# Consensus Development for Thrust Joint Manipulation Acquisition Assessment in Entry-Level Doctor of Physical Therapy Students: A Modified Delphi Study.

William H. Kolb, PT, DPT

A capstone project submitted to the faculty of Radford University

in partial fulfillment of the requirements for the degree of

Doctor of Health Sciences

April 22nd, 2025



| | |
|---|---|
| Kevin Chui | 4/24/2025 |
| Kevin Chui, PT, DPT, PhD, MBA, GCS, OCS, CEEAA, FAAOMPT (Chair) | Date |
| Laura Newsome, Ph.D., ACSM-CEP, FCEPA | 4/23/2025 |
| Laura J. Newsome, PhD, ACSM-CEP, EIM, FCEPA | Date |
| PT DPT PhD | 4/22/2025 |
| Emilio Puentedura, PT, DPT, PhD, OCS, FAAOMPT | Date |
| Amy Mcdevitt | 4/23/2025 |
| Amy McDevitt, PT, DPT, PhD, OCS, FAAOMPT | Date |

**Abstract**

**Introduction:** The purpose of this Delphi study was to develop by expert consensus a new universal measure to provide formative feedback for thrust joint manipulation (TJM) tasks. TJM is a component of orthopedic manual physical therapy (OMPT) and is a complex procedural skill to teach. Although teaching resources for TJM have existed for decades and Delphi studies have described what paradigms to teach for OMPT, student surveys consistently report low confidence and limited use of TJM during clinical experiences.

**Literature Review:** When learning a new task, formative feedback enhances the learners' strategies especially in the early practice phase. However, a review of nine studies with rubrics for OMPT techniques reveals a disparity of definitions for similar task criteria with seven rubrics used for summative purposes. There is a lack of consensus in physical therapy education for the essential criteria to perform most TJM tasks.

**Subjects:** 480 emails were sent to OMPT teaching experts inviting them to participate in the Delphi process.

**Methods:** A workgroup developed the quick psychomotor operator performance assessment (QPOPA) for TJM tasks based on a literature review, national presentation and publication. A modified Delphi used a content validation index (CVI) to assess consensus of the QPOPA. Round One used CVI and open text suggestions to review TJM task elements of setup and thrust as criteria for technique assessment. Round Two asked participants to select the best option for wording of criteria and reviewed a quality rating scale.

**Results:** The number of respondents was 66 for Round One and of these 44 completed Round Two. Consensus was reached on all assessment criteria and quality scale items.

Total scale CVI scored 99% for relevance and 92% for clarity for the five thrust and three setup criteria.

**Discussion and Conclusion:** The QPOPA was created to provide universal feedback for novice learners of TJM tasks and has achieved content validation by an expert consensus method. Potential benefits of a universal TJM task assessment include alignment of teaching, learning, and assessment of complex procedural skills across body regions. The QPOPA needs further testing to acquire additional validation evidence. A universal measure such as the QPOPA could support multi-site studies to improve the scholarship of TJM teaching strategies.

**Key words:** learning, physical therapy, procedural skills, professional education

William H. Kolb, PT, DPT

Department of Physical Therapy, 2025

Radford University

## Dedication

First, I praise God for giving me this opportunity to advance my education including the pursuit of my passion for researching teaching methods for physical therapy skills.

> *Whatever you do, work at it with all your heart, as working for the Lord, not for human masters, since you know that you will receive an inheritance from the Lord as a reward. It is the Lord Christ you are serving.* Colossians 3:23-24.

The capstone research work and the DHSc courses have been an incredible journey of personal and professional growth that would not be possible without the support of family members who made me who I am today. I would like to thank both of my school teaching parents. For my mother, whose constant support and encouragement that you can "learn to do anything if you put your mind to it," gave me a growth mindset before we know what that even was and kept me going. For my father who modeled living as a loving Christian father while working two jobs teaching me the "keep going when it gets tough" work ethics. For my oldest son Caleb who I have been able to teach on his journey to become a physical therapist, which has given me unique insights on teaching and learning. For my youngest son Sam, who was always willing to discuss the difficulties of our work and share a cheerful word of encouragement. Finally for my loving, supportive and understanding wife Cindy who worked alone way too many days on the goat farm yet still encouraged me to keep going. Words cannot express my appreciation for the support of each of you.

> *Now to Him who is able to do exceedingly abundantly above all that we ask or think, according to the power that works in us.* Ephesians 3:20.

May God bless and keep you all.

**Acknowledgements**

Many thanks to the members of my Capstone committee (Dr. Chui, Dr. Newsome, Dr. Puentedura and Dr. McDevitt) for sharing your time to provide guidance and direction throughout the Capstone process. A note of appreciation as well to the members of my Delphi workgroup including Michelle Dolphin, Kyle Adams and both Dr. McDevitt and Dr. Puentedura who did double duty as both committee members and workgroup members.

Many additional people supported this endeavor, and I could not possibly list each of them here. For the DHSc faculty, including the research course instructors Dr. Everhart and Dr Blavos for pushing me outside of my comfort zone and spending so much time reviewing my rough drafts. I'm indebted as well to those DHSc instructors who created courses that helped me dig deeper into the how, why and when of teaching, including Dr. Jenny Hall, Dr. Lisa Alison-Jones and Dr. David Halpin. Many thanks as well to Dr. Frank Dane for taking my many Delphi methodology questions and his overall words of encouragement. And yes, Jamie Price for providing many insights for the search strategies as reference librarian and who is actually faster than the Millennium Falcon.

This work would not have been possible without each of you!

All the best,

WK

**Table of Contents**

**List of Tables**

**List of Figures**

## List of Abbreviations

APTA……………… .American Physical Therapy Association

AAOMPT……….......American Academy of Orthopedic Manual Therapists

C1-C2…....................First and Second Cervical Vertebrae

CASE……. ................Cystectomy assessment and surgical evaluation

CLT………..................Cognitive Load Theory

CVI……......................Content Validation Index

DHSc……..................Doctor of Health Sciences

EPA…. ......................Entrustable Professional Activity

HPE… .......................Health Professional Education

MT…........................Manual Therapy

NTI…… ....................Neonatal Tracheal Intubation

OSCE… ...................Objective Structured Clinical Examination

OMPT….. .................Orthopaedic Manual Physical Therapy

PI….........................Primary Investigator

POCUS……..............Point of Care Ultrasound

PT….........................Physical Therapists

QPOPA…. ................Quick Psychomotor Operator Procedural Assessment

ROC….. ....................Receiver Operator Characteristic (ROC)

S-CVI…… ................Scale-Content Validation Index

TJM……….................Thrust Joint Manipulation

# Chapter One Introduction

Teaching procedural skills in health professional education (HPE) is challenging and complex. Due to differences in clinical tasks and contexts among health professions, there are few agreed-upon methods of teaching or assessing procedural skills (Nicholls et al., 2016). These differences were confirmed by a recent scoping review noting low levels of evidence for methods to teach or assess procedural skills (Bourassa et al., 2024). These evidence gaps are important as pre-clinical instruction in HPE aims to ensure competency in skills prior to clinical work (Sawyer et al., 2015; Timmerberg et al., 2022). Thrust joint manipulation (TJM) is a complex procedural skill to learn for physical therapists (PTs) and is recommended as a first-level treatment for musculoskeletal pain conditions in both physician and PT practice guidelines (Blanpied et al., 2017; George et al., 2021; Martin et al., 2021; Qaseem et al., 2017). However, despite published guidelines to teach TJM, evidence exists that PT students' ability to transfer TJM skills from the classroom to the clinic is limited (Corkery et al., 2020; Puentedura et al., 2017; Struessel et al., 2012).

The aim of this line of research is to construct and validate a universal measure to assess many different TJM tasks. The ideal use for the new measure would be to provide formative feedback to novice learners and establish universal assessment criteria for psychomotor performance for future research on teaching methods. Based on principles used in medical education for assessment development and validation, there are two overall steps to create a new measure (Cook et al., 2014; Cook & Lineberry, 2016; Yudkowsky et al., 2020).

1) Develop the measure using an expert consensus process for content validation.

2) Test the measure to accumulate additional validity evidence.

Therefore, the aim of this exploratory study was to create a new TJM assessment, as no measures currently exist. The first step to developing the measure is to use an expert consensus process such as a modified Delphi technique, which was the scope of this capstone project and final paper. A proposed model for the TJM assessment rubric has been published (Adams et al., 2024) based on a literature review and prior work done by the author (WK) which served as the framework for the first Delphi survey round. The literature review for this study examined theories of motor learning and cognitive load, assessment rubrics with a focus on TJM, and the Delphi consensus process for validation of new assessment methods in HPE.

**1.1. Background**

The traditional outcomes-based approach to education consists of three components: first learning outcomes, basically what do we want the students to learn, second, teaching and learning activities commonly referred to as instructional methods, and third, assessment (Chan, 2022). Although the evidence is growing on teaching methods for procedural skills in HPE the literature is sparse on assessment of procedural skills (Bourassa et al., 2024). A potential consequence of limited assessment literature in HPE includes a reduction in assessment literacy described as "knowledge about the basic principles of sound assessment practice, including its terminology, the development and use of assessment methodologies and techniques, and familiarity with standards of quality in assessment" (Chan, 2022, p. 50). The lack of validated assessments for procedural skill research has a broad impact including overreliance on theory versus evidence for instructional design thereby reducing the overall skill use and proficiency during clinical experiences for the novice healthcare student.

Pre-professional, also known as pre-licensure HPE aims to ensure competence as some level of safety for the public before performing procedures on patients during clinical placements (Sawyer et al., 2015; Timmerberg et al., 2022). Pre-clinical instruction usually begins with the theory and then the skill instruction with practice and assessment for safety prior to sending students out for clinical experiences supervised by licensed instructors. Compounding the problem for HPE is an ever-growing list of skills required to be taught during pre-clinical education, which shortens the time for deliberate practice of each task prior to clinical experiences (Thibault, 2020). Consequences of less deliberate practice time impacts learners in that those who have not reached a competent level of performance in the classroom are then given reduced practice privileges from clinical instructors (Bannister et al., 2003; McGaghie, 2015; Vogel & Harendza, 2016).

A common teaching method is to deconstruct a complex clinical task into cognitive, communication, technical, and psychomotor skill subcomponents to be taught as a procedural skill (Sattelmayer, K., Jagadamma, & Sattelmayer, F. et al., 2020; Sawyer et al., 2015). Brown and colleagues (2014) provide a working definition of learning as “acquiring knowledge and skills and having them readily available from memory so you can make sense of future problems and opportunities” (p. 2). Brown’s definition of learning includes skills, which aligns with the motor learning definition of Schmidt et al. (2019) as “a set of processes associated with practice or experience leading to relatively permanent changes in the capability for skilled movements” (p. 283). Motor learning theory provides a framework for how to structure the teaching of complex procedural skills (Lewthwaite & Wulf, 2017). Because motor learning processes are not directly observable and must be assumed (Schmidt et al., 2019), teaching clinical tasks in HPE relies on theories and sequenced approaches to teaching procedural skills (Giacomino et al., 2020).

Several authors identify the need to expand the evidence base for teaching complex procedural skills in entry-level PT education (Cecilio-Fernandes et al., 2023; Garcia-Ros et al., 2024; Sattelmayer et al., 2016), including TJM (Giacomino et al., 2020; Gradl-Dietsch et al., 2016). TJM, also known as thrust manipulation, is a component of orthopedic manual physical therapy (OMPT), which is described as the integration of clinical reasoning with hands-on skill development driven by the evidence and the biopsychosocial framework of each individual patient (Silvernail et al., 2024). Within the context of PT education, TJM can be defined as the skilled application of a high-velocity, low-amplitude movement to the articular surfaces to create cavitation within the synovial fluid of the joint (Evans & Lucas, 2023; Griswold et al., 2018). The procedural task components of TJM include clinical aspects of communication, cognitive recognition of indications and contraindications, and sequencing of motor skill subcomponents, which partly explain the complexity and challenge of teaching the novice learner.

### *1.1.1. Theoretical Frameworks for Procedural Skill Teaching*

Although many theories and sequenced approaches lack evidence to support their use for instructional design (Nicholls et al., 2016), the theories provide helpful frameworks for procedural skill teaching. Often cited frameworks to guide instructional design include Miller's pyramid (1990), Fitts and Posner's (1967) developmental phases of motor learning, the Dreyfus (2004) stages of expert practice development, and Ericsson's deliberate practice theory (2015).

Miller's (1990) pyramid provides a framework to deconstruct the teaching of procedural skills into the cognitive, i.e., "knows" and "knows how," and psychomotor domains of "shows how." The top level in Miller's (1990) framework, i.e., "does," is an assessment of competency prior to the performance on patients. Specific to the teaching of psychomotor skills, Fitts and Posner (1967) described the developmental phases of motor learning as 1) Cognitive, 2) Associative, and 3) Autonomous. A strength of the Fitts and Posner phases is that the three phases align directly with what aspects of teaching the skill benefit the learner most.

The Dreyfus (2004) stages of skill acquisition are a sequenced approach that describes a learner's potential capability to perform a task. Per Dreyfus, with practice of the task, the learner progresses through the stages of novice, advanced beginner, competent, proficient, and expert. Deliberate practice is a form of structured practice described by Ericsson that, with feedback, assists learners in their progression from novice to expert (2015). The key aspect of deliberate practice is for the teacher to help the learner develop internal representations of performing the task to improve their performance (Higgins et al., 2021). The theoretical frameworks described above provide the foundation for many of the sequenced skill teaching methods used in HPE.

#### *1.1.2. Skill Teaching Methods*

Many variations of sequenced skill teaching methods are used in HPE. Variations used for procedural skill teaching include the simple Halsted model of "see one, do one, teach one" to more complex four-step (Walker, 1998), five-step methods (George & Doto, 2001), and upwards of eleven different steps (Nicholls et al., 2016). A key consideration for instructional design includes managing the learner's cognitive load, especially for complex skills, to prepare for real-world transfer. To manage the cognitive load for procedural skills like TJM, an optimal instructional design must balance skill complexity, prior learner experience, and context (high or low fidelity) of the skill (Leppink & van den Heuvel, 2015; Young et al., 2014). Fidelity is the term used to describe the context of the skill performance, with low fidelity being simulated practice with peers or patient actors and high fidelity referring to actual patients (Leppink & van den Heuvel, 2015).

For skills of higher complexity, the Peyton teaching approach step of demonstration-verbalization is helpful as the learner talks the instructor through the skill steps, thus reinforcing their image of correct skill performance prior to actual practice (Giacomino et al., 2020; Rossettini et al., 2017). Upon the completion of the pre-practice phase, deliberate practice for complex skills like TJM allows skill development and immediate formative feedback to correct movement errors, build the learner's understanding of movement strategies, and reflection in action to optimize deliberate practice needed for autonomous performance (Ericsson, 2015; Fitts & Posner, 1967; Leech et al., 2022).

