Perceived Accuracy of Doctor of Physical Therapy Students' Self-Assessment Using Two Clinical Performance Assessment Tools

Tracy Porter

Concordia University–Portland

College of Education

Dissertation submitted to the Faculty of the College of Education

in partial fulfillment of the requirements for the degree of

Doctor of Education in

Educational Leadership

Donna Graham, Ph.D., Faculty Chair Dissertation Committee

Dana Shelton, PhD., Content Specialist

Michael Hollis, Ph.D., Content Reader

Concordia University–Portland

2020

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Concordia University–Portland

College of Education

Doctorate of Education Program

WE, THE UNDERSIGNED MEMBERS OF THE DISSERTATION COMMITTEE

CERTIFY THAT WE HAVE READ AND APPROVE THE DISSERTATION OF

Tracy Porter

CANDIDATE FOR THE DEGREE OF DOCTOR OF EDUCATION

Donna Graham, Ph.D., Faculty Chair Dissertation Committee

Dana Shelton, Ph.D., Content Specialist

Michael Hollis, Ph.D., Content Reader

Abstract

Clinical education plays an important role in entry-level, doctoral training for physical therapists. While completing clinical education experiences, Doctor of Physical Therapy (DPT) students self-assess clinical performance and are also assessed by their assigned clinical instructors. Student self-assessment accuracy has not been widely studied in DPT education in the United States. Insight into DPT student self-assessment patterns is needed to inform best practice in DPT clinical performance assessment. The National Consortium of Clinical Educators identified the need to study clinical performance assessment tools for DPT clinical education. This study investigated the perceived accuracy of student self-assessment of clinical performance in DPT education as measured by two assessment tools and congruence with clinical instructor ratings. Guided by adult learning theory, this quasi-experimental, quantitative design compared DPT student self-assessment accuracy between the Physical Therapist Clinical Performance Instrument (PTCPI) and the Clinical Internship Evaluation Tool (CIET). The manipulated independent variable was the clinical performance assessment tool. Subjects were not randomly assigned to groups, but assigned based on cohort. The clinical performance assessment tool utilized by the control group represented the current assessment standard for DPT education. Results revealed DPT student were accurate in self-assessment of clinical performance in all but the control group PTCPI final evaluation (n = 52). At the final assessment, DPT students in the control group rated themselves significantly lower than their clinical instructors. Student self-ratings in the experimental group were congruent with clinical instructor ratings at midterm and final (n = 51). There was improved DPT student self-assessment accuracy observed in the experimental group evaluated using the CIET.

Keywords: clinical performance assessment, physical therapy education

Dedication

I dedicate this dissertation to my parents for instilling in me a love of learning, to my husband and sons for supporting me through my educational endeavors, and to my professional mentors and dissertation chair for providing valuable guidance.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Introduction to the Problem

There are over 200 Doctor of Physical Therapy (DPT) programs in the United States. Clinical education plays a significant role in entry-level, doctoral training for physical therapists. DPT students must complete a minimum of 30 weeks of full-time clinical education experiences in a variety of practice settings as part of a clinical doctoral program (Commission on Accreditation in Physical Therapy Education, 2018). Clinical education experiences account for up to one third of DPT program curricula. While completing clinical education experiences, DPT students formally self-assess clinical performance and are also assessed by their assigned clinical instructors. Formal self and clinical instructor assessments occur at the midpoint and end of clinical education experiences (Mori, Brooks, Norman, Herold, & Beaton, 2015).

Accurate student self-assessment of clinical performance is a desired learning outcome of health professions education (Pawluk, Zolezzi, & Rainkie, 2018; Poirier, Pailden, Jhala, Ronald, Wilhelm, & Fan, 2017). The ability of students to achieve this learning outcome is not well reported in the literature (Lo, Osadnik, Leonard, & Maloney, 2016). Student self-assessment accuracy has not been widely studied specific to DPT education in the United States. Insights into DPT students' self-assessment accuracy, rating patterns, and assessment influences can inform best practice in DPT clinical performance assessment according to the American Physical Therapy Association (2017).

Recently, the American Physical Therapy Association formed an Education Leadership Partnership. This partnership includes representatives from the National Consortium of Clinical Educators, the Academy of Physical Therapy Education, the Clinical Education Special Interest Group, and the Physical Therapist Assistant Special Interest Group. The Education Leadership

Partnership was expressly tasked with identifying best practices for DPT education. Optimizing clinical performance assessment is one goal of the partnership. This study investigated the perceived accuracy of student self-assessment of clinical performance in DPT education and compared the congruence of student and clinical instructor ratings between the Physical Therapist Clinical Performance Instrument (PTCPI) and the Clinical Internship Evaluation Tool (CIET). This design was selected to not only add to the body of knowledge specific to DPT student self-assessment accuracy in the United States, but also to examine the influence of tool design on assessment accuracy.

Background, Context, History, and Conceptual Framework for the Problem

Physical therapy education has evolved significantly since the early 1990s. The entry-level clinical degree requirement for licensure in the United States has increased. Beginning in 2015 the Commission for Accreditation of Physical Therapy Education mandated that all physical therapy education programs in the United States award a clinical doctorate. A clinical doctorate differs from traditional doctor of philosophy degrees. Similar to other clinical doctorates such as doctor of optometry (OD), medical doctor (MD), and doctor of physical therapy (DPT), the emphasis is on clinical reasoning and autonomous evaluation and management of patients and clients. Clinical doctoral programs typically include education on critical consumption of literature rather than the production of independent research, and focus on clinical practice.

The advancement in required education for licensure eligibility was a response to the changing healthcare landscape and progression of the physical therapy profession. Physical therapists have evolved from technicians to autonomous movement experts. Most states, for example, have some form of direct access. This means that physical therapists serve as an entry

point into the healthcare system. The clinical doctorate (DPT) trains physical therapists to evaluate and manage patients without a physician referral, to screen for the need for referral to other healthcare providers, and to function as part of a collaborative healthcare team.

In 2014, the American Physical Therapy Association created the Best Practices in Clinical Education Task force. The formation of the task force represented the initiation of a focused effort to study and advance physical therapy clinical education practices to best support the development of physical therapists to meet the demands of current and future practice. Additional research is needed in all areas of physical therapy clinical education, including effective and efficient clinical performance assessment. This study provided information regarding the accuracy of DPT student self-assessment and compared metrics between two clinical performance assessment tools, the PTCPI and CIET.

DPT students are guided by self and clinical instructor assessment of their clinical performance. The accuracy of DPT student self-assessment and the influence of assessment tool design has not been clearly established in DPT education in the United States. Adult learning theory emphasizes the value of self-direction and self-assessment for learning (Knowles, 1984). Self-assessment involves identifying attributes of performance and using criteria to compare one's performance to the desired standard (Boud, 2003). The desired standards of clinical performance are defined within a clinical performance assessment tool. Self-assessment is utilized for many purposes in education including informal self-monitoring, diagnosis and remediation of deficits, and performance improvement and may be an independent process or combined with outside assessment sources (Boud, 2003). The development of accurate self-assessment skills as DPT students may assist in continued professional growth once licensed.

Health professions students are often required to participate in self-assessment activities. In DPT education, students are required to formally self-assess clinical performance at midterm and final when completing clinical education experiences. Student's assigned clinical instructors also rate performance using the same assessment tool. Over 90% of physical therapy programs in the United States utilize the PTCPI. DPT students and their clinical instructors review each other's assessments, providing the opportunity for students to receive formative feedback as they progress toward the desired standard of performance (Boud, 2003).

Accuracy of DPT student self-assessment has been explored in Australia (Lo et al., 2015). Congruence of student self-assessment with clinical instructor assessment of clinical performance was compared using the Assessment of Physiotherapy Practice tool. The Assessment of Physiotherapy Practice tool shares characteristics with both the PTCPI and the CIET, which are validated DPT clinical performance assessment tools utilized in the United States. The Assessment of Physiotherapy Practice contains seven performance domains including: professional behavior, communication, assessment, analysis and planning, intervention, evidence based practice, and risk management. Twenty total specific skills are assessed, each categorized into one of the performance domains. Each skill is rated on a numeric scale from, and student performance is compared to that of an entry-level clinician (Lo et al., 2015).

The PTCPI and CIET assess performance domains similar to those assessed by the Assessment of Physiotherapy Practice. The rating scales and definitions differ between them, however. The PTCPI rates performance on a continuum from beginner to beyond entry-level in comparison to an entry-level clinician. Ratings are assigned based on the amount of supervision and guidance the student requires in providing patient care as well as the quality, efficiency, and

consistency of student performance. The CIET rates professional behaviors by how often they are displayed and patient management items from well below to above in relation to the skills of a competent clinician. Ratings are assigned based on the amount of supervision and guidance required by the DPT student, as well as the complexity of the patient caseload the DPT student is capable of managing.

Lo et al. (2015) concluded that student and clinical instructor clinical performance assessments differed significantly. Similar student self-assessment inaccuracy is reported in the literature studying other health professions. The purpose of this study was to examine the accuracy of DPT student self-assessment of clinical performance in the United States. Student self-assessment accuracy was compared between two clinical performance assessment tools to identify the influence of assessment tool design. The PTCPI (2006) and the CIET are both validated clinical performance assessment tools for DPT education (Fitzgerald, Delitto, & Irrgang, 2007).

While the PTCPI is most widely used in DPT education in the United States, other validated tools are available (National Consortium of Clinical Educators, 2019). The CIET is an alternative developed at the University of Pittsburgh in 1998. The tool was designed to increase assessment efficiency and decrease the number of specific, individual performance items being assessed. Assessment tools with less complexity and lower number of individual items to evaluate are associated with higher accuracy and inter-reliability (Muhamad, Henry, & Ramli, 2016; Tavares & Eva, 2014). The CIET also has a global rating of clinical competence item which is conducive to the trend that experts tend to assess student performance based on global or holistic impressions (Byrne, Tweed, & Halligan, 2014; Klamen, Williams, Roberts, & Cianciolo, 2016).

In October 2019, the National Consortium of Clinical Educators presented survey results related to clinical performance assessment tools and their impact on physical therapy clinical education. Approximately 75% of DPT programs responded to the survey with all geographic regions of the United States represented. Relevant findings included that nearly 50% of programs are considering moving away from using the PTCPI. Common reasons cited include the length of time required to complete the assessment and the number of individual performance domains assessed. In addition, the ability of DPT students to accurately assess their performance using the PTCPI is not known. As research is conducted to identify improved tools for DPT clinical performance assessment, understanding the influence of tool design on assessment accuracy is critical.

Clinical instructors in Canada also noted the PTCPI did not apply across all practice settings (Mori et al., 2015). The Canadian National Association for Clinical Education in Physiotherapy identified that a clinical performance assessment tool should be psychometrically sound, competency based, user friendly, and relevant to current physiotherapy practice (Mori et al., 2015). The CIET is a validated tool which aligns with many of the recommendations of the Mori et al. (2015) study. Programs in the United States are also exploring the use of alternate assessment tools, including the CIET, which may be more efficient to complete and applicable to emerging models of physical therapy clinical education (National Consortium of Clinical Educators, 2019). This study contributes to the body of knowledge in physical therapy clinical performance assessment by examining DPT student self-assessment accuracy using two assessment tools, the PTCPI and CIET.

Conceptual Framework

The conceptual framework for this study was adult learning theory (Knowles, 1984). Adult learning theory assumes certain characteristics of learners. Adult learners are actively involved in the planning of their learning experiences. DPT students are involved in selecting clinical education placements, and may select clinical placements based on curricular requirements and professional interests. In general, adult learners are less dependent and more self-directed.

Adult learners possess broader experiences with which to connect new learning. DPT students are required to complete observations in physical therapy practice settings prior to entering graduate school. These observational experiences, combined with the integrated clinical experiences in DPT curricula, provide opportunities to connect new learning. Adult learners are interested in learning related to a specific role and prefer the ability to immediately apply new knowledge. Finally, adult learners are intrinsically motivated to learn (Knowles, 1984).

Adult learners have been defined by other characteristics including age, whether being a student is their primary or secondary role, or whether their needs differ from traditional students. DPT students in the United States may be classified as adult learners by age, but more importantly by their desire for learning related to a specific role. DPT students enroll in an academic program with the desired outcome of becoming a licensed physical therapist.

Self-assessment and reflection is a primary skill required to optimize new learning relative to prior experiences. Intentional self-assessment and reflection are common learning strategies aligned with adult learning theory (Malik, 2015). DPT students participate in formal self-assessment and reflection on clinical performance as they complete clinical education experiences. Gaining proficiency as a DPT student is associated with enhanced professional

development beyond the educational environment. Schon (1983) expanded adult learning theory identifying the value of reflective practice for professional development. Reflective practice includes: knowing-in-action, reflection-on-action, and reflection-in-action (Ganni, Botden, Schaap, Verhoeven, Goossens, & Jakimowicz, 2018). The formal self-assessment opportunities embedded in DPT clinical education experiences represent reflection-on-action. Identifying attributes of desired performance and using criteria to compare one's performance to the desired standard were added as elements of self-assessment (Boud, 2003). DPT students are provided with clinical performance assessment tools which provide performance criteria and the desired performance benchmarks. Self-assessment of clinical performance informs DPT students' progress toward their specific goal of becoming licensed physical therapists. Pastore (2017) noted research designs studying self-assessment in higher education are insufficient, however. While self-assessment is "recognized as one of the most important learning skills that students need to become self-regulated learners" (p. 259), its validity and reliability are challenged.

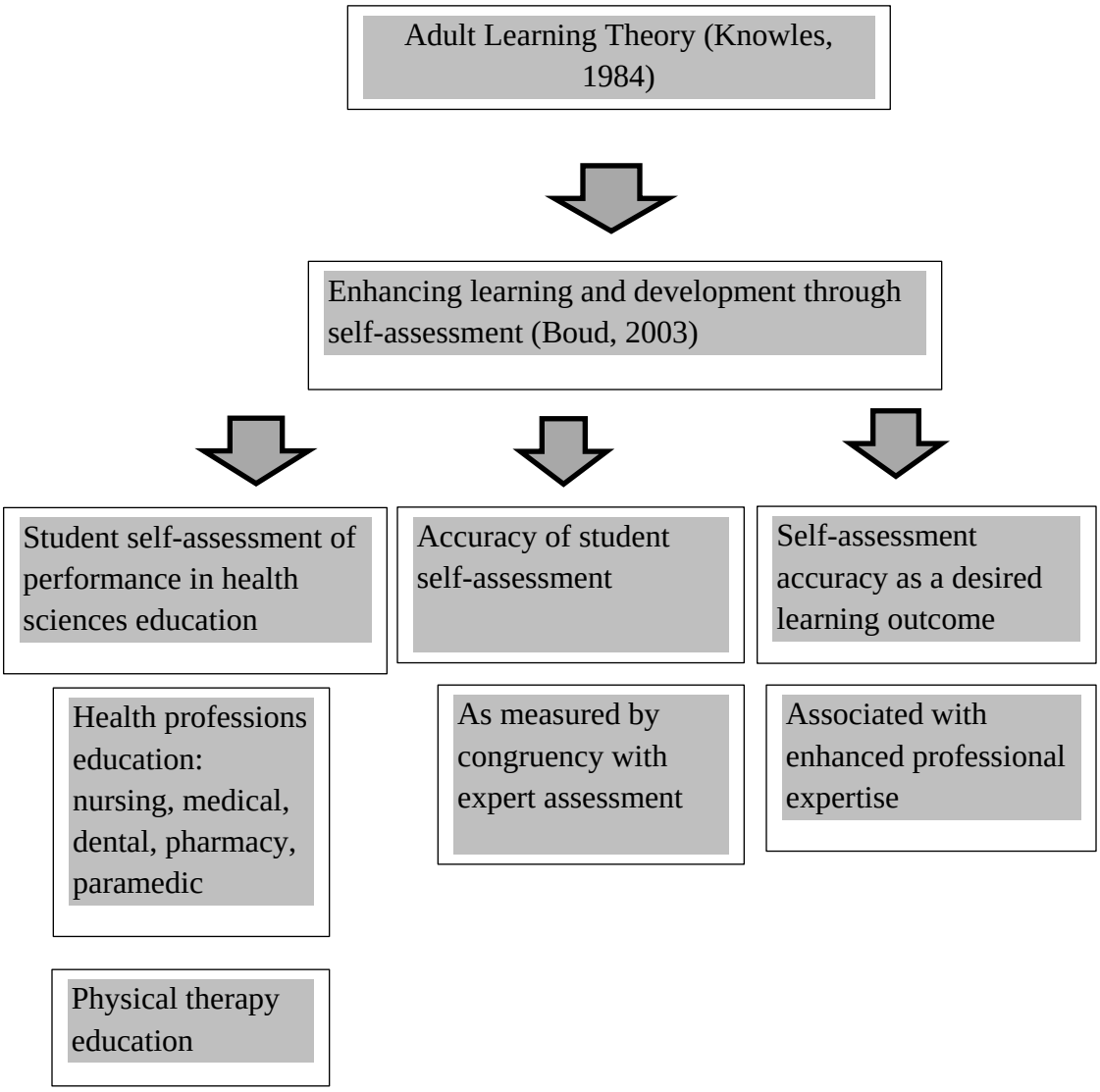


Figure 1. Conceptual framework.

Statement of the Problem

Accurate DPT student self-assessment of clinical performance is necessary in preparation for entry-level physical therapy practice and continued professional development. The assessment accuracy of DPT students in the United States is not widely reported.

Purpose of the Study

The purpose of this study was to investigate the perceived accuracy of student self-assessment of clinical performance in DPT education as measured by two assessment tools and clinical instructor (expert) ratings. This study compared the congruence of student and clinical instructor ratings between the PTCPI and the CIET. The choice to compare DPT students' self-assessment accuracy between to validated clinical performance assessment tools adds to the study's purpose. The PTCPI and the CIET have different designs. They differ in number of individual items assessed, and the complexity of assessment criteria. Comparing self-assessment accuracy between two clinical performance assessment tools allows the influence of assessment tool design to be explored in order to inform clinical performance assessment best practices.

Research Questions

RQ 1. To what extent is DPT student self-assessment of clinical performance accurate as measured by congruence with clinical instructor assessment ratings?

RQ 2. To what extent does the accuracy of student self-assessment of clinical performance differ between the Physical Therapist Clinical Performance Instrument and the CIET as measured by student and clinical instructor rating congruence?

Rationale, Relevance, and Significance of the Study

This study contributes to the body of knowledge related to DPT students' ability to accurately self-assess their clinical performance. Insufficient research is available on self-assessment accuracy specific to DPT education in the United States. The results of this study may be utilized by DPT program chairs and Directors of Clinical Education to inform curriculum. In addition, this study contributes to the ongoing work of the National Consortium of Clinical Educators in identifying optimal clinical performance assessment tools. DPT students

and clinical instructors will also benefit from insights on student clinical performance self-assessment accuracy and impact of specific clinical performance assessment tools on assessment accuracy.

Definition of Terms

Clinical education: Supervised experiential (clinical) learning, focused on the development of patient/client-centered skills and professional behaviors (American Council of Academic Physical Therapy, 2017)

Clinical Performance Assessment Tool: A validated and reliable tool used to determine whether a student meets established objectives during clinical education experiences (American Council of Academic Physical Therapy, 2017).

Clinical instructor: The physical therapist responsible for directly instructing, guiding, supervising, and formally assessing the DPT student during clinical education experiences (American Council of Academic Physical Therapy, 2017).

Doctor of Physical Therapy (DPT): The entry-level degree required for physical therapy licensure in the United States (Commission on Accreditation in Physical Therapy Education, 2018).

Assumptions, Delimitations, and Limitations

The primary assumption of this study was that clinical instructors accurately assess students learning outcomes in physical therapy practice. This assumption is common in the literature. Congruence with expert rating is most often used as a measure of student self-assessment accuracy (Ganni et al., 2017; Hadid, 2016; Lo et al., 2016; Adrian, Zeszotarski, & Ma, 2015; Kachingwe, Phillips, & Beling, 2015). This assumption is further influenced by the range in expertise of clinical instructors measured by years in practice, advanced practice

certifications, or years serving as a clinical instructor. While out of the scope of this study, methods to assess clinical instructor assessment accuracy include comparing narrative comments to benchmark ratings or assigning multiple clinical instructors to evaluate the same student.

