR performance, the null hypothesis in research question two was retained.

Research Question 3

To what degree will the level of knowledge demonstration predict study participant level of clinical reasoning performance in practice? Simple linear regression analysis was used to evaluate the degree to which knowledge demonstration represented as the *Triple Jump* was statistically significant in predicting study participant CR performance (CIET) in research question three. The assumptions of linear regression analysis were addressed and satisfied through statistical means (independence of error; normality of residuals; influential outliers) and visual inspection of scatter plots (linearity; homoscedasticity).

The predictive model was non-statistically significant (F (1,82) = 0.05, p = .82, R^2 = .00), indicating that knowledge demonstration (*Triple Jump*) did not explain a significant proportion of variation in study participant CR performance (CIET). As a result, the model's independent predictor variable of knowledge demonstration was not further evaluated. Table 7 contains a summary of the predictive model used to evaluate the predictive ability of knowledge demonstration (*Triple Jump*) for study participant CR performance (CIET) in research question three.

Table 7

Linear Regression Summary Table: Knowledge Demonstration (Triple Jump) Predicting Study Participant CR Performance (CIET)

Model	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	186.02	12.00	[162.15, 209.88]	0.00	15.51	< .001
Knowledge Demonstration	0.03	0.14	[-0.24, 0.30]	0.02	0.22	.82

***H_0 3*:**

The level of demonstration will not significantly predict the level of clinical reasoning performance in practice. Considering the non-statistically significant predictive relationship between demonstration expressed as the *Triple Jump* and study participant CR performance, the null hypothesis in research question three was retained.

Research Question 4

Considering study participant level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, which is most predictive of study participant level of clinical reasoning performance in practice? Multiple linear regression (MLR) analysis was used to evaluate the degree to which the confluence of knowing, application of knowledge, and demonstration was statistically significant in predicting study participant CR performance (CIET) in research question four. The assumptions of MLR analysis were addressed and satisfied through statistical means (independence of error; normality of residuals; influential outliers; multicollinearity) and visual inspection of scatter plots (linearity; homoscedasticity).

The MLR predictive model was non-statistically significant, (F (3,80) = 0.33, p = .81, R^2 = .01), indicating that the confluence of the model's independent variables (knowing; application of knowledge; and knowledge demonstration) did not explain a statistically significant proportion of variation in study participant CR performance (CIET). Considering the non-statistical significance of the predictive model in research question four, the individual predictors in the model were not evaluated further or interpreted. Table 8 contains a summary of finding for the study participant clinical reasoning level predicted by the confluence of the three independent variables (knowing; application of knowledge; and demonstration of knowledge) in research question four.

Table 8

Multiple Linear Regression Model: Predicting CR Performance (CIET) by Knowing, Application of Knowledge, and Knowledge Demonstration

Model	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	193.70	26.62	[140.72, 246.69]	0.00	7.28	< .001
Knowing	0.15	0.17	[-0.18, 0.48]	0.12	0.91	.36
Knowledge Application	-0.19	0.33	[-0.85, 0.46]	-0.08	-0.59	.56
Knowledge Demonstration	0.005	0.16	[-0.31, 0.32]	0.004	0.03	.97

***H_0 4*:**

Of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, none will significantly predict level of CR performance in practice. Considering the non-statistically significant predictive values for each of the three independent variables, the null hypothesis is retained for research question four.

$H_1$4:

The combination of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, will significantly predict clinical reasoning performance in practice. Considering the non-statistically predictive finding for the confluence of the variables of knowing, application of knowledge, and demonstration in predicting study participant clinical reasoning, the alternative hypothesis in research question four was rejected.

Follow-up Ancillary Analysis

A follow-up, ancillary analysis was conducted to evaluate the ability of study participant cohort rotation (whether the student's outpatient CE was first, second, or third in sequence) in predicting subsequent clinical reasoning level. Simple linear regression represented the statistical technique used in the predictive modeling process. As a result, the predictive model used in the ancillary analysis was statistically significant (F (2,81) = 2.15, p = .04, R^2 = .05). Study participant cohort rotation was statistically significant in predicting subsequent CR performance level (*B* = 3.57; *p* = .04; 95% CI [0.14; 7.00]) indicating that for every full unit of increase in study participant rotational level, there is a 3.57 unit on average predicted increase in the value of CR performance (CIET). Table 9 contains a summary of findings for the predictive model used to evaluate respective cohort of study participant and subsequent CR performance level (CIET).

Table 9

Ancillary Analysis Summary Table: Predicting Study Participant CR Performance (CIET) by Cohort Rotation

Model	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	183.09	2.92	[177.29, 188.89]	0.00	62.79	< .001
Cohort Rotation	3.57	1.73	[0.14, 7.00]	0.22	2.07	.04*

*$p < .05$

Summary

The study's findings were presented and reported in Chapter IV. The study's independent variables of knowing (biomechanics MCQ exams), application of knowledge (clinical skills practical exams), and demonstration of knowledge (*Triple Jump*) were non-statistically significant in individually predicting study participant CR performance (CIET). The confluence of the three independent variables were, moreover, non-statistically significant in predicting study participant CR performance using a MLR predictive model. Although not a formal research question posed in the study, participant cohort rotation was statistically significant in predicting subsequent CR performance level in a follow-up, ancillary analysis. Chapter V contains a discussion of the study's findings as presented in Chapter IV.