Although multiple theories, models, and steps are proposed, skill instruction essentially consists of two phases: pre-skill and skill practice (Bilyeu et al., 2023). The goal of the pre-skill phase is to prepare the learner for successful practice considering skill complexity, context and learner level for how instruction is delivered. Next, the goal of the practice phase is to create an optimal learning environment to enhance learner expectancies and motivation by supporting deliberate practice (Leech et al., 2022; Lewthwaite & Wulf, 2017). Key to initial skill acquisition is providing constructive and timely feedback on skill performance, also known as formative feedback (Bilyeu et al., 2023). Formative feedback has been described as "information communicated to the learner that is intended to modify his or her thinking or behavior to improve learning" (Shute, 2008, p. 154).

In summary, the critical components of instructional design for complex procedural skills like TJM are optimizing practice for the learner and developing an accurate assessment of skill performance. Traditional teaching methods often used in HPE, such as see-one-do-one, skip the support and formative assessment needed for the learner in the early practice phase, which, according to motor learning theory is where the most variation occurs in performance (Dreyfus, 2004; Schmidt et al., 2019). Early practice for the novice is also when instructor feedback is most beneficial to assist the learner in forming representations of the ideal skill performance (Ericsson, 2015). As reviewed above, many theories, approaches, and methods have been written to guide procedural skill classroom teaching. However, the literature lacks feedback and assessment tools for the learner, especially in the early practice phase for complex tasks like TJM (Adams et al., 2024).

### ***1.1.3. Manipulation Education Teaching Resources***

The PT profession has had resources to support teaching TJM, including the American Physical Therapy Associations (APTA) Manipulation Education Manual for Physical Therapist Professional Degree Programs (Manipulation Education Committee of the APTA Manipulation Task Force, 2004), commentaries on teaching manipulation skills (Flynn et al., 2006), and standardized terminology (Mintken et al., 2008). Prior OMPT expert consensus research has described what to teach for competency development (Sizer et al., 2008) and what paradigms of manual therapy to teach (Keter et al., 2023). In addition, a recent Delphi identified what characteristics are important for therapist and patient during side-lying neutral gap manipulation (O'Donnell et al., 2016). However, these task elements were not listed as criteria and paired with quality levels or scales for performance assessment. Despite these teaching resources, no agreed-upon assessment criteria exist to provide feedback for a novice learner in the skill practice phase.

Potential reasons limited evidence exists to guide the teaching of complex skills like TJM include variation in expert practice standards and the need for validated assessment tools. By necessity, experts have learned many variations of TJM skills to adapt to unique practice situations. Experts, by definition, organize and apply their knowledge differently than novices (Ambrose et al., 2010). The concept of expert blind spots refers to how expert instructors unknowingly leave out parts of a complex skill, confusing the novice learner (Ambrose et al., 2010). One strategy for teaching a novice TJM would be developing a checklist or formative skill assessment early in the practice phase. However, a barrier to creating a checklist is that multiple variations of TJM skills exist, which is typical for expert practitioners. Multiple checklists may therefore exist, each with unique variations and nuances according to each expert instructor and program. Regardless, the lack of a validated assessment of TJM is a barrier to the novice learner and higher-quality educational research to inform teaching practices. Because a validated TJM assessment does not exist, most studies of TJM teaching methods report outcomes of perception and attitude versus skill acquisition or transfer (Bourassa et al., 2024).

**1.2. Significance**

Although guidelines recommend OMPT and TJM interventions to reduce musculoskeletal pain (George et al., 2021; Martin et al., 2021; Qaseem et al., 2017) and TJM skills are being taught in PT programs, cross-sectional surveys of PT students report low levels of TJM use in clinical settings (Corkery et al., 2020; Puentedura et al., 2017; Struessel et al., 2012). Many studies have been completed on different teaching methods for procedural skills in HPE; however, few report skill acquisition or retention outcomes, and even fewer use validated assessments (Bourassa et al., 2024). Of nine studies specific to manual therapy skill rubrics reviewed for this proposal, only four completed some assessment validation, one of which only examined face validity. Currently, a bottleneck exists due to the lack of validated assessment tools to progress the evidence base for teaching procedural skills like TJM.

The importance of developing an assessment of TJM is to reduce variation in teaching, increase the quality of educational research, and support the shift of PT education to competency-based education (CBE) (Jensen et al., 2020; Timmerberg et al., 2022). Proponents of CBE in PT education recognize the unmet need to develop feasible and validated assessments (Chesbro et al., 2018; Jensen et al., 2020; Timmerberg et al., 2022). Arguments supporting the development of a more generalized universal measure center around increased utility, including adaptability to different body regions, and reducing instructor-specific variations in TJM feedback in large lab settings. Recently, a TJM assessment framework for entry-level PT students has been proposed (Adams et al., 2024). However, the assessment tool has not been validated.

### 1.3. Purpose

The purpose of this exploratory study was to develop by expert consensus a new universal assessment tool for most TJM techniques. The expert consensus process was the first step used to support content validation and accumulate evidence of the new measure's overall validity.

### 1.4. Research Questions

Primary Aim: Develop a validated universal assessment measure for the psychomotor performance of most TJM techniques. Based on the primary aim, the following research questions were developed.

Research Questions:

RQ1. Can consensus be reached on the criteria necessary to evaluate the psychomotor performance for most TJMs?

RQ2. Can consensus be reached on the rating scale to evaluate the quality of criteria for the psychomotor performance of novice learners for most TJMs?

RQ3. Can consensus be reached on a global rating scale to evaluate the overall psychomotor performance for most TJMs?

RQ4. Can consensus be reached on a list of TJMs that the newly created measure is most appropriate to use?

Hypothesis:

H1. Elements of what criteria to assess for a TJM will be established by expert consensus of agreement above the *a priori* threshold.

H2. A rating scale to evaluate the quality of the criteria to assess for a TJM will be established by expert consensus of agreement above the *a priori* threshold.

H3. A global rating scale to evaluate the overall performance of TJM will be established by expert consensus above the *a priori* threshold.

H4. A list of thrust joint manipulations for which the new measure is most appropriate to use will be established by expert consensus above the *a priori* threshold.

**1.5. Definitions**

The terms psychomotor and procedural are often used interchangeably in the literature. One possible reason proposed by Nicholls is that the traditional teaching of procedural skills has been linked to a "sequenced and stepped" approach adopted from psychomotor learning theories (2016). An example of the sequenced approach for procedural skill development comes from Sawyer and colleagues, who cite the taxonomy of the psychomotor domain of Simpson and Harrow as a framework for medical education (Sawyer et al., 2015). Another explanation for the confusion between terms is the multidisciplinary nature of skills shared between and among professions, yet some procedural skills are distinct and unique to a single health profession (Burgess et al., 2020). Procedure skills can be referred to as either diagnostic or intervention and vary in complexity (Nicholls et al., 2016; Sattelmayer, M., Jagadamma, & Sattelmayer, F. et al., 2020). Examples of procedural skills in HPE include taking a health history, handwashing, resuscitation, performing a bed-to-chair transfer, performing a bed-side dysphagia evaluation (Burgess et al., 2020), changing a wound-vac, inserting a peripheral IV, soft tissue mobilization in manual therapy, or a computed tomography assisted needle biopsy. A unique aspect of procedural skills is that they must be adapted to the individual patient needs and context adding to their complexity for teachers and learners (Cutrer et al., 2017).

Definitions of learning, skills, and tasks are necessary to clarify the use of these terms in this paper. Learning has been defined by Brown and colleagues (2014) as “acquiring knowledge and skills and having them readily available from memory so you can make sense of future problems and opportunities” (p. 2). Brown's definition of learning includes skills, which aligns with the motor learning definition of Schmidt et al., (2019) as “a set of processes associated with practice or experience leading to relatively permanent changes in the capability for skilled movements” (p. 283).

A skill is a term used to describe an individual’s potential capability to perform a specific task. A task is then defined as an activity to be performed, such as throwing a baseball. In general teaching literature, the term skill can imply the ability to perform cognitive tasks such as completing a word problem or an algebraic equation. However, for context in this paper, the term skill will imply motor or psychomotor skill unless otherwise stated. Per Magill, motor skills are activities or tasks that require voluntary head, body, and/or limb movement to achieve a goal (2014, p. 3). For example, in conversation, we say “The learner has a high level of skill to throw a baseball” using the term skill to note the performance capability of the person performing the task (Magill, 2014, p. 5).

Procedural skills are defined for this project as the psychomotor, cognitive, communication, and technical aspects, all of which must be adapted to complete a clinical task. This definition combines the learning of “mental and motor activities” from

Sawyer et al. (2015, p. 1026) and the “psychomotor skill” terminology of Nicholls et al. (2016) within the context of “clinical performance” from Hibbert et al. (2013, p. 2). Inherent in this definition of procedural skills, in the understanding of steps and proficiency in multiple domains required to complete the task. Defining procedural skills as above is important for pedagogical reasons which is expanded upon in the literature review in the next section.

# Chapter Two Literature Review

This literature review takes the shape of an hourglass. The review begins broadly with theories of teaching procedural skills as applied to HPE, then narrows in on rubrics for manual therapy assessment and concludes with a general review of methods for educational assessment development and validation. The literature review of methods for manual therapy skill assessment used an exhaustive systematic review process to allow for specific reporting. However, to summarize the topics of teaching procedural skills and assessment development, a critical review method was used as described with the purpose of integrating several schools of thought to summarize topics and provide a foundation for future scholarly work (Grant & Booth, 2009). As teaching procedural skills is an immense topic, an attempt was made to critically review the literature as it pertains to application in the classroom.

## 2.1. Theories of Teaching

A critical review of the literature was completed for the theories of constructivism, motor learning and cognitive load as they apply to teaching procedural skills. Databases used for the search included Medline, Cumulative Index of Nursing and Allied Health Literature and Google Scholar. Constructivist learning theory begins the theoretical review of teaching methods, as the implications of constructivism are to assimilate multiple theories and approaches for adult learning. Next, motor learning, followed by cognitive load theory, is described with definitions critical to teaching procedural skills and a review of traditional skill teaching in HPE. These theories provide the background for the systematic review of the literature for assessment rubrics used for procedural skills in PT education specific to manual therapy.

### *2.1.1. Constructivist Learning Theory*

Constructivist learning theory can be simply described as a process whereby the learner constructs new knowledge by making sense of their prior experiences (Amineh & Asl, 2015; Dennick, 2012). The origins of constructivist learning theory have been credited to Piaget, Dewey, and Vygotsky (Amineh & Asl, 2015; Dennick, 2012). A key distinction of constructivist theory is that it "describes how structures, language, activity, and meaning making come about, rather than one that simply characterizes the structures and stages of thought, or one that isolates behaviors learned through reinforcement" (Fosnot, 2005, p. 34). As such, learning can be viewed as development that is learner-led by questions and investigations, resulting in creating their own meaning of big ideas which can be generalized across experiences (Fosnot, 2005). Implications of constructivism theory are that it assimilates multiple theories of adult learning including, but not limited to, behaviorist, cognitive, and experiential learning theories (Amineh & Asl, 2015; Dennick, 2012).

Pedagogical methods attributed to constructivist learning theory include assessing prior knowledge of the learner, e.g., through questioning, and developing conceptual understanding through debate and sharing of meanings. The role of the teacher then becomes that of developing a level of cognitive conflict, also known as dissonance, for the students' prior understandings, followed by facilitating the application of new theories to create shared meanings. These activities of exploration and creation of shared meanings between the learner, other learners, and the teacher are originally described by Vygotsky as social constructivism (Taylor & Hamdy, 2013).

Specific to HPE, constructivist theory is directly connected to the development of the hypothetico-deductive process of clinical reasoning (Dennick, 2012; Taylor & Hamdy, 2013). Essential for HPE is the use of constructivism to develop not only the learners' cognitive frameworks but also their procedural knowledge and skill development. Physical therapy students' clinical reasoning has been described as consisting of three components: content knowledge (identification of facts), procedural knowledge (psychomotor skills), and conceptual reasoning (metacognitive abilities to reflect on action and in action) (Furze et al., 2015). The importance of integrating clinical reasoning with hands-on skill development has been reinforced by an updated description and definition of OMPT (Silvernail et al., 2024). Therefore, constructivist theory can serve as a framework to develop the novice PTs cognitive knowledge and their ability to integrate procedural skills in an adaptive clinical reasoning model that improves through guided practice and experience.

The Kolb experiential learning theory is often used as a theoretical foundation to support skill instruction in HPE. Many sources place experiential learning within constructivist theory (Dennick, 2016; Lockey et al., 2021; Mukhalalati & Taylor, 2019; Taylor & Hamdy, 2013). The view of experiential learning as essentially being constructivist theory is supported by Kolb and Kolb's description of learning as being the "process whereby knowledge is created through the transformation of experience" (2005, p. 194). The Kolb learning cycle describes four learning styles of concrete experience (feeling), reflective observation, abstract conceptualization, and active experimentation as types of transformative experience whereby learners construct knowledge (Taylor & Hamdy, 2013). Although experiential learning and constructivism theories generally support procedural skill instruction in HPE, a critical literature review was completed for the application of updated models of instructional design phases of pre-skill, practice, and assessment for this scholarly work.

***2.1.2. Motor Learning Theory***

Motor learning theory is often cited as the foundation for psychomotor skill instruction and practice strategies. However, the theoretical terms used in motor learning represent a broad blanket of 27 principles and approaches lacking operationalization (Kafri & Atun-Einy, 2019). This review first describes classic motor learning definitions and then summarizes recent theoretical developments as a foundation for the scholarly work in this paper. A central tenet of motor learning theory is that the actual learning of a motor skill is not directly measurable but must be inferred by performance (Schmidt et al., 2019). Skill acquisition is the term used to describe a time point after an educational intervention designed to benefit the learner, for which the learner's performance meets a set standard or outcome for the skill. Retention and transfer tests are used to determine more lasting learning effects after time passes from the initial skill acquisition. Retention tests differ from transfer in that retention evaluates the same skill as practiced during the initial acquisition phase after time has passed. In contrast, transfer tests evaluate a variation of the skill practiced during the initial acquisition phase (Schmidt et al., 2019).

The conundrum of instructional design is that classroom conditions that support initial skill acquisition, i.e., comfortable practice, may limit the learner's ability to transfer the skill during highly variable conditions in the clinic (Schmidt et al., 2019). Ericsson (2015) describes deliberate practice as a distinct form of teacher-led practice in which learners must develop three types of internal representations to improve their skill performance. The role of the instructor is to present an expert representation of the skill and guide feedback so the learner can: 1) imagine the desired expert-level performance, 2) represent strategies to execute the performance, and 3) during performance reflectively monitor for a potential gap compared to the expert standard (Ericsson, 2015). A critical distinction between deliberate vs. repeated practice is the role of the instructor to provide feedback to assist the learner in developing internal skill representations (Higgins et al., 2021). Therefore, multiple skill practice variables, learner level, skill complexity, and practice schedules must be considered when designing instruction for complex procedural skills.

Additional complications of motor learning theory for teaching design is that multiple practice variables exist, including repeated task-specific part-versus whole-practice and many derivations of practice schedules, i.e., blocked, sequential, and random (Magill, 2014; Schmidt et al., 2019). Research surrounding how practice variables impact skill acquisition has expanded into additional disciplines but remains largely isolated in each respective research field (Roemmich & Bastian, 2018). Recently, Leech and colleagues (2022) have summarized the motor learning literature for implementation into PT clinical practice; however, one could surmise the same strategies apply to teaching psychomotor skills in PT education. Leech et al. (2022) describe four overall motor learning mechanisms: 1) use-dependent, 2) instructive, 3) reinforcement, and 4) sensorimotor adaptation-based.