A second assumption is that clinical instructors and students receive sufficient training in self-assessment. For the PTCPI clinical instructors and students complete a free online training module and assessment which requires approximately two hours to complete. Training for the CIET is less standardized. A ten-minute video is available from the authors of the tool; however, no assessment is required. Within DPT curricula, students are provided with varying amounts of self-assessment content and practice. Examples include self-assessment of practical examination performance, self-assessment of professional behaviors, and self-assessment of selected skills such as patient interviewing or patient education.

Two delimitations are present in this study design. The first is that Clinical Education Experience 1 represented the first full-time clinical education within the curriculum. Clinical Education 1 was the first opportunity for students to formally self-assess their clinical performance. By choosing to compare self-assessment accuracy between clinical performance assessment tools during Clinical Education Experience 1, neither the control or experimental groups had prior experience with either clinical performance assessment tool.

A second delimitation was the use of a sample of convenience. This study was conducted using two cohorts of DPT students at a single, private, graduate health sciences university. The sample size was a third delimitation. A convenience sample of 103 second year DPT students attending a private, graduate health sciences university in the Midwest were utilized as subjects. Fifty-two students from the 2020 cohort represented the control group and 51 students from the 2021 cohort the experimental group. Sample size analysis was not required as the 52 students in

the control group and 51 students in the experimental group represented the entire cohorts completing Clinical Education Experience 1.

Cohort data for the control group included an average age at matriculation of 23 years and average matriculation grade point (GPA) of 3.68. Cohort data for the experimental group included an average age at matriculation of 22 years and average matriculation GPA of 3.67. The control group was 75% female while the experimental group was 52% female. Clinical instructors for the control group had practiced physical therapy for an average of 10.872 years and had served as clinical instructors for an average of 7.717 years. Clinical instructors for the experimental group had practiced physical therapy for an average of 12.407 years and served as clinical instructors for an average of 9.017 years.

The sample of convenience was representative of DPT students enrolled in all programs in the United States. Physical Therapist Centralized Application Service data indicates the mean age of DPT students in the United States is 23.57 and the mean percentage of female students is 61.4%. The average undergraduate grade point average of students accepted into DPT programs in the United States is 3.57 (American Physical Therapy Association, 2019). The students in each group represented the entire cohort completing Clinical Education Experience 1.

A limitation of this study was that students and clinical instructors require training prior to utilizing their assigned clinical performance assessment tool. This limitation was addressed by requiring completion of established training courses for the PTCPI and CIET. A second limitation was that many of the clinical instructors assessing students in the experimental group had prior experience using the PTCPI as it is the most commonly utilized clinical performance assessment tool in DPT education in the United States (Fitzgerald et al., 2007).

Chapter 1 Summary

Accurate student self-assessment of performance is a desired outcome of many health professions curricula (Pawluk et al., 2018; Poirier et al., 2017). Self-assessment is beneficial for continued professional development and lifelong learning. This quasi-experimental study investigated the perceived accuracy of student self-assessment in DPT education using quantitaive methodology to compare student self-assessment ratings with clinical instructor ratings of clinical performance using two different assessment tools.

Chapter 2: Literature Review

Introduction to the Literature Review

Clinical education plays a significant role in the entry-level doctoral training for physical therapists. DPT students complete a minimum of 30 weeks of full-time clinical education experiences in a variety of practice settings (Commission on Accreditation in Physical Therapy Education, 2018). During clinical education experiences, students are required to formally self-assess clinical performance and are simultaneously assessed by their assigned clinical instructors at the midpoint and end of the clinical education experiences.

Accurate self-assessment is considered an important skill for health professions students, allowing them to "determine their level of knowledge and identify knowledge gaps to remain current and safe in practice" (Hadid, 2016, p. 70). DPT students often have their first opportunity for formal self-assessment during clinical education experiences. The accuracy of DPT student self-assessment accuracy has not been widely studied in the United States. Concerns have been identified regarding the accuracy of student self-assessment in other health professions in the United States, and in physical therapy education in other countries (Pawluk et al., 2018; Hadid, 2016). In this literature review, the researcher explored the current body of knowledge related to student self-assessment accuracy with an emphasis on health professions, and more specifically, DPT students.

Proficient student self-assessment is associated with improved learning outcomes and higher academic achievement (Panadero, Brown, & Strijbos, 2016). Students who more accurately compare their performance to a desired standard or benchmark can better identify development activities toward the achievement of specific outcomes. Opportunities for student self-assessment of clinical performance are embedded in health professions education (Recker-

Hughes, Padial, Becker, & Becker, 2016). DPT programs require self-assessment of clinical performance during each clinical education experience.

An underlying assumption exists that DPT students are equipped to accurately self-assess their clinical performance (Commission on Accreditation in Physical Therapy Education, 2018). This assumption is not widely supported in the literature from other health professions or from physical therapy education in other countries. While limitations exist in the literature regarding student and expert assessment congruence, “such methods remain the predominant basis for evaluating the attainment of clinical skill competencies” (Lo et al., 2016, p. 12). Additional research is needed to identify whether student self-assessment is indeed accurate in DPT education in the United States.

The influence of clinical performance assessment tool design on assessment accuracy has also not been widely studied in DPT education in the United States. Lo et al. (2015) examined the congruence of student self-assessment with clinical instructor assessment of clinical performance in physical therapy education using the Assessment of Physiotherapy Practice (APP). The Assessment of Physiotherapy Practice is a clinical performance assessment tool primarily utilized in Australia. Significant differences were found between student and clinical instructor performance ratings at both the midpoint and end of clinical experiences. The purpose of this study was to compare the accuracy of DPT student self-assessment of clinical performance in the United States using the PTCPI (2006) and the CIET (Fitzgerald et al., 2007). The PTCPI was selected as it represents the most commonly used DPT clinical performance assessment tool in the United States.

Fifty percent of DPT programs in the United States are considering moving away from the PTCPI (National Consortium of Clinical Educators, 2019). Multiple reasons to explore the

use of alternative clinical performance assessment tool have been cited. These include the need for more efficient evaluation, less complex rating benchmark criteria, and potentially improvement in accuracy. The ability of DPT students to accurately assess their performance using the PTCPI is not known. As research is conducted to identify improved tools for DPT clinical performance assessment, understanding the influence of tool design on assessment accuracy will be critical. The CIET is a validated tool representing a viable alternative. Hence, it was selected for comparison in this study.

This chapter will include details of the literature search strategies, inclusion/exclusion criteria, and results. The conceptual and theoretical framework for examining student self-assessment will be defined, and synthesis of the literature will present emerging themes. Finally, the methodology of current research in self-assessment accuracy will be critiqued with gaps in the current body of knowledge presented. The literature review was conducted using the following databases: Eric (ProQuest), EBSCO host, Education Source, and Cinahl with full-text.

Conceptual Framework

DPT students in the United States are classified as adult learners by age and other characteristics. While being a student is typically their primary role, DPT students are classified as adult learners by their desire for learning related to a specific role. DPT students enroll in an academic program toward the specific outcome of becoming a licensed physical therapist. DPT curricula emphasize learning toward that specific goal and include immediate and practical application of content (Knowles, 1984). Clinical education experiences represent one aspect of the immediate and practical content application.

DPT students regularly self-assess and reflect on clinical performance. While participating in clinical education experiences, assessment tools are provided to guide self-

assessment. These tools identify attributes of desired performance and establish specific performance criteria. Guided self-assessment is utilized for self-monitoring, diagnosing and remediating deficits, and improving performance. Self-assessment may be an independent process or combined with outside assessment sources such as peers or experts (Boud, 2003). The self-assessment of DPT students during clinical education experiences is combined with assessment by their clinical instructors. The accuracy of student self-assessment is most commonly measured as congruence with clinical instructor assessment.

While self-assessment is "recognized as one of the most important learning skills that students need to become self-regulated learners" (Pastore, 2017, p. 259), its validity and reliability are challenged. This literature review and study further examined the body of knowledge related to self-assessment accuracy. The attributes of health professions and DPT education served as important inclusion criteria, however key studies from other professions were included.

Review of Research Literature and Methodological Literature

Two primary methods of measuring student self-assessment accuracy were found in the literature. The first was comparison of student self-assessment to an objective measure of performance, such as a written examination (Thawabieh, 2017; Lo et al., 2016, 2015). Regardless of measurement method, health professions students in large are inaccurate in their self-assessment of performance (Oh, Liberman, & Mishler, 2018). While limitations exist in the literature regarding student and expert assessment congruence, "such methods remain the predominant basis for evaluating the attainment of clinical skill competencies" (Lo et al., 2016, p. 12). Selecting the PTCPI and CIET for comparison in the study was beneficial for several reasons. The PTCPI is currently the most commonly used tool. Recently, multiple areas for

potential improvement have been identified. These include reducing the length of the tool, the number of performance areas measured, and reducing its rating complexity. Assessment tool design can influence assessment accuracy; therefore, the CIET was selected for specific comparison of DPT student self-assessment accuracy. It is a validated tool currently utilized by a small number of DPT programs in the United States as an alternative to the PTCPI. The CIET is designed for greater efficiency and increased alignment with competency-based education. Studying the accuracy of student self-assessment using the CIET further informs improvements in DPT clinical education, including identification of an optimal assessment tool.

Various influences on self-assessment accuracy have been identified in studies comparing student self-assessment accuracy to expert assessment (Lo et al., 2015). Student experience and academic proficiency influence self-assessment accuracy. Less experienced and lower performing students commonly overrate their performance level when compared with their expert evaluators (Panadero et al., 2016). More experienced and higher performing students often underrate their level of clinical competence when compared with their expert evaluators (Oh et al., 2018). Other studies support that students either underrate or overrate performance, but do not identify specific patterns of or influences on self-rating (Lo et al., 2015; Pawluk et al., 2018; Poirier et al., 2017).

Demographic factors have been identified as influences on self-assessment accuracy. Age is negatively correlated with self-rating. Older health professions students self-assess their performance more critically than younger students (Hadid, 2016). Gender also influences self-assessment. Female health professions students underrate performance in comparison with their male counterparts (Madrazo, Lee, McConnell, & Khamisa, 2018). The findings related to gender are consistent with those from other science, technology, engineering, and math (STEM) fields.

The conclusions in the literature highlight the need for further understanding as to whether DPT student self-assessment accuracy is influenced similarly.

Reasons for the contributions of age and gender on self-assessment accuracy is not well understood in health professions education (Lo et al., 2015). Several hypotheses are proposed. Students may overrate their clinical performance level to protect against a perceived negative impact of a lower self-assessment, and to protect self-image (Adrian et al., 2015; Pawluk et al., 2018). Adrian et al. (2015) suggested student overrating during self-assessment resulted from a fear of negative consequences, such as negative instructor bias.

Student self-assessment may differ from expert assessment due to lack of confidence or experience (Greenfield, Bridges, Carter, Barefoot, Dobson, Eldridge, & Phillips, 2017). This may be especially influential during early health professions educational experiences which represent the first opportunity students have to compare their clinical skills performance against established standards. Additionally, students and experts may weigh different elements of performance (Greenfield et al., 2017; Ibrahim, MacPhail, Chadwick, & Jeffcott, 2014). For example, students may focus on skill specific elements while an expert emphasized clinical reasoning, professional behavior, or safety during clinical activities.

The lack of congruence between student and expert assessment may also be influenced the differing levels of professional experience between students and experts. Experts tend to assess student performance based on global or holistic impressions as opposed to distinct, individual performance attributes (Byrne et al., 2014; Klamen et al., 2016). Expert practitioners more easily identify student performance as competent versus incompetent, but may weight different salient factors of student performance differently in their assessment (Gingerich, Ramlo, van der Vleuten, Eva, & Regehr, 2017).

Clinical instructors in health professions education engage with students extensively over the course of the clinical education experiences. Clinical instructors are typically invested in the success of the student and accept personal responsibility for student success on clinical education experiences. In physical therapy education, clinical faculty may overrate student performance because they are hesitant to fail the student. DPT program directors of clinical education are the course coordinators for clinical education experiences. In that role, the directors of clinical education assign grades for clinical education experiences based on a variety of factors, including the student and clinical instructor assessments of performance. Although clinical instructors do not assign grades, they may perceive that a below benchmark rating will jeopardize student progression through a DPT curriculum. Clinical instructors may also make inferences regarding student performance versus strictly assessing objective behavior (Lo et al., 2015). They may assume that a student arrived at a clinical conclusion based on a similar strategy of reasoning to their own. Clinical instructors are encouraged to ask students to articulate clinical reasoning to minimize the likelihood of false inferences.

The structure and design of clinical performance assessment tools can influence congruence between student and expert assessments. Assessment tool design has not been widely studied in DPT education in the United States. Clinical performance assessment tools with a higher number of individual items to assess have lower assessment inter-rater reliability (Tavares & Eva, 2014). Muhamad et al. (2016) identified the use of a more global rating scale as preferred to a more complex assessment tool. This recommendation is consistent with the finding that experts tend to assess more holistically (Byrne et al., 2014; Klamen et al., 2016) and is an important consideration in this study examining student self-assessment accuracy between two different tools, the PTCPI and the CIET.

Byrne et al. (2014) cited a combination of assessor factors and assessment tool design as contributing to assessment inaccuracy. Expert assessors experience high mental workload evaluating students, especially when utilizing complex assessment tools. Expert assessors demonstrate more inter-rater reliability in identifying excellent or substandard performance and less reliability identifying adequate performance (Kirwan, Clark, And Daltran, 2019). While clinical instructors consistently identified similar core attributes impacting their ratings of physiotherapy student performance, attributes related to safety were the most influential in the assignment of performance rating. Other factors such as technical skill and confidence were also frequently reported as important in rating decisions. The authors hypothesized a possible rationale for the variability in differentiating between adequate and good or excellent performance. While some components of student performance may have been performed at an excellent level, other components within the same scenario may have been less than adequate. As a result, the clinical instructor's individual bias may be the key determining factor in the overall assessment rating in such cases. Further research is recommended to determine whether assessor training or changes in assessment tool psychometrics can reduce mental workload for the evaluator and improve assessment accuracy (Byrne et al., 2014).

Strategies have been suggested in the literature to improve student self-assessment accuracy. Examples include providing opportunities to practice self-assessment, increasing familiarity with performance criteria, involving students in criteria development, providing opportunities to compare assessment with experts, and emphasizing reflective narrative (Falender, & Shafranske, 2017; Greenfield et al., 2017; Oh et al., 2018). Both the PTCPI and CIET included in this study require reflective narrative, although the quantity differs. Armstrong and Jarriel (2016) cited the importance of training and practice with an assessment tool to

improve inter-rater reliability. Dudek and Dojeiji (2014) developed specific training tips to improve assessment such as providing behavior-based, detailed narrative comments, aligning numeric ratings with comments and including data from other sources including supervisors and peers (Dudek & Dojeiji 2014).

Encouraging students to complete self-assessment prior to skill performance is also associated with improved assessment accuracy and overall task performance. Ganni et al. (2017) provided one group of students with education on self-assessment and reviewed the assessment tool with students prior to use. In addition, students were allowed to complete a practice assessment on a video recorded laparoscopic procedure. Students in this experimental group had higher scores on the actual laparoscopic skill test and their self-ratings were more closely aligned with faculty ratings. It was hypothesized that increased exposure to and practice with the assessment tool assisted students in identifying the key performance expectations for the laparoscopic skill and therefore perform at a higher level. Greenfield et al. (2017) recommended the use of a framework to guide student self-assessment, specifically when qualitative reflection is required. In addition, Huhn (2017) saw improvements in Self Reflection and Insight Scale scores after students completed a specific clinical reasoning course. The findings of Greenfield et al. (2017) and Huhn (2017) support the value of structured education on and practice in self-assessment in health professions and DPT curricula.

DPT students participate in self-assessment during each full-time clinical education experience at midterm and final. The most commonly utilized tool for DPT student clinical assessment in the United States is the PTCPI. The PTCPI is a proprietary tool endorsed by the American Physical Therapy Association (Physical Therapist Clinical Performance Instrument, 2006). The assessment includes 18 individual performance domains. Examples include safety,

clinical reasoning, and patient education. Under each of the 18 performance domains are sample behaviors representing the domain. Ratings are on a continuum between six anchors from beginning performance to beyond entry-level (Physical Therapist Clinical Performance Instrument, 2006). When assigning a rating, assessors considered the amount of supervision and guidance required, the complexity of patient cases, and the consistency, quality, and efficiency of student performance (Physical Therapist Clinical Performance Instrument, 2006).

Along with the ratings assigned for each performance dimension, qualitative narrative descriptions of clinical performance are required. The narrative comments aid in student and clinical instructor rating of performance. The PTCPI requires clinical instructor and student self-assessment at midterm and final timeframes of each clinical experience. This model is beneficial since inaccuracy in student self-assessment makes self-assessment alone inadequate for measuring competence (Kritek, 2015). The PTCPI incorporates quantitative and qualitative elements with clearly identifying expected performance benchmarks as recommended by Boscardin, Fergus, Hellevig, and Hauer (2018) for competency based medical education.

Although currently most commonly utilized, the PTCPI is not the only validated clinical performance assessment tool available for DPT programs in the United States. The CIET represents a validated alternative. The CIET is a clinical performance assessment tool measuring student performance against that of a competent clinician. It was created to improve the efficiency of clinical performance assessment and to measure performance against a defined set of competencies. It contains four items related to professionalism which are measured on a 5-point rating scale with 1 being “never displays the behavior” and 5 being “always displays the behavior” (Fitzgerald et al., 2007). Four items assess patient management and are rated on a 5-point scale with 1 being “well below” and 5 being “well above” (Fitzgerald et al., 2007).

The CIET does not include qualitative narrative for each item assessed, however, comments are required for a limited set of specific skills such as examination, evaluation, and interventions. A summative global rating of clinical competence is provided, as well as a question as to whether the student is performing at a level satisfactory for his or her level of education (Fitzgerald et al., 2007). The purpose of this study was to compare the accuracy of student self-assessment in DPT education in the United States using the PTCPI versus the CIET.

There are limited studies measuring the accuracy of DPT student self-assessment of clinical performance. Lo et al. (2015) examined the congruence of student self-assessment with clinical instructor assessment of clinical performance in physical therapy education using the Assessment of Physiotherapy Practice (APP). Significant differences between student and clinical instructor performance ratings existed at the midpoint and end of the clinical experience. Additional research is needed to investigate DPT student self-assessment accuracy in the United States, and to better understand the influence of clinical performance assessment tool design on self-assessment accuracy.

Review of Methodological Issues

Studies were selected for methodological review, representing experimental, quasi-experimental, and nonexperimental designs. Nonexperimental studies were included if a systematic or comprehensive review process was utilized. Methodological review identified five primary research themes: accuracy of student self-assessment as compared to faculty assessment, accuracy of student self-assessment compared to objective performance, influences on student self-assessment, recommendations to improve student self-assessment, and student perceptions of self-assessment.

Accuracy of student self-assessment as measured by congruence with faculty or expert assessment emerged as the most common methodologic approach (Ganni et al., 2017; Lo et al., 2016 & 2015). This measure of accuracy has been applied in studies of a variety of health professions including pharmacy, nursing, and medicine (Adrian et al., 2015; Ganni et al., 2017). Fewer studies exist specific to physical therapy education. Kachingwe et al. (2015) compared physical therapy student self-assessment ratings with faculty ratings on a practical examination. The experimental group completed their self-assessment with the aid of an available video recording of their performance. Although the sample size was small, student self-assessment accuracy was not improved with the availability of a recording of performance. Lo et al. (2015, 2016) compared physical therapy student and clinical instructor ratings on the Assessment of Physiotherapy Practice. Results indicated physical therapy student self-ratings were not congruent with clinical instructor ratings.

Student self-assessment accuracy is also measured via comparison with an objective measure (Thawabieh, 2017). Lo et al. (2016) compared the congruence of student and clinical instructor scores against final clinical performance rating. Students who underrated their performance as compared with their clinical instructors' rating often received higher overall performance scores on the clinical education experience. This may be related to the correlation between higher performing students and the tendency to underrate. Additionally, a more critical self-assessment of performance at midterm may motivate students to intentionally develop specific skills ultimately leading to higher clinical performance at the conclusion of the experience. Sami et al. (2016) and Thawabieh (2017) compared student self-assessment accuracy against performance on a written examination finding students were unable to predict exam performance accurately.