V. DISCUSSION

The purpose of this study was to determine to what degree CR performance in didactic coursework can predict CR performance in CEs for DPT students. A retrospective quantitative exploratory observational design was used for this study. Miller's (1990) pyramid served as a conceptual framework for the study design. The lower three tiers of the pyramid equated to didactic assessments and served as independent variables, while the top tier equated to clinical assessment and served as the dependent variable.

Review of Methodology

Archived assessment data from a 4-year hybrid DPT program were used to fulfill each tier of Miller's (1990) pyramid. First-year biomechanics MCQ exam scores were used for the bottom tier to reflect knowledge-level assessment. Second-year clinical skills practical exams were used for the second tier to reflect application-level assessment. Third-year musculoskeletal *Triple Jump* scores were used for the third tier to reflect competence-level assessment. Fourth-year CIET scores for outpatient CEs were used for the top tier to reflect performance-level assessment. Archived assessment data for graduates of the classes of 2022 and 2023 (N = 84) were collected and analyzed using tests of statistical regression.

The class of 2022 had 44 graduates, and the class of 2023 had 40 graduates. The majority of participants were female, and the majority of participants were age 28 or younger (see Table 1). Most participants completed their outpatient CE first (n = 42) in the sequence of three CEs.

Many participants completed their outpatient CE second ($n = 36$), and very few participants completed their outpatient CE last ($n = 6$). No demographic data for participants, except sequence of outpatient CE, had significant correlation to the dependent variable of CIET scores. No significant difference was found between the classes of 2022 and 2023 for any of the assessment scores analyzed in this study (see Table 3).

Discussion by Research Question

Overall, no significant findings resulted from regression analyses for any research question. However, a follow-up ancillary analysis found that cohort rotation for the sequence of outpatient CE was statistically significant in predicting CR in clinical practice.

Research Question 1

To what degree will the level of knowing (factual recall of multiple-choice question assessments) predict study participant level of clinical reasoning performance in practice?

H_0 1: The level of knowing will not significantly predict the level of clinical reasoning performance in practice.

No statistical significance was found using simple linear regression analysis for biomechanics MCQ exams and CIET scores. Therefore, the null hypothesis for research question one was retained. MCQ exams may assess parts of CR, such as developing a leading diagnosis and management or treatment of patients (Daniel et al., 2019). MCQ exams also reveal some CR processes such as non-analytical and analytical reasoning, emotional reactions, pattern recognition, heuristics, intuition, and identifying key features (Surry et al., 2017). Due to the nature of MCQ exams reflecting part-task CR processes, rather than being a closer representation of the complex whole-task of CR, MCQ exams alone may not be a good predictor of CR in clinical practice.

Research Question 2

To what degree will the level of interpretation and application of knowledge be predictive of study participant level of clinical reasoning performance in practice?

H_0 2: The level of interpretation and application of knowledge will not significantly predict the level of clinical reasoning performance in practice.

No statistical significance was found using simple linear regression analysis for clinical skills practical exams and CIET scores. Therefore, the null hypothesis for research question two was retained. Application-level assessments incorporate some clinical context and procedural psychomotor skills into CR processes. However, previous literature involving application-level assessments usually combined them with other levels of assessment when examining CR (Furze et al., 2015; McDevitt et al., 2019; Reilly et al., 2020). Research examining practical exams in isolation and correlation to CR in clinical practice could not be found during the literature search for this study. Like MCQ exams, the practical exams used for the second tier of assessments in this study address parts of task for CR. It is not surprising, then, that application-level assessments alone were not good predictors of CR performance in clinical practice.

Research Question 3

To what degree will the level of knowledge demonstration predict study participant level of clinical reasoning performance in practice?

H_0 3: The level of demonstration will not significantly predict the level of clinical reasoning performance in practice.

No statistical significance was found using simple linear regression analysis for musculoskeletal *Triple Jump* scores and CIET scores. Therefore, the null hypothesis for research question three was retained. The *Triple Jump* was more comprehensive in nature including clinical context and all aspects of the CR concept as described by Huhn et al. (2019). The *Triple Jump* was similar to an OSCE used in medical education programs. Literature examining OSCEs and CR in clinical practice showed that OSCEs in isolation were poor predictors of clinical readiness, and low to moderate predictors of CE performance (Pelland et al., 2022). The findings for OSCEs by Pelland et al. (2022) corroborate the non-significant finding for the *Triple Jump,* in isolation, to predict CR in clinical practice.

Research Question 4

Considering study participant level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, which is most predictive of study participant level of clinical reasoning performance in practice?

H_0 4: Of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, none will significantly predict level of CR performance in practice.

$H_1$4: The combination of the three variables, level of fact recall, interpretation and application of knowledge, and demonstration of knowledge, will significantly predict clinical reasoning performance in practice.

No statistical significance was found using multiple linear regression analysis for all three independent variables and the dependent variable of CIET scores. Therefore, the null hypothesis for research question four was retained, and the alternative hypothesis was rejected. The non-statistically significant finding for research question four contradicts previous literature that explored combinations of assessments to predict student clinical performance. Terry et al. (2020) combined the OSCE with written exams and seminar presentations and found that the model moderately predicted CE performance for medical students. Reilly et al. (2020) created a model for teaching CR skills that included multiple assessments and resulted in successful outcomes, such as students passing each assessment in the model, meeting expectations on the CPI during CEs, and 100% licensure exam pass rates. Triangulation of multiple assessments has successfully predicted clinical performance in previous studies; therefore, the alternative hypothesis for research question four was directional toward a positive prediction of students' CR performance in clinic based on combining multiple assessments of CR at lower levels of education. However, this study did not find successful prediction of CR clinical performance from grouping the independent variables, and confounding factors may have played a role in this result.