**2.1.2.1. Use-dependent.** The use-dependent mechanism refers to repetition-based practice strategies with performance improvement curves directly related to the volume of practice by the learner, referred to as experience-dependent neuroplasticity (Leech et al., 2022). The primary consideration of use-dependent mechanisms is that although repetitions are less cognitively demanding, the learner's level of cognitive engagement directly affects skill improvement (Leech et al., 2022; Lewthwaite & Wulf, 2017). The disadvantage of use-dependent mechanisms is that lasting improvements take a long time to accumulate (Leech et al., 2022).

**2.1.2.2. Instructive Based.** The instructive motor learning mechanism is activated when the learner receives specific external feedback about a movement error or performance (Leech et al., 2022). A vital component of the instructive motor learning mechanism is that the learner can conceptualize the error-reducing movement strategy provided by the feedback, which requires a higher cognitive focus than the use-dependent mechanism. Related components to describe the ability to conceptualize the movement strategy are termed internal representations by Ericsson’s deliberate practice theory (2015) and explicit strategy-based learning by others (Schmidt et al., 2010).

**2.1.2.3. Reinforcement-based Motor Learning.** The reinforcement-based motor learning method has been defined by Leech et al. as “improvement in motor behavior that is driven by binary outcome-based feedback” (2022, p. 4). Essentially, knowledge of a successful outcome prompts the learner to reinforce the successful movement strategies and avoid the strategies associated with non-success (Spampinato & Celnik, 2021). Potential disadvantages of reinforcement-based motor learning are that the strategy takes more time and cognitive processing compared to the use-dependent or sensory-motor adaptation methods (Abe et al., 2011; Vassiliadis et al., 2021). However, evidence suggests reinforcement-based motor learning is associated with longer retention of acquired movements (Abe et al., 2011; Leech et al., 2022).

**2.1.2.4. Sensorimotor Adaptation-Based.** The sensorimotor adaptation-based type of motor learning is driven by sensory prediction errors (Bastian, 2008; Tsay et al., 2022). Sensorimotor adaptation is mediated by the cerebellum when a task or environmental task demands change and is detected, requiring automatic updates to the motor program to adapt the movement and reduce the magnitude of the prediction error (Tanaka et al., 2020). Gentile's taxonomy of tasks is one framework for sensorimotor adaptation often used by therapists in rehabilitation to vary practice conditions of environment and body movement (Gentile et al., 1987). Sensorimotor adaptation occurs on the fastest time scale with the lowest cognitive load compared to the other three motor learning methods as movements are adjusted quickly with each repetition based on changes in task demands (Bastian, 2008; Leech et al., 2022). As reviewed above, each motor learning mechanism is believed to have different cognitive load demands. Therefore, cognitive load theory is reviewed next.

### *2.1.3. Cognitive Load Theory*

Cognitive load theory (CLT) was developed with the overall goal to optimize learning ability by proposing a framework to explain how memory works (Sweller et al., 1998). CLT was developed from experiments designed to explain the differences between novices and experts by describing how memory works in problem-solving. CLT provides guidance for the use of teaching frameworks and task deconstruction depending on the learner's level, context and task complexity (Leppink & van den Heuvel, 2015; Schilling, 2017; van Merriënboer & Sweller, 2010).

Cognitive load can be described simply as the mental effort or total working memory available to learn and solve problems (Leppink & van den Heuvel, 2015).

Working memory is the conscious information processing in the form of organizing, contrasting and comparing elements simultaneously (Sweller et al., 1998) and is limited by duration of attention and the number of elements considered (Leppink & van den Heuvel, 2015). CLT assumes that working memory can hold from five to nine information elements and actively process no more than four elements simultaneously (van Merriënboer & Sweller, 2010). Schemas assist the development of human expertise as multiple elements become organized to be automatically available from long-term memory when needed. Therefore, expertise comes from how schemas organize knowledge for retrieval from long-term memory versus cognitive overload from working with multiple new elements simultaneously (van Merriënboer & Sweller, 2010).

Implications for general teaching include encouraging schema development and automation to recognize similar elements across tasks. The goal of instructional design becomes how to present information to assist working memory's ability to construct schemas for long-term memory (van Merriënboer & Sweller, 2010). Working memory capacity includes both the intrinsic load of the task itself and the extrinsic load of the environment and how the information is presented to avoid overload (Howie et al., 2023; Leppink & van den Heuvel, 2015). Germane load is the cognitive resources dedicated to learning a new task (Howie et al., 2023; Leppink & van den Heuvel, 2015) specific to the intrinsic cognitive load of the task in consideration of complexity and amount of prior learner experience with a similar task (van Merriënboer & Sweller, 2010). To summarize, schema development is facilitated when instruction avoids overload of extrinsic factors, e.g. limiting number of new elements presented, organization of elements (Leppink & van den Heuvel, 2015; van Merriënboer & Sweller, 2010) and engaging prior knowledge i.e. constructivism. Cognitive load as the intrinsic load can be managed by matching task complexity and context to the learner (Howie et al., 2023; Young et al., 2014) which effectively optimizes the germane load for the learner (Howie et al., 2023; Leppink & van den Heuvel, 2015; van Merriënboer & Sweller, 2010).

CLT implications for teaching procedural skills to novice learners are that when task complexity increases, so does the intrinsic cognitive load (Howie et al., 2023; Nicholls et al., 2016), which limits capacity for memory and learning (van Merriënboer & Sweller, 2010). Proposed strategies to manage cognitive load include limiting the number of new concepts being taught, breaking complex tasks into smaller elements (Sweller, 1998), and using familiar stepwise approaches to teaching (Bilyeu et al., 2023; Nicholls et al., 2016). Therefore, each domain of procedural skills, e.g., cognitive, psychomotor, technical, and communication, should be considered alongside prior learner experience for optimal instructional design.

**2.2. Assessment of Procedural Skills**

An exhaustive literature search was used specifically for assessment of procedural skills in physical therapy education as the main emphasis of the research study. The objective of the search was to locate resources at two levels: level one, a general procedural skill assessment for PT and level two, a more specific manual therapy intervention task assessment. Search methods were adapted with the assistance of a reference librarian from the authors' (WK) prior work on a scoping review for the best methods to teach psychomotor skills (Bourassa et al., 2024). Adaptations for the search for this project included adding the term physical therapy to the prior three search terms of procedural skills, learning, and professional education. Inclusion criteria were publication in the English language in the last 10 years, teaching or assessing procedural skills primarily in PT, and human subjects. Theoretical papers or reviews were accepted if the goal was to inform procedural skill assessment. Exclusion criteria for the search were articles without procedural skill teaching or assessment, e.g. solely professional behaviors or communication skills, or not of the PT profession unless meeting the criteria specific to search level two.

Using these updated search terms in the Medline database yielded 372 articles, which were reduced to 40 by title and abstract screening. Of the 40 articles for full text review only one was not able to be retrieved resulting in 39 articles. Then using full text review 22 articles were excluded mostly for lack of teaching or assessing a procedural skill, with two articles excluded for an emphasis on instrumented only type assessment versus classroom teaching resulting in 15 articles. See Appendix A for a PRISMA flow diagram for the procedural skills assessment and TJM literature search.

For these 15 articles, citation searching was next completed along with the "cited by" feature of Google Scholar resulting in 21 additional records retrieved for review. After full-text review of these additional 21 records, six were excluded, resulting in an additional 15 included resources for 30 total level one general PT procedural skill assessment articles. Level two criteria were next applied to narrow the inclusion criteria to be only physical therapy or manual therapy classroom teaching with assessment of psychomotor skills, which yielded 12 articles, of which three more were excluded for not being MT specific for classroom skill assessment resulting in a final MT specific assessment list of 9 references. The reasons for the three being excluded in level two were for instrument assessment type only (Horbacewicz, 2018), workplace-based global performance assessment (Cunningham & McFelea, 2017), and survey-only data (Corkery et al., 2020).

### *2.2.1. Overview of Rubrics for Procedural Skill Assessment*

A review of general principles of assessment is first described to provide background and definitions for the more focused discussion of manual therapy rubrics below. One of the goals for assessment in pre-professional healthcare education is to determine learner competency and safety before practicing on patients (Yudkowsky et al., 2020). The traditional recommendation in assessment development is to align the type of instrument with the learner's expected level of proficiency according to Miller's pyramid (1990). The higher two levels of Miller's pyramid, i.e., "shows how" and "does" are typically assessed with behavioral types of tests which require a demonstration of performance. For example, at the "shows how" level, learners are typically assessed with practical skill exams, task checks, objective structured clinical examinations (OSCEs), and simulation, versus the highest "does" level, which needs to be assessed in the workplace (Miller, 1990). The difficulty in constructing procedural skill assessments is that unlike objects that can be directly counted or knowledge that has a right or wrong answer, educational constructs such as procedural skills are latent variables that cannot be directly measured (Yudkowsky et al., 2020). Therefore, the key concept in assessment development is first to consider what the intended purpose of the new test will be, so the test construction can adequately align.

To objectively assess the performance of procedural tasks in HPE, test instruments are often created as rubrics. A "rubric" is a scoring guide used to evaluate student responses (Popham, 1997). Brookhart provides more detail by defining rubrics as “a coherent set of criteria for students’ work that includes descriptions of the level of performance quality on the criteria” (2013, p. 4.). For comparison, Yudkowsky et al. describes a rubric as a detailed guide that shows raters how to standardize the assignment of numeric scores and reduce the subjectivity of human bias (2020). Standardization of assessment is the underlying principle of rubrics by matching the performance to the description versus judging it (Brookhart, 2013).

Rubrics are usually constructed by pairing a list of the task's sub-elements as criteria with rating scales for measurement for comparison to the fidelity of the criterion (Yudkowsky et al., 2020). For clarity in this paper, the term criteria will be used to refer to the elements of the task being assessed, and scales will refer to the levels of quality of performance that are scored. Although rubrics are often thought of mostly for summative purposes, Brookhart makes the point that rubrics help coordinate both instruction and assessment (2013), a view that is echoed by others in HPE (Garcia-Ros et al., 2024; Pérez-Guillén et al., 2022). Therefore, well-written rubrics can have both a formative and a summative role for the learner and instructor (Brookhart, 2013; Popham, 1997).

Because rubrics take many forms, some additional definitions are appropriate. Formative assessment is feedback for the learner to improve their capability to perform a task. Formative feedback has been described as “information communicated to the learner that is intended to modify his or her thinking or behavior to improve learning” (Shute, 2008, p.154). Formative assessments should take place during the learner's course of study (Yudkowsky et al., 2020) to develop self-awareness and critiquing abilities (Stenberg et al., 2021). Examples of formative assessments include short quizzes, written reflections, task trainers, peer assessments and low-stakes skill checks. The benefits of formative feedback are for both learner and teacher to assess how the class is constructing knowledge, i.e., social constructivism (Dennick, 2016), and as an evaluation towards meeting the module objectives (Shute, 2008). Summative assessment is typically used at the end of a course of study with the purpose of measuring the overall achievement of the learner and providing a rating or grade of performance. Most summative assessments can be considered high-stakes and should be of sufficient quality and defendable when used for the determination of final grades or progression to clinical practice (Yudkowsky et al., 2020).

Rubrics can be further classified as global or analytic and general versus task-specific (Brookhart, 2013). Global rubrics, also referred to as holistic, are designed to capture the entire performance of a task. Global rubrics have been adapted more recently to rate entrustable professional activities (EPAs) scored as entrustment levels (Ilgen et al., 2015). Analytic rubrics are designed to evaluate separate criteria or components for each dimension of an overall task (Yudkowsky et al., 2020). General rubrics use broader criteria and descriptions of performance to allow their use across multiple similar tasks. In contrast, task-specific rubrics are designed to assess a single task, improving their inter-rater reliability versus general rubrics (Brookhart, 2013). Task-specific rubrics are similar to procedural skill checklists, which benefit novice learners by providing a process to follow or duplicate when first learning a new task (Yudkowsky et al., 2020).

Rating scales used for global or analytic rubrics can employ a range of response items to indicate quality of performance versus checklist rubrics, which only evaluate what was done or undone (Yudkowsky et al., 2020). Although checklist rubrics with dichotomous ratings have advantages of reliability due to ease of scoring (Johnston et al.,

2022) they fall short when assessing more complex task behaviors of skilled performance (Cook & Hatala, 2016; McKinley et al., 2008). For example, advanced clinicians may score low on checklist items for taking a history as they use pattern recognition and perform a more abbreviated history and physical (Yudkowsky et al., 2014).

Rubrics that use behaviorally anchored rating scales are recommended because of the ability of raters to exercise expert judgment on the quality of the learner's performance (Yudkowsky et al., 2020). An example of behaviorally scaled anchors would be teamwork skills rated on a 5-frequency scale from never to always (Brookhart, 2018). Benefits of rating scales that incorporate qualitative descriptions which are behaviorally anchored are recommended for the improvement of agreement between raters (Yudkowsky et al., 2020) and to provide formative feedback on learning (Brookhart, 2013).

### *2.2.2. Rubrics for Manual Therapy Task Assessment*

Although recommendations for TJM assessment have existed, e.g. Manipulation Education Manual for Physical Therapist Professional Degree Programs (Manipulation Education Committee of the APTA Manipulation Task Force, 2004), specific criteria for assessment have not been published. For example, in 2006 Flynn et al. described TJM assessment as consisting of two components, setup and thrust, however criteria for assessment of these components were not specified. Therefore, the second level of the literature search was to locate references with published examples of manual therapy assessment rubrics as models for a universal assessment of TJM.

The quality of the literature is reviewed first, followed by comparisons of rubric type, criteria, and scales used for assessment. The inclusion criteria were met by nine articles which described an assessment rubric for student learners of an OMPT intervention. Subjects for all nine articles were pre-professional, and pre-licensure, including seven articles of physical therapy students, and one each for medical (Gradl-Dietsch et al., 2016) and osteopathic students (Seals et al., 2016). Overall, study designs were strong for the purpose of comparing teaching methods with six of the studies using randomized comparisons, two studies which compared cohorts or teaching sections and only one study using a cross-sectional survey. In addition, seven of the studies used an assessor who was blinded to group allocation. Refer to Appendix B for comparison of included MT rubrics.

Although the study designs were strong for the nine studies with OMPT assessment rubrics, only four reported some type of attempted assessment validation with only one study completing reliability before the study began. Washmuth et al. (2020) reported consulting other academic faculty, both internal and external to their institution to revise a general TJM measure and improve its face validity. Luedtke et al. (2023) used a rubric agreed upon by other internal faculty for face validity with inter-rater reliability between two video raters for 56 students averaged across three tasks with Cohen's kappa of k=0.418. Rossettini et al. (2017) report developing a new measure specific for the first and second cervical vertebrae (C1-C2) mobilization technique, which was validated for face and content validity with five experts in teaching manual therapy and then testing with 10 students for inter-rater reliability resulting in an ICC of 0.97. Manton et al. (2023) report using a prior validated measure for lumbar manipulation with unpublished validation results and completed intra-rater reliability on 14 videos resulting in an ICC of 0.988.