Hadid (2016) and Yan (2018) examined a variety of factors and influences on student self-assessment. Hadid (2016) studied factors in nursing students while Yan (2018) examined similar factors in primary and secondary grade students. Both looked at individual factors including gender, age, religion, degree of task importance, and self-efficacy. Hadid (2016) found no significant difference in self-assessment based on gender. In this particular study, the number of male nursing students may have been insufficient to determine statistical significance. Older students tended to rate their performance more critically. Interestingly, a strong correlation was found between self-rating and religion. Druse then Muslim students rated themselves higher. Jewish students rated themselves lowest. There was a significant difference between expert assessment rating and student self-rating with students rating themselves higher. Motivation and self-efficacy were positively correlated with self-rating. It was outside of the scope of many of the studies to examine why certain student characteristics influenced self-assessment accuracy.

Several studies identified recommendations for improving student self-assessment. Barton, Schofield, McAleer, & Ajjawi (2016) applied the interACT process for curricular development to improve student self-assessment accuracy. Boscardin et al. (2018) applied best practices from the literature related to the use of student dashboards to promote self-monitoring and assessment. Ganni et al. (2017) and Huhn (2017) explored the impact of a self-assessment course provided to students. Ganni et al. (2017) found a self-assessment training course improved congruence of student ratings with expert ratings. The test group also performed better overall on the simulated laparoscopic procedure. “Reflection-before-practice” and training assists students in identifying strengths and weaknesses in advance of the assessment, and improves performance. Kachingwe et al. (2015) examined the influence of video review on self-assessment, but found access to recorded performance did not improve student self-assessment

accuracy. Kritek (2015), Panadero et al. (2016), and Pastore (2017) synthesized best practices identified in a review of self-assessment literature. Student perceptions related to self-assessment was a common theme identified in self-assessment literature. Ibrahim et al. (2014) conducted semi-structured interviews with medical interns regarding their experiences and perceptions with self and peer assessment. Interns reported assessing themselves and their peers based on the completion of tasks efficiently and consistently. Interns reported lack of confidence in self-assessment and preferred formal feedback from their supervisor. Ndoye (2017) explored graduate social sciences students' perception of the value of self-assessment specific to their learning. Students perceived self-assessment as enhancing learning and promoting personal responsibility for the learning process. Actionable feedback, collaboration, and a supportive environment were identified as enhancing the value of self-assessment. Schoo, Lawn, Rudnik, and Litt (2015) examined whether self-assessment and reflection led to transformative learning.

A variety of research designs existed among the reviewed studies. The majority were experimental, followed by quasi-experimental, then nonexperimental. Quantitative designs were most common (Pawluk et al., 2018; Yan, 2018). Comparing numeric student and faculty ratings of performance was a common design to measure student self-assessment accuracy as congruence with expert assessment (Ganni et al., 2017; Hadid, 2016; Lo et al., 2016, 2015; Adrian et al., 2015). Lee, Tsai, Chiu, & Ho (2016) compared numeric self-assessment values with peer assessments of performance as opposed to comparison with expert assessment. Quantitative methodology was used by Byrne et al. (2014) and Tavares and Eva (2014) to measure rater demands on assessment of student performance. Quantitative values of workload were measured using the NASA Task Load Index. Sami et al. (2016) compared student self-assessments of preparedness with numeric exam scores.

Literature reviews followed quantitative designs in frequency (Boscardin et al., 2018; Pastore, 2017). Kritek (2015) reviewed self-assessment literature to identify best practices toward improving healthcare student self-assessment accuracy. Panadero et al. (2016) completed a comprehensive literature review on the current body of knowledge on student self-assessment. The results of the literature review included that students trained in self-assessment demonstrate improved academic performance. Additionally, accuracy of student self-assessment was associated with higher learning outcomes. Factors influencing self-assessment included the performance assessment tool and methodology, the timing of self-assessment in the learning process, the purpose of self-assessment, and the amount of self-assessment training provided to students. Interventions which improved self-assessment accuracy include: clear definitions and expectations, student involvement in developing the assessment criteria, lower stakes assessments, self-assessment training, and modeling from experts (Panadero et al., 2016; Kritek, 2015).

Mixed methods designs were conducted by Poirier et al. (2017) and Schoo et al. (2015). Poirier et al. (2017) studied nursing, pharmacy, and dental students participating in a simulated experience which emphasized interprofessional error disclosure. Students and faculty assessed performance before and after viewing a video of their performance. Assessment ratings were compared, and students and faculty were interviewed to gather perceptions of the experience. Schoo et al. (2015) studied 36 physical and occupational therapy students conducting Motivational Interviews. The students were provided with a rubric to evaluate their motivational interview, however, still scored themselves higher than did their faculty assessors.

Fewer researchers chose qualitative designs. Greenfield et al. (2017) instructed 20 DPT students in the Gibb's model of self-assessment and qualitatively evaluated their self-assessment

narratives. The Gibb's model provides a self-assessment framework using a series of questions such as: What was the central issue you encountered? What confused you about the issue/case? What feelings did you experience during this issue? What did you learn about yourself from this issue/case/encounter? What would you do differently if you encountered this issue again? By implementing the Gibb's model into student self-assessment narrative, higher levels of reflection occurred. Students tended to be more self-focused vs. patient centered in their reflections, however, likely due to lack of confidence. Ibrahim et al. (2014) conducted semi-structured interviews with medical interns to identify their perceptions on the value of self-assessment. Only one case study was included in the reviewed studies (Barton et al., 2016). This case study followed the application of interACT processes to programmatic feedback and assessment practices in an online medical education program.

Data was collected from a variety of sources. Internally developed data collection tools were most commonly used. Surveys and narrative reflections served as data sources (Adrian et al., 2015; Greenfield et al., 2017; Ndoye, 2017). Specific course rubrics were used to collect self-assessment data (Pawluk et al., 2018; Poirier et al., 2017; Hadid, 2016; Kachingwe et al., 2015). Several studies utilized data from externally developed tools such as the NASA Task Load Index, Competency Assessment Tool, and Mini Clinical Evaluation Exercise (Ganni et al., 2017; Gingerich et al., 2017; Tavares & Eva, 2014). Lo et al. (2016 & 2015) utilized a validated clinical performance assessment tool, the Assessment of Physiotherapy Practice (APP).

Study designs varied in the literature reviewed. Only one study utilized a randomized control trial design (Kachingwe et al., 2015). Ganni et al. (2017) conducted a non-randomized control trial. Pretest/posttest designs were common (Huhn, 2017; Poirier et al., 2017). Literature reviews were not systematic and did not provide search strategies or inclusion/exclusion criteria

(Boscardin et al., 2018; Panadero et al., 2016). Qualitative studies utilized sound methodology, but were limited by sample size. Greenfield et al. (2017) utilized an iterative approach, thematic analysis. Ndoye (2017) also performed a thematic analysis using axial coding. Schoo et al. (2015) completed content analyses of collected qualitative data.

In many cases, the quality of the literature reviewed was limited by small sample sizes. Ten subjects were included in the studies by Barton et al. (2016) and Byrne et al. (2014). Ndoye (2017) and Greenfield et al. (2017) had sample sizes of 16 and 20 respectively. Ibrahim et al. (2014) and Pawluk et al. (2018) had just over 20 subjects. Review of methodological issues revealed limited statistically powerful quantitative studies using externally validated tools, systematic reviews, and randomized control trials in self-assessment literature.

Synthesis of Previous Research

Adult learning theory emphasizes the value of self-direction and self-assessment for learning (Knowles, 1984). Adult learning theory was expanded to include the value of reflective practice for learning and professional development. Reflective practice and self-assessment involve identifying attributes of desired performance and using criteria to compare one's performance to the desired standard (Schon, 1983). Student self-assessment is associated with improved learning outcomes and higher academic achievement (Panadero et al., 2016). The framework of adult learning theory and emphasis on reflective practice guides the synthesis of previous research. Opportunities for self-assessment are common in health professions education curricula. Medical, nursing, pharmacy, and physical therapy curricula rely on student self-assessment as one measure of clinical competency (Recker-Hughes et al., 2016). The assumption that students are equipped to accurately assess their performance, however, is challenged (Pastore, 2017).

Measuring student self-assessment accuracy. Self-assessment accuracy is most frequently measured as congruence with expert assessment. Such comparisons have been explored in medical, nursing, pharmacy, and physical therapy education (Adrian et al., 2015; Ganni et al., 2017; Lo et al., 2015). Student self-assessments differ significantly from expert or faculty assessments. The magnitude and direction of variance, as well as possible contributing factors, are explored in subsequent sections.

The accuracy of student self-assessment may also be determined by comparison to objective measures. The California Critical Thinking and Disposition Inventory and Self Reflection and Insight Scale were used by Huhn (2017) to examine students' self-assessment accuracy. Summative performance evaluations or written examinations may also be used to measure student self-assessment accuracy (Lo et al., 2016; Shaban, Aburawi, Elzubeir, Elango, & El-Zubeir, 2016). In most cases, student self-assessment was not found to be congruent with objective performance.

Self-assessment rating patterns. Patterns have been identified in student self-assessment as compared to expert assessment. Less experienced and lower performing students overrate performance while more experienced and higher performing students underrate when compared to expert ratings (Oh et al., 2018). Hadid (2016) found no significant difference in self-assessment based on gender, however Madrazo et al. (2018) cited multiple sources in medical student self-assessment literature which indicate female health professions students underrate performance in comparison with their male counterparts.

Self-assessment patterns are influenced by age. Older students tend to rate their performance more critically. Religion has also been hypothesized to influence self-assessment. Hadid (2016) found that Druse and Muslim nursing students rated their performance consistently

higher than students of other religious backgrounds. Jewish students rate themselves lowest. Personal motivation and self-efficacy are positively correlated with self-rating (Hadid, 2016).

It is unclear why specific student characteristics influence self-assessment accuracy (Lo et al., 2015). Several hypotheses have been proposed. Students may overrate to protect against a perceived negative impact of a lower self-assessment, and to protect self-image (Pawluk et al., 2018). Student self-assessment may differ from expert assessment due to lack of confidence or experience (Greenfield et al., 2017). Additionally, students and experts may weigh different elements of performance (Greenfield et al., 2017; Ibrahim et al., 2014). For example, students may focus on skill specific elements while an expert may emphasize clinical reasoning or professional behavior during a clinical activity.

Influence of self-assessment tool design. Assessment tool design can influence self-assessment accuracy. Assessment tools with a higher number of individual performance items to assess have lower assessment reliability (Tavares & Eva, 2014). Muhamad et al. (2016) identified the use of a global rating scale as preferred to a more complex assessment tool. More complex assessment tools with a higher number of individual performance items to assess contribute to higher assessor mental workload. High mental workload may contribute to errors in accuracy (Byrne et al., 2014)

Strategies to improve student self-assessment accuracy. Further research is recommended to determine whether training or changes in assessment tool design might reduce mental workload and improve assessment accuracy (Byrne et al., 2014). Panadero et al. (2016) recommend further study of self-assessment accuracy exploring an interaction of multiple variables of influence. “We should stop studying accuracy in isolation and start exploring the effects of the various factors reviewed” (Panadero et al., 2016, p. 817).

Student-directed strategies to improve self-assessment were common in the reviewed literature. Recommended strategies include: providing students with opportunities to practice self-assessment, increasing familiarity with performance criteria, involving students in criteria development, providing opportunities to compare self-assessment with expert assessment, and emphasizing reflective narratives to encourage reflection (Falender & Shafranske, 2017; Greenfield et al., 2017; Oh et al., 2018).

Armstrong and Jarriel (2016) cited the importance of training on the criteria and utilization of specific assessment tools. Training is required prior to the use of both the PTCPI and CIET in DPT programs. Dudek and Dojeiji (2014) provide specific training tips to improve assessment accuracy. The use of behavior-based, detailed narrative comments was recommended with a focus on aligning numeric ratings with narrative comments. It is also recommended to include data from a variety of assessment sources.

Encouraging medical students to complete and practice self-assessment in advance of skill performance was also associated with improved assessment accuracy and overall task performance. Ganni et al. (2017) provided the experimental group of students with education on self-assessment, reviewed the assessment tool with students, and allowed students to complete a practice assessment on a video recorded laparoscopic procedure. Students in the experimental group had higher scores on the actual laparoscopic skill test. Their self-ratings were more closely aligned with faculty ratings. Greenfield et al. (2017) recommended the use of a framework to guide student self-assessment narrative writing. A model guiding students to connect “specific experiences with their thoughts and feelings about those experiences” (p. 49) was associated with higher levels of self-reflection. Huhn (2017) saw improvements in Self Reflection and Insight Scale scores after students completed a specific clinical reasoning course.

Self-Assessment in DPT Education

DPT students participate in self-assessment throughout their curricula including during full-time clinical experiences. The PTCPI is a common tool utilized for self-assessment and expert assessment of clinical skills. The PTCPI is a validated, proprietary tool endorsed by the American Physical Therapy Association (PTCPI, 2006). The assessment includes 18 individual performance domains. Examples include safety, clinical reasoning, and patient education. When assigning a rating, assessors considered the amount of supervision and guidance required, the complexity of patient cases, and the consistency, quality, and efficiency of student performance. Ratings are on a continuum between six anchors from beginning performance to beyond entry-level (PTCPI, 2006).

Along with the ratings assigned for each performance dimension are required qualitative narrative descriptions of clinical performance, which aid in student self-assessment and clinical instructor rating of performance. The PTCPI requires clinical instructor and student self-assessment at midterm and final timeframes of each clinical experience. This model is beneficial since inaccuracy in student self-assessment makes self-assessment alone inadequate for measuring competence (Kritek, 2015). The PTCPI incorporates quantitative and qualitative elements with clearly identifying expected performance benchmarks as recommended by Boscardin et al. (2018) for competency based medical education.

The PTCPI is not the only validated clinical performance assessment tool available. The CIET represents an alternative. The CIET is a validated clinical performance assessment tool which measures student performance against that of a competent clinician. It was created to improve the efficiency of clinical performance assessment and to measure performance against a defined set of competencies. It contains four items related to professionalism which are measured

on a 5-point rating scale with 1 being “never displays the behavior” and 5 being “always displays the behavior” (Fitzgerald et al., 2007). Four items assess patient management and are rated on a 5-point scale with 1 being “well below” and 5 being “well above.”

The CIET does not require qualitative narrative for each item assessed, however, comments are provided for a limited set of specific skills such as examination, evaluation, and interventions. A summative global rating of clinical competence is provided, as well as a question as to whether the student is performing at a level satisfactory for his or her level of education (Fitzgerald et al., 2007).

There are limited studies measuring the accuracy of DPT student self-assessment of clinical performance in the United States using the PTCPI or the CIET. Lo et al. (2015) examined the congruence of student self-assessment with clinical instructor assessment of clinical performance in physical therapy education using the Assessment of Physiotherapy Practice. The Assessment of Physiotherapy Practice is a 20-item questionnaire rating physiotherapy students from “not adequate” (0) to “excellent in practice” (4). Students and clinical instructors must mark only one rating on this scale, not permitting scores in-between rating marks. As with the PTCPI and CIET, students in Australia are assessed at midterm and final. Lo et al. (2051) found significant differences between student and clinical instructor performance ratings at the midpoint and end of clinical experiences. This study provides a framework to determine if the same discrepancy exists among DPT students in the United States.

Critique of Previous Research

Literature on student self-assessment accuracy in health professions education is relatively limited and represents a wide variety of health professions. Studies examining self-assessment accuracy in medical and nursing students are most common while other health

professions are less represented in current self-assessment literature (Barton et al., 2016; Byrne et al., 2014; Ganni et al., 2017; Hadid, 2016). Only one systematic review on student self-assessment accuracy was found. Experimental and quasi-experimental study designs are common; however, few utilize randomized control trial designs. Other experimental designs in current self-assessment literature include non-randomized trials and single cohort designs. A limited number of studies utilize mixed methods or qualitative designs. Literature reviews are common, but the majority unstructured and non-systematic. Most do not outline the specific databases and search terms utilized, or the inclusion/exclusion criteria.

Pastore (2017) completed a systematic content analysis focused on self-assessment in higher education. The review methodology aligns with a theoretical framework following three research lines: self-assessment efficacy, perspectives on and experiences with self-assessment, and congruency between student self-assessment and expert assessment. Journals and articles were selected for the systematic review using a Christie and Fleisher flowchart. Clear inclusion criteria were provided for the content analysis. Few articles in higher education journals on self-assessment met the inclusion criteria. Most studies included in the systematic content analysis were nonexperimental. Many were case studies or other lower forms of evidence. While self-assessment is "recognized as one of the most important learning skills that students need to become self-regulated learners" (p. 259), its validity and reliability are challenged.

Tavares and Eva (2014) conducted a randomized control trial in which subjects were grouped into one of four conditions. Subjects were asked to rate a clinical scenario using an assessment tool with either seven or two dimensions. One group was also exposed to extraneous distraction during the assessment. The total sample size (n = 44) was small when divided into four groups. The sample size was sufficient, however, for statistical power using an ANOVA.

The study design was based on a theoretical framework and variables chosen matched the statistical analyses performed.

A non-randomized control trial was conducted by Ganni et al. (2017). Boud's theory of reflective practice guided the intervention as the experimental group was provided with a systematic self-assessment course focused on reflection before action. The study had a clear research question. The sample size sufficient for statistical power with 30 subjects in both control and experimental groups. The statistical analyses were appropriate for the variables studied.

Kachingwe et al. (2015) completed a randomized repeated measures study. The sample size was sufficient for statistical power with 51 total subjects. Twenty-four students were enrolled in an introductory assessment course and 27 in an orthopedics course. Subjects were randomly assigned to either video or non-video groups. Among other findings, the authors concluded the use of video recording and analysis did not improve student self-assessment accuracy. Lo et al. (2016, 2015) employed retrospective cohort designs with large sample sizes of 100 and 101 respectively. The study designs were consistent with the research question and the statistical analyses appropriate.

Sami et al. (2016) conducted a single cohort design. The initial sample size was large (n = 471), however, only 59 completed both pre-and post-assessment surveys. Among subjects who completed the study, gender and self-reported preparation for a written examination influenced self-assessment accuracy. Adrian et al. (2015) also used a single cohort design with a sample size of n = 175. The grading rubric utilized included communication elements indicated in a widely used communication textbook in Pharmacy education, however was not validated.

Poirier et al. (2017) completed a mixed method design to examine student and faculty assessment of performance in an interprofessional scenario. The sample size was sufficient at *n*=233. The sample was representative of a variety of health disciplines including dental, nursing, and pharmacy students. Schoo et al. (2015) also completed a mixed method design to examine student self-assessment of motivational interviewing skill. Thirty-six physical and occupational therapy students participated with a final *n* = 22 completing both the quantitative and qualitative study elements. A valid and reliable tool, the Motivational Interviewing Treatment Integrity tool was used to self-assess interview quality. Focus groups and written reflections captured qualitative comments related to the experience and use of the Motivational Interviewing Treatment Integrity tool.

Greenfield et al. (2017) employed qualitative methodology to examine the efficacy of a reflective framework in guiding DPT student self-assessment during early clinical experiences. The sample size was relatively small at *n* = 20. A systematic process was used to categorize student narratives into five themes including patient-centered care, professional role, ethical issues, critical thinking, and student and clinical instructor relationship. Ibrahim et al. (2014) collected qualitative data on student perceptions of performance feedback through surveys and semi-structured interviews. The sample size was small at *n* = 21. While a thematic framework was utilized to evaluate responses, the interviews were not consistently administered. Ndoye (2017) conducted a survey of 31 students enrolled in a Social Science course. Of the 31 students, only 16 responded to the survey. Emerging axial coding was utilized to categorize student responses into themes related to their perception of the value of self and peer assessment.

Literature reviews were largely unstructured including those by Boscardin et al. (2018), Dudek and Dojeiji (2014), and Kritek (2015). Panadero et al. (2016) conducted a conceptual

synthesis of the field of student self-assessment research. Topics requiring further research include: types of student self-assessment, accuracy of student self-assessment, the role of expertise in assessment, teaching and curricular expectations, and the impact of self-assessment based on student characteristics.

Critique of previous literature is helpful to identify gaps, as well as inform research design. A quantitative design with sufficient sample size will be important to add to the body of literature on congruence of DPT student self-assessment with expert assessment. Alignment with an identified conceptual framework is recommended. In addition, comparison of student and expert assessment congruence between two different clinical performance assessment tools will inform of the influence of assessment tool design.