The classes of 2022 and 2023 were assessed at lower levels of the pyramid during the time of the COVID-19 pandemic. Although the DPT program utilized in this study was already a hybrid program, students in each cohort lost the few face-to-face educational encounters during COVID-19 shutdowns that they otherwise would have had. The change in educational format for hands-on skills that would have occurred in person, but instead occurred online until the university was allowed to reopen its doors, may have caused lower scores on early assessments. Logically, lower scores on assessments at the lower tiers of the pyramid should have correlated to lower scores on the CIET; however, this study found no such correlation. It may be that any skills or knowledge lost, or never learned in the first place, were regained during the remaining time of didactic coursework, or were regained or learned well during full-time CEs with one-to-one face to face instruction from CIs.

Other confounding variables could be characteristics of the students in the classes of 2022 and 2023. Age and graduation year were explored for potential factor impact on results; however, no significant correlations or predictions were found for these variables (see Tables 3 and 4). Some students entering DPT education programs have prior licenses or certifications such as physical therapy assistant, athletic trainer, personal trainer, massage therapist, etc. One variable not explored was whether a student had previous clinical experience from working under a prior certification. It is possible that students with prior certifications may have performed well in clinical environments in which they were comfortable from previous work experience, however, did not perform as well on didactic coursework assessments at lower tiers of the pyramid. The variable of prior certification may have confounded results when predicting clinical CR performance from didactic assessments.

CIET Nuances

Other confounding variables that potentially had a major impact on results were nuances of using the CIET instrument. The CIET incorporates CR across multiple categories throughout the instrument. However, because it is a newer instrument for DPT clinical education, some nuances of using the instrument may have played a role in the outcomes for this study. North and Sharp (2020) reported CI uncertainty in understanding performance expectations using the CIET. Birkmeier et al. (2022) also found that CIs needed more training with the CIET. These findings highlight the importance of CI training on the CIET instrument. The DPT program utilized in this study did provide training for CIs, and this training was standardized according to clinical education best practices. CIs were required to do a one-time training consisting of a 10-minute video and a 20-question quiz. CIs had to achieve at least 80% on the quiz to participate in assessing students with the CIET. Therefore, CI misunderstanding of instrument use and performance expectations was not considered a factor in the outcomes of this study.

For the category of Professional Behaviors on the CIET instrument, a subcategory of Communication exists. For the subcategory of Communication, CIs had the option to mark *Not Observed* (0) if the task or item was not observed by the CI during the student's CE. The researcher was unaware of the option to mark zero in the subcategory of Communication until it was discovered during the time of data collection. The Director of Clinical Education for the DPT program utilized in this study confirmed that the zero option for the Communication subcategory was not something the DPT program allowed in isolation, but rather was a standard part of the scoring for the CIET instrument overall. Therefore, the researcher recorded all instances of CI selection of zero within the Communication subcategory. In total, 24 instances of a score of zero were tallied for line items in the Communication subcategory. Three line items, specifically, were selected by CIs using the zero score:

- Communicates verbally with precise and appropriate terminology and in a timely manner with health care professionals (e.g., MD's, nurses, insurance carriers, OT, SLT, etc.); (n = 18);
- Communicates in writing with precise and appropriate terminology and in a timely manner when completing documentation to professionals (e.g., plans of care, physician letters, etc.); (n = 4);
- Communicates in writing with precise and appropriate terminology and in a timely manner with patients and families/caregivers when creating home programs, patient instructions, etc.; (n = 2).

The option to mark zero added a 6th point to the Likert scale for the Communication subcategory only. With 24 instances of CIs selecting zero for line items in this category, it is possible results may have been skewed because zero scores lowered CIET total scores that were used for data analysis. However, the score of zero for these line items did not necessarily mean that a student lacked CR for the subcategory of Communication; rather, it was simply not observed. If the 24 instances of zero scores had been excluded from data analysis, statistical power would have decreased. Therefore, the 24 instances of zero scores were included in the data set.

The intention of the line items for which CIs scored zeros was to examine interprofessional communication or communication with patients' family members or caregivers. In many outpatient clinical settings, interprofessional communication seldom occurs due to the physical separation of the stand-alone clinic to providers who refer to PT or providers who may work alongside PTs in other settings such as hospitals or nursing homes. Also, patients who come to outpatient clinics generally do not require care from family members or caregivers like they do in other PT clinical settings. Therefore, the line items listed above may not have applied to the outpatient CE setting and may be a flaw of the CIET instrument that needs to be examined or explored by future research.