All nine of the rubrics developed for MT assessment were analytic with multiple elements of the task listed as criteria. None of the nine MT rubrics assessed global ratings of performance or used an entrustment scale. Six of the rubrics could be described as general and five of these were used to assess multiple manual therapy techniques in their respective studies (Luedtke et al., 2023; McCallister et al., 2023; Pérez-Guillén et al., 2022; Seals et al., 2016; Washmuth et al., 2020). Two of the rubrics were task specific using checklist type of criteria including a C1-C2 mobilization (Rossettini et al., 2017), and lumbar manipulation (Manton et al., 2023). One article blended general with task specific criteria to create four separate but similar rubrics for a more universal assessment of TJM for the cervical, thoracic, lumbar and sacroiliac regions (Gradl-Dietsch et al, 2016).

Common criteria shared between the nine manual therapy assessment rubrics included patient position and some measure of intended effect. However the intended effect criteria was often described differently; for example grades or direction of movement and rate (Luedtke et al., 2023; McCallister et al., 2023), correct application of parameters (Cuenca-Martínez et al., 2023), application of principles (Seals et al., 2016) or simple manipulation (Gradl-Dietsch et al., 2016). All nine rubrics had some type of criteria to describe either therapists (operators) position, hand placement or both. Three of the rubrics used specific criteria for thrust including force direction, speed and / or force intensity (Manton et al., 2023; McCallister et al., 2023; Washmuth et al., 2020). Only three of the rubrics included criteria for taking up the tissue slack pre-thrust (Gradl-Dietsch et al., 2016; Manton et al., 2023; Seals et al., 2016). Most of the emphasis of the rubrics was on delivery of the intervention; however, three also included some type of criteria pre-assessment (Gradl-Dietsch et al., 2016; Seals et al., 2016), post-assessment (Seals et al., 2016), or clinical reasoning (Pérez-Guillén et al., 2022).

Scales used for the manual therapy assessment rubrics varied from dichotomous correct/incorrect to upwards of five levels. Most of the scales used global descriptors to distinguish between levels of criteria such as Novel, Competent, Proficient, and Expert (Pérez-Guillén et al., 2022) or Student, Entry level graduate, and Seasoned clinician (Washmuth et al., 2020). One of the rubrics used a scale adapted from levels of assistance often used in the PT field, that of Maximum-Moderate-Minimal-Verbal Cues and Independent (McCallister et al., 2023). The two checklist rubrics used rating scales of observable behaviors specific to the individual components of the task being assessed. For example, from the Rossettini et al. checklist rubric, task specific observable behaviors for criteria E, titled "Detection of C2" (2017) are listed below.

- The physical therapists caudal hand detects the spinous process of C2 correctly (**Score 2**)
- The physical therapists caudal hand detects the spinous process of C2 incorrectly (one level error in detecting C2) (**Score 1**)
- The physical therapists caudal hand detects the spinous process of C2 incorrectly (two or more level error in detecting C2) (**Score 0**)

### ***2.2.3. Non-Manual Therapy Rubrics for Physical Therapy***

Other high-quality rubrics developed for procedural skill assessment in PT contain similar features as those rubrics to assess manual therapy skills. Based on the updated literature review for this paper, the best examples for comparison include rubrics developed for general rehabilitation skills (Sattelmayer, K., Jagadamma, & Hilfiker,

2020) and neurologic techniques (Garcia-Ros et al., 2024; Gittinger et al., 2022). Refer to Appendix B for comparisons of the criteria used in these PT skill assessment rubrics.

Sattelmayer et al. developed and validated a general rubric to be used for multiple rehabilitation skills with six overall criteria: preparation, knowledge & decision making, safety, communication, procedure execution, and comfort each with 3-7 subcomponents and a five-level scale: very poor, poor, acceptable, good, and very good (Sattelmayer, K., Jagadamma, & Hilfiker, 2020). Garcia-Ros et al. developed a neurologic PT rubric with five criteria: physical therapist position, patient position, verbal facilitation, holds, and execution with a four-level scale of inadequate, needs improvement, adequate, and advanced. Garcia-Ros et al. report the rubric was validated by Delphi study methods on 19 different neurologic PT tasks, however publication is pending (Garcia-Ros et al., 2024; Gittinger et al., 2022). These two articles, along with the MT specific rubrics that attempted validation, suggest PT education is beginning to catch up with assessment development in medical and nursing education required for CBE.

**2.3. Assessment Validation in Medical Education**

The development of feasible and validated assessments has been recognized as an unmet need by proponents of CBE in PT (Chesbro et al., 2018; Jensen et al., 2020; Timmerberg et al., 2022). To evaluate improvement in learner outcomes in CBE an essential component is the ongoing development of validated assessments of performance. Validated performance assessments are essential to provide learner feedback (Holmboe et al., 2010) and to support reproducible and defensible decisions about learner preparation for clinical practice (Cook et al., 2014). The next section provides an overview of the recommended steps to create and validate a new measure followed by examples in medical education.

A critical review of the literature was completed for the topic of procedural skill assessment development and validation. Databases used for the search included Medline, Cumulative Index of Nursing and Allied Health Literature and Google Scholar. In the book Assessment in Health Professions Education, Lineberry makes the point that even the validity experts such as Messick, Kane and others don't agree on how to even define validity (Yudkowsky et al., 2020, p. 17). Lineberry goes on to describe the first steps in assessment creation are to describe the purpose and planned uses which then formulate a logical argument for choice of assessment method (Yudkowsky et al., 2020). Once the initial questions about an assessment are answered for purpose and planned use then the next steps are to create the measure and accumulate evidence, i.e. develop support for the argument that validates the measures intended use (Cook & Lineberry, 2016).

### *2.3.1. Five Types of Validation Evidence*

To validate a new measure, five types of evidence have been identified as content, internal structure, relationship with other variables, response process, and consequences evidence (Cook et al., 2014; Cook & Lineberry, 2016). The first three types (content, internal structure, relationship with other variables) connect with classic methods of validity and reliability formerly known as criterion, correlational, and construct validity, respectively (Cook & Beckman, 2006). Because of the recent changes the updated framework from the work of David Cook and Matthew Lineberry for developing validity evidence for a new measure are described below (Cook, 2014; Cook & Lineberry, 2016).

Content evidence describes the steps taken to ensure the assessment developed measures the construct. Content evidence can be accumulated by recording the steps taken to develop the measure such as expert review, adapting a prior measure, clinical guidelines and/or a test blueprint (Cook et al., 2014; Cook & Lineberry, 2016). Internal structure evidence evaluates relations among assessment items and how they relate to the construct. Evidence of internal structure includes item reproducibility (reliability), difficulty and factor analysis (Cook et al., 2014; Cook & Lineberry, 2016). Relations with other variables can be described as the statistical associations between scores on the new measure and other measures or connections with theoretical features. Examples of relations with other variables include correlations of scores with parallel measures, learner prior grades or performance, novice vs. experts or learner confidence (Cook et al., 2014; Cook & Lineberry, 2016).

Both response process and consequences evidence are important considerations owing to the intended purpose of the new measure according to context for evaluation of quality and feasibility (Yudkowsky et al., 2020). Response process evidence includes analysis of how raters or examinees actions align with the intended construct including thought processes, response systems and test security. Evidence of response process includes think-aloud protocols, rater training perspectives and testing procedures (Cook & Lineberry, 2016). Consequences type evidence is the impact of the measure itself and any factors that influence the rigor of decisions made from the measure. Although important as a source of validity evidence, consequences evidence is rarely reported in assessment development (Cook & Lineberry, 2016). Examples of consequences evidence include high-stakes examinations used with rigorous cut points that require standard setting procedures, e.g. Angoff method or receiver operator characteristic (ROC) curve to set cut-points for pass/fail (Cook et al., 2014) or so-called low-stakes feedback over multiple domains, e.g. professional behaviors, self-efficacy, self-directed learning that are used multiple times affecting students' development over an entire year (Cook & Lineberry, 2016).

***2.3.2. Assessment Development Case Example***

A case example in medical education is the development of a simulation assessment for neonatal tracheal intubation (NTI) from the work of Johnston and colleagues (Johnston et al., 2019). The purpose of the new assessment was to improve the skill of neonatal intubation by creating an assessment for feedback and deliberate practice (Sawyer et al., 2015) in the simulation setting (Johnston et al., 2019). The initial tool was developed by literature search to combine a published procedural task checklist with a global skills assessment and a 5-point EPA level (Johnston et al., 2019). The second step used a modified Delphi process with subject matter experts skilled in the procedure to review each of the proposed items and revise the measure. The literature search and Delphi study provided evidence for content validity (Johnston et al., 2019).

After the NTI measures development and content validation was completed, collection of additional validity evidence commenced. For internal structure (reliability), consequences and relations to other variables validation, the measure was tested by video recording 23 different participants with a wide range of experience level performing an NTI. For response process validity four blinded raters were trained and then scored each of the 23 video recordings. Relations to other variables validity evidence was demonstrated by correlations with experience level of the participants and within the measure between the checklist, global skills assessment and EPA ratings. Consequences validation was also reported by setting specific standards for the intertester reliability and developing a cut-point checklist score for entrustment of NTI (Johnston et al., 2019). Additional studies using the same measure confirmed its multidimensional validity on 58 participants of different experience levels (Whalen et al., 2022) and determined a two or three level scale for the checklist were both effective (Johnston et al., 2022).

**2.4. The Delphi Method**

A critical iterative review of the literature was completed for Delphi methods in three areas. First, an overall search of the Delphi methodology literature, second, a search for PT education specific Delphi examples and third, a search for case examples and methods for use of the content validation index (CVI) for assessment development. Databases used for the Delphi topical search included Medline, Cumulative Index of Nursing and Allied Health Literature and Google Scholar.

The term Delphi has been widely used in the health sciences literature to refer to many similar but different methods of developing expert consensus (Hasson et al., 2000; Humphrey-Murto et al., 2017). Development of the Delphi process has historically been attributed to the RAND corporation, which was a think tank used during the Cold War to provide advice to the United States Army Air Corps (Shang, 2023). The RAND corporation developed the consensus method of recruiting several military experts, asking each for anonymous feedback and repeating the process until agreement was reached (Helmer, 1967).

Although updated reporting guidelines have been recommended (Diamond et al., 2014; Jünger et al., 2017), there are no agreed-upon methods for completing a Delphi study in the health science literature (Humphrey-Murto et al., 2017; Nasa et al., 2021; Shang, 2023). Generally, methods for a “classic” Delphi study involve a first round to develop initial statements of a concept, followed by subsequent rounds to achieve consensus (Shang, 2023). The term “eDelphi” is inconsistently used to refer to an electronic version of a classic Delphi study (Shang, 2023). For clarification, in this paper, the term Delphi will be used to refer to a classic Delphi study that is completed electronically.

In a classical Delphi the first round begins with open ended questions to capture more qualitative responses which are coded into themes for expert review in round two. However in a modified Delphi the literature is first reviewed by a steering group (Nasa et al., 2021) to summarize professional guidelines and other relevant sources to develop constructs for the first round of expert review (Keeney et al., 2011; McKenna, 1994). The modified Delphi design is typical of studies that explore elements for HPE assessments where professional guidelines already exist; however, consensus is valuable to develop more specific criteria. Examples using a modified Delphi in HPE include use of two rounds to refine a pre-developed tool on NTI (Johnston et al., 2019), two rounds to clarify criteria for professional behavior assessment for an OSCE (Davies et al., 2017), and three rounds of health promotion competencies for PT based on summarized themes from professional guidelines (Magnusson et al., 2020).

The Conducting and REporting DElphi Studies (CREDES) framework provides guidelines to conduct and report Delphi studies (Jünger et al., 2017). Additional reviews (Diamond et al., 2014; Hasson et al., 2000; Humphrey-Murto et al., 2017; Nasa et al., 2021; Shang, 2023; Spranger et al., 2022) list common Delphi study components. A review of each of these five Delphi components is listed below with detail in the text that follows.

1) Identification and justification of the research problem

2) Selection and recruitment of subject matter experts

3) Managing the iterative feedback process

4) Defining consensus

5) Reporting

Identification and justification of the research problem include completing a literature search and a preliminary description of the topic to be studied (Hasson et al., 2000; Jünger et al., 2017). Selection and recruitment of the expert panel uses purposive sampling to identify group members who can contribute the most to the topic of interest. Criteria to define expertise for the panel typically include educational attainment, years of experience, and authorship of a peer-reviewed presentation/publication (Diamond et al., 2014; Nasa et al., 2021; Shang, 2023). The number of panelists for the review rounds has been reported to be greater than 100; however, at least 20-30 participants are recommended to balance response stability and research workload (Hasson et al., 2000; Nasa et al., 2021; Shang, 2023). Anonymity can be ensured by the use of electronic survey methods to hide respondents' data and prevent potentially more senior experts' opinions from undue influence of responses (Humphrey-Murto et al., 2017; Jünger et al., 2017; Shang, 2023).

Recommendations for managing the iterative feedback process include the use of a study guidance group, also referred to as a workgroup (McDevitt et al., 2022), responsible for developing and piloting questions and collating responses between survey rounds (Jünger et al., 2017; Nasa et al., 2021; Spranger et al., 2022). Recommendations of how to start round one varies in the literature; however, usually, the study guidance group interprets the results of the literature search and creates more open-ended qualitative-type questions (Hasson et al., 2000; Jünger et al., 2017; Spranger et al., 2022). Responses to these questions are then coded to identify themes formed into statements proposed in subsequent survey rounds using agree-disagree response scales (Hasson et al., 2000; Jünger et al., 2017). Delphi studies include at least two, and upwards of 10 survey rounds depending on the definition of consensus and stopping point (Humphrey-Murto et al., 2017).

The definition of consensus includes methodological considerations for agreement or non-agreement and stability of responses (Diamond et al., 2014; Jünger et al., 2017). Recommendations include determining an *a priori* threshold level, e.g., 75%, depending on the implications of the outcome, at which the expert responses agree (Shang, 2023). A key point is that the threshold level also applies to disagreement of items, which would result in the item being removed (Jünger et al., 2017; Nasa et al., 2021). The term stability has been used to refer both to the level of agreement of items and the overall determination of consensus for when to terminate the Delphi study (Diamond et al., 2014; Nasa et al., 2021; Shang, 2023). Statistical methods to determine agreement include percentage of agreement, measures of central tendency, and interquartile range (Shang, 2023). Higher-quality Delphi studies used the determination of consensus versus a predetermined number of rounds to determine when to terminate the study (Diamond et al., 2014).

### *2.4.1. Delphi Examples in Physical Therapy Education*

Specific to PT, the Delphi method has been used to establish criteria for measurement of professional behaviors during OSCEs in Canada (Davies et al., 2017) and for determining competencies of health promotion and wellness (Magnusson et al., 2020). Specific to manual therapy, prior consensus documents have been published on recommended instructions for teaching manipulation, e.g., the Manipulation Education Manual for Physical Therapist Professional Degree Programs (Manipulation Education Committee of the APTA Manipulation Task Force, 2004), what to teach for competency development (Sizer et al., 2008) and what paradigms of MT to teach (Keter et al., 2023). In addition, a recent Delphi identified what characteristics are important for therapist and patient during side-lying neutral gap manipulation, however these task elements were not listed as criteria and paired with quality levels or scales for performance assessment (O’Donnell et al., 2016). Currently to the author’s knowledge there are no consensus documents for criteria and quality scales to be used for TJM skill acquisition assessment rubric in physical therapy students.