Chapter 2 Summary

Adult learning theory includes reflection on experience as a key element. DPT students are required to formally reflect on and self-assess clinical performance at the midpoint and end of each clinical education experience. It is unclear whether DPT students can assess clinical performance accurately (Pawluk et al., 2018; Hadid, 2016). This chapter reviewed literature related to student self-assessment accuracy with an emphasis on health professions.

Accuracy of student self-assessment has not been widely studied in DPT education. This study investigated the accuracy of DPT student self-assessment of clinical performance as measured by congruence with clinical instructor (expert) assessment. This study also compared the congruence of student and clinical instructor ratings between the PTCPI and the CIET.

Chapter 3: Methodology

Introduction

This study investigated the perceived accuracy of student self-assessment of clinical performance in Doctor of Physical Therapy education as measured by two assessment tools and clinical instructor (expert) ratings. Student self-assessment of clinical performance is a common curricular element of health professions education, with accurate self-assessment as a desired learning outcome (Pawluk et al., 2018; Poirier et al., 2017). DPT students complete self-assessment of clinical performance during full-time clinical education experiences. Typically, clinical performance assessments are completed at the midpoint and end of each experience by both the student and clinical instructor. For this study, the clinical performance assessments were completed at midterm (week 5) and final (week 10) of the first, full-time clinical education experience.

Self-assessment accuracy was measured by comparing student self-ratings to clinical instructor (expert) ratings. This accuracy measurement method most common in self-assessment accuracy literature (Adrian et al., 2015; Lo et al., 2015). A quasi-experimental design was utilized. The control group was the cohort of DPT students graduating in May 2020 from a private, graduate health sciences university in the Midwest. The experimental group was the cohort graduating in May 2021. The 2020 cohort utilized the PTCPI for clinical performance assessment during their first, full-time clinical education experience (Clinical Education Experience 1). The evaluation tool represents the current or control state. The PTCPI is the most commonly used assessment tool utilized in the United States. In addition, it was the tool utilized by the sample graduate health sciences University for 20 years. Data for the control group were collected and analyzed retrospectively.

The 2021 cohort utilized the CIET for clinical performance assessment during Clinical Education Experience 1. The CIET is a validated alternative to the PTCPI. It was selected as the experimental clinical performance assessment tool for its reported efficiency and better alignment with competency-based education. The selection of Clinical Education Experience 1 for this study was beneficial as it represented the first exposure of both the control and experimental groups to their designated clinical performance assessment tool. Therefore, neither group of students had familiarity and practice with one clinical performance assessment tool over the other.

The DPT curriculum represented in this study is a 34-month, 126.5 credit program which includes four full-time clinical education experiences. Students are required to complete one outpatient and one inpatient clinical education experience. During the remaining two clinical education experiences in the curriculum, students may choose to explore an area of professional interest such as sports medicine or pediatrics. Students may also choose to explore areas of practice considered elective. Examples of electives include skilled nursing facilities or home health.

The first clinical education experience occurs during the fall term of year two and is 10 weeks in length. All students complete this clinical education experience in an outpatient setting with an emphasis on orthopedic practice. Prior to the first clinical education experience, the didactic curriculum is focused on orthopedic evaluation and management. The second clinical education experience occurs during the summer term between years two and three and is also 10 weeks in length. Students may complete their second clinical education experience in a variety of settings. Prior to this second clinical education experience, the foundational didactic curriculum is delivered including orthopedic, neurological, and cardiopulmonary evaluation and

management. Two terminal, full-time clinical education experiences occur during spring term of year three. The final two clinical education experiences are eight weeks in length and may occur in a variety of settings. Prior to the final clinical education experiences, students complete a final didactic term emphasizing specialty practice and advance manual therapy skills.

For this study, Clinical Education Experience 1 was chosen as it represented the first exposure of students to clinical performance self-assessment. This study contributed to the current body of knowledge related to student self-assessment accuracy of clinical performance. Studies examining student self-assessment accuracy exist in other health professions, but few relate specifically to DPT education. Additionally, no studies have compared self-assessment accuracy between clinical performance assessment tools (PTCPI and CIET).

Purpose of the Study

The purpose of this study was to investigate the perceived accuracy of student self-assessment of clinical performance in Doctor of Physical Therapy education as measured by two assessment tools and clinical instructor (expert) ratings. A variable of clinical performance assessment tool was chosen to examine the influence of assessment tool design on assessment accuracy. Assessment tool design has been shown to influence assessment accuracy, for example, assessment tools with less complexity and lower number of individual items to evaluate are associated with higher accuracy and inter-reliability (Muhamad et al., 2016; Tavares & Eva, 2014).

The PTCPI the most widely clinical performance assessment tool used in DPT education in the United States (American Physical Therapy Association, 2017). The CIET represents an alternative and was developed at the University of Pittsburgh in 1998. Because the American

Physical Therapy Association endorsed and invested funding in cloud-based infrastructure for the PTCPI, fewer programs currently utilize the CIET.

The CIET was designed to increase clinical performance assessment efficiency. It contains fewer specific, individual performance items for assessment. Assessment tools with less complexity and lower number of individual items to evaluate are associated with higher accuracy and inter-reliability (Muhamad et al., 2016; Tavares & Eva, 2014). The CIET also includes a global rating of clinical competence item which is conducive to experts tending to assess student performance based on global or holistic impressions (Byrne et al., 2014; Klamen et al., 2016).

In October 2019, the National Consortium of Clinical Educators presented survey results related to clinical performance assessment tools and their influence on physical therapy clinical education. Approximately 75% responded to the survey with all geographic regions of the United States represented. Relevant findings include that nearly 50% of programs are considering moving away from using the PTCPI. Programs are considering alternate assessment tools which may be more efficient to complete and be applicable to emerging models of physical therapy clinical education (National Consortium of Clinical Educators, 2019). This study contributed to the body of knowledge in physical therapy clinical performance assessment by examining student self-assessment accuracy using two assessment tools.

Comparing student self-assessment accuracy between similar cohorts of DPT students on similar clinical education experiences provided insight into the influence of clinical performance assessment tool design on assessment accuracy. Cohort demographics including age, gender, and matriculation grade point average were compared to identify homogeneity between the control and experimental groups. Data on the clinical instructors' year of clinical practice and years of clinical instruction were gathered from completed clinical site evaluations. This data was

analyzed and compared between groups identify homogeneity between clinical instructor level of expertise in the control and experimental groups.

Retrospective Clinical Education Experience 1 clinical performance assessment data gathered from the 2020 cohort represented the control group. The 2020 cohort utilized the PTCPI for clinical performance assessment. The 2021 cohort of DPT students represented the experimental group. The CIET was utilized to assess clinical performance during Clinical Education Experience 1 for the experimental group.

Research Questions

RQ 1. To what extent is DPT student self-assessment of clinical performance accurate as measured by congruence with clinical instructor assessment ratings?

RQ 2. To what extent does the accuracy of student self-assessment of clinical performance differ between the Physical Therapist Clinical Performance Instrument and the CIET as measured by student and clinical instructor rating congruence?

Hypotheses

The following hypotheses are associated with research question 1:

H_{o1}. DPT student self-assessment accuracy of clinical performance is accurate as measured by congruence with clinical instructor assessment.

H_{a1}. DPT student self-assessment accuracy of clinical performance is inaccurate as measured by congruence with clinical instructor assessment.

The following hypotheses are associated with research question 2:

H_{02}. There is no significant difference in the perceived accuracy of student self-assessment between the PTCPI and CIET as measured by student and clinical instructor rating congruence.

H_{a2}. There is a significant difference in the perceived accuracy of student self-assessment between the PTCPI and CIET as measured by student and clinical instructor rating congruence.

Research Design

The study design was quasi-experimental. The control group was identified as such because the assessment condition, use of the PTCPI, represented current and standard assessment practice. The PTCPIs has been utilized by the specific DPT program in this study for over 20 years. In addition, the PTCPI is also used by over 90% of DPT education programs in the United States. The experimental group was exposed to a novel clinical performance assessment tool for this DPT program, the CIET. A quasi-experimental, quantitative design was appropriate as performance ratings were converted to numeric interval ratings and compared between the student and associated clinical instructor (Pawluk et al., 2018; Yan, 2018). A retrospective cohort graduating in May 2020 represented the control group and a cohort graduating in May 2021 the experimental group.

Target Population, Sampling Method (power) and Related Procedures

A convenience sample of 103 second-year DPT students attending a private, graduate health sciences university in the Midwest were utilized as subjects. Fifty-two students from the 2020 cohort represented the control group and 51 students from the 2021 cohort the experimental group. Sample size analysis was not required as the 52 students in the control group and 51 students in the experimental group represented the entire cohorts completing Clinical Education Experience 1 during their respective time frames.

Cohort data for the control group included an average age at matriculation of 23 years and average matriculation grade point (GPA) of 3.68. Cohort data for the experimental group included an average age at matriculation of 22 years and average matriculation GPA of 3.67. The

control group was 75% female while the experimental group was 52% female. Clinical instructors for the control group had practiced physical therapy for an average of 10.872 years and had served as clinical instructors for an average of 7.717 years. Clinical instructors for the experimental group had practiced physical therapy for an average of 12.407 years and served as clinical instructors for an average of 9.017 years.

The sample of convenience is representative of DPT students enrolled in all programs throughout the United States. Physical Therapist Centralized Application Service data identifies the mean age of DPT students in the United States as 23.57 and the mean percentage of female students per cohort as 61.4%. The mean undergraduate grade point average of students accepted into DPT programs in the United States is 3.57 (American Physical Therapy Association, 2019).

Instrumentation

The PTCPI is a validated, proprietary tool endorsed by the American Physical Therapy Association (Physical Therapist Clinical Performance Instrument, 2006). Written permission was obtained from the American Physical Therapy Association to utilize the PTCPI for this research. It is utilized by over 90% of DPT clinical education programs in the United States (Fitzgerald et al., 2007). The assessment includes 18 physical therapist clinical performance domains and is completed by the student and clinical instructor at the midpoint and end of a clinical education experience. When assigning a rating, the student and clinical instructor must consider the amount of supervision and guidance required, complexity of patient cases, and consistency, quality, and efficiency of student performance (PTCPI, 2006). Rating occurs on a continuum with six anchors ranging from beginning performance to beyond entry-level (PTCPI, 2006). Each performance dimension has associated sample behaviors. Operational definitions are provided for each rating anchor on the continuum. Qualitative narrative comments are required in addition to the rating of

each performance dimension. Qualitative remarks are intended to aid in student self-assessment and clinical instructor rating of performance. Prior to completing the PTCPI, students and clinical instructors are required to complete an online training module developed by the American Physical Therapy Association.

A recent survey conducted by the National Consortium of Clinical Educators identified over 50% of DPT education programs are considering alternate assessment tools for clinical education experiences. There are multiple factors contributing to this exploration. One factor is the length of time required to complete the PTCPI. As productivity demand increase for physical therapist clinical instructors, some are citing the length of the assessment tool as a barrier to offering clinical education opportunities for DPT students. The PTCPI also has a considerable number of individual domains to assess and has more complex operational definitions for each rating anchor.

The CIET is a validated assessment tool which measures student performance against that of a competent clinician. It was developed by faculty at the University of Pittsburgh as a more efficient, less complex clinical performance assessment tool (Fitzgerald et al., 2007). It assesses four performance dimensions related to professionalism. These are rated based on the frequency with which they are displayed by the DPT student. Professionalism domains are measured on a 5-point rating scale with 1 being “never displays the behavior” and 5 being “always displays the behavior” (Fitzgerald et al., 2007). Four patient management items are assessed and are rated on a 5-point scale with 1 being “well below” and 5 being “well above” (Fitzgerald et al., 2007). When rating patient management items, DPT students are compared to a competent clinician in their ability to manage familiar and complex patients.

The CIET does not require qualitative narrative for each item assessed. Instead, narrative comments are required for select significant performance elements. Narrative text boxes are included for a limited set of skills such as examination, evaluation, and interventions. The opportunity to provide a summative global rating of clinical competence is provided. Clinical instructors are also asked to comment as to whether the student is performing at a level satisfactory for his or her level of education (Clinical Internship Evaluation Tool, 1998).

Data Collection

This study was conducted under IRB approval from Concordia University–Portland and with approval of the university and physical therapy program in which the subjects were students. Participants in the experimental group received written and verbal advisement that inclusion of their data in this study was voluntary and they may withdraw their data from the study at any time. Cohort demographic data specific to age, gender, and matriculation grade point average was gathered from aggregated and summarized admissions reports. Clinical instructor years of experience in physical therapy practice and clinical instruction were gathered from the students' completed site evaluations.

Students and their assigned clinical instructors completed either the PTCPI or the CIET at the midpoint (week five) and end (week 10) of Clinical Education Experience 1. Students and clinical instructors completed training prior to utilizing both the PTCPI and the CIET. Training for the PTCPI consisted of a 60-minute recorded presentation followed by a brief written assessment of competence in using the tool. Training for the CIET consisted of viewing a 10-minute video and supplemental written materials. Student and clinical instructor data were automatically paired for both clinical performance assessment tools. PTCPI evaluations were completed via a proprietary, web-based system. The system was password protected and students

and clinical instructors were provided access to only their assigned evaluations. The CIET was completed via Qualtrics licensed by the researcher's university. Access to assigned evaluations was granted via an individualized link sent to students and clinical instructors. Both collection systems allowed for the export of data for analysis to Excel and SPSS.

Clinical performance assessment data from the PTCPI for the class of 2020 (control group) Clinical Education Experience 1 was exported to Excel via the web-based system's report generating function. The exported data remained paired for students and assigned clinical instructors, but were de-identified. Clinical performance assessment data from the CIET for the class of 2021 (experimental group) Clinical Education Experience 1 were exported to Excel via Qualtrics report generating function. The exported data remained paired for students and assigned clinical instructors, but were de-identified. The Excel exports for both the control and experimental groups were stored on a password protected computer in a locked office. Only the researcher, co-course coordinator, and clinical education academic assistant had access to the raw data.

Interval ratings were assigned to the PTCPI ranging from "beginner" (1) to "beyond entry-level" (11). The CIET used a 5-point Likert scale ranging from "never displays the behavior" (1) and "always displays the behavior" (5) for professionalism dimensions. A 5-point Likert scale ranging from "well below" (1) to "well above" (5) was used for patient management items (Fitzgerald et al., 2007). Demographic data for each cohort was gathered from admission records including age, incoming grade point average, and gender. Data on clinical instructor years of experience in practice and clinical instruction were gathered from completed site evaluation reports.

Operationalization of Variables

The independent variables in this study include the clinical performance assessment tool (PTCPI or CIET), the timing of the clinical performance assessment (midterm or final), and the assessor (DPT student or clinical instructor). While the assessment tools share similar performance domains, their structure and ratings systems differ. The dependent variables in the study are the student self-assessment ratings and the clinical instructor ratings of performance in the areas of professionalism, evaluation, and interventions.

The PTCPI and CIET have external and internal validity. The PTCPI assesses 18 clinical performance dimensions. Six are categorized under professional practice: safety, professional behavior, accountability, communication, cultural competence, and professional development. Twelve are categorized under patient management: clinical reasoning, screening, examination, evaluation, diagnosis and prognosis, plan of care, procedural interventions, educational interventions, documentation, outcomes assessment, financial resources, and direction and supervision of personnel. Construct validity is established. The performance domains measured represent the behaviors and skills required for entry-level physical therapist practice. The variables are further operationalized by assigning interval, numeric values to the ratings of performance ranging from beginner (1) to beyond entry-level (11).

The CIET assesses the professional behaviors of safety, professional ethics, initiative, and communication. Patient management skills assessed include: examination, evaluation, diagnosis/prognosis, and intervention. Construct validity is established. The variables measured are representative of the behaviors and skills required for competent physical therapy practice. The variables are further operationalized via Likert scale interval ratings of performance. Professional behaviors are measured based on the frequency at which they are demonstrated by

the DPT student. Patient management skills are measured in comparison to a competent clinician's ability to manage familiar and complex patients.

Data Analysis Procedures

Cohort demographic data including gender, matriculation grade point average, and age were compared descriptively as the data were only available to this research in aggregate form for each cohort. Average years of clinical instructor practice experience were compared between the control and experimental group using a paired *t*-test. Years of clinical instructor practice experience and clinical instruction served as a measure of clinical instructor expertise.

Data were exported from each clinical performance assessment tool via the associated reporting functions and de-identified. Student ratings of clinical performance were compared with the ratings provided by their assigned clinical instructor. Congruence of the student and clinical instructor ratings of clinical performance was used as the measure of student self-assessment accuracy. Clinical performance assessment data from the PTCPI and CIET were analyzed by treating the performance domain rating measurements as interval data. Three performance domains were selected for analysis from each clinical performance assessment tool. Analysis was limited to three performance domains from each tool as correlations exist between DPT student and clinical instructor rating patterns throughout the assessment (Porter, 2016). The domains were matched by descriptors and selected to represent affective, cognitive, and psychomotor performance domains for each clinical performance assessment tool.

Professional behavior, evaluation, and procedural interventions were the domains analyzed from the PTCPI. Professional ethics item one, evaluation item one, and interventions item two were analyzed from the CIET. When rating professional behavior using the PTCPI, 13 sample behaviors are provided and includes such things as maintaining productive working

relationships with patients and others, displaying compassion, and maintaining integrity in practice. The entire domain with its sample behaviors is given one rating ranging from beginner to entry-level on a 10-anchor scale.

The corresponding professional ethics domain is structured differently on the CIET. Seven individual behaviors are listed and include such things practicing in accordance with professional and legal guidelines and demonstrating positive regard for patients and colleagues. Item one was analyzed, and corresponds with the affective performance domain selected from the control tool in its focus on integrity in practice. Each listed professional behavior on the CIET is rated separately on a 5-point scale based on the frequency with which the behavior is demonstrated.

Similar assessment tool design differences exist for the evaluation performance domains selected for analysis from each clinical performance assessment tool. The PTCPI includes four sample behaviors under evaluation ranging focused on the student's ability to make sound and efficient clinical decisions based on examination findings. The corresponding evaluation domain is structured on the CIET with three individual behaviors listed which closely match the sample behaviors in the PTCPI. The three listed behaviors are rated separately on a 5-point scale in comparison to a competent clinician.

The procedural interventions domain on the PTCPI contains 10 sample behaviors focused on the student's ability to perform interventions safely and effectively and to modify interventions based on patient response. The intervention domain on the CIET contains eight individual items with examples including: "applies effective treatment using appropriate psychomotor skills" and "modifies intervention according to patient/client's response to treatment" which align well with the associated sample behaviors on the PTCPI.

A two-factor, repeated measures design was used to analyze clinical performance data. Factor one represented the performance evaluations time (midterm and final). Factor two represented the assessor group (student and clinical instructor). Paired *t* tests were performed on each data set to compare midterm and final ratings by assessor groups and between assessor groups. A within factors design paired each student with four ratings: self-midterm and final and clinical instructor midterm and final. Because of this design, a more rigorous significance level was applied ($p \leq .0125$ as opposed to $p \leq 0.05$).

Limitations and Delimitations of the Research Design

Three primary limitations were present in this research design. The first was the assumption that clinical instructors assess students accurately as experts in physical therapy practice and clinical instruction. This assumption is supported in the literature. Congruence with expert rating is the most common measure of student self-assessment accuracy (Ganni et al., 2017; Hadid, 2016; Lo et al., 2016; Adrian et al., 2015; Kachingwe et al., 2015). A second limitation was the assumption that clinical instructors and students have sufficient proficiency in using the assigned clinical performance assessment tool. This limitation was addressed by requiring completion of established training courses for the PTCPI and CIET. A third limitation was that many of the clinical instructors assessing students in the experimental group had prior experience using the PTCPI as it is the most commonly utilized clinical performance assessment tool in DPT education in the United States (Fitzgerald et al., 2007). Physical therapist clinical instructors often serve in that role for multiple DPT students from multiple DPT programs over time.

Two delimitations were present in this study design. The first was that Clinical Education Experience 1 represented the first full-time clinical education within the curriculum and the first

opportunity for students to formally self-assess clinical performance. Experience and practice are known to improve student self-assessment accuracy (Panadero et al., 2016; Pastore, 2017). Choosing to compare self-assessment accuracy between clinical performance assessment tools during Clinical Education Experience 1 was intentional, however. Using this design, neither the control nor experimental groups of DPT students had prior experience with either clinical performance assessment tool at the time of the study.