Follow-up Ancillary Analysis

Given the statistically non-significant findings for all four research questions, other factors were explored during a follow-up ancillary analysis, such as participant age, graduation cohort, and outpatient CE rotation cohort during the sequence of three CEs. Of the additional factors explored, only the outpatient CE rotation cohort had a significant correlation to CIET scores. The later in sequence that a student completed the outpatient CE, the better the CIET total score. No studies could be found during a search of the literature that specifically address sequence of CE and CR performance during CEs. However, it is possible that CR performance improved over time from the mere fact that students were receiving direct one to one in-person instruction from the CI in clinical context regardless of the type of setting for the CE. In other words, it is possible that CR performance in practice was going to improve anyway as a student progressed through the sequence of the three full-time CEs regardless of which setting the student completed first, second, or third. Furze et al. (2015) and McDevitt et al. (2019) found that students' CR developed over time during didactic coursework, and that curricular content at certain time points may have influenced students' concurrent CR processes. Gilliland (2017) also found that students develop CR overtime during their didactic coursework and clinical education. Studies examining PT residents also show improvement in CR during the course of mentored residency (Baker et al., 2017; Cunningham et al., 2019). CR appears to improve over time and in clinical context. The in-person factor of full-time CEs may have affected CR performance for the participants of this study because until their full-time CEs, the majority of their learning was done online. The DPT program utilized in this study had a hybrid format, with the majority of teaching and learning occurring asynchronously online. The change to synchronous in-person learning with one-to-one attention from a CI may have affected CR growth and performance for participants in this study. Future research is needed to explore whether the change to in-person situational learning played a role in CR development.

Significance and Implications for Future Practice

Clinical reasoning is a major topic of educational research in PT because it is difficult to teach, learn, and assess. Lack of CR has been linked with failure of full-time outpatient CEs. If a methodology, framework, or paradigm for assessing CR in PT education could be found that predicted CE success for PT students, then faculty would have a way to identify and remediate struggling students before they reach their first full-time CE. The aim of this study was to determine whether and to what degree assessments during didactic coursework in a DPT program could predict CR performance during full-time CEs in the final year of the DPT program. Miller's (1990) pyramidal framework was utilized to organize data points and align assessments within one domain of PT practice.

Given the lack of significant findings in this study, one conclusion that may be drawn is that Miller's (1990) pyramidal framework may not be the best model for assessing and predicting CR for PT students. Miller's (1990) framework was selected for this study based not only on the seminal work by Miller (1990) using medical students, but also from recent work by Fu (2015) who used Miller's framework to explore the development of the TASPE, which is similar to the *Triple Jump* used in this study. Fu (2015) suggested that future research examine whether Miller's (1990) pyramid could serve as a predictor of clinical reasoning given assessments at the lower three tiers. For this study, Miller's (1990) pyramid was not able to predict CR performance in the clinical setting, at least for one domain of PT practice. The musculoskeletal domain of PT practice was chosen for this study due to Silberman et al.'s (2018) finding that CR was the most prevalent cause of failure for outpatient CEs, Fu's (2015) use of the musculoskeletal domain in using Miller's (1990) pyramid for development of the TASPE assessment, and the expertise of the author of this study. Future research may examine predictive validity of Miller's (1990) pyramid in different domains of PT practice, such as neuromuscular or acute care settings.

Another avenue for future research may be to abandon Miller's (1990) pyramid for more recently developed models. Reilly et al. (2022) developed a CR learner blueprint for DPT education. Reilly et al.'s (2022) blueprint was not chosen for this study due to its complexity, its malalignment to the archived data, and its focus on PT education beyond entry-level, such as residencies and fellowship. However, the blueprint has merit in its design and great potential for future use. Research is needed to examine use of the blueprint in DPT education as well as its predictive validity for CR in clinical practice.

Study Limitations

Limitations of this study include generalizability, instrument and CE placement infidelity, and statistical power. Only one hybrid DPT program at one university was utilized for this study due to convenience and access to archived data. Use of multiple DPT programs in multiple locations around the United States would increase generalizability of results. Issues with CIET scoring were previously discussed as the addition of a zero-score point to one subsection of the instrument may have skewed results.

Another aspect of the data that may have skewed results is the patient population for the outpatient clinics where students completed CEs. Some of the outpatient affiliated clinical sites did not treat solely orthopedic or musculoskeletal conditions. Some sites included patients with neurological conditions and were listed as "general" outpatient (rather than outpatient "orthopedic") on the clinical education placement document from which data were gleaned. For the class of 2022, six sites were listed as "general" outpatient, and for the class of 2023, four sites were listed as such. In total, 10 students completed outpatient CEs at a "general" outpatient clinic rather than one treating solely musculoskeletal conditions. The sites' patient populations, however, were outside the control of the researcher and the DPT program. All students were required to complete one full-time outpatient CE, per CAPTE accreditation requirements. However, because of site and CI availability, the DPT program had to place students in some clinics that did not treat patients solely in the musculoskeletal domain.

Statistical power was another limitation for this study. Although an a priori power analysis showed that the number of participants for this study was sufficient for strong power at a medium effect size, the effect sizes resulting from data analyses were too small for statistical power. The model for the multiple linear regression analysis for research question four had a very small effect size (R^2 = .01), which would have required a sample size ranging from 300-600 to reach significant power. Similarly, the predictive effects for each independent variable in the simple linear regression analyses for all research questions were trivial ($R^2 < .01$). These effect sizes also would have required a sample size closer to 600 participants to reach statistically significant power.