### *2.4.2. Delphi Use in Assessment Development*

Examples of HPE procedural task assessments developed with the Delphi method include tasks of neonatal tracheal intubation (NTI) (Johnston et al., 2019), a cystectomy assessment and surgical evaluation (CASE) (Hussein et al., 2018), and a point of care ultrasound (POCUS) checklist (Soni et al., 2022). All three of these skill assessments used similar methods for development and accumulation of multiple types of validity evidence (Cook et al., 2014; Cook & Lineberry, 2016; Messick, 1995).

For the NTI skill a literature review of professional standards was conducted and with guidance from a core group to develop the initial assessment (Johnston et al., 2019). For the CASE and POCUS tools a small group of experts first created lists of criteria. Next a modified Delphi technique with an expert panel of 12 for NTI (Johnston et al., 2019), 10 for the CASE (Hussein et al., 2018) and 14 for POCUS (Soni et al., 2022) provided feedback to review and revise the measure. Level of agreement was reported as greater or equal to 75% for the CASE (Hussein et al., 2018), 80% for the POCUS assessment (Soni, 2022), however was not reported for the NTI assessment (Johnston et al., 2019). After consensus was reached the measures were pilot tested using video recordings of students and doctors of different experience levels for review by multiple examiners for inter-rater reliability and relations to other variables validity evidence (Cook & Hatala, 2016).

### *2.4.3. Content Validation Index in Assessment Development*

An important distinction between the Delphi methods to develop the NTI and POCUS instruments are that for the CASE tool, a CVI was used for agreement (Hussein et al., 2018). The CVI has been described as a measure of relevance of items to the research goal and so can have different rating criteria depending on the needs for the expert review (Madadizadeh & Bahariniya, 2023). Although the authors of the CASE tool do not specify specific criteria used for CVI, others have reported up to five separate criteria to rate each item of a proposed tool (Madadizadeh & Bahariniya, 2023; Torres-Narváez et al., 2018). For example, in the expert review of a clinical assessment developed for PT students in Spanish speaking countries a five criteria CVI scale was used for relevance, sufficiency, pertinence, coherence and clarity. However, these constructs were not defined or referenced in the article (Torres-Narváez et al., 2018).

A review of CVI use reports that most toolmakers have reduced to a three criteria scale for clarity, simplicity and relevance (Madadizadeh & Bahariniya, 2023). For example, a recent Delphi to develop a patient reported experience measure incorporated three criteria: clarity, importance and relevance (Bull et al., 2022). To determine if an item was dropped from the next round, an 80% agreement level was used based on importance and relevance, however if lower than 80% for clarity or comments were provided the item was kept and revised for the next round. CVI's have also been averaged to develop a total scale CVI referred to as S-CVI (Polit et al., 2007). In any case, the examples provided above underscore the latitude provided in Delphi study designs to align with the needs of the researchers goals for accumulating validity evidence (Cook & Hatala, 2016).

# Chapter Three Methods

## 3.1. Study Design

An expert consensus method using a multiple round Delphi was used to collect validation evidence for the overall goal to create a universal assessment of TJM task performance in novice learners. A Delphi has been described as an exploratory survey study design collecting data with both quantitative and qualitative approaches within an interpretative paradigm of social constructivism (Hanafin, 2004; Keeney et al., 2011). Consensus methods such as a Delphi study are justified for step one of the overall aims of this research line to develop a validated universal TJM assessment since no current assessment exists.

A proposed assessment for TJM tasks was developed based on an educational model, informed by a review of the literature as well as feedback from a national presentation and publication (Adams et al., 2024). The proposed universal TJM assessment was named the Quick Psychomotor Operator Procedural Assessment for TJM (QPOPA TJM or QPOPA for short), reflecting its intended use as a short formative assessment for a novice learner during the early practice phase of a new TJM task. The QPOPA consisted of two overall categories termed setup and thrust, with five and three individual components respectively thought to represent the critical criteria for feedback on psychomotor performance for most TJM tasks.

The guidance on conducting and reporting Delphi studies (CREDES) framework (Jünger et al., 2017) and additional resources (Nasa et al., 2021; Spranger et al., 2022) were used to design the content validation Delphi study for the QPOPA on these five steps: 1) identification and justification of the research problem, 2) selection and recruitment of subject matter experts, 3) managing the iterative feedback process, 4) defining consensus, and 5) reporting including the statistical analysis plan. Each of these five steps are described below.

### ***3.1.1. Identification and Justification of the Research Problem***

Although Delphi studies have been completed to determine content for OMPT competencies (Sizer et al., 2008), what paradigms to teach (Keter et al., 2023), and lists of manipulation characteristics (O'Donnell et al., 2016), no agreed upon assessment currently exists for TJM tasks. In addition, the literature review for the new measure revealed a wide range of assessment criteria and scales that lack definition or consensus. Lack of a validated assessment limits the ability to compare teaching methods for TJM which restricts the future scholarship of teaching for OMPT.

### ***3.1.2. Sampling: Subject Matter Experts and Recruitment***

Subject matter experts were identified as those who teach TJM tasks to entry-level PT students and novices. For consistency subject matter experts were recruited that teach or live within the United States to ensure similar expectations of the educational system. To develop the recruitment email list potential subject matter experts must have met at least one of the following criteria:

1) Faculty who currently teach OMPT content in entry-level, resident, or fellowship programs.
2) Presenters at national level PT conferences including the American Academy of Orthopaedic Manual Physical Therapists (AAOMPT) or the APTA Combined Sections Meeting in the last five years.
3) Contributing author of TJM teaching methods or assessment publications within the last five years.

The primary means of recruitment was by email. Emails were sent to specific addresses where available to experts who met one of three criteria above and by emailing the chair of orthopedic residency and MT fellowship programs to forward it to the faculty responsible for teaching OMPT content. Verification of inclusion in one of three subject matter expert criteria was completed by cross reference with participant's demographic data reported as part of the electronic survey in round one. The goal for recruitment and retention was to have at least 20 participants complete the final round. Based on a 25% response rate from Delphi studies recruiting similar OMPT experts, at least 118 invites needed to be sent for Round One. This would yield 30 participants for Round One at a 25% response rate and provide a buffer for decay to upwards of three rounds. Retention of experts for subsequent rounds was promoted by reminding participants of the purpose for the research and keeping the time that elapses between rounds to one month or less.

Informed consent was provided for each round of the Delphi surveys. The research questions and consent information were provided as the first viewable survey page once a potential respondent clicked the link provided to the survey in the recruitment email. After reading the informed consent cover page of the survey, the potential respondent needed to click “I Agree - Continue to Survey” to proceed to the survey questions for each round.

### *3.1.3. Instruments and Managing the Feedback Process*

The modified Delphi process was managed by the primary investigator (PI) and guided by the core workgroup. All members of the core workgroup meet the criteria as subject matter experts outlined above. The core workgroup consisted of five full-time OMPT faculty, including two who were co-presenters and authors along with the primary investigator of the background article (Adams et al., 2024). The draft survey instrument for Round One was developed from the published background article with feedback from the workgroup.

Round One survey questions were created prior to recruitment of subject matter experts. To frame the expert review for Round One, subject matter experts were asked to consider the proposed framework as a whole and how the individual components could be used to provide feedback to a novice learner of most TJM techniques. The overall category descriptions (setup and thrust) and individual components were rated for content validity, relevance and clarity on a 4-point agree-disagree scale, followed by an open text box for suggested amendments to the descriptions. At the end of the setup and thrust sections review questions were added to rate the overall relevance and clarity and provide comments to evaluate the entire category. The survey included demographic questions to summarize the expertise of the survey participants. The final Round One survey consisted of 12 items for content validity followed by 11 demographic questions.

To execute Round One the web platform Qualtrics was used to host the electronic survey. The PI (WK) was responsible for the logistics of the Delphi study including entering the survey questions into Qualtrics, managing email reminders, de-identification and downloading response data. Written responses to revise or suggest new items were collated and then thematically coded by the PI, and one of the members of the workgroup (MD). Any disagreements were arbitrated by a third reviewer (KA). Three of the five members of the workgroup (KA, MD, WK) wrote questions for Round Two with review from the other two workgroup members (AM, LP). Survey questions and email communications were sent to an external review group for feedback prior to sending out as second round to be reviewed by the experts. The time between rounds was four weeks allowing for thematic coding of responses and revision of rubric criteria. Only subjects that responded to Round One and submitted their email address were included in the Round Two survey.

Delphi Round Two was planned by the workgroup prior to the recruitment of subject matter experts. The purpose of Round Two was to review the results from Round One and consider a proposed quality rating scale for the QPOPA. Holding the review of the quality rating scale until Round Two was proposed to make the Round One survey shorter and more manageable for the expert reviewers. The proposed three-level quality rating scale was constructed by the workgroup to score each criterion of the QPOPA. The three-level rating scale is described below.

0 = Not present or insufficient so that the intended aspect would be ineffective

1 = Partially present, the skill aspect is evolving

2 = Present and competent, the skill is adequately demonstrated

To assess the proposed three-level quality rating scale, the workgroup kept the original 75% threshold for CVI. The Round Two survey was tested and reviewed by all members of the workgroup before being sent to the expert review panel.

### *3.1.4. Defining Consensus*

Based on the research aim to develop a universal assessment that can be used to assess TJM techniques, and the expected difference in opinion for assessing student performance of a TJM task, the level of consensus for this study was set *a priori* at 75% or greater for agreement. A four-point scale was designed without a neutral, e.g. strongly agree, agree, disagree and strongly disagree. Two measures of content validation were used to assess agreement: 1) relevance and 2) clarity. The criteria of relevance was defined as the relative importance of assessing the item for performance of a TJM. Clarity was defined as the adequacy of the description provided for the assessment criteria of the TJM to provide feedback to a novice learner of the task.

For individual items to remain after each round of the Delphi, both relevance and clarity levels of agreement were specified before the start of the study to be greater than or equal to 75%. If an item scored on both relevance and clarity from 74-26% the item needed to be revised, and if < or equal to 25% the item was to be removed. In addition, if an item scored 75% or higher for relevance but lower for clarity, or vice versa, the item was to be revised by the workgroup for consideration in the next round.

#### ***3.1.5. Statistical Analysis***

The demographic data from the subject matter experts was analyzed with descriptive statistics. Narrative responses from participants for round one and round two were coded to analyze themes based on constructivist theory. CVI was calculated as a ratio by the number of raters giving an item either agree or strongly agree divided by the total number of raters (Madadizadeh & Bahariniya, 2023). Polit notes that a 75% level of agreement using the CVI construct corresponds to a kappa value considered excellent with at least ten expert raters (2007). The overall CVI, also known as the scale CVI (S-CVI) was calculated by averaging all the individual item CVI's (Polit, 2007).

### **3.2. Institutional Review Board**

The Radford University Institutional review board (IRB) approved the study (Number 2024-116). Informed consent was provided electronically as the first page viewable for each survey round. All data was de-identified by the PI and protected by storing it on a Radford University password protected computer. Any data shared with the co-investigators and used for publication was de-identified and shared en masse to prevent the identity of the survey participants.

# Chapter Four Results

## 4.1. Participants Data

The response rate was 13.8% for Round One, with 66 respondents of 480 invites sent who agreed to continue by reporting their email for future rounds. Qualifications as subject matter experts were established by meeting at least one of the three inclusion criteria. The demographic data to describe the characteristics of the respondents is presented in table 1 below. The means and standard deviations (SD) of demographic data were calculated by using standard formulas in the Microsoft® Excel® (version 2501) 64-bit application. Respondents represented all five geographical regions of the United States, with the highest number reporting teaching in the Southeast 21 (31.8%), followed by the Midwest 15 (22.7%) and the Northeast 13 (19.7%). The respondents had practiced PT for a mean of 19.9 years (SD = 10.9) and taught PT for a mean of 10.4 years (SD = 9.8). Over 75% had attained the clinical doctorate of PT, and over 56% achieved a terminal academic doctoral degree, e.g., PhD or EdD. Participants possessed post-professional training in manual therapy, as reported by 95.3% completing advanced certifications and 84.2% having completed an orthopedic residency. Demographic data of the respondents are presented in Table 1.

**Table 1**

*Characteristics of Subject Matter Experts*

| **Panel Characteristics** | | n | (%) |
|---|---|---|---|
| **Region where teaching** | | 66 | |
| | Northeast | 13 | 19.7 |
| | Southeast | 21 | 31.8 |
| | Midwest | 15 | 22.7 |
| | Southwest | 8 | 12.1 |
| | West | 9 | 13.6 |
| **Clinical Degree** | | 66 | |

| | | | |
|---|---|---|---|
| | Bachelors | 5 | 7.6 |
| | Masters | 6 | 9.1 |
| | DPT | 50 | 75.8 |
| | Other | 5 | 7.6 |
| **Education Terminal Degree** | | 66 | |
| | PhD | 18 | 27.3 |
| | DHS | 1 | 1.5 |
| | DS | 7 | 10.6 |
| | EdD | 3 | 4.5 |
| | ScD | 2 | 3.0 |
| | Other | 4 | 6.1 |
| | No terminal degree | 31 | 47.0 |
| **Primary Employment** | | 66 | |
| | Faculty | 50 | 75.8 |
| | Clinician | 7 | 10.6 |
| | Business Owner | 6 | 9.1 |
| **APTA Residency** | | 57 | |
| | Orthopedics | 48 | 84.2 |
| | Sports | 5 | 8.8 |
| **Post Professional Education** | | 64 | |
| | Manual Therapy | 61 | 95.3 |
| | Dry Needling | 28 | 43.8 |
| | Chronic Pain | 9 | 14.1 |

Note: Percentages of some panel characteristics are representative of data reported versus being reported in total.