A second delimitation was the use of a sample of convenience. This study was conducted using two cohorts of DPT students at a single, private, graduate health sciences university. The sample size of 103 subjects with 52 in the control group and 51 in the experimental group was a delimitation. The 52 students in the control group and the 51 in the experimental group represented the entire cohorts completing Clinical Education Experience 1 during their respective time frames. While the sample included cohorts DPT students from only one university, the demographic and academic characteristics of the sample population were comparable to national DPT student demographics reported in the United States.

Table 1

Sample and National DPT Student Demographics

	Control Group	Experimental Group	US DPT Students
Number of students	52	51	NA
Mean age at matriculation	23	22	24
Gender (female %)	75	52	61
Mean GPA at matriculation	3.68	3.67	3.57

Internal and External Validity

The PTCPI was endorsed by the American Physical Therapy Association in 1997. Over 90% of physical therapy education programs in the United States utilize the PTCPI to assess DPT student clinical performance. The PTCPI assesses student performance as compared to the level of performance required by an entry-level clinician. The 1997 version had adequate psychometrics including moderate interrater reliability and content validity.

An ad hoc committee was tasked by the American Physical Therapy Association to review the tool based on variability in its use and prior to the assessment being transitioned from paper based to a web-based platform. The PTCPI was revised in 2006 with the number of assessed performance criteria decreasing from 24 to 18. The visual analog rating scale was converted to the current categorical rating scale. The 2006 version demonstrates high internal consistency (Cronbach alpha 0.99) and construct validity. Interrater reliability has not been determined (Roach, Frost, Francis, Giles, Nordrum, & Delitto, 2012).

The CIET is a validated assessment tool measuring student performance against that of a competent clinician. Developed in 1998 and first administered in 1999, the CIET was designed to assess clinical competence in alignment with guiding physical therapy standards of practice. The CIET was also designed to be less complex and more efficient to complete than the PTCPI. The CIET demonstrates high internal consistency (Cronbach alpha 0.98) and a correlation of 0.76 between the two measures of clinical competence (Fitzgerald et al., 2007).

Expected Findings

Based on student self-assessment literature for other health professions, it was expected that DPT students will not accurately self-assess clinical performance. It was unclear whether they would consistently over or underrate performance relative to their clinical instructors, or

whether student self-assessment rating patterns would be influenced by individual factors such as academic proficiency, age, or gender. DPT students are typically high academic achievers. According to findings in the literature, DPT students therefore may underrate clinical performance compared to their clinical instructors (Oh et al., 2018).

Based on literature specific to assessment tool design, it was expected that student's self-assessments would more accurately align with clinical instructor assessments using the CIET versus the PTCPI. Assessment tools with less complexity and a lower number of individual items to evaluate are associated with higher accuracy and inter-reliability (Muhamad et al., 2016; Tavares & Eva, 2014). The CIET had fewer individual performance dimensions to evaluate and used a simpler Likert rating scale with simpler operational definitions.

Ethical Issues in the Study

This study was conducted under IRB approval from Concordia University Portland and with approval of the university and Department of Physical Therapy for which this researcher is employed. Maintaining privacy of student academic data represented an ethical issue in the study. This issue was mitigated in the collection, storage, and processing of data. Participants in the experimental group received written and verbal advisement that inclusion of their data in the study was voluntary and they could choose to withdraw their data from the study at any time. Informed consent was not required as clinical performance assessment is a persistent course requirement. Cohort demographic data specific to age, gender, and matriculation grade point average were gathered from aggregated and summarized admissions reports. Clinical instructor years of practice experience and clinical instruction were gathered from completed site evaluation reports.

PTCPI evaluations were completed via a password protected, web-based system. Students and clinical instructors were only able to view their assigned evaluations. The CIET was completed via Qualtrics licensed by the researcher's university. While the CIET represented a modification of assessment method for a clinical course, its validation and use by other DPT programs mitigated that ethical concern. Access to assigned evaluations was granted via an individualized link sent to students and clinical instructors. Both collection systems allowed for the export of de-identified data for analysis to Excel and SPSS. The Excel exports were stored on a password protected computer in a locked office. Only the researcher, co-course coordinator, and clinical education academic assistant had access to the data.

Chapter 3 Summary

Adult learning theory includes reflection on experience and self-assessment of performance as key elements (Schon, 1983). DPT students self-assess clinical performance formally at both the midpoint and end of clinical education experiences. It is unclear whether DPT students assess clinical performance accurately (Pawluk et al., 2018; Hadid, 2016). This study used a quasi-experimental design to investigate the accuracy of DPT student self-assessment of clinical performance as measured by congruence with clinical instructor (expert) assessment. This study also compared the congruence of student and clinical instructor ratings between the PTCPI and the CIET. This study contributed to the current body of knowledge related to the accuracy of DPT student self-assessment of clinical performance. Additionally, it provided insight into the potential influence of clinical performance assessment tool design on student self-assessment accuracy and informed potential benefits of one assessment tool over the other.

Chapter 4: Data Analysis and Results

Introduction

Clinical education accounts for up to one third of DPT education curricula (Commission on Accreditation in Physical Therapy Education, 2018). Within clinical education, student self-assessment of clinical performance is a common requirement preparing students to be reflective practitioners. Accurate student self-assessment is associated with improved learning outcomes and higher academic achievement (Panadero et al., 2016). The most common measure of student self-assessment accuracy is congruence with faculty or expert rating. The purpose of this study was to investigate the perceived accuracy of DPT student self-assessment of clinical performance at a private, graduate health sciences university as measured by two assessment tools and clinical instructor (expert) ratings.

The variable of clinical performance assessment tool was chosen to examine the influence of assessment tool design on assessment accuracy. Simpler assessment tools with a lower number of individual items to evaluate are associated with higher accuracy and inter-rater reliability (Muhamad et al., 2016; Tavares & Eva, 2014). The assessment tools selected for this study were the PTCPI (2006) and the CIET (Fitzgerald et al., 2007). The PTCPI is utilized by over 90% of DPT education programs in the United States (National Consortium of Clinical Educators, 2019). It is a proprietary tool endorsed by the American Physical Therapy Association.

The PTCPI includes 18 individual performance domains. Examples include safety, professional behavior, clinical reasoning, evaluation, and procedural interventions. Ratings are assigned on a continuum between six anchors based on the amount of supervision and guidance

required, complexity of patient cases, and consistency, quality, and efficiency of student performance (PTCPI, 2006).

The CIET is an alternative and validated tool for DPT student clinical performance assessment developed at the University of Pittsburgh in 1998. It was created to improve the efficiency of clinical performance assessment and to measure performance against a defined set of competencies. The CIET contains four items related to professionalism measured on a 5-point rating scale with 1 being “never displays the behavior” and 5 being “always displays the behavior” and four items assessing patient management rated on a 5-point scale with 1 being “well below” and 5 being “well above” (Fitzgerald et al., 2007). A summative global rating of clinical competence is provided, as well as a question as to whether students are performing at a level satisfactory for their level of education (Fitzgerald et al., 2007).

A quasi-experimental design was utilized to measure student self-assessment accuracy during the first 10-week, full-time clinical education experience in the DPT curriculum in comparison to clinical instructor (expert) ratings. The control group was the cohort of DPT students graduating in May 2020. The experimental group was the cohort graduating in May 2021. The 2020 cohort utilized the PTCPI for clinical performance assessment during their first, full-time clinical education experience (Clinical Education Experience 1). The data were analyzed retrospectively. The 2021 cohort utilized the CIET for clinical performance assessment during Clinical Education Experience 1. Clinical Education Experience 1 was chosen as it represents the first exposure of students to clinical performance self-assessment.

The instruments used to collect data (the PTCPI and CIET) have been validated. The PTCPI demonstrates high internal consistency (Cronbach alpha 0.99) and construct validity. Interrater reliability has not been determined (Roach et al., 2012). The CIET demonstrates high

internal consistency (Cronbach alpha 0.98) and a correlation of 0.76 between the two measures of clinical competence (Fitzgerald et al., 2007).

Cohort demographic data specific to age, gender, and matriculation grade point average were gathered from aggregated admissions reports. Clinical instructor years of physical therapy and clinical instruction experience were gathered from the students' completed clinical instructor evaluations. For the control group, the PTCPI was utilized to assess clinical performance during Clinical Education Experience 1 midpoint (week five) and final (week 10). Clinical Education Experience 1 was completed by this group during the fall of 2018. PTCPI evaluations were completed via a proprietary, password protected, web-based platform.

For the experimental group, the CIET was used to collect midterm (week five) and final (week 10) assessment ratings from both students and clinical instructors completing Clinical Education Experience I during the fall of 2019. The CIET was delivered via Qualtrics to students and their assigned clinical instructors. Access to assigned evaluations was granted via an individualized link. Data from both assessment tools were exported and de-identified into Excel for analysis using SPSS software.

Three performance domains were selected for analysis from each clinical performance assessment tool. Analysis was limited to three performance domains from each tool as correlations exist between DPT student and clinical instructor rating patterns throughout the assessment (Porter, 2016). The domains were matched by descriptors and selected to represent affective, cognitive, and psychomotor performance domains for each clinical performance assessment tool. Professional behavior, evaluation, and procedural interventions were the domains analyzed from the PTCPI. Professional ethics item one, evaluation item one, and interventions item two were analyzed from the CIET.

When rating professional behavior using the PTCPI, 13 sample behaviors are provided and includes such things as maintaining productive working relationships with patients and others, displaying compassion, and maintaining integrity in practice. The entire domain with its sample behaviors is given one rating ranging from beginner to entry-level on a 10-anchor scale. The corresponding professional ethics domain is structured differently on the CIET. Seven individual behaviors are listed and include such things practicing in accordance with professional and legal guidelines and demonstrating positive regard for patients and colleagues. Item one was analyzed, and corresponds with the affective performance domain selected from the control tool in its focus on integrity in practice. Each listed professional behavior on the CIET is rated separately on a 5-point scale based on the frequency with which the behavior is demonstrated.

Similar assessment tool design differences exist for the evaluation and intervention related performance domains selected for analysis from each clinical performance assessment tool. The PTCPI includes four sample behaviors under evaluation ranging focused on the student's ability to make sound and efficient clinical decisions based on examination findings. The corresponding evaluation domain is structured on the CIET with three individual behaviors listed which closely match the sample behaviors in the PTCPI. The three listed behaviors are rated separately on a 5-point scale in comparison to a competent clinician. The procedural interventions domain on the PTCPI contains 10 sample behaviors focused on the student's ability to perform interventions safely and effectively and to modify interventions based on patient response. The intervention domain on the CIET contains eight individual items. Item two: "applies effective treatment using appropriate psychomotor skills" was analyzed for this study.

A two-factor, repeated measures design was used to analyze clinical performance data. Factor one represented the performance evaluations time (midterm and final). Factor two

represented the assessor group (student and clinical instructor). Paired *t*-tests were performed on each data set to compare midterm and final ratings by assessor groups and between assessor groups. A within factors design paired each student with four ratings: self-midterm and final and clinical instructor midterm and final. Because of this design, a more rigorous significance level was applied ($p \leq 0.0125$ as opposed to $p \leq 0.05$). To help guide this research, the following questions were asked:

RQ 1. To what extent is DPT student self-assessment of clinical performance accurate as measured by congruence with clinical instructor assessment ratings?

RQ 2. To what extent does the accuracy of student self-assessment of clinical performance differ between the Physical Therapist Clinical Performance Instrument and the CIET as measured by student and clinical instructor rating congruence?

Description of the Sample

This research was conducted using a convenience sample of two cohorts of DPT students at a single, private, graduate health sciences university and their assigned clinical instructors during the first full-time clinical education experience. There were 52 students in the control group representing the entire cohort completing Clinical Education Experience I during the fall of 2018. Performance ratings were collected retrospectively from the student and clinical instructor assessments at midterm and final. There were 51 students in the experimental group representing the entire cohort completing Clinical Education Experience I during the fall of 2019. Cohort size is typically 52 students; however, one student did not successfully complete the pre-requisites required to advance to Clinical Education Experience I. Aggregate student demographic data for each group are provided below (Table 2). Statistical analysis could not be performed as only aggregate data were available.

Table 2

Student Characteristics

	Control Group	Experimental Group
Number of students	52	51
Age at matriculation	23	22
Gender (female %)	75	52
Mean GPA at matriculation	3.68	3.67

Clinical instructor data including the years of practice as a physical therapists and years serving as a clinical instructor were gathered from completed student clinical instructor evaluations. The data were de-identified and exported into Excel for analysis using SPSS software. Clinical instructors for the control group had practiced physical therapy for an average of 10.872 years and had served as clinical instructors for an average of 7.717 years. Clinical instructors for the experimental group had practiced physical therapy for an average of 12.407 years and served as clinical instructors for an average of 9.017 years. Assigned clinical instructor data for the experimental and control groups are summarized in Table 3. No significant differences were observed between groups for the years of physical therapy practice ($p = 0.284$) or years of clinical instruction experience ($p = 0.934$).

Table 3

Clinical Instructor Characteristics

	Statistic	Control Group	Experimental Group
Years of physical therapy practice	Range	1–30	2–31
	Mean	10.872	12.407
	Std. Deviation	7.717	9.017
Years as clinical instructor	Range	0–24	0–30
	Mean	7.145	6.704
	Std. Deviation	7.944	7.230

Summary of the Results

The control and experimental student groups were compared descriptively as only aggregate data were available. Average age and matriculation grade point average were similar between the control and experimental groups. The most notable difference between DPT student cohorts was gender (75% and 51% female respectively). No significant differences were observed between clinical instructors associated with the control and experimental groups relative to the years of physical therapy practice ($p = 0.284$) and years as clinical instructor ($p = 0.934$)

Clinical performance assessment data from the PTCPI and CIET were analyzed by treating the performance domain rating measurements as interval data. Three performance domains were selected for analysis from each clinical performance assessment tool. Professional behavior, evaluation, and procedural interventions were the domains analyzed from the PTCPI. Professional ethics item one, evaluation item one, and interventions item two were analyzed from the CIET. The domains were matched by descriptors and represented affective, cognitive, and psychomotor domains for each clinical performance assessment tool.

A two-factor, repeated measures design was used to analyze clinical performance data. Factor one represented the performance evaluations time (midterm and final). Factor two represented the assessor group (student and clinical instructor). Paired *t* tests were performed on each data set to compare midterm and final ratings by assessor groups and between assessor groups. A within factors design paired each student with four ratings: self-midterm and final and clinical instructor midterm and final. Because of this design, a more rigorous significance level was applied ($p \leq 0.0125$ as opposed to $p \leq 0.05$).

PTCPI (control group). The effect of time (midterm and final) was significant for the domains of professional behavior, evaluation, and procedural interventions ($p = 0.000$). Students rated themselves significantly higher for professional behavior ($p = 0.001$), evaluation ($p = 0.000$), and interventions ($p = 0.000$) from midterm to final. Clinical instructors rated the students significantly higher for each of the three performance domains from midterm to final ($p = 0.000$).

At midterm, there were no significant differences between student and clinical instructor ratings for professional behavior ($p = 0.127$). At final, students rated themselves significantly lower than their clinical instructors for professional behavior ($p = 0.000$). At midterm there were no significant differences between student and clinical instructor ratings for Evaluation ($p = 0.055$). At final, students rated themselves significantly lower than their clinical instructors for Evaluation ($p = 0.000$). At midterm, there were no significant differences between student and clinical instructor ratings for procedural interventions ($p = 0.617$). At final, students rated themselves significantly lower that their clinical instructors for procedural interventions ($p = 0.000$).

CIET (experimental group). The effect of time (midterm and final) was not significant for the domain of professional ethics ($p = 0.059$). The effect of time (midterm and final) was significant for the domains of evaluation ($p = 0.000$) and interventions ($p = 0.000$). Students rated themselves significantly higher for each of the three performance domains from midterm to final: professional ethics ($p = 0.005$), evaluation ($p = 0.003$), and interventions ($p = 0.002$). There were no significant differences in clinical instructor ratings for professional ethics from midterm to final ($p = 0.663$). Clinical instructors rated the students significantly higher for evaluation ($p = 0.000$) and interventions ($p = 0.000$) from midterm to final.

At both midterm and final, there were no significant differences between student and clinical instructor ratings for professional ethics ($p = 1.000$ and $p = 0.159$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for Evaluation ($p = 0.322$ and $p = 0.261$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for interventions ($p = 0.057$ and $p = 0.485$).

Detailed Analysis

PTCPI (control group). A two-factor, repeated measures ANOVA demonstrated a significant effect of time (midterm and final) for professional behavior ($p = 0.000$). A paired *t*-test was applied to compare midterm and final ratings for professional behavior by group. Students and clinical instructors rated professional behavior significantly higher from midterm to final ($p = 0.000$). At midterm, there were no significant differences between student and clinical instructor ratings for professional behavior ($p = 0.127$). At final, students rated themselves significantly lower than their clinical instructors for professional behavior ($p = 0.000$). Results for professional behavior are summarized in Table 4 and Figure 2.

Table 4

PTCPI: Professional Behavior

Time	Group	Mean Rating	Std. Deviation
Midterm	Student	3.904	1.774
Midterm	Clinical Instructor	4.423	2.199
Final	Student	5.019	1.799
Final	Clinical Instructor	5.981	1.852

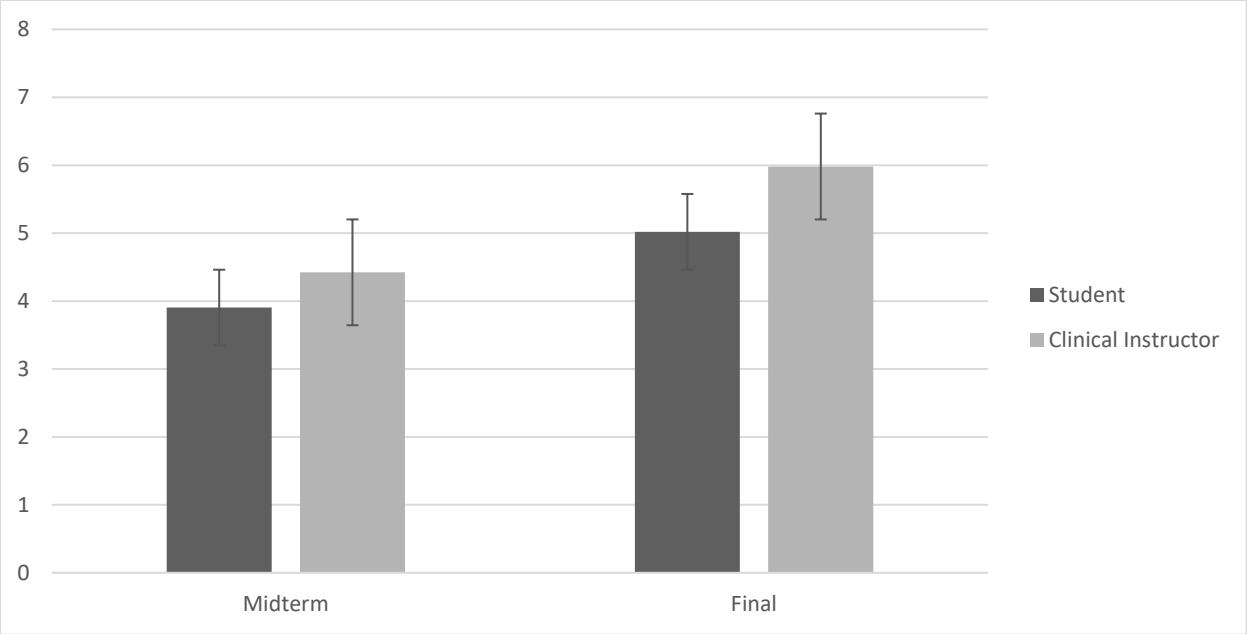


Figure 2. PTCPI: Professional behavior.

A two-way, repeated measures ANOVA demonstrated a significant effect of time (midterm and final) for Evaluation ($p = 0.000$). A paired *t* test was applied to compare midterm and final ratings for Evaluation by group. Students and clinical instructors rated Evaluation significantly higher from midterm to final ($p = 0.000$). At midterm there were no significant differences between student and clinical instructor ratings for Evaluation ($p = 0.055$). At final, students rated themselves significantly lower than their clinical instructors for Evaluation ($p = 0.000$). Results for Evaluation are summarized in Table 5 and Figure 3.