Recommendations for Future Research

Although this study did not result in any statistically significant findings for the research questions proposed, the study here can be used as a platform for future research. Replication of this study could be done with multiple DPT programs at multiple universities to increase sample size, generalizability, and statistical power. The DPT program used in this study was a hybrid program. Future research could explore differences between hybrid and traditional (face to face) DPT programs. Miller's (1990) pyramid could be used to examine prediction of CR in practice for students who learned in these different formats during didactic coursework. Another programmatic and curricular difference that could be explored is the number of full-time CEs and part-time CEs. The curriculum for the DPT program in this study included a one-week Integrated Clinical Experience in an outpatient setting during year three, after completion of the musculoskeletal series of courses. The only other CEs for the DPT program in this study were the three full-time CEs that occurred during year four. Other DPT programs may have more part-time CEs during didactic coursework, and more or less full-time CEs at different time points during their curricula.

Another avenue for future research could be to replicate this study using a different domain of PT practice, such as the neuromuscular or acute care setting domains. CAPTE requirements include full-time CEs in the inpatient clinical setting, as well, which includes acute hospital-type settings and skill nursing facilities. Patients with neurological conditions tend to be more prevalent in inpatient settings. Replicating this study with assessments throughout the DPT curriculum pertaining to the neuromuscular system or the acute care setting may yield significant predictive results. Other assessments could be utilized at the four tiers of Miller's (1990) pyramid as well. Replication of this study, even remaining in the musculoskeletal domain, could be done with other types of assessments. Free-response tests and oral tests focusing on verbal repetition of facts could be used to assess the Knows level (Miller, 1990; Pugh et al., 2019). Written exams and assignments, case-based tests, Script Concordance Tests, Key Features Tests, virtual patients, patient management problems, and clinical reasoning problems could be used to assess the Knows How level (Fu, 2020). A practical exam or a TASPE with a standardized patient could be used for the Shows How level (Christensen et al., 2017; Fu, 2015). The CIET or the CPI could be used for the Does level.

Future research could include qualitative aspects of assessing CR. This study focused on quantitative scores at each level of Miller's (1990) pyramid. However, the CIET includes space for CI comments within each subsection and at the end of the instrument under the Global Rating Scale. The CPI also includes space for CI comments for each line item. Lower-level assessments could be changed to include qualitative data as well. A qualitative study examining aspects of CR that may be more apparent in verbal or written format could shed more light on the connection between didactic CR assessments and clinical CR assessment. Furthermore, a study using a mixed methods design could explore the best of both worlds, quantitative and qualitative, to provide a more comprehensive picture of CR development and performance.

Finally, a different structure and design for using Miller's (1990) pyramid has been suggested by Fu (2020). This study used each tier of Miller's (1990) pyramid to assess a separate year in the DPT curriculum. However, Fu (2020) suggested an alternative design using repeated measures of all four tiers of the pyramid at different time points in the curriculum within the same content domain. For example, in a first-year biomechanics course, an assessment could be done using all four tiers of Miller's (1990) pyramid within that single assessment or using multiple assessments to cover all tiers within the biomechanics course. Then during a second-year clinical skills course, the same pattern could be repeated. A single assessment utilizing all four levels of the pyramid could be done perhaps as a culminating comprehensive exam, or multiple assessments addressing each tier separately could be done within the clinical skills course. This pattern could repeat during a third-year musculoskeletal course as well. By using all four tiers of Miller's (1990) pyramid at different time points in a curriculum, faculty may gain a more complete picture of students' CR throughout didactic coursework. Faculty may be able to identify the level and time point for a student struggling in CR and remediate appropriately before the student progresses in the curriculum. Fu's (2020) suggested structure of using Miller's (1990) pyramid in this way is similar to Reilly et al.'s (2022) suggested CR learner blueprint. Each model presents assessment of students' CR at multiple levels and repeated intervals during a DPT curriculum.

Conclusion

Clinical reasoning in PT education is a difficult concept to teach and assess. PT students exhibit difficulty in developing CR during didactic coursework as well as during clinical education. Lack of CR is a major reason for failure of outpatient CEs (Silberman et al., 2018). Faculty need a model for assessing CR throughout a DPT curriculum to identify struggling students and remediate deficits to prevent CE failures. Miller's (1990) pyramid was used as a conceptual framework for this study to examine assessments at progressive levels of a DPT education curriculum. Huhn et al. (2019) created a concept analysis for CR that was used in this study as a basis for selection of CR assessments. The purpose of this study was to determine to what degree CR performance in didactic coursework could predict CR performance in CEs for DPT students.

Using a retrospective quantitative exploratory observational design, archived data from a hybrid DPT program were collected and analyzed using statistical tests of regression. Two cohorts, the classes of 2022 and 2023, were utilized for this study ($N = 84$). Assessments for each level of Miller's (1990) pyramid were biomechanics MCQ exams for the Knows level, clinical skills practical exams for the Knows How level, MSK3 *Triple Jump* scores for the Shows How level, and CIET scores for the Does level. Simple linear regression analyses were conducted to examine whether each of the three lower levels of the pyramid could predict the top level (CIET). Multiple linear regression analysis was conducted to decipher whether the combination of the lower three levels of the pyramid could predict the top level. All tests resulted in statistically non-significant findings. However, follow-up ancillary analysis revealed that cohort CE rotation did have a significant correlation to CIET scores.