### 4.2. Results Round One

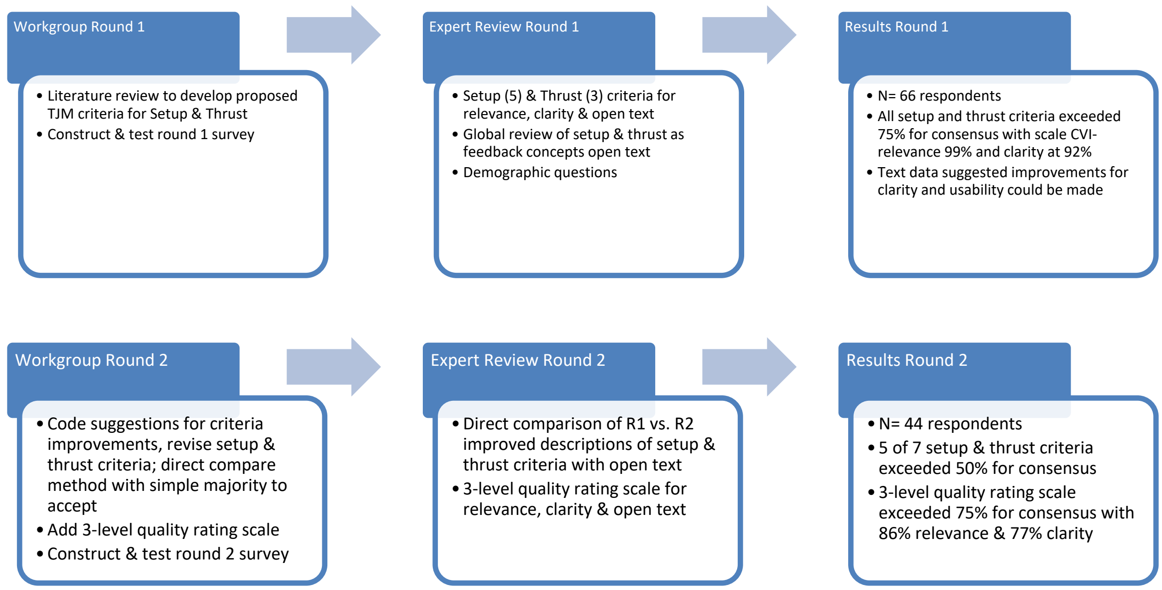
The percentage agreement for CVI of all 12 items for TJM criteria in Round One was 94% or higher for relevance and 83% or higher for clarity. Total scale CVI scored 99% for relevance and 92% for clarity the five thrust and three setup criteria. Total scale CVI was determined by taking the average scores of all individual items for the final measure not including the overall review questions. The hypothesis for RQ1 is accepted as all statements for setup and thrust exceeded the 75% threshold for agreement.

Although the percentage agreement of CVI exceeded 75%, the workgroup noted several themes from the expert comments that warranted improvement. Therefore, the workgroup decided Round Two of expert review would be applicable as an attempt to improve clarity and ease of use for the measure. Comments were independently coded for themes to improve clarity by two members of the workgroup (MD, WK), and any disagreements were resolved by a third reviewer (KA). Based on these themes, changes in wording were proposed to improve the measure in subsequent rounds. For example, the term “skin lock” was updated to “pre-thrust tissue tensioning” as potentially a more global description that could apply to most TJM techniques. Other changes proposed for Round Two included replacing the terms “learner” or “student” with “operator” for consistency throughout the measure and to encompass possible use in the pre-and post-professional settings.

To construct the survey for Round Two the workgroup used a direct comparison method, asking expert reviewers to select from two options with the prior agreed-upon statement first, followed by the proposed improved description. Because all statements from Round One exceeded the 75% threshold for CVI on the assessment concepts, a simple majority, e.g., >50%, to accept wording changes were agreed upon by the workgroup for the possible improvements in Round Two.  A summary of the iterative Delphi process used for the two rounds is presented in Figure 1.

**Figure 1**

*Delphi Process and Summary of Results by Round*




Note: legend for abbreviations include content validity index (CVI), Number (N), Round one (R1), Round two (R2), thrust joint manipulation (TJM).

### 4.3. Results Round Two

There were 44 subject matter experts that responded and completed the second round. All but two of the proposed changes to setup and thrust criteria exceeded the simple majority, i.e., > 50%, to accept wording changes agreed upon by the workgroup for potential improvements in Round Two. The percentage acceptance for the Round Two wording changes are presented in Table 2 below.

**Table 2**

*Comparison of Round 1 Versus Round 2 Descriptions and Percentage of Acceptance or Non-Acceptance for Round 2*

| Question Number | Version A (Round 1) | Version B (Proposed Round 2) | Percentage Acceptance for Version B (Round 2) |
|---|---|---|---|
| **Q1** | Setup: all of the psychomotor steps that occur before the thrust. | Setup: all of the psychomotor steps that occur before and may continue during the thrust. | 0.5 |
| **Q2** | Setup consists of 5 criteria: Body Mechanics, Spatial Orientation, Point of Application, Skin Lock and Pre-Manipulative Load/Hold. | Setup consists of 5 criteria: Body Mechanics, Spatial Orientation, Point of Application, Pre-Thrust Tension, and Pre-Manipulative Load/Hold. | NG |
| **Q3** | Body Mechanics: The term used to describe how someone moves during activities. When someone moves or carries their body improperly, undue stress may be put on various parts of the body, possibly leading to impaired safety with those activities. | Body Mechanics: The manner in which the operator moves to maximize operator comfort, efficiency, and effectiveness during the TJM. | 0.977 |
| **Q4** | Spatial Orientation: The ability to align the learner's body and maintain orientation relative to that of the patient prior to and during the TJM. | Spatial Orientation: The ability to align the operator's body and maintain orientation relative to that of the appropriately positioned patient prior to and during the TJM. | 0.682 |
| **Q5** | Point of Application: The point of contact that occurs between the learner and the patient during manual therapy techniques. It may be the learner's hand, sternum/ribs, forearm, etc. | Point of Application: The point of contact that occurs between the operator and the patient during manual therapy techniques. It may be the operator's hand, sternum/ribs, forearm, etc. | NG |
| **Q6** | Skin lock: The process of taking up slack in the soft tissues at the point of contact when performing a TJM. | Pre-Thrust Tissue Tensioning: The process of taking up the slack to achieve proper tension and localization of the target tissues. | 0.75 |
| **Q7** | Pre-manipulative Load/Hold: The initial gradual application of force that is administered prior to a TJM. | Pre-manipulative Load/Hold: Holding a patient's joint(s) in the position of the TJM for a period of time while monitoring the patient's response. | 0.659 |
| **Q8** | Thrust can be described as: the psychomotor components to optimize forces delivered when performing a TJM. | | NG |

| | | | |
|---|---|---|---|
| **Q9** | Thrust consists of 3 criteria: Direction of Force, Amplitude of Force and Rate of Force. | | NG |
| **Q10** | Direction of Force: The course along which a force vector or TJM is provided. For example, anterior to posterior, medial to lateral, upglide, or long axis distraction. | Direction of Force: The course along which a force vector is applied for a successful TJM while avoiding a pre-thrust back-off. For example, anterior to posterior, medial to lateral, upglide, or long axis distraction. | 0.364 |
| **Q11** | Amplitude of Force: The displacement, magnitude, extent or size of the force that occurs during a TJM. | | NG |
| **Q12** | Rate of Force: The rate or velocity at which the force is applied. | Rate of Force: The speed at which the force is applied. | 0.727 |

Note: Not given (NG), Thrust joint manipulation (TJM)

The proposed change (see Table 2, question 1) to describe setup as continuing during thrust was 50% suggesting non-consensus; therefore, the workgroup decided to keep the original definition from Round One. The attempt to improve the item direction of force (see Table 2, question 10) by adding "avoiding a pre-thrust back-off" only received 36% in Round Two. The workgroup noted that the original definition of the direction of force implies avoiding a pre-thrust back-off, and the original Round One description was accepted.

For the three-level quality rating scale, the CVI scored 86% for relevance and 77% for clarity. The hypothesis for RQ2 was accepted as all statements exceeded the 75% threshold for consensus for the quality rating scale. The workgroup decided to postpone RQ3, global rating scale, and RQ4 to list the appropriateness of the new tool for specific TJMs to the next phase of the validation where implementation would provide examples of needed improvements versus the use of the Delphi survey method. The final version of the QPOPA combining the setup and thrust criteria along with the quality rating scale, is presented in Table 3 below.

**Table 3**

*QPOPA Version 1 as Developed by Kolb, W. et al., in Pursuit of the DHSc*

| **TJM Feedback Criteria** | **TJM Quality Scale** | | |
|---|---|---|---|
| **Setup: all of the psychomotor steps that occur before the thrust.** | 0 = Not present or insufficient so that the intended aspect would be ineffective | 1 = Partially present, the skill aspect is evolving | 2 = Present and competent, the skill is adequately demonstrated |
| Body Mechanics: The manner in which the operator moves to maximize operator comfort, efficiency, and effectiveness during the TJM. | | | |
| Spatial Orientation: The ability to align the operators body and maintain orientation relative to that of the patient prior to and during the TJM. | | | |
| Point of Application: The point of contact that occurs between the operator and the patient during manual therapy techniques. It may be the operator's hand, sternum/ribs, forearm, etc. | | | |
| Pre-Thrust Tissue Tensioning: The process of taking up the slack to achieve proper tension and localization of the target tissues. | | | |
| Pre-manipulative Load/Hold: Holding a patient's joint(s) in the position of the TJM for a period of time while monitoring the patient's response. | | | |
| **Thrust: the psychomotor components to optimize forces delivered when performing a TJM.** | 0 = Not present or insufficient so that the intended aspect would be ineffective | 1 = Partially present, the skill aspect is evolving | 2 = Present and competent, the skill is adequately demonstrated |
| Direction of Force: The course along which a force vector or TJM is provided. For example, anterior to posterior, medial to lateral, upglide, or long axis distraction. | | | |
| Amplitude of Force: The displacement, magnitude, extent or size of the force that occurs during a TJM. | | | |
| Rate of Force: The speed at which the force is applied. | | | |

Note: Final version of the Delphi consensus formative assessment for thrust joint manipulation (TJM) with criteria and quality rating scale combined.

## Chapter Five Discussion

This study aimed to develop universal criteria to provide formative feedback for the psychomotor aspects of TJM tasks for learners in the early skill-practice phase. The modified Delphi process was used with consensus from experts in manual therapy to develop content validity as the first phase of establishing validity evidence for the QPOPA. Although many MT rubrics have been created to assess effectiveness for single studies of teaching methods, e.g., Peyton's (Gradl-Dietsch et al., 2016; Rossettini et al., 2017) or pre-practice video watching (Luedtke et al., 2023; Seals et al., 2016), this line of research takes a broader view to construct and validate assessment criteria first. The goal of standardizing TJM assessment criteria is to establish a foundation for multi-site educational research on the effectiveness of future teaching methods.

Assessment literacy considers standards and purposes for the selection of assessments, including alignment with learning outcomes and how to provide the most significant impact on learning (Chan, 2022). Based on models of procedural skill teaching (Bilyeu et al., 2023; Nicholls et al., 2016), the QPOPA was developed to standardize formative feedback for novice learners early in the skill-practice, i.e., when psychomotor performance is highly variable and the learner can benefit greatly from formative feedback. The Delphi process was used to develop consensus for a set of coherent criteria and a quality rating scale for performance levels to use the QPOPA as a general feedback rubric (Brookhart, 2013). The QPOPA is designed to fill a void in TJM instruction as a formative feedback tool that makes explicit the aspects of performing a TJM task, i.e., setup and thrust criteria, to help learners make sense of information and enhance their own learning strategies.

**5.1. Relationship With Other Assessments**

Similar to other reported general manual therapy assessments, the newly constructed QPOPA incorporates a motor learning theoretical framework and describes several criteria as components of the overall task (Cuenca-Martínez et al., 2023; Luedtke et al., 2023; McCallister et al., 2023; Pérez-Guillén et al., 2022; Seals et al., 2016; Washmuth et al., 2020). The QPOPA is different as a universal rubric to provide feedback for most TJM motor tasks, compared to other previously created rubrics formulated as checklists for specific manual therapy techniques; for example, a C1-C2 mobilization (Rossettini et al., 2017) or lumbar manipulation (Manton et al., 2023). Each MT skills rubric is optimally used for feedback, summative assessment, or both, benefiting learners, teachers, and institutions. While task-specific checklists and rubrics with predefined criteria enhance reliability and internal validity for raters, their dichotomous outcomes limit the ability to assess the adaptive quality of more complex skills (Cook & Hatala, 2016; Johnston et al., 2022; Yudkowsky et al., 2020).

The iterative nature of a new measure's creation is illustrated by this study as the first phase of gathering validation evidence to inform content validity (Cook & Lineberry, 2016; Kane, 2001). The modified Delphi process strengthened content validation for the QPOPA, including the workgroups literature review, presentation and publication which informed the first and second rounds' refinement of the prior accepted criteria to improve usability. The consensus method was based on the CVI concept, with the QPOPA achieving an overall or total S-CVI of 99% for relevance and 92% for clarity in Round One. These ratings were determined for five setup and three thrust criteria by a large sample of 66 manual therapy experts. Overall CVI-scale scores of .90 or higher are considered excellent agreement, suggesting strong conceptual development, well-written items and clear instructions to experts (Polit et al., 2007). The workgroup interpreted the high CVI scores for setup and thrust criteria in round one as evidence of strong conceptual development. Given the iterative nature of assessment creation, the workgroup took an additional step to refine the wording for improved clarity and ease of use as feedback during the early practice phase. Results from the second round produced some desirable improvements such as simplifying the description for “body mechanics” and changing “skin lock” to a more global term of “pre-thrust” tissue tension. The workgroup determined that a third Delphi round was not necessary as the goal of developing consensus for content validity evidence was achieved following the first two rounds with greater than 90% CVI agreement for setup and thrust criteria and consensus above threshold for the three-level quality scale, making the feedback for the measure complete. The future steps in the process of developing validation evidence include reliability testing, response process and consequences (Cook & Hatala, 2016) and are needed to continue to refine the measure.

Future uses of the QPOPA include standardization of feedback for instructional laboratory teaching, incorporating the measure into longer summative assessments, and use as a universal tool for multi-site educational research of psychomotor/procedural teaching methods in both pre- and post-professional settings. The primary purpose of this measure's development is to standardize feedback on the common components of TJM techniques. Advantages of a universal TJM assessment like the QPOPA include standardized criteria and terminology for both learners and teachers, reducing cognitive load and enhancing feedback effectiveness when introducing a new TJM technique in a different body region. Results from a prior study of a global rubric for neurological PT skills noted that advantages include alignment of feedback during practice and the explicit nature of future summative assessment (Garcia-Ros et al., 2024). Although the intended use of the QPOPA is to provide formative feedback during the early practice phase, these criteria could be combined with other domains, e.g. safety, communication, or clinical reasoning as part of a more comprehensive assessment rubric to be used for summative purposes. Universal expert developed assessments of procedural skills have served as the foundation for future educational research in healthcare, and are needed in physical therapy (Jensen et al., 2020). Universally validated expert criteria for procedural skills in medical education may provide a professional competency based assessment for learners across different sites (Hussein et al., 2018; Soni et al., 2022). These criteria also provide a foundation for future research on teaching and assessment methods including simulation (Johnston et al., 2022; Whalen et al., 2022).

**5.2. Limitations**

Several limitations for the proposed TJM measure should be noted which may be attributed to the framing of the measures (QPOPA) intended use and which are inherent to content validation using a modified Delphi process. First, the nature of the Delphi process has inherent limitations, including the development of consensus from subject matter experts. In general, experts unknowingly have blind spots related to their performance, including their unique viewpoints on skill instruction (Ambrose et al., 2010), which could hinder the usefulness of TJM feedback criteria for the novice operator for which the measure is intended. Although a broad recruitment strategy was used to improve the diversity of MT expert respondents, no attempt was made to record MT fellowship programs or associated patient management philosophies. It is possible that the MT respondents represent a limited sample of MT fellowship programs and, therefore, may limit generalizability to others. Second, framing the content validation for Round One was based on the prior literature review, presentation and publication by three of five members of the workgroup and may have biased the Delphi process. However, testing survey questions by all members of the workgroup and the high CVI results from Round One reduce this possibility. A third limitation is based on the method of task deconstruction for procedural skill instruction and the QPOPA's intended use, i.e., to place emphasis on the learning process for a novice learner in the early practice phase of skill development. Criteria developed for the QPOPA focus on the psychomotor aspects of the task by removing communication and clinical reasoning, which are essential for higher fidelity or workplace-based assessments. Finally, the QPOPA has not yet been assessed for reliability, external validation, or educational consequences, which are components of the next phases of gathering additional evidence with piloted implementation.