Table 5

PTCPI: Evaluation

Time	Group	Mean	Std. Deviation
Midterm	Student	2.327	1.080
Midterm	Clinical Instructor	2.769	1.352
Final	Student	3.596	1.089
Final	Clinical Instructor	4.962	1.521

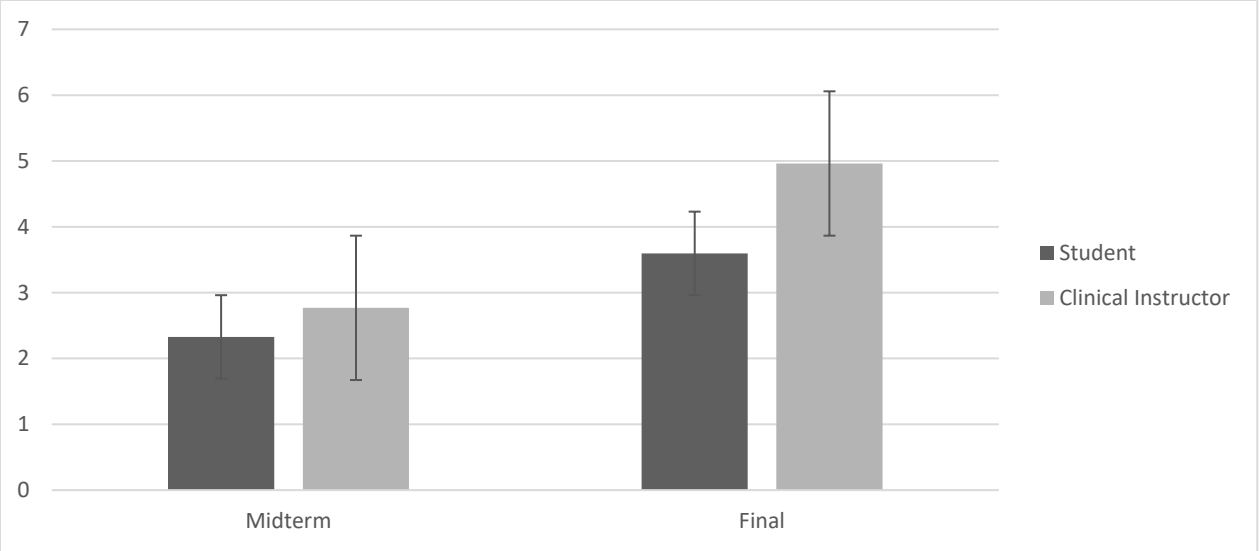


Figure 3. PTCPI: Evaluation.

A two-way, repeated measures ANOVA demonstrated a significant effect of time (midterm and final) for procedural interventions ($p = 0.000$). A paired *t* test was applied to compare midterm and final ratings for procedural interventions by group. Students and clinical instructors rated procedural interventions significantly higher from midterm to final ($p = 0.000$). At midterm, there were no significant differences between student and clinical instructor ratings for procedural interventions ($p = 0.617$). At final, students rated themselves significantly lower that their clinical instructors for procedural interventions ($p = 0.000$). Results for procedural interventions are summarized in Table 6 and Figure 4.

Table 6

PTCPI: Procedural interventions

Time	Group	Mean	Std. Deviation
Midterm	Student	2.846	1.091
Midterm	Clinical Instructor	2.962	1.468
Final	Student	4.212	1.377
Final	Clinical Instructor	5.269	1.523

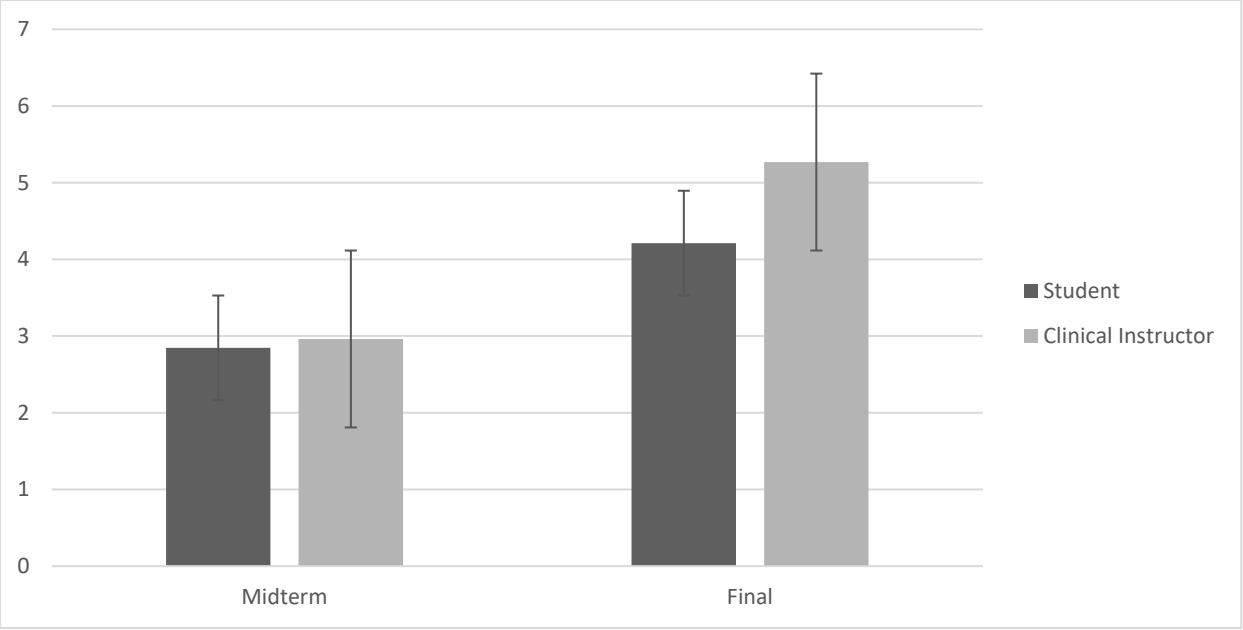


Figure 4. PTCPI: Procedural interventions.

CIET (experimental group). A two-way, repeated measures ANOVA did not demonstrate a significant effect of time (midterm and final) for professional ethics ($p = 0.059$). A paired *t* test was applied to compare midterm and final ratings for each performance dimension by group. Students rated their professional ethics significantly higher from midterm to final ($p = 0.005$). Clinical instructors did not rate professional ethics significantly higher from midterm to final ($p = 0.663$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for Professional ethics ($p = 1.000$ and $p = 0.159$). Results for professional ethics are summarized in Table 7 and Figure 5.

Table 7

CIET: Professional ethics

Time	Group	Mean	Std. Deviation
Midterm	Student	4.882	0.325
Midterm	Clinical Instructor	4.882	0.475
Final	Student	4.980	0.140
Final	Clinical Instructor	4.941	0.238

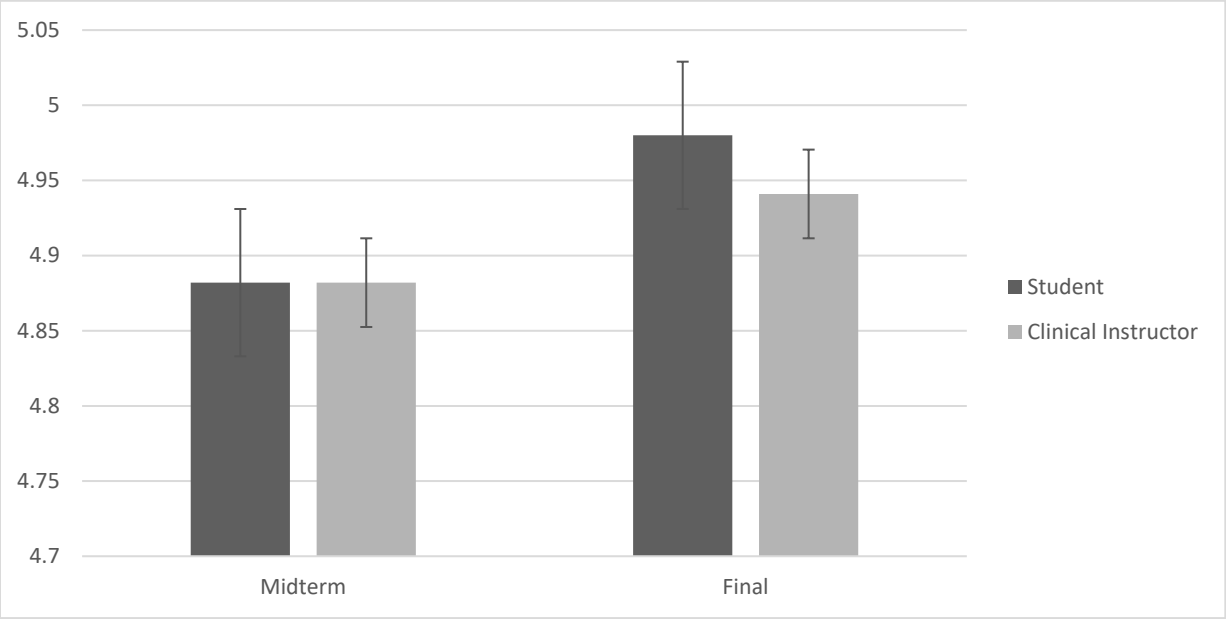


Figure 5. CIET: Professional ethics.

A two-way, repeated measures ANOVA demonstrated a significant effect of time (midterm and final) for evaluation ($p = 0.000$). A paired *t* test was applied to compare midterm and final ratings for each performance dimension by group. Students and clinical instructors rated evaluation significantly higher from midterm to final ($p = 0.003$ and $p = 0.000$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for evaluation ($p = 0.322$ and $p = 0.261$). Results for evaluation are summarized in Table 8 and Figure 6.

Table 8

CIET: Evaluation

Time	Group	Mean	Std. Deviation
Midterm	Student	2.471	0.644
Midterm	Clinical Instructor	2.608	0.695
Final	Student	2.882	0.588
Final	Clinical Instructor	3.000	0.721

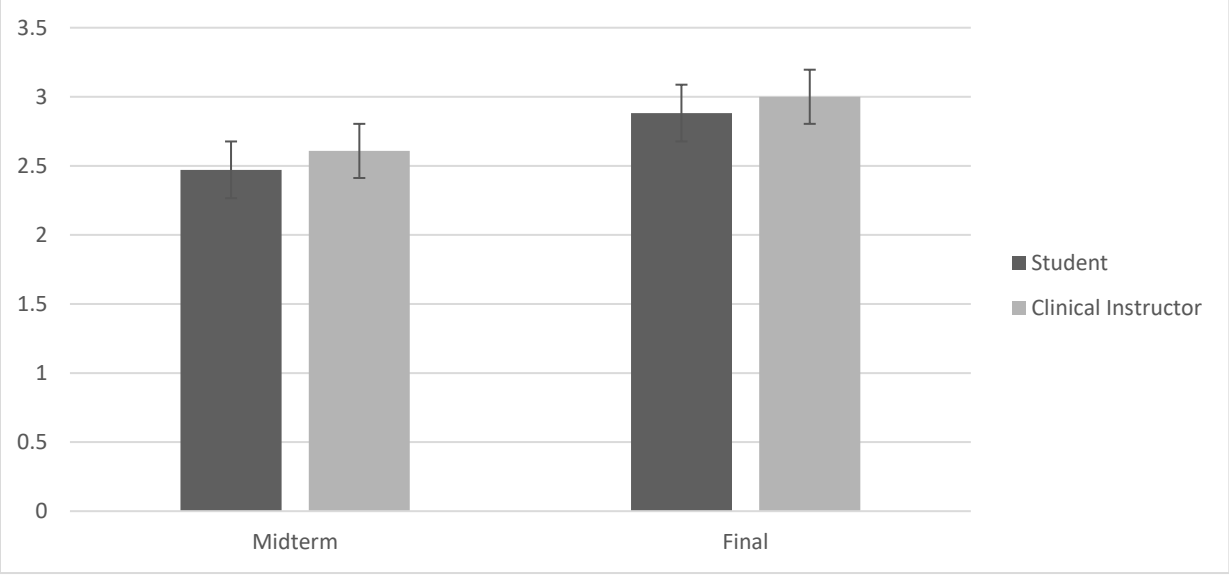


Figure 6. CIET: Evaluation.

A two-way, repeated measures ANOVA demonstrated a significant effect of time (midterm and final) for interventions ($p = 0.000$). A paired *t* test was applied to compare midterm and final ratings for each performance dimension by group. Students and clinical instructors rated interventions significantly higher from midterm to final ($p = 0.002$ and $p = 0.000$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for professional ethics ($p = 1.000$ and $p = 0.159$). At both midterm and final, there were no significant differences between student and clinical instructor ratings for interventions ($p = 0.057$ and $p = 0.485$). Results for intervention are summarized in Table 9 and Figure 7.

Table 9

CIET: Interventions

Time	Group	Mean	Std. Deviation
Midterm	Student	2.667	0.683
Midterm	Clinical Instructor	2.921	0. 796
Final	Student	3.255	0.796
Final	Clinical Instructor	3.333	0.840

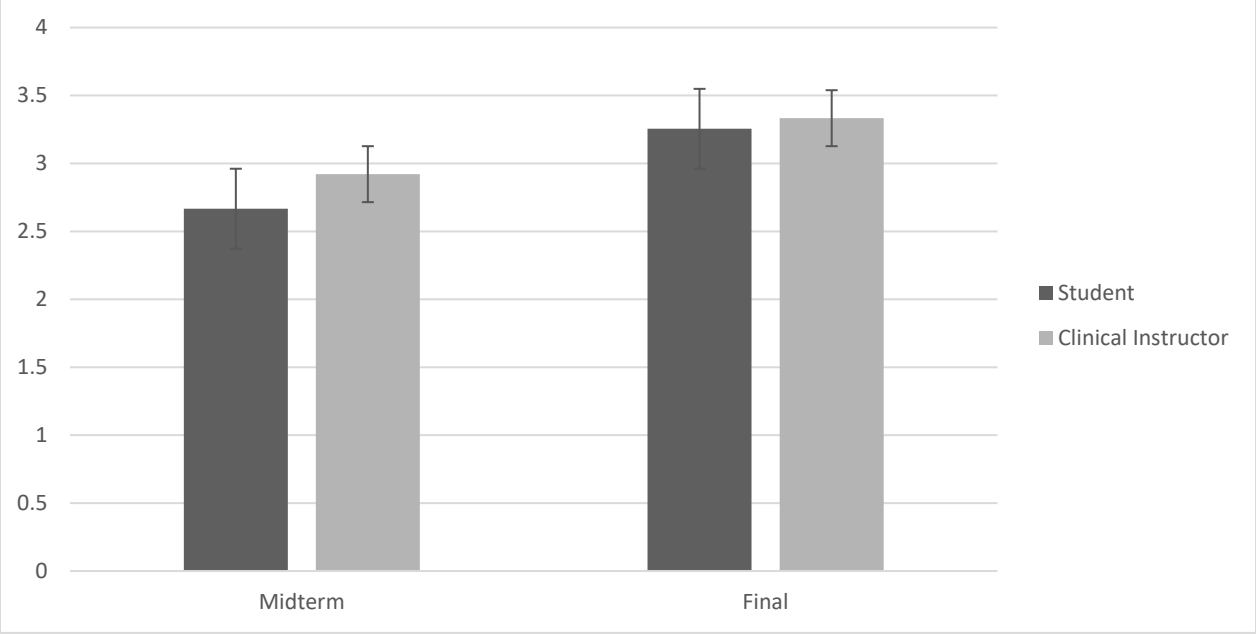


Figure 7. CIET: Interventions.

Chapter 4 Summary

This quasi-experimental, quantitative research model analyzed the perceived extent of DPT student clinical performance assessment accuracy as measured by congruence with clinical instructor ratings using two different clinical performance assessment tools, the PTCPI and the CIET. In addition, the ability of each tool to measure student clinical performance over time was included. The researcher used Excel and SPSS statistical software to complete the analyses.

For the control group, no significant differences were seen between student and clinical instructor ratings at midterm. Significant differences were found, however, between student and clinical instructor ratings in the final evaluations for professional behavior, evaluation, and procedural interventions. For the experimental group, there were no significant differences between student and clinical instructor ratings for professional ethics, evaluation, or interventions at either the midterm or final timeframes. Therefore, the null hypothesis is rejected. DPT student self-assessment accuracy of clinical performance was accurate as measured by congruence with clinical instructor assessment in all but the final performance domain evaluations using the PTCPI.

Significant differences were found between student and clinical instructor final performance ratings for professional behavior, evaluation, and procedural interventions in the control group ($p = 0.000$). Student and clinical instructor performance ratings in the experimental group did not differ significantly at midterm or final for professional ethics, evaluation, or interventions. Therefore, the null hypothesis is rejected, and the alternative hypothesis accepted. There was a significant difference in the perceived accuracy of student self-assessment between the PTCPI and CIET as measured by student and clinical instructor rating congruence.

In all but one case, significant differences in clinical performance ratings were observed from midterm to final. In the control group, both students and clinical instructors rated performance significantly higher from midterm to final for professional behavior, evaluation, and interventions. In the experimental group, both students and clinical instructors rated performance significantly higher from midterm to final for evaluation and interventions. Both the PTCPI and the CIET effectively measured student clinical performance over time.

Chapter 5: Discussion and Conclusion

Introduction

Clinical education accounts for approximately one third of DPT educational programs. Accurate assessment of clinical performance is therefore critical in the preparation of competent physical therapists. DPT students are required to formally self-assess clinical performance at the midpoint and end of full-time clinical education experiences. At the same time points, DPT students are also assessed by their assigned clinical instructors. These self and clinical instructor performance assessments are one component used to determine successful completion of a clinical education experience. DPT clinical education experiences are coordinated by core DPT program faculty who ultimately assign grades based on the clinical performance assessments and other data sources. Understanding the accuracy of DPT student assessment while on clinical education experiences is important to inform the development of optimal clinical performance assessment tools, to facilitate DPT student clinical skill development, and to ensure accurate grading.

The results of this study may be utilized by DPT program chairs and Directors of Clinical Education to inform both didactic and clinical education elements of the curriculum. Regarding the didactic curriculum, opportunities for additional content specific to effective self-assessment may be identified. Regarding the clinical education curriculum, this work will inform selection of an optimal clinical performance assessment tool. In addition, this study will contribute to the ongoing work of the National Consortium of Clinical Educators working to improve clinical performance assessment. DPT students and clinical instructors will also benefit from insights gained on the accuracy of DPT student self-assessment accuracy and its relationship to the clinical performance assessment tool utilized.

The objective of this dissertation was to investigate the perceived accuracy of DPT student self-assessment of clinical performance as measured congruence with clinical instructor ratings and by two assessment tools. In addition, this dissertation examined the effectiveness of two clinical performance assessment tools in measuring DPT student clinical performance over time. This study compared the congruence of student and clinical instructor ratings at midterm and final during a 10-week, full-time clinical education experience. Rating congruence was compared using both the PTCPI and the CIET. Assessment ratings for specific affective, cognitive, and psychomotor performance domains were selected for analysis from each clinical performance assessment tool.

DPT students participate in formal self-assessment and reflection on clinical performance. DPT students complete self-assessment of clinical performance using selected clinical performance assessment tools which provide performance criteria and the desired performance benchmarks. Identifying attributes of desired performance and using criteria to compare performance to the desired standard are key elements of self-assessment (Boud, 2003). The value of self-assessment in DPT education aligns with adult learning theory, which served as the conceptual framework for this dissertation (Malik, 2015; Knowles, 1984).

The American Physical Therapy Association recently formed an Education Leadership Partnership consisting of representation from the National Consortium of Clinical Educators, Academy of Physical Therapy Education, Clinical Education Special Interest Group, and Physical Therapist Assistant Special Interest Group. The Education Leadership Partnership is expressly tasked with identifying best practices for DPT Education. Optimizing clinical performance assessment is a goal of the partnership. This research contributes to the body of knowledge in physical therapy clinical performance assessment by examining student self-

assessment accuracy using two assessment tools. Chapter 5 offers a summary of the results and their correlations to current literature. Implications of the results for future practice, policy, and theory will be examined. Finally, limitations and recommendations for future research will conclude this dissertation.

Summary of the Results

This quasi-experimental, quantitative research study analyzed the perceived accuracy of DPT student clinical performance assessment as measured by congruence with clinical instructor ratings using two different clinical performance assessment tools, the PTCPI and the CIET. In addition, the ability of each tool to measure student clinical performance over time was included. The PTCPI was selected to represent the control. It was the current tool utilized by the DPT sample population in this study and the tool utilized by over 90% of DPT education programs in the United States. The CIET was selected for use by the experimental group. It was a new tool for the DPT sample population in this study and is less commonly utilized by DPT education programs in the United States. The CIET was identified as a validated alternative to the PTCPI and is designed for increased assessment efficiency.