Miller's (1990) pyramid proved to be a poor predictor of CR performance in clinical practice for this study. However, many avenues for future research exist and can be developed from the platform provided by this study. Many variables may be manipulated to explore other assessments of CR, clinical practice domains, DPT student populations, or DPT program formats that may reveal different results for predicting CR in practice. Future research may also take a different route entirely by exploring the merits and uses of a new framework, such as the CR learner blueprint developed by Reilly et al. (2022). CR is a complex yet paramount topic for educators in DPT programs to teach and assess. Equally crucial is students' learning of CR and implementation in clinical practice. Educational research will continue to pave the way toward a cohesive paradigm that will serve faculty and students in the pursuit of excellence in clinical reasoning.

References

American Council of Academic Physical Therapy. (2018). Physical therapist clinical education glossary. In: *Reference Manual for Site Coordinators of Clinical Education*. https://aptaeducation.org/special-interest-group/clinical-education-faculty-sig/.

American Council of Academic Physical Therapy. (2022). Excellence framework for physical therapist education. www.acapt.org/excellence.

American Physical Therapy Association. (n.d.). Guide to physical therapist practice 3.0. https://guide.apta.org/.

Baker, S. E., Painter, E. E., Morgan, B. C., Kaus, A. L., Peterson, E. J., Allen, C. S., Deyle, G. D., & Jensen, G. M. (2017). Systematic clinical reasoning in physical therapy (SCRIPT): Tool for the purposeful practice of clinical reasoning in orthopedic manual physical therapy. *Phys. Ther., 97*(1):61-70.

Birkmeier, M., Wheeler, E., McGregor Garske, H., Gorman, S. L., Richards, H., Wolff-Burke, M., & Bureau, M. R. (2022). Feasibility of use of the Clinical Internship Evaluation Tool in full-time clinical education experiences: A multi-institutional study. *J. of Phys. Ther. Ed.,36*(3):263-271.

Borleffs, J. C. C., Custers, E. J. F. M., van Gijn, J., & ten Cate, O. T. J. (2003). "Clinical reasoning theater": A new approach to clinical reasoning education. Acad. Med., 78(3):322-325.

Commission on Accreditation for Physical Therapy Education. (2016). CAPTE aggregate program data. http://www.capteonline.org/AggregateProgramData/.

Commission on Accreditation for Physical Therapy Education. (2017). Standards and required elements for accreditation of physical therapist education programs.

Christensen, N., Black, L., Furze, J., Huhn, K., Vendrely, A., & Wainwright, S. (2017). Clinical reasoning: Survey of teaching methods, integration, and assessment in entry-level physical therapist academic education. *Phys. Ther. 97*:175-186.

Cunningham, S., Litwin, B., Fernandez-Fernandez, A., & Canbek, J. (2019). Influence of residency training on the clinical reasoning development of Kenyan physiotherapists. *J. of Manual & Manipulative Therapy, 27*(4):237-244.

Daniel, M., Rencic, J., Durning, S., Holmboe, E. Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R., Sergio da Silva, A., Cleary, T., Stojan, J., Gruppen, L. D. (2019). Clinical reasoning assessment methods: A scoping review and practical guidance. *Acad. Med., 94*(6):902-912.

Delany, C. & Golding, C. (2014). Teaching clinical reasoning by making thinking visible: an action research project with allied health clinical educators. *BMC Medical Education, 14(*20):1-10.

Edwards, I., Jones, M., Carr, J., Braunack-Mayer, A., & Jensen, G. M. (2004). Clinical reasoning strategies in physical therapy. *Phys. Ther. 84*(4):312-330.

Fitzgerald, L. M., Delitto, A., & Irrgang, J. J. (2007). Validation of the Clinical Internship Evaluation Tool. *Phys. Ther., 87*(7): 844-860.

Fu, W. (2015). Development of an innovative tool to assess student physical therapists' clinical reasoning competency. *J. of Phys. Ther. Ed. 29*(4):14-26.

Fu, W. (2020). Using a comprehensive framework to assess clinical reasoning. In: Musolino, G., & Jensen, G. M. *Clinical Reasoning and Decision-Making in Physical Therapy: Facilitation, Assessment, and Implementation*. Slack Inc.

Furze, J., Black, L., Hoffman, J., Barr, J. B., Cochran, T. M., & Jensen, G. M. (2015). Exploration of students' clinical reasoning development in professional physical therapy education. *J. of Phys. Ther. Ed., 29*(3):22-33.

Furze, J. A., Black, L., McDevitt, A. W., Kobal, K. L., Durning, S. J., & Jensen, G. M. (2022). Clinical reasoning: The missing core competency in physical therapist education and practice. *Phys. Ther. 102*:1-4.

Gilliland, S. (2017). Physical therapist students' development of diagnostic reasoning: A longitudinal study. *J. of Phys. Ther. Ed. 31*(1): 31-48.

Gilliland, S. & Wainwright, S. F. (2017). Patterns of clinical reasoning in physical therapist students. *Phys. Ther., 97*(5):499-511.

Gilliland, S. & Wainwright, S. F. (2020). Perspectives and practice: Physical therapist students' clinical reasoning. *J. of Phys. Ther. Ed., 34*(2):150-159.

Gruppen, L. D. (2017). Clinical reasoning: Defining it, teaching it, assessing it, studying it. *West. J. Emerg. Med. 18*(1):4-7. Doi: 10.5811/westjem.2016.11.33191.

Horwitz, A., Reilly, M., Tichenor, C. J., & Andersen, M. (2021). The functional loss clinical reasoning framework for the neuromusculoskeletal patient. *Orthopaedic Practice, 33*(4):210-218.