**5.3. Conclusion**

A panel of expert manual therapy educators validated a new TJM assessment (QPOPA) to provide universal feedback criteria for novice learners in the early practice phase of psychomotor skill development. The benefits of universal feedback criteria for the psychomotor aspects of TJM tasks include the alignment of teaching, learning, and assessment of complex procedural skills such as TJM techniques across body regions. The benefits of standardized QPOPA TJM criteria include assisting students by providing explicit descriptions of criteria for self-directed learning and enabling faculty to assess the impacts of diverse teaching methods. Potential benefits of standardized TJM assessment criteria for the PT profession include multi-site studies to investigate teaching strategies and improved TJM skills from didactic training to increase TJM use in the context of clinical experiences.

## References

Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., & Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology: CB*, *21*(7), 557–562.

Adams, K. R., Kolb, W. H., Geiser, M. B., & Dolphin, M. (2024). A novel model for developing thrust joint manipulation skills: A teaching and learning perspective. *The Journal of Manual & Manipulative Therapy*, 1–9.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.

Amineh, R. J., & Asl, H. D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages*.

Bannister, S. L., Hilliard, R. I., Regehr, G., & Lingard, L. (2003). Technical skills in paediatrics: A qualitative study of acquisition, attitudes and assumptions in the neonatal intensive care unit. *Medical Education*, *37*(12), 1082–1090.

Bastian, A. J. (2008). Understanding sensorimotor adaptation and learning for rehabilitation. *Current Opinion in Neurology*, *21*(6), 628–633.

Blanpied, P. R., Gross, A. R., Elliott, J. M., Devaney, L. L., Clewley, D., Walton, D. M., Sparks, C., & Robertson, E.K. (2017). Neck pain: Revision 2017: Clinical practice guidelines linked to the international classification of functioning, disability and health from the orthopaedic section of the American Physical Therapy Association. *Journal of Orthopaedic & Sports Physical Therapy*, *47*(7), A1-A83.

Bilyeu, C. A., McDevitt, A. W., & Judd, D. L. (2023). A blended approach to developing psychomotor skills in novice learners in a Doctor of Physical Therapy curriculum. *Medical Teacher*, 1–8.

Bourassa, M., Kolb, W. H., Cunningham, S., Neurauter, V., Bourassa, A., Grieshaber, E., Loyd, K., Walden, R. R., & Wassinger, C. (2024). Psychomotor skills assessment in healthcare education: A scoping review. *Physical Therapy Reviews: PTR*, 1–23.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, *3*. https://doi.org/10.3389/feduc.2018.00022

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Harvard University Press.

Bull, C., Crilly, J., Latimer, S., & Gillespie, B. M. (2022). Establishing the content validity of a new emergency department patient-reported experience measure (ED PREM): A Delphi study. *BMC Emergency Medicine*, *22*(1), 65.

Burgess, A., van Diggele, C., Roberts, C., & Mellis, C. (2020). Tips for teaching procedural skills. *BMC Medical Education*, *20*(Suppl 2), 458.

Cecilio-Fernandes, D., Medina-Ramírez, R., Sandars, J., & Costa, M. J. (2023). Self-regulated learning processes across different physiotherapy clinical procedural skills and time intervals: A SRL microanalysis study. *Medical Teacher*, *45*(10), 1170–1176.

Chan, C. K. Y. (2022). *Assessment for experiential learning*. Routlege, Taylor & Francis

Group.

Chesbro, S. B., Jensen, G. M., & Boissonnault, W. G. (2018). Entrustable professional activities as a framework for continued professional competence: Is now the time? *Physical Therapy*, *98*(1), 3–7.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*(2), 166.e7–e16.

Cook, D. A., & Hatala, R. (2016). Validation of educational assessments: A primer for simulation and beyond. *Advances in Simulation (London, England)*, *1*, 31.

Cook, D. A., & Lineberry, M. (2016). Consequences validity evidence: Evaluating the impact of educational assessments. *Academic Medicine: Journal of the Association of American Medical Colleges*, *91*(6), 785–795.

Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education: Theory and Practice*, *19*(2), 233–250.

Corkery, M. B., Hensley, C. P., Cesario, C., Yen, S.-C., Chui, K., & Courtney, C. (2020). Use of thrust joint manipulation by student physical therapists in the United States during clinical education experiences. *The Journal of Manual & Manipulative Therapy*, *28*(5), 266–274.

Cuenca-Martínez, F., Suso-Martí, L., Peréz-Domínguez, B., Calatayud, J., López-Bueno, R., Gargallo, P., Blanco-Díaz, M., & Casaña, J. (2023). Movement representation strategies as a tool for educational innovation in physiotherapy students: A randomized single-blind controlled-pilot trial. *International Journal of Environmental Research and Public Health*, *20*(5). https://doi.org/10.3390/ijerph20054473

Cunningham, S., & McFelea, J. (2017). The influence of an orthopedic, manual therapy residency program on improved knowledge, psychomotor skills, and clinical reasoning in Nairobi, Kenya. *Frontiers in Public Health*, *5*, 55.

Cutrer, W. B., Miller, B., Pusic, M. V., Mejicano, G., Mangrulkar, R. S., Gruppen, L. D., Hawkins, R. E., Skochelak, S. E., & Moore, D. E., Jr. (2017). Fostering the development of master adaptive learners: A conceptual model to guide skill acquisition in medical education. *Academic Medicine: Journal of the Association of American Medical Colleges*, *92*(1), 70–75.

Davies, R., Ellerton, C., & Evans, C. (2017). Reaching consensus on measuring professional behaviour in physical therapy objective structured clinical examinations. *Physiotherapy Canada. Physiotherapie Canada*, *69*(1), 65–72.

Dennick, R. (2012). Twelve tips for incorporating educational theory into teaching practices. *Medical Teacher*, *34*(8), 618–624.

Dennick, R. (2016). Constructivism: Reflections on twenty five years teaching the constructivist approach in medical education. *The Journal of the International Association of Medical Science Educators: JIAMSE*, *7*, 200–205.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society*, *24*(3), 177–181.

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine: Journal of the Association of American Medical Colleges*, *90*(11), 1471–1486.

Evans, D. W., & Lucas, N. (2023). What is manipulation? A new definition. *BMC Musculoskeletal Disorders*, *24*(1), 194.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. https://psycnet.apa.org/fulltext/1967-35040-000.pdf

Flynn, T. W., Wainner, R. S., & Fritz, J. M. (2006). Spinal manipulation in physical therapist professional degree education: A model for teaching and integration into clinical practice. *The Journal of Orthopaedic and Sports Physical Therapy*, *36*(8), 577–587.

Fosnot, C. T. (Ed.). (2005). *Constructivism: Theory, perspectives, and practice* (2nd ed.). Teachers College Press.

Furze, J., Gale, J. R., Black, L., Cochran, T. M., & Jensen, G. M. (2015). Clinical reasoning: Development of a grading rubric for student assessment. *Journal of Physical Therapy Education*, *29*(3), 34.

Garcia-Ros, R., Ruescas-Nicolau, M.-A., Cezón-Serrano, N., Flor-Rufino, C., Martin-Valenzuela, C. S., & Sánchez-Sánchez, M. L. (2024). Improving assessment of procedural skills in health sciences education: a validation study of a rubrics system in neurophysiotherapy. *BMC Psychology*, *12*(1), 147.

Gentile, A., Carr, J., & Shepherd, R. (1987). Movement science: Foundations for physical therapy in rehabilitation. *Skill Acquisition: Action, Movement and Neuromotor Processes. Rockville: Aspen*.

George, J. H., & Doto, F. X. (2001). A simple five-step method for teaching clinical skills. *Family Medicine*, *33*(8), 577–578.

George, S. Z., Fritz, J. M., Silfies, S. P., Schneider, M. J., Beneciuk, J. M., Lentz, T. A., Gilliam, J. R., Hendren, S., & Norman, K. S. (2021). Interventions for the management of acute and chronic low back pain: Revision 2021. *The Journal of Orthopaedic and Sports Physical Therapy*, *51*(11), CPG1–CPG60.

Giacomino, K., Caliesch, R., & Sattelmayer, K. M. (2020). The effectiveness of the Peyton's 4-step teaching approach on skill acquisition of procedures in health professions education: A systematic review and meta-analysis with integrated meta-regression. *PeerJ*, *8*, e10129.

Gittinger, F. P., Lemos, M., Neumann, J. L., Förster, J., Dohmen, D., Berke, B., Olmeo, A., Lucas, G., & Jonas, S. M. (2022). Interrater reliability in the assessment of physiotherapy students. *BMC Medical Education*, *22*(1), 186.

Gradl-Dietsch, G., Lübke, C., Horst, K., Simon, M., Modabber, A., Sönmez, T. T., Münker, R., Nebelung, S., & Knobe, M. (2016). Peyton's four-step approach for teaching complex spinal manipulation techniques--a prospective randomized trial. *BMC Medical Education*, *16*(1), 284.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108.

Griswold, D., Learman, K., Kolber, M. J., O'Halloran, B., & Cleland, J. A. (2018). Pragmatically applied cervical and thoracic nonthrust manipulation versus thrust manipulation for patients with mechanical neck pain: A multicenter randomized clinical trial. *The Journal of Orthopaedic and Sports Physical Therapy*, *48*(3), 137–145.

Hanafin, S. (2004). Review of literature on the Delphi Technique. *Dublin: National Children's Office*, 1–51.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015.

Helmer, O. (1967). *Analysis of the future: The Delphi method*. Rand Corporation Santa Monica. https://apps.dtic.mil/sti/citations/AD0649640

Hibbert, E. J., Lambert, T., Carter, J. N., Learoyd, D. L., Twigg, S., & Clarke, S. (2013). A randomized controlled pilot trial comparing the impact of access to clinical endocrinology video demonstrations with access to usual revision resources on medical student performance of clinical endocrinology skills. *BMC Medical Education*, *13*, 135.

Higgins, M., Madan, C. R., & Patel, R. (2021). Deliberate practice in simulation-based surgical skills training: A scoping review. *Journal of Surgical Education*, *78*(4), 1328–1339.

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Medical Teacher*, *32*(8), 676–682.

Horbacewicz, J. (2018). Effect of blocked versus random practice on physical therapy students' manual force modulation. *Perceptual and Motor Skills*, *125*(6), 1173–1185.

Howie, E. E., Dharanikota, H., Gunn, E., Ambler, O., Dias, R., Wigmore, S. J., Skipworth, R. J. E., & Yule, S. (2023). Cognitive load management: An invaluable tool for safe and effective surgical training. *Journal of Surgical Education*, *80*(3), 311–322.

Humphrey-Murto, S., Varpio, L., Gonsalves, C., & Wood, T. J. (2017). Using consensus group methods such as Delphi and Nominal Group in medical education research. *Medical Teacher*, *39*(1), 14–19.

Hussein, A. A., Sexton, K. J., May, P. R., Meng, M. V., Hosseini, A., Eun, D. D., Daneshmand, S., Bochner, B. H., Peabody, J. O., Abaza, R., Skinner, E. C., Hautmann, R. E., & Guru, K. A. (2018). Development and validation of surgical training tool: cystectomy assessment and surgical evaluation (CASE) for robot-assisted radical cystectomy for men. *Surgical Endoscopy*, *32*(11), 4458–4464.

Ilgen, J. S., Ma, I. W. Y., Hatala, R., & Cook, D. A. (2015). A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Medical Education*, *49*(2), 161–173.

Jensen, G. M., Jette, D. U., Timmerberg, J. F., Chesbro, S. B., Dole, R. L., Kapasi, Z., & Lotshaw, A. (2020). Competency-based education in physical therapy: Developing a framework for education research. *Journal of Physical Therapy Education*, *36*(4), 334–340.

Johnston, L., Sawyer, T., Nishisaki, A., Whitfill, T., Ades, A., French, H., Glass, K., Dadiz, R., Bruno, C., Levit, O., & Auerbach, M. (2022). Comparison of a dichotomous versus trichotomous checklist for neonatal intubation. *BMC Medical Education*, *22*(1), 645.

Johnston, L., Sawyer, T., Nishisaki, A., Whitfill, T., Ades, A., French, H., Glass, K., Dadiz, R., Bruno, C., Levit, O., Gangadharan, S., Scherzer, D., Moussa, A., Auerbach, M., & INSPIRE Research Network. (2019). Neonatal intubation competency assessment tool: Development and validation. *Academic Pediatrics*, *19*(2), 157–164.

Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, *31*(8), 684–706.

Kafri, M., & Atun-Einy, O. (2019). From motor learning theory to practice: A scoping review of conceptual frameworks for applying knowledge in motor learning to physical therapist practice. *Physical Therapy*, *99*(12), 1628–1643.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342.

Keeney, S., McKenna, H. A., & Hasson, F. (2011). *The Delphi technique in nursing and health research*. John Wiley & Sons.

Keter, D., Griswold, D., Learman, K., & Cook, C. (2023). Priorities in updating training paradigms in orthopedic manual therapy: An international Delphi study. *Journal of Educational Evaluation for Health Professions*, *20*, 4.

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and*

*Education*, *4*(2), 193–212.

Leech, K. A., Roemmich, R. T., Gordon, J., Reisman, D. S., & Cherry-Allen, K. M. (2022). Updates in motor learning: Implications for physical therapist practice and education. *Physical Therapy*, *102*(1). https://doi.org/10.1093/ptj/pzab250

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education*, *4*(3), 119–127.

Lewthwaite, R., & Wulf, G. (2017). Optimizing motivation and attention for motor performance and learning. *Current Opinion in Psychology*, *16*, 38–42.

Lockey, A., Conaghan, P., Bland, A., & Astin, F. (2021). Educational theory and its application to advanced life support courses: A narrative review. *Resuscitation Plus*, *5*, 100053.

Luedtke, K., Luebke, L., Elizagaray-Garcia, I., Schindler, O., & Szikszay, T. M. (2023). Effectiveness of online teaching during the COVID-19 pandemic on practical manual therapy skills of undergraduate physiotherapy students. *The Journal of Manual & Manipulative Therapy*, *31*(5), 349–357.

Madadizadeh, F., & Bahariniya, S. (2023). Tutorial on how to calculating content validity of scales in medical research. *Perioperative Care and Operating Room Management*, *31*, 100315.

Magill, R. (2014). *Motor learning and control* (10th ed.). New York: McGraw-Hill Publishing.

Magnusson, D. M., Rethorn, Z. D., Bradford, E. H., Maxwell, J., Ingman, M. S., Davenport, T. E., & Bezner, J. R. (2020). Population health, prevention, health promotion, and wellness competencies in physical therapist professional education: Results of a modified Delphi study. *Physical Therapy*, *100*(9), 1645–1658.

Manipulation Education Committee of the APTA Manipulation Task Force. (2004). *Manipulation education manual for physical therapist professional degree programs*. American Physical Therapy Association.

Manton, C., Conley, Q., Cleland, J. A., & Puentedura, E. J. (2023). The effectiveness of instruction on lumbar spine thrust joint manipulation task performance. *Journal of Physical Therapy Education*, *38*(3) 212-220. https://doi.org/10.1097/JTE.0000000000000321

Martin, R. L., Davenport, T. E., Fraser, J. J., Sawdon-Bea, J., Carcia, C. R., Carroll, L. A., Kivlan, B.R., & Carreira, D. (2021). Ankle stability and movement coordination impairments: Lateral ankle ligament sprains revision 2021: Clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy of the American physical therapy association. *Journal of Orthopaedic & Sports Physical Therapy*, *51*(4), CPG1-CPG80.

McCallister, E., Bernard, A., Flowers, D. W., & Rachal Sant, L. (2023). Effects of a blocked versus full-length semester design on doctor of physical therapy student performance and retention. *Journal, Physical Therapy Education*, *37*(3), 186–192.

McDevitt, A. W., Cleland, J. A., Addison, S., Calderon, L., & Snodgrass, S. (2022). Physical therapy interventions for the management of biceps tendinopathy: An international Delphi study. *International Journal of Sports Physical Therapy*, *17*(4), 677–694.

McGaghie, W. C. (2015). Mastery learning: It is time for medical education to join the 21st century. *Academic Medicine: Journal of the Association of American Medical Colleges*, *90*(11), 1438–1441.

McKenna, H. P. (1994). The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing*, *19*(6), 1221–1225.

McKinley, R. K., Strand, J., Ward, L., Gray, T., Alun-Jones, T., & Miller, H. (2008). Checklists for assessment and certification of clinical procedural skills omit essential competencies: A systematic review. *Medical Education*, *42*(4), 338–349.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, *50*(9), 741–749.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine: Journal of the Association of American Medical Colleges*, *65*(9 Suppl), S63–S67.

Mintken, P. E., Derosa, C., Little, T., Smith, B., & American Academy of Orthopaedic Manual Physical Therapists. (2008). A model for standardizing manipulation terminology in physical therapy practice. *The Journal of Manual & Manipulative Therapy*, *16*(1), 50–56.

Mukhalalati, B. A., & Taylor, A. (2019). Adult learning theories in context: A quick guide for healthcare professional educators. *Journal of Medical Education and Curricular Development*, *6*, 2382120519840332.

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, *11*(4), 116–129.

Nicholls, D., Sweet, L., Muller, A., & Hyett, J. (2016). Teaching psychomotor skills in the twenty-first century: Revisiting and reviewing instructional approaches through the lens of contemporary literature. *Medical Teacher*, *38*(10), 1056–1063.

O'Donnell, M., Smith, J. A., Abzug, A., & Kulig, K. (2016). How should we teach lumbar manipulation? A consensus study. *Manual Therapy*, *25*, 1–10.

Pérez-Guillén, S., Carrasco-Uribarren, A., Celis, C. L., González-Rueda, V., Rodríguez-Rubio, P. R., & Cabanillas-Barea, S. (2022). Students' perceptions, engagement and satisfaction with the use of an e-rubric for the assessment of manual skills in physiotherapy. *BMC Medical Education*, *22*(1), 623.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459–467.

Popham, W. (1997). What's wrong--and what's right--with rubrics. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A.*, *55*, 72–75.

Puentedura, E. J., Slaughter, R., Reilly, S., Ventura, E., & Young, D. (2017). Thrust joint manipulation utilization by U.S. physical therapists. *The Journal of Manual & Manipulative Therapy*, *25*(2), 74–82.

Qaseem, A., Wilt, T. J., McLean, R. M., Forciea, M. A., Clinical Guidelines Committee of the American College of Physicians, Denberg, T. D., Barry, M. J., Boyd, C., Chow, R. D., Fitterman, N., Harris, R. P., Humphrey, L. L., & Vijan, S. (2017). Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the American College of Physicians. *Annals of Internal Medicine, 166*(7), 514–530.

Roemmich, R. T., & Bastian, A. J. (2018). Closing the loop: From motor neuroscience to neurorehabilitation. *Annual Review of Neuroscience*, *41*, 415–429.

Rossettini, G., Rondoni, A., Palese, A., Cecchetto, S., Vicentini, M., Bettale, F., Furri, L., & Testa, M. (2017). Effective teaching of manual skills to physiotherapy students: A randomised clinical trial. *Medical Education*, *51*(8), 826–838.

Sattelmayer, K. M., Jagadamma, K. C., Hilfiker, R., & Baer, G. (2020). Acquisition of procedural skills in preregistration physiotherapy education comparing mental practice against no mental practice: The learning of procedures in physiotherapy education trial - A development of concept study. *Journal of Medical Education and Curricular Development*, *7*, 2382120520927382.

Sattelmayer, K. M., Jagadamma, K. C., Sattelmayer, F., Hilfiker, R., & Baer, G. (2020). The assessment of procedural skills in physiotherapy education: a measurement study using the Rasch model. *Archives of Physiotherapy*, *10*, 9.

Sattelmayer, M., Elsig, S., Hilfiker, R., & Baer, G. (2016). A systematic review and meta-analysis of selected motor learning principles in physiotherapy and medical education. *BMC Medical Education*, *16*, 15.

Sawyer, T., White, M., Zaveri, P., Chang, T., Ades, A., French, H., Anderson, J., Auerbach, M., Johnston, L., & Kessler, D. (2015). Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Academic Medicine: Journal of the Association of American Medical Colleges*, *90*(8), 1025–1033.

Schilling, J. (2017). In respect to the cognitive load theory: Adjusting instructional guidance with student expertise. *Journal of Allied Health*, *46*(1), e25–e30.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., & Zelaznik, H. N. (2019). *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics.

Seals, R., Gustowski, S. M., Kominski, C., & Li, F. (2016). Does replacing live demonstration with instructional videos improve student satisfaction and osteopathic manipulative treatment examination performance? *The Journal of the American Osteopathic Association*, *116*(11), 726–734.

Shang, Z. (2023). Use of Delphi in health sciences research: A narrative review. *Medicine*, *102*(7), e32829.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Silvernail, J. L., Deyle, G. D., Jensen, G. M., Chaconas, E., Cleland, J., Cook, C., Courtney, C. A., Fritz, J., Mintken, P., & Lonnemann, E. (2024). Orthopaedic manual physical therapy: A modern definition and description. *Physical Therapy*, *104*(6). https://doi.org/10.1093/ptj/pzae036

Sizer, P., Sawyer, S., Felstehausen, V., Couch, S., Dornier, L., & Cook, C. (2008). Intrinsic and extrinsic factors important to manual therapy competency development: A Delphi investigation. *The Journal of Manual & Manipulative Therapy*, *16*(1), e9–e19.

Soni, N. J., Nathanson, R., Andreae, M., Khosla, R., Vadamalai, K., Kode, K., Boyd, J. S., LoPresti, C. M., Resop, D., Basrai, Z., Williams, J., Bales, B., Sauthoff, H., Wetherbee, E., Haro, E. K., Smith, N., Mader, M. J., Pugh, J., Finley, E. P., &

Schott, C. K. (2022). Development of a multisystem point of care ultrasound skills assessment checklist. *The Ultrasound Journal*, *14*(1), 17.

Spampinato, D., & Celnik, P. (2021). Multiple motor learning processes in humans: Defining their neurophysiological bases. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *27*(3), 246–267.

Spranger, J., Homberg, A., Sonnberger, M., & Niederberger, M. (2022). Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*, *172*, 1–11.

Stenberg, M., Mangrio, E., Bengtsson, M., & Carlson, E. (2021). Formative peer assessment in higher healthcare education programmes: A scoping review. *BMJ Open*, *11*(2), e045345.

Struessel, T. S., Carpenter, K. J., May, J. R., Weitzenkamp, D. A., Sampey, E., & Mintken, P. E. (2012). Student perception of applying joint manipulation skills during physical therapist clinical education: Identification of barriers. *Journal of Physical Therapy Education*, *26*(2), 19.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Tanaka, H., Ishikawa, T., Lee, J., & Kakei, S. (2020). The cerebro-cerebellum as a locus of forward model: A review. *Frontiers in Systems Neuroscience*, *14*. https://doi.org/10.3389/fnsys.2020.00019

Taylor, D. C. M., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical Teacher*, *35*(11), e1561–e1572.

Thibault, G. E. (2020). The future of health professions education: Emerging trends in the United States. *FASEB bioAdvances*, *2*(12), 685–694.

Timmerberg, J. F., Chesbro, S. B., Jensen, G. M., Dole, R. L., & Jette, D. U. (2022). Competency-based education and practice in physical therapy: It's time to act! *Physical Therapy*, *102*(5). https://doi.org/10.1093/ptj/pzac018

Torres-Narváez, M.-R., Vargas-Pinilla, O.-C., & Rodríguez-Grande, E.-I. (2018). Validity and reproducibility of a tool for assessing clinical competencies in physical therapy students. *BMC Medical Education*, *18*(1), 280.

Tsay, J. S., Haith, A. M., Ivry, R. B., & Kim, H. E. (2022). Interactions between sensory prediction error and task error during implicit motor learning. *PLoS Computational Biology*, *18*(3), e1010005.

Turoff, M. (1970). The design of a policy Delphi. *Technological Forecasting and Social Change*, *2*(2), 149–171.

van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education*, *44*(1), 85–93.

Vassiliadis, P., Derosiere, G., Dubuc, C., Lete, A., Crevecoeur, F., Hummel, F. C., & Duque, J. (2021). Reward boosts reinforcement-based motor learning. *iScience*, *24*(7), 102821.

Vogel, D., & Harendza, S. (2016). Basic practical skills teaching and learning in undergraduate medical education - A review on methodological evidence. *GMS Journal for Medical Education*, *33*(4), Doc64.

Walker M, P. J. (1998). Teaching in the theatre. In J. W. R. Peyton (Ed.), *Teaching and learning in medical practice* (pp. 171–180). Rickmansworth: Manticone Publishing House.

Washmuth, N. B., Ross, S., & Bowens, A. N. (2020). Effect of motor learning theory-assisted instruction versus traditional demonstration on student learning of spinal joint manipulation. *Health Professions Education*, *6*(1), 72–84.

Whalen, A. M., Merves, M. H., Kharayat, P., Barry, J. S., Glass, K. M., Berg, R. A., Sawyer, T., Nadkarni, V., Boyer, D. L., & Nishisaki, A. (2022). Validity evidence for a novel, comprehensive bag–mask ventilation assessment tool. *The Journal of Pediatrics*, *245*, 165–171.e13.

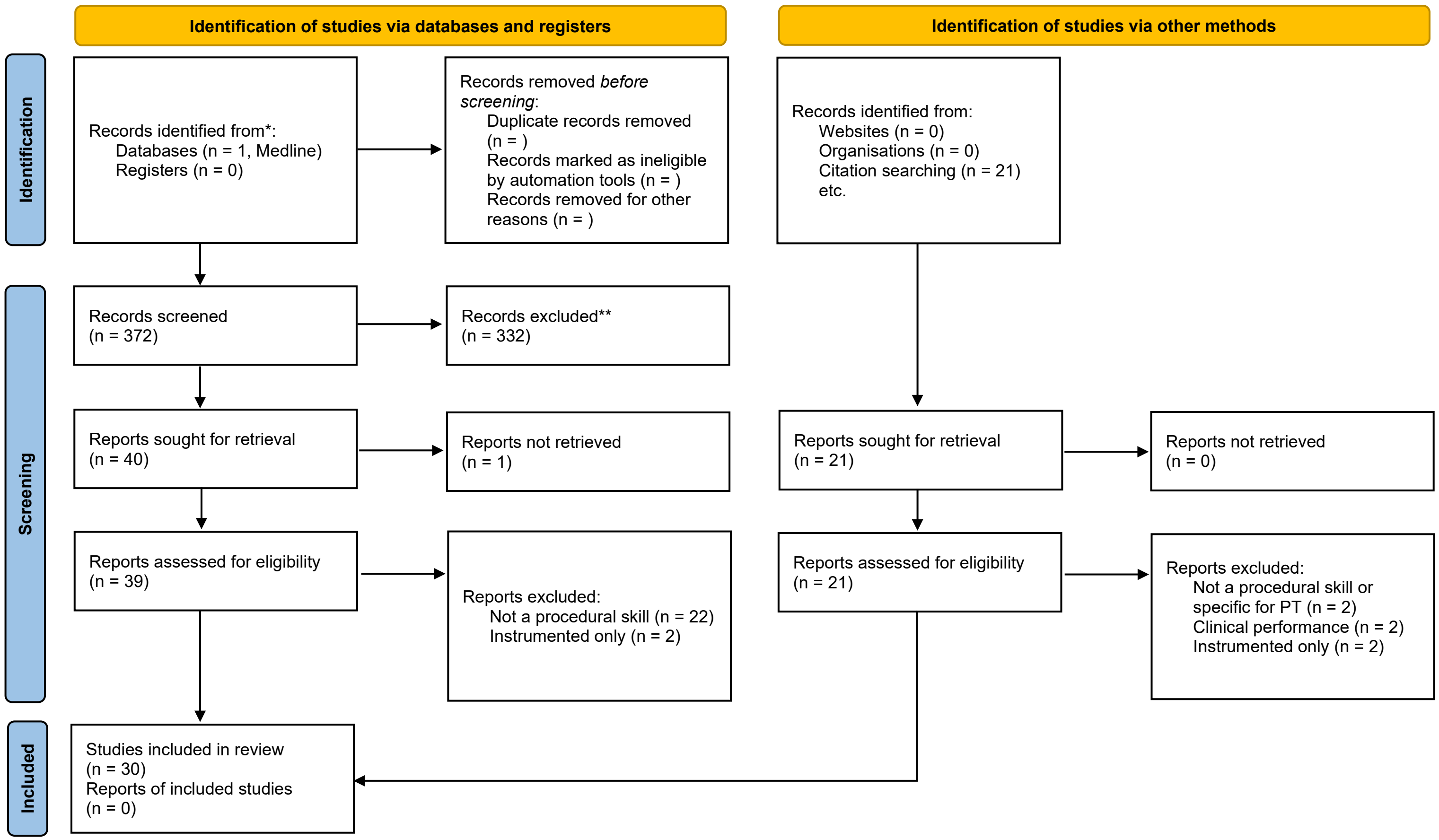
Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE Guide No. 86. *Medical Teacher*, *36*(5), 371–384.

Yudkowsky, R., Park, Y. S., & Downing, S. M. (2020). *Assessment in health professions education* (2nd ed.). Routledge.

Yudkowsky, R., Park, Y. S., Riddle, J., Palladino, C., & Bordage, G. (2014). Clinically discriminating checklists versus thoroughness checklists: Improving the validity of performance test scores. *Academic Medicine: Journal of the Association of American Medical Colleges*, *89*(7), 1057–1062.

# Appendix A

PRISMA Flow Diagram for Procedural Skill Assessment and Thrust Joint Manipulation Literature




Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

# Appendix B

## Literature Review for Comparison of Manual Therapy Rubrics

| Author Last Name, Year | Profession | Skill Assessed | Study Design | Theoretical