The control group in this study was the cohort of 52 DPT students graduating in May 2020. The 2020 cohort utilized the PTCPI for clinical performance assessment during their first 10-week, full-time clinical education experience (Clinical Education Experience 1). The control group completed Clinical Education Experience 1 during the fall term of 2018. The clinical performance assessment data were analyzed retrospectively. The experimental group was the cohort of 51 DPT students graduating in May 2021. The 2021 cohort utilized the CIET during their first 10-week, full-time clinical education experience (Clinical Education Experience 1) which was completed during the fall term of 2019.

A two-factor, repeated measures design was used to analyze clinical performance data. Factor one represented the time of clinical performance assessment (midterm and final). Factor two represented the assessor group (student and clinical instructor). Paired *t* tests were performed on each data set to compare midterm and final ratings by assessor groups and between assessor groups. A within factors design paired each student with four ratings: self-midterm and final and clinical instructor midterm and final. Because of this design, a more rigorous significance level was applied ($p \leq 0.0125$ as opposed to $p \leq 0.05$). To help guide this research, the following questions were asked:

RQ 1. To what extent is DPT student self-assessment of clinical performance accurate as measured by congruence with clinical instructor assessment ratings?

RQ 2. To what extent does the accuracy of student self-assessment of clinical performance differ between the Physical Therapist Clinical Performance Instrument and the CIET as measured by student and clinical instructor rating congruence?

For the control group, no significant differences were seen between student and clinical instructor ratings at midterm. Significant differences were present, however, between student and clinical instructor ratings in the final evaluations for professional behavior, evaluation, and procedural interventions. For the experimental group, there were no significant differences between student and clinical instructor ratings for professional ethics, evaluation, or interventions at either the midterm or final timeframes. DPT student self-assessment accuracy of clinical performance was accurate as measured by congruence with clinical instructor assessment in all but the final evaluation using the PTCPI.

Significant differences were found between student and clinical instructor final performance ratings for professional behavior, evaluation, and procedural interventions in the

control group (p = 0.000). Student and clinical instructor performance ratings in the experimental group did not differ significantly at midterm or final for professional ethics, evaluation, or interventions. A difference was identified in the perceived accuracy of student self-assessment between the PTCPI and CIET. The CIET was associated with greater rating congruence.

In all but one case, significant differences in clinical performance ratings were observed from midterm to final. In the control group, both students and clinical instructors rated performance significantly higher from midterm to final for professional behavior, evaluation, and interventions. In the experimental group, both students and clinical instructors rated performance significantly higher from midterm to final for evaluation and interventions. Both the PTCPI and the CIET effectively measured student clinical performance over time, which is important as DPT students progress through clinical education experiences.

Discussion of the Results

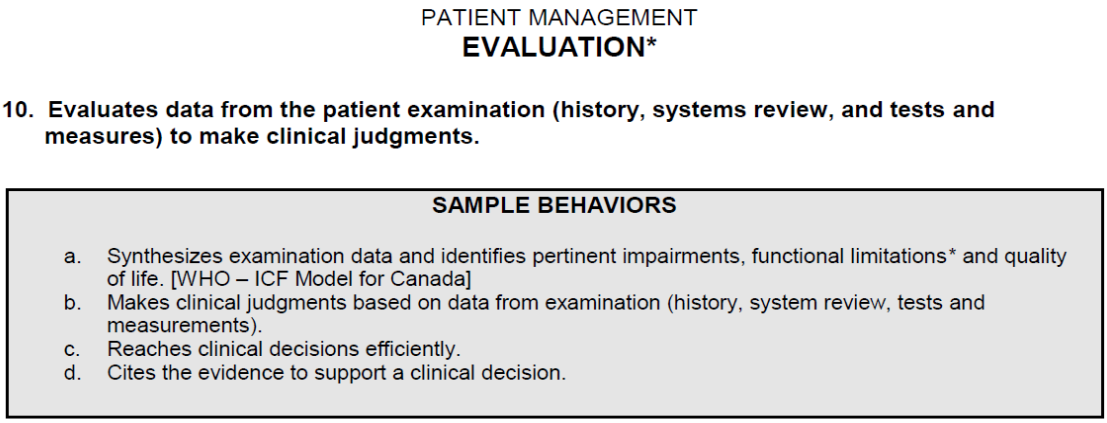
One purpose of this dissertation was to examine the accuracy of DPT student self-assessment of clinical performance. Studies on the self-assessment accuracy of health professions students suggest they do not rate themselves accurately when compared with faculty or expert assessments (Pawluk et al., 2018; Hadid, 2016). Similar studies have been completed in other countries with other clinical performance assessment tools, however, DPT student self-assessment accuracy had not been examined in the United States. Lo et al. (2015) examined the congruence of student self-assessment with clinical instructor assessment of clinical performance in Australian physical therapy education using the Assessment of Physiotherapy Practice and found physical therapy students underrated performance relative to their clinical instructor assessments.

The PTCPI is the clinical performance assessment tool used by over 90% of DPT programs in the United States. Results from this research indicate that DPT students were not able to accurately self-assess clinical performance using the PTCPI at the end of Clinical Education Experience 1. At the end of the clinical education experience, DPT students rated themselves significantly lower than their clinical instructors for professional behavior, evaluation, and procedural interventions ($p = 0.000$). This finding suggests DPT students' self-assessment of clinical performance is not accurate using the PTCPI as measured by congruence with clinical instructors' ratings.

A second purpose of this dissertation was to examine the effect of the assessment tool on DPT students' clinical performance self-assessment accuracy. Both the PTCPI and CIET are applicable to a broad range of clinical settings and can be used throughout the continuum of clinical education experiences. The PTCPI requires that performance domains be rated on student performance relative to entry-level. Specific to each clinical setting, it is necessary to identify how the sample behaviors would be demonstrated at entry-level by students. The CIET compares student performance to that of a competent clinician able to skillfully manage patients in an efficient manner while achieving an effective outcome. Authors of the tool identified this benchmark as more consistent when compared to the individualized definition of entry-level used in the PTCPI.

The PTCPI and CIET assess some of the same or similar performance domains. They differ, however, in the number of individual items, complexity of the rating scales, and assessment criteria. The PTCPIs assesses 18 performance domains including: safety, professional behavior, accountability, communication, cultural competence, professional development, clinical reasoning, screening, examination, evaluation, diagnosis and prognosis, plan of care,

procedural interventions, educational interventions, documentation, outcomes assessment, financial resources, and direction and supervision of personnel. Each domain includes a list of sample behaviors. Figure 8 shows the evaluation domain and its sample behaviors.



PATIENT MANAGEMENT

EVALUATION*

10. Evaluates data from the patient examination (history, systems review, and tests and measures) to make clinical judgments.

SAMPLE BEHAVIORS

a. Synthesizes examination data and identifies pertinent impairments, functional limitations* and quality of life. [WHO – ICF Model for Canada]
b. Makes clinical judgments based on data from examination (history, system review, tests and measurements).
c. Reaches clinical decisions efficiently.
d. Cites the evidence to support a clinical decision.

Figure 8. Evaluation domain, PTCPI.

Performance is rated on a continuum from beginner (1) to beyond entry-level (11). Each anchor rating is defined and considers patient complexity, supervision and guidance required, consistency, and efficiency.

The CIET measures professional behaviors by identifying four categories: safety, professional ethics, initiative, and communication. Four categories are also identified related to patient management: examination, evaluation, diagnosis/prognosis, and intervention. The professional behavior categories are measured on a 5-point rating scale based on the frequency with which they are demonstrated by the DPT student. Patient management items are rated on a 5-point scale in relation to the skills of a competent clinician. Under each category, sample behaviors are scored individually (Fitzgerald et al., 2007). Figure 9 shows the evaluation category.

EVALUATION					
1. Makes correct clinical decisions based on the data gathered in the examination (confirms/disconfirms initial and alternative hypotheses)					
2. Identifies impairments in body structure and function; activity limitations; and participation restrictions					
3. Administers further tests and measures as needed for appropriate clinical decision making					

Figure 9. Evaluation domain, CIET.

In this study, greater congruence was observed between student and clinical instructor clinical performance ratings in the experimental group using the CIET. No significant differences were found in the experimental group between student and clinical instructor ratings at midterm or final for professional ethics, evaluation, or intervention. Students in the control group using the PTCPI rated themselves significantly lower at the final evaluation than did their clinical instructors for professional behavior, evaluation, and procedural interventions (p = 0.000) while midterm and final ratings using the CIET were not significantly different between students and clinical instructors.

The variance in assessment tool design may have influenced the level of congruence between student and clinical instructor ratings. For example, when rating professional behavior using the PTCPI, 13 sample behaviors are provided ranging from “maintains productive working relationships with patients, families, CI, and others” to “exhibits caring, compassion, and empathy in providing services to patients” with caring, compassion, and empathy representing core values of the physical therapy profession. The entire domain with its sample behaviors is given one rating ranging from beginner to entry-level on a 10-anchor scale.

The corresponding professional ethics domain is structured differently on the CIET. Seven individual behaviors are listed ranging from “demonstrates positive regard for patients/peers during interactions” to “adheres to ethical and legal standards of practice,

including Practice Act and APTA Code of Ethics." Each listed behavior is rated separately on a 5-point scale based on the frequency with which the behavior is demonstrated. Once each behavior is rated, students and clinical instructors denote whether the performance level in each domain met the expected benchmark for the student's level within their DPT curriculum. The rating of behaviors individually within each domain may have contributed to the increased congruence between students and clinical instructors.

Similar assessment tool design differences exist for the evaluation performance domains. The PTCPI includes four sample behaviors under evaluation ranging from "makes clinical judgments based on data from examination" to "reaches clinical decisions efficiently." The corresponding evaluation domain is structured differently on the CIET. Three individual behaviors are listed ranging from "makes correct clinical decisions based on the data gathered in the examination" to "identifies impairments in body structure and function; activity limitations; and participation restrictions." Each listed behavior is rated separately on a 5-point scale in comparison to a competent clinician. The procedural interventions domain on the PTCPI contains 10 sample behaviors ranging from "performs interventions safely, effectively, efficiently, fluidly, and in a coordinated and technically competent manner" to "assesses patient response to interventions and adjusts accordingly." The intervention domain on the CIET contains eight individual items with examples including: "applies effective treatment using appropriate psychomotor skills" and "modifies intervention according to patient/client's response to treatment."

A third purpose of this dissertation was to determine whether the PTCPI and CIET could measure student clinical performance over time. Significant differences were seen between the midterm and final ratings by both students and clinical instructors using the PTCPI for

professional behavior, evaluation, and procedural interventions ($p = 0.000$). Significant differences were seen between the midterm and final ratings by students using the CIET for professional ethics ($p = 0.005$), evaluation ($p = 0.003$) and interventions ($p = 0.002$). Clinical instructors did not rate professional ethics significantly higher from midterm to final ($p = 0.663$). Significant differences were seen between midterm and final ratings by clinical instructors for evaluation ($p = 0.000$) and interventions ($p = 0.000$). These findings suggest that for the cognitive and psychomotor domains, both tools can measure significant changes in student clinical performance over time.

In only one case, clinical instructor assessment of professional ethics, no significant change in performance was noted from midterm to final. This finding may relate to the rating scale applied to professional ethics in the Clinical Instructor Evaluation Tool. The 5-point scale rates the frequency with which a student "demonstrates positive regard for patients/peers during interactions." A DPT student may be likely to display professional ethics "most of the time" or "always" regardless of the clinical education experience or the assessment timeframe (midterm or final).

Discussion of the Results in Relation to the Literature

The most common method of measuring student self-assessment accuracy found in the literature was comparison with expert assessment (Oh, Liberman, & Mishler, 2018; Lo et al. 2016). This study compared student self-ratings with those of their clinical instructors (experts) and examined this congruence between two different clinical performance assessment tools validated for DPT education. Clinical performance assessment tool was selected as an independent variable the structure and design of clinical performance assessment tools can influence congruence between student and expert assessments. Clinical performance assessment

tools with a higher number of individual items to assess have lower assessment inter-rater reliability (Tavares & Eva, 2014). Muhamad et al. (2016) identified the use of a more global rating scale as preferred to a more complex assessment tool. This recommendation is consistent with the finding that experts tend to assess more holistically (Byrne et al., 2014; Klamen et al., 2016) and is an important consideration in this study examining student self-assessment accuracy between two different tools, the PTCPI and the CIET.

Because demographic factors can influence self-assessment, it was important to compare demographics of the control and experimental groups. Specifically, age and gender were compared. Age is negatively correlated with self-rating. Older health professions students self-assess their performance more critically than younger students (Hadid, 2016). Gender also influences self-assessment. Female health professions students underrate performance in comparison with their male counterparts (Madrazo, Lee, McConnell, & Khamisa, 2018). The findings related to gender are consistent with those from other science, technology, engineering, and math (STEM) fields. The control and experimental groups were homogenous in age, however, varied in percentage of female subjects. The control group was 75% female, while the experimental group was 52% female.

Academic proficiency is an additional influence on self-assessment accuracy. Lower performing students commonly overrate their performance level when compared with their expert evaluators (Panadero et al., 2016). Higher performing students often underrate their level of clinical competence when compared with their expert evaluators (Oh et al., 2018). It was important to compare matriculation GPA between the control and experimental groups as a representation of academic proficiency. The control and experimental groups were homogenous for matriculation GPA.

Student self-assessment may differ from expert assessment due to lack of confidence or experience (Greenfield, Bridges, Carter, Barefoot, Dobson, Eldridge, & Phillips, 2017). This may be especially influential during early health professions educational experiences which represent the first opportunity students have to compare their clinical skills performance against established standards. In this study, the subjects in both groups were relatively inexperienced in formal clinical skill self-assessment.

Accurate self-assessment is considered an important skill for students entering health professions, including DPT students, allowing them to "determine their level of knowledge and identify knowledge gaps to remain current and safe in practice" (Hadid, 2016, p. 70). DPT students often complete their first formal self-assessment during clinical education experiences. In this study, Clinical Education Experience I indeed represented the first comprehensive self-assessment opportunity for the students in both the control and experimental groups.

Strategies have been suggested in the literature to improve student self-assessment accuracy. Examples include providing opportunities to practice self-assessment, increasing familiarity with performance criteria, involving students in criteria development, providing opportunities to compare assessment with experts, and emphasizing reflective narrative (Falender, & Shafranske, 2017; Oh et al., 2018). Greenfield et al. (2017) recommended the use of a framework to guide student self-assessment, specifically when qualitative reflection is required. Huhn (2017) saw improvements in Self Reflection and Insight Scale scores after students completed a specific clinical reasoning course. The findings of Greenfield et al. (2017) and Huhn (2017) support the value of structured education on and practice in self-assessment in health professions and DPT curricula.

Within the didactic terms prior to Clinical Education Experience 1, DPT students received content on the value of and best practices for self-assessment. Examples of best practices discussed included focusing on observable behavior and supporting assessment with specific examples. DPT students were practiced formative self-assessment on professional behaviors, but did not practice a comprehensive self-assessment of clinical skills. During lab experiences, DPT students were encouraged to self-assess their performance of discrete skills. Following practical examinations, DPT students were required to view the video recording of the examination and self-assess their performance during the simulated patient encounter.

The results of this dissertation were not entirely consistent with the findings of the study by Lo et al. (2015) which examined the congruence of student self-assessment with clinical instructor assessment of clinical performance in Australian physical therapy education using the Assessment of Physiotherapy Practice. Lo et al. (2015) found significant differences between student and clinical instructor performance ratings at both the midpoint and end of clinical experiences, with students rating themselves lower than their clinical instructors. In this study, DPT students significantly underrated their performance in comparison to their clinical instructors' ratings in the control group, but only at the final assessment time point. Midterm ratings were congruent between students and clinical instructors using the PTCPI. Both midterm and final assessment ratings were congruent between students and clinical instructors using the CIET.

The difference in findings may be attributed to the design of the selected clinical performance assessment tool. The Assessment of Physiotherapy Practice in the study by Lo et al. (2015) shares characteristics with both the PTCPI and CIET used in this study. The Assessment of Physiotherapy Practice assesses 20 performance domains, similar to those in the PTCPI

(control group tool). It rates performance using a 5-point rating system, similar to the CIET (experimental group tool).

The difference in findings may also relate to the sample. The study by Lo et al. (2015) included a sample of 101 undergraduate physiotherapy students. The students were completing terminal clinical education experiences within the undergraduate curriculum. In this study, DPT students were completing their first clinical education experience in a graduate curriculum. The sample in this dissertation were older and had completed undergraduate training prior to entering the DPT program.

Age and academic experience correlate positively with self-assessment accuracy (Panadero et al., 2016; Oh et al., 2018). Gender also influences self-assessment accuracy. While the control and experimental groups in this dissertation were homogenous in terms of average age and matriculation GPA, a notable difference in the percentage of female students was present between groups. The control group was 75% females and the experimental group 52%. Female health professions students tend to underrate performance when compared to their male counterparts (Madrazo et al., 2018). The larger percentage of female students in the control group may have contribute to the lower self-ratings at final as compared to clinical instructor ratings.

Student self-assessment may also differ from expert assessment due to lack of confidence or experience (Greenfield et al., 2017). This may be especially influential during early health professions educational experiences which represent the first opportunity students have to compare their clinical skills performance against established standards. Based on this hypothesis, it would be expected that student and clinical instructor ratings would become more congruent from midterm to final. This was not observed in this dissertation. In the control group (using the

PTCPI), the variance between student and clinical instructor ratings actually increased from midterm to final. In the experimental group (using the CIET), variance between student and clinical instructor ratings remained constant and was not significant at either time point.

Limitations

This dissertation was limited by the assumption that clinical instructors assess students accurately as experts in physical therapy practice. This assumption is supported in the literature, and congruence with expert rating is a commonly used measure of student self-assessment accuracy (Ganni et al., 2017; Hadid, 2016; Lo et al., 2016; Adrian et al., 2015; Kachingwe et al., 2015). This assumption is further influenced by the range in expertise of clinical instructors measured by years in practice, advanced practice certifications, or years serving as a clinical instructor. The years of practice and clinical instruction experience were homogenous in the study between the control and experimental groups. Clinical instructors for the control group had practiced physical therapy for an average of 10.872 years and served as clinical instructors for an average of 7.717 years. Clinical instructors for the experimental group had practiced physical therapy for an average of 12.407 years and served as clinical instructors for an average of 9.017 years. While out of the scope of this study, methods to assess clinical instructor assessment accuracy include comparing narrative comments to benchmark ratings or assigning multiple clinical instructors to evaluate the same student.

A second limitation was the assumption is that clinical instructors and students receive sufficient training in self-assessment. For the PTCPI clinical instructors and students complete a free online training module and assessment which requires approximately two hours to complete. Training for the CIET is less standardized. A 10-minute video is available from the authors of the tool; however, no assessment is required. Within DPT curricula, students are provided with

varying amounts of self-assessment content and practice. Examples include self-assessment of practical examination performance, self-assessment of professional behaviors, and self-assessment of selected skills such as patient interviewing or patient education.

A third limitation of this study was that many of the clinical instructors assessing students in the experimental group had prior experience using the PTCPI as it is the most commonly utilized clinical performance assessment tool in DPT education in the United States (Fitzgerald et al., 2007). It is likely many clinical instructors in the experimental group were familiar with the PTCPI, but did not have prior experience using the CIET.

Two delimitations are present in this study design. The first is that Clinical Education Experience 1 represents the first full-time clinical education within the curriculum. Clinical Education 1 is the first opportunity for students to formally self-assess their clinical performance. By choosing to compare self-assessment accuracy between clinical performance assessment tools during Clinical Education Experience 1, neither the control nor experimental groups had prior experience with either clinical performance assessment tool.

A second delimitation is the use of a convenience sample. This study was conducted using two cohorts of DPT students at a single, private, graduate health sciences university. This sample, however, is representative of DPT students in the United States. Physical Therapist Centralized Application Service data indicates the mean age of DPT students in the United States is 23.57 and the mean percentage of female students is 61.4%. The average undergraduate grade point average of students accepted into DPT programs in the United States is 3.57 (American Physical Therapy Association, 2019). Convenience sample data and national data are summarized in Table 10.