Huhn, K. (2017). Effectiveness of a clinical reasoning course on willingness to think critically and skills of self-reflection. *J. of Phys. Ther. Ed., 31*(4):59-63.

Huhn, K., Black, L., Christensen, N., Furze, J., Vendrely, A., & Wainwright, S. (2018). Clinical reasoning: Survey of teaching methods and assessment in entry-level physical therapist clinical education. *J. of Phys. Ther. Ed., 32*(3):241-247.

Huhn, K., Gilliland, S. J., Black, L. L., Wainwright, S. F., & Christensen, N. (2019). Clinical reasoning in physical therapy: A concept analysis. *Phys. Ther. 99*:440-456.

Huhn, K. (2020). The role of standardized tests in assessing clinical reasoning and critical thinking. In: Musolino, G., & Jensen, G.M. *Clinical Reasoning and Decision-Making in Physical Therapy: Facilitation, Assessment, and Implementation*. Slack Inc.

Jensen, G. M., Hack, L. M., Nordstrom, T., Gwyer, J., & Mostrom, E. (2017). National study of excellence and innovation in physical therapist education: Part 2 – a call to reform. *Phys. Ther. 97*(9):875-888.

Jette, D. U., Brechter, J. H., Heath, A. E., & Tschoepe, B. (2020). Defining excellence in physical therapist education: A comparison of 2 theoretical perspectives. *J. of Phys. Ther. Ed. 34*(3):198-205.

Kobal, K. L., Rubertone, P. P., Kelly, S. P., & Finley, M. (2021). A comparison of instructional methods on clinical reasoning in entry-level physical therapy students: A pilot study. J. of Phys. Ther. Ed. 35(2):138-145.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad. Med., 65*(9 Supplemental):S63-S67.

McDevitt, A., Rapport, M. J., Jensen, G., & Furze, J. (2019). Utilization of the Clinical Reasoning Assessment Tool across a physical therapy curriculum: Application for teaching, learning, and assessment. *J. of Phys. Ther. Ed., 33*(4):335-342.

North, S., & Sharp, A. (2020). Embracing change in the pursuit of excellence: Transitioning to the Clinical Internship Evaluation Tool for student clinical performance assessment. *J. of Phys. Ther. Ed., 34*(4):313-320.

Øberg, G. K., Normann, B., & Gallagher, S. (2015). Embodied-enactive clinical reasoning in physical therapy. *Physiother. Theory Pract., 31*(4):244-252. Doi: 10.3109/09593985.2014.1002873.

Pelland, L., Kolomitro, K., Hopkins-Rosseel, D., Durando, P. (2022). Rigor of the objective structured examination for competency assessment in health sciences education: A systematic review. *J. of Phys. Ther. Ed., 36*(4):322-333.

Pugh, D., de Champlain, A., & Touchie, C. (2019). Plus ça change, plus c'est pareil: Making a continued case for the use of MCQs in medical education. *Med. Teacher, 41*(5): 569-577.

Reilly, M., Beran-Shepler, K., & Paschal, K. A. (2020). Pedagogy for effective learning of clinical skills: An integrated laboratory model. *J. of Phys. Ther. Ed., 34*(3):234-241.

Reilly, M., Furze, J., Black, L., Knight, H., Niski, J., Peterson, J., & Jensen, G. (2022). Development of a clinical reasoning learner blueprint: A guide for teaching, learning, and assessment. *J. of Phys. Ther. Ed., 36*(1):43-50.

Richards, J.B., Hayes, M.M., & Schwartzstein, R.M. (2020). Teaching clinical reasoning and critical thinking: From cognitive theory to practical application. *Chest, 158*(4):1617-1628.

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med. Educ., 41*(12):1133-1139.

Sebelski, C. A., Hoogenboom, B. J., Hayes, A. M., Bradford, E. H., Wainwright, S. F., & Huhn, K. (2020). The intersection of movement and clinical reasoning: Embodying "body as a teacher" to advance the profession and practice. *Phys. Ther. 100*(2):201-204.

Silberman, N., LaFay, V., Hansen, R. L., & Fay, P. (2018). Physical therapist student difficulty in clinical education settings: incidence and outcomes. *J. of Phys. Ther. Ed., 32*(2):175-182.

Stickley, L. A. (2018). Creating significant learning experiences for clinical reasoning by physical therapist students. *Internet Journal of Allied Health Sciences and Practice, 17*(1), Article 4.

Surry, L. T., Torre, D., & Durning, S. J. (2017). Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Med. Educ., 51*:1075-1085.

Terry, R., Hing, W., Orr, R., & Milne, N. (2020). Relationships between pre-clinical summative assessment scores and the clinical performance of physiotherapy students. *J. of Allied Health, 49*(1):e13-e19.

Timmerberg, J. F., Dole, R., Silberman, N., Goffar, S. L., Mathur, D., Miller, A., Murray, L., Pelletier, D., Simpson, M. S., Stolfi, A., Thompson, A., & Utzman, R. (2019). Physical therapist student readiness for entrance into the first full-time clinical experience: A delphi study. *Phys. Ther. 99*(2):131-146.

Wainwright, S. F., Shepard, K. F., Harman, L. B., & Stephens, J. (2011). Factors that influence the clinical decision making of novice and experienced physical therapists. *Phys. Ther. 91*(1):87-101.

Wainwright, S. F. (2020). Andragogy: Health professions clinical reasoning transitioning from novice to expert. In: Musolino, G. M. & Jensen, G. M. (2020). *Clinical reasoning and decision-making in physical therapy: Facilitation, assessment, and implementation.* SLACK Inc. pp. 71-88.

World Health Organization. (n.d.). International classification of functioning, disability and health. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health.

Appendix

Clinical Internship Evaluation Tool (CIET)

(Fitzgerald et al., 2007)

Part I. Standards for Professional Behaviors

Professional Behaviors: SAFETY

1. Follows health and safety precautions (e.g., universal standard precautions).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

2. Takes appropriate measures to minimize risk of injury to self (e.g., appropriate body mechanics).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

3. Takes appropriate measures to minimize risk of injury to patient (e.g., choose correct level of assist).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

COMMENTS:

Professional Behaviors: PROFESSIONAL ETHICS

1. Demonstrates compliance with all regulations regarding patient privacy, confidentiality, and security (e.g., HIPAA, DOH, State Practice Act).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

2. Demonstrates positive regard for patients/peers during interactions.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

3. Demonstrates cultural competence (e.g., shows tolerance of and sensitivity to individual differences).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

4. Adheres to ethical and legal standards of practice, including Practice Act and APTA Code of Ethics.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

5. Maintains appropriate appearance and attire in accordance with the facility's dress code.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

6. Maintains appropriate professional conduct and demeanor as per the Code of Professional Conduct.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

7. Demonstrates awareness of patients' rights and responsibilities.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

COMMENTS:

Professional Behaviors: INITIATIVE

1. Recognizes and maximizes opportunity for learning.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

2. Implements constructive criticism.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

3. Utilizes available resources for problem solving.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

4. Is a positive contributor to the efficient operation of the clinic through the demonstration of teamwork and flexibility.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always

COMMENTS:

Professional Behaviors: COMMUNICATION SKILLS

1. Communicates verbally with precise and appropriate terminology and in a timely manner with patients and families/caregivers.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always___ Not observed

2. Communicates verbally with precise and appropriate terminology and in a timely manner with health care professionals (e.g., MD's, nurses, insurance carriers, OT, SLT, etc.).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always___ Not observed

3. Communicates in writing with precise and appropriate terminology and in a timely manner when completing patient documentation.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always___ Not observed

4. Communicates in writing with precise and appropriate terminology and in a timely manner when completing documentation to professionals (e.g., plans of care, physician letters, etc.).
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always___ Not observed

5. Communicates in writing with precise and appropriate terminology and in a timely manner with patients and families/caregivers when creating home programs, patient instructions, etc.
   Never ___ Rarely ___ Sometimes ___ Most of the time ___ Always___ Not observed

COMMENTS:

Part II. Standards for Patient Management

Patient Management: EXAMINATION

1. Obtains an accurate history of current problem.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

2. Identifies problems related to activity limitations and participation restrictions using standardized outcomes instruments when available.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

3. Performs systems review and incorporates relevant past medical history.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

4. Generates an initial hypothesis.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

5. Generates alternative hypotheses (i.e., list of differential diagnosis).
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

6. Selects evidence-based tests and measures to confirm or disconfirm hypotheses.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

7. Recognizes contraindications for further tests and measures.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

8. Demonstrates appropriate psychomotor skills when performing tests and measures.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

COMMENTS:

Patient Management: EVALUATION

1. Makes correct clinical decisions based on the data gathered in the examination (i.e., confirmed/disconfirms initial and alternative hypotheses).
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

2. Identifies impairments in body structure and function (i.e., activity limitations, participation restrictions).
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

3. Administers further tests and measures as needed for appropriate clinical decision making.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

COMMENTS:

Patient Management: DIAGNOSIS/PROGNOSIS

1. Determines a diagnosis for physical therapy management of the patient.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

2. Determines expected outcomes (using standardized indices of activity limitations and participation restrictions where applicable) of physical therapy interventions (goals).
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

3. Selects appropriate physical therapy interventions or makes appropriate consultations or referrals.
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

4. Determines appropriate duration and frequency of intervention (e.g., considers cost effectiveness).
   Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

5. Determines criteria for discharge.

Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

COMMENTS:

Patient Management: INTERVENTION

1. Adheres to evidence during treatment selection.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

2. Applied effective treatment using appropriate psychomotor skills.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

3. Incorporates patient/family education into treatment.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

4. Incorporates discharge planning into treatment.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

5. Assesses progress of patient using appropriate measures.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

6. Modifies intervention according to patient/client's response to treatment.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

7. Recognizes when expected outcome has been reached & makes appropriate recommendations.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

8. Recognizes psychosocial influences on patient management.
Well below __ Below __ At that level for familiar patients __ At that level for all patients __ Above

COMMENTS:

Part III. Global Rating of Student Clinical Competence

On a scale from 0-10, how does the student compare to a competent clinician who is able to skillfully manage patients in an efficient manner to achieve effective patient/client outcomes?

0 = Well below a competent clinician → 5 = Competent Clinician w/familiar pts → 10 = Above a competent clinician

0 ___ 1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10

Is the student performing at a level that is satisfactory for his/her current level of education?

___ Yes ___ No

<u>COMMENTS:</u>