Table 10

Sample and National DPT Student Demographics

	Control Group	Experimental Group	US DPT Students
Number of students	52	51	NA
Age at matriculation	23	22	24
Gender (female %)	75	52	61
Mean GPA at matriculation	3.68	3.67	3.57

Implication of the Results for Practice, Policy, and Theory

The results of this dissertation inform practice in the area of DPT clinical performance assessment. The results are especially significant to the work being conducted by the Education Leadership Partnership of the American Physical Therapy Association. This partnership, made up of representatives from the National Consortium of Clinical Educators, Academy of Physical Therapy Education, Clinical Education Special Interest Group, and Physical Therapist Assistant Special Interest Group are focused on identifying best practices for DPT clinical education and optimal clinical performance assessment tools and methods.

The findings of this study suggest the PTCPI and CIET can detect growth in student performance between midterm and final assessments within a clinical education experience. For all performance domains analyzed in both the control and experimental groups, significant improvements in clinical performance were detected from midterm to final, except for the professional ethics domain included in the CIET. It is hypothesized that affective performance domains may develop more quickly, while cognitive and psychomotor skills progress in a more linear manner over the course of clinical education experiences.

Survey results presented by the National Consortium of Clinical Educators reveal approximately 50% of DPT education programs are considering a change in clinical performance

assessment tools. Recommendations are proposed to revise the control tool (PTCPI) or to consider an alternate tool such as the CIET. DPT education programs are seeking assessment tools which are more efficient to complete and applicable to emerging models of physical therapy clinical education (National Consortium of Clinical Educators, 2019). It will be valuable to align clinical performance assessment tools as education models emerge. For example, competency based education would be better supported by the CIET as its rating structure compares student performance to that of a competent physical therapist.

The CIET demonstrates a higher level of congruence between student and clinical instructor clinical performance ratings than the PTCPI. This congruence may indicate a higher level of associated assessment accuracy. The design of clinical performance assessment tools is suggested to influence interrater reliability. According to the literature, simpler assessment tools with fewer discrete performance domains are associated with improved interrater congruence. DPT students develop clinical knowledge, skills, and attitudes through self and clinical instructor assessment of their clinical performance. The accuracy of DPT student self-assessment had not been clearly established in the United States. This research contributes to the body of knowledge specific to DPT student assessment accuracy in the United States and highlights the important interplay between assessment tool design and assessment accuracy.

Recommendations for Further Research

The accuracy of clinical instructors' assessment of student clinical performance is an area for further research. The assumption that clinical instructors are proficient in DPT student clinical performance assessment was identified as a limitation in this study. Experts tend to assess student performance based on global or holistic impressions as opposed to distinct, individual performance attributes (Byrne et al., 2014; Klamen et al., 2016). Clinical instructors in

DPT education may feel personal responsibility for student success on clinical education experiences. In physical therapy education, clinical faculty may overrate student performance because they are hesitant to fail the student. Although clinical instructors do not officially assign grades, they may perceive that a below benchmark rating will jeopardize student progression through a DPT curriculum. Clinical instructors may also make inferences regarding student performance versus strictly assessing objective behavior (Lo et al., 2015). They may assume that a student arrived at a clinical conclusion based on a similar strategy of reasoning to their own. Clinical instructors are encouraged to ask students to articulate clinical reasoning to minimize the likelihood of false inferences.

While experts may demonstrate high inter-rater reliability classifying exceptional or inadequate performance, less agreement exists on what constitutes adequate performance. Clinical instructor evaluations of student performance may also be impacted by variance in the weighting of specific performance elements. Clinical instructors may also evaluate student performance based on inferences versus observed behavior (Byrne et al., 2014; Klamen et al., 2016). Accuracy of clinical instructor ratings may further be influenced by the range in the expertise of clinical instructors which may be measured by years in practice, advanced practice certifications, or years serving as a clinical instructor. Studies of clinical instructor assessment accuracy may include comparing narrative comments to benchmark ratings or assigning multiple clinical instructors to evaluate the same student.

Repeating this study design across multiple DPT education programs is also recommended to increase sample size and diversify sample representation. This research was conducted using two cohorts at a single site. The cohorts were relatively matched by age and

matriculation GPA, however, differed in gender representation. Increasing the sample size and using intentional demographic matching is recommended.

Accuracy of student self-assessment is thought to increase with practice and experience. Further research is recommended following DPT student clinical performance and self-assessment accuracy through multiple clinical education experiences. Finally, additional research on the influence of assessment tool design on rater congruence is recommended. Similar efforts are underway in other countries including Australia and Canada (Mori et al., 2015). It is important for the Education Leadership Partnership of the American Physical Therapy Association and other researchers to identify best practices for DPT clinical performance assessment tool design.

Conclusion

Previous studies identified a variety of factors influencing student self-assessment accuracy including assessment tool design and methodology, the timing of self-assessment in the learning process, the articulated purpose of self-assessment, and the amount and quality of self-assessment training provided. The first objective of this dissertation was to investigate the perceived accuracy of DPT student self-assessment accuracy of clinical performance, as this has not been widely studied in the United States. DPT students in the control group were accurate in self-rating at midterm. DPT students in the experimental group were accurate in self-rating at midterm and final.

As a second objective, this dissertation compared the influence of clinical performance assessment tool design on DPT student self-assessment accuracy. This study compared the congruence of student and clinical instructor ratings between the PTCPI and the CIET for selected affective, cognitive, and psychomotor performance domains. Improved congruence

between student and clinical instructor ratings was identified in the experimental group using the CIET.

The factor of assessment timing was also examined in this study to determine the influence of assessment experience and the ability of two clinical performance assessment tools to measure physical therapy students' clinical performance over time. DPT student self-assessment accuracy was measured at both midterm and final time points during Clinical Education Experience I. Self-assessment accuracy did not consistently improve from midterm to final. Both clinical performance assessment tools appeared to measure students' growth in performance over time.

The results of this study may be utilized by DPT program chairs and Directors of Clinical Education to inform curriculum. In addition, this study will contribute to the ongoing work of the National Consortium of Clinical Educators in identifying optimal clinical performance assessment tools. In the control group using the PTCPI, no significant differences were seen between student and clinical instructor ratings at midterm. Significant differences between student and clinical instructor ratings were present, however, in the final evaluations for professional behavior, evaluation, and procedural interventions. This finding suggests there are opportunities to improve student self-assessment accuracy in DPT education in the United States.

In the experimental group using the CIET, there were no significant differences between student and clinical instructor ratings for professional ethics, evaluation, or interventions at either the midterm or final timeframes demonstrating a higher level of congruence. The design of the CIET utilized a 5-point rating scale as opposed to the 10-point scale in the PTCPI. Criteria for assigning ratings are less complex in the CIET. Additionally, behaviors are rated individually

within each performance domain. These design elements of the CIET may have contributed to increased student self-assessment accuracy.

In all but one case, significant differences in clinical performance ratings were observed from midterm to final. In the control group, both students and clinical instructors rated performance significantly higher from midterm to final for professional behavior, evaluation, and interventions. In the experimental group, both students and clinical instructors rated performance significantly higher from midterm to final for evaluation and interventions. Both the PTCPI and CIET appear to effectively measure changes in student clinical performance over time.

References

Adrian, J., Zeszotarski, P., Ma, C. (2015). Developing pharmacy student communication skills through role-playing and active learning. *American Journal of Pharmaceutical Education*, *79*(3), 1–8. doi: 10.5688/ajpe79344

American Council of Academic Physical Therapy. (2017) *Common Terminology Glossary for Physical Therapist Clinical Education.* Retrieved from https://acapt.org/docs/default-source/pdfs/common-terminology-glossary-for-physical-therapist-clinical-education.pdf?sfvrsn=25978bd8_2

American Physical Therapy Association. (2019). *Physical therapist centralized application service 2017-2018 applicant data report.* Retrieved from http://www.ptcas.org/uploadedFiles/PTCASorg/About_PTCAS/PTCASApplicantDataRpt.pdf

American Physical Therapy Association. (2017). *Best practices for physical therapist clinical education task force report.* Retrieved from http://www.apta.org/Educators/TaskForceReport/PTClinicalEducation/

Armstrong, K. J., & Jarriel, A. J. (2016). Standardized patients provide a reliable assessment of athletic training students' clinical skills. *Athletic Training Education Journal*, *11*(2), 88–94. doi: 10.4085/110288

Boscardin, C., Fergus, K., Hellevig, B., & Hauer, K. (2018). Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Medical Teacher*, *40*(8), 855–861. doi: 10.1080/0412159X.2017.1396306

Boud, D. (2003). *Enhancing learning through self-assessment*. New York, NY: Routledge/Falmer.

Byrne, A., Tweed, N., & Halligan, C. (2014). A pilot study of the mental workload of objective structure clinical examination examiners. *Medical Education, 48,* 262–267. doi: 10.1111/medu.12387

Commission on Accreditation in Physical Therapy Education. (2018). Retrieved from http://www.capteonline.org/home.aspx

Dudek, N. & Dojeiji, S. (2014). Twelve tips for completing quality in-training evaluation reports. *Medical Teacher, 36,* 1038–1052 .doi: 10.3109/0142159X.2014.932897

Falender, C. & Shafranske, E. (2017). Competency-based clinical supervision: Status, opportunities, tensions, and the future. *Australian Psychologist, 52,* 86–93. doi:10.1111/ap.122655

Fitzgerald, L., Delitto, A., & Irrgang, J. (2007) Validation of the clinical internship evaluation tool. *Physical Therapy, 87*(7), 855–860 . doi.org/10.2522/ptj.20060054

Ganni, S., Botden, S., Schaap, D., Verhoeven, B Goossens, R, and Jakimowicz, J. (2017). "Reflection-before-practice" improves self-assessment and end-performance in laparoscopic surgical skills training. *Journal of Surgical Education, 75*(2), 527–533. doi: 10.1016/j.jsurg.2017.07.030

Gingerich, A., Ramlo, S., Cees, P., van der Vleuten, K., Eva, K., & Regehr, G. (2017). Inter-rater variability as mutual disagreement: Identifying raters' divergent points of view. *Adv in Health Sci Educ, 22*, 819–838 . doi:10.1007/s10459-016-9711-8

Greenfield, B., Bridges, P., Carter, V., Barefoot, T., Dobson, A., Eldridge, E., & Phillips, T. (2017). Reflective narratives by student physical therapists on their early clinical

experiences: An inductive and deductive analysis-the use of a reflective framework. *Journal of Physical Therapy Education, 31*(4), 49–58 . Retrieved from https://journals.lww.com/jopte/toc/2017/31040#415664911

Hadid, S. (2016). Factors influencing nursing student self-assessment in relation to instructor assessment. *Journal of Nurs Educ, 56*(2), 70–76. doi:10.3928/01484834-20170123-03

Huhn K. (2017). Effectiveness of a clinical reasoning course on willingness to think critically and skills of self-reflection. *Journal of Physical Therapy Education, 31*(4), 59–63. doi: 10.1097/JTE.0000000000000007

Ibrahim, J., MacPhail, A., Chadwick, L. & Jeffcott, S. (2014). Interns' perceptions of performance feedback. *Medical Education, 48*, 417–429 . doi: 10.1111/medu.12381

Kachingwe, A., Phillips, B., & Beling, J. (2015). Videotaping practical examinations in physical therapist education: Does it foster student performance, self-assessment, professionalism, and improve instructor grading? *Journal of Physical Therapy Education, 29* (1), 25–33. doi:10.1097/00001416-201529010-00006

Kirwan, G., Clark, C., & Dalton, M. (2019). Rating of physiotherapy student clinical performance: is it possible to gain assessor consistency? *BMC Medical Education, 19*(32), 1–8. doi: 10.1186/s12909-19-1459-4

Klamen, D., Williams, R., Roberts, N. & Cianciolo, A. (2016). Competencies, milestones, and EPAs - Are those who ignore the past condemned to repeat it? *Medical Teacher, 38*, 904–910. doi:10.3109/0142159x.2015.1132831

Knowles, M. S. (1984). *Andragogy in Action: Applying modern principles of adult education.* San Francisco, CA: Jossey Bass.

Kritek, P. (2015). Strategies for effective feedback. *Seminars for Educators*, *12*(4), 557–560. doi:10.1513/AnnalsATS.201411-524FR

Lee, K., Tsai, S., Chiu, Y., and Ho, M. (2016). Can student self-ratings be compared with peer ratings? A study of measurement invariance of multisource feedback. *Adv in Health Sci Educ*, *21*, 401–413. doi: 10.1007/s10459-015-9638-5

Lo, K., Osadnik, C. Leonard, M, & Maloney, S. (2016). Student-clinician agreement in clinical competence as a predictor of clinical placement performance in Australian undergraduate physiotherapy students. *Physiotherapy Theory and Practice*, *32*(1), 63–68 . doi: 10.3109/09593985.2015.1092622

Lo, K., Osadnik, C., Leonard, M, & Maloney, S. (2015). Differences in student and clinician perceptions of clinical competency in undergraduate physiotherapy. *New Zealand Journal of Physiotherapy*, *42*(1), 11–15. doi: 10.15619INZJPI43.1.02

Madrazo, L., Lee, C., McConnell, M., & Khamisa, K. (2018). Self-assessment differences between genders in a low-stakes objective structured clinical examination (OSCE). *BMC Research Notes*, *11*, 393–398. doi: 10.1186/s13104-018-3494-3.

Malik, M. (2016). Assessment of a professional development program on adult learning theory. *Libraries and the Academy*, *16*(1), 47–70. doi: 10.1353/pla.2016.0007

Mori, B., Brooks, D., Norman, K., Herold, J., & Beaton, D. (2015). Development of the Canadian Physiotherapy Assessment of Clinical Performance: A new tool to assess physiotherapy students' performance in clinical education. *Physiotherapy Canada*, *67*(3), 281–289. doi: 10.3138/ptc.2014-29E

Muhamad, Z., Henry, L., Ramli, A. (2016). A narrative review of assessment appraisal in

clinical performance of physiotherapy students. *International Medical Journal*, *23*(2), 148–152. ISSN: 1341-2051

National Consortium of Clinical Educators (2019, October). *Survey on Assessment Tools and Impact on Clinical Education: Key Findings.* Presentation at the American Physical Therapy Association Education Leadership Conference, Bellevue, WA.

Oh, S., Liberman, L., & Mishler, O. (2018). Faculty calibration and students' self-assessments using an instructional rubric in preparation for practical examination. *Eur J Dental Educ*, *22*, 400–407. doi: 10.1111/eje.12318

Panadero, E., Brown, G., Strijbos, J. (2016). The future of student self-assessment: a review of known unknowns and potential directions. *Educ Psychol Rev*, *28*, 803–830. doi: 10.1007/s10648-015-9350-2

Pastore, S. (2017). Research designs and methods in self-assessment studies: A content analysis. *International Journal of Evaluation and Research in Education*, *6*(4), 257–264. ISSN: 2252-8822

Pawluk, S., Zolezzi, M., & Rainkie, D. (2018). Comparing student self-assessments of global communication with trained faculty and standardized patient assessments. *Currents in Pharmacy Teaching and Learning*, *20*, 779–784. doi: 10.1016/j.cptl.2018.03.012

PTCPI. (2006) Alexandria, VA: *American Physical Therapy Association*. Retrieved from http://www.AmericanPhysicalTherapy Association.org/PTCPI/

Poirier, T., Pailden, J., Jhala, R., Ronald, K., Wilhelm, M., & Fan, J. (2017). Instructional design and assessment student self-assessment and faculty assessment of performance in interprofessional error disclosure simulation training program. *American Journal of*

Pharmaceutical Education, *81*(3), 1–8. doi: 10.5688/ajpe81354

Porter, T. (2016, October). *Analysis of "beyond entry-level" performance ratings on the physical therapist clinical performance instrument*. Platform presentation at the American Physical Therapy Association Education Leadership Conference, Phoenix, AZ.

Recker-Hughes, C., Padial, C., Becker, E., & Becker, M. (2016). Clinical site directors' perspectives on clinical education. *Journal of Physical Therapy Education*, *30*(3), 21–27. Retrieved from https://aptaeducation.org/members/jopte/

Roach, K., Frost, J., Francis, N., Giles, S., Nordrum, J., Delitto, A. (2012). Validation of the revised PTCPI: version 2006. *Physical Therapy*, *92*(3), 416–428. doi: 10.2522/ptj.20110129

Schon, D. (1983). *The reflective practitioner: How professionals think in action*. London, England:Temple Smith

Schoo, A., Lawn, S., Rudnik, E., and Litt, J. (2015). Teaching health sciences students foundation motivational interviewing skills: use of motivational interviewing treatment integrity and self-reflection to approach transformative learning. *BMC Medical Education*, *12*, 228–238. doi: 10.1186/s12909-015-0152-1

Tavares, W. & Eva, K. (2014). Impact of rating demands on rater-based assessments of clinical competence. *Education for Primary Care*, *25*, 308–318. doi:10.1080/14739879.2014.11730760

Thawabieh, A. (2017). A comparison between students' self-assessment and teachers' assessment. *Journal of Curriculum and Teaching*, *6*(1), 14–20. doi: 10.5430/jct.v6n1p14

Appendix A: IRB Approval Letter

CONCORDIA UNIVERSITY
-PORTLAND, OREGON-

DATE:	August 16, 2019
TO:	Tracy Porter
FROM:	Concordia University - Portland IRB (CU IRB)
PROJECT TITLE:	[1472641-1] Perceived accuracy of Doctor of Physical Therapy students' self-assessment of clinical performance using the PTCPI and the CIET
REFERENCE #:	EDD-20190729-Graham-Porter
SUBMISSION TYPE:	New Project
ACTION:	APPROVED
APPROVAL DATE:	August 16, 2019
REVIEW TYPE:	Limited Review

Thank you for your submission of New Project materials for this project. The Concordia University - Portland IRB (CU IRB) has APPROVED that your submission fits the requirements for Limited Review. This project is EXEMPT from further CU IRB review according to federal regulations.

We will retain a copy of this correspondence within our records. Please keep this correspondence within your records.

At the time when you need to demonstrate that you have closed out your project, you can provide your faculty chair with a copy of this letter, explaining that you are exempt from needing to close out your project.

The researcher is responsible to conduct research, even if it is exempt, with integrity and care. You are encouraged to continue to work with the CU IRB Office and involve others at Concordia University as necessary and prudent in your research.

If you have any questions, please contact Amon Johnson at (503) 280-8127 or amjohnson@cu-portland.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within Concordia University - Portland IRB (CU IRB)'s records. August 16, 2019

Appendix B: Statement of Original Work

The Concordia University Doctorate of Education Program is a collaborative community of scholar-practitioners, who seek to transform society by pursuing ethically-informed, rigorously-researched, inquiry-based projects that benefit professional, institutional, and local educational contexts. Each member of the community affirms throughout their program of study, adherence to the principles and standards outlined in the Concordia University Academic Integrity Policy. This policy states the following:

> **Statement of academic integrity.**
>
> As a member of the Concordia University community, I will neither engage in fraudulent or unauthorized behaviors in the presentation and completion of my work, nor will I provide unauthorized assistance to others.
>
> **Explanations:**
>
> ***What does "fraudulent" mean?***
>
> "Fraudulent" work is any material submitted for evaluation that is falsely or improperly presented as one's own. This includes, but is not limited to texts, graphics and other multi-media files appropriated from any source, including another individual, that are intentionally presented as all or part of a candidate's final work without full and complete documentation.
>
> ***What is "unauthorized" assistance?***
>
> "Unauthorized assistance" refers to any support candidates solicit in the completion of their work, that has not been either explicitly specified as appropriate by the instructor, or any assistance that is understood in the class context as inappropriate. This can include, but is not limited to:
>
> - Use of unauthorized notes or another's work during an online test
> - Use of unauthorized notes or personal assistance in an online exam setting
> - Inappropriate collaboration in preparation and/or completion of a project
> - Unauthorized solicitation of professional resources for the completion of the work.

Statement of Original Work (continued)

I attest that:

1. I have read, understood, and complied with all aspects of the Concordia University–Portland Academic Integrity Policy during the development and writing of this dissertation.

2. Where information and/or materials from outside sources has been used in the production of this dissertation, all information and/or materials from outside sources has been properly referenced and all permissions required for use of the information and/or materials have been obtained, in accordance with research standards outlined in *The Publication Manual of The American Psychological Association*.

Tracy Porter

Digital Signature

Tracy Porter

Name (Typed)

3/1/2020

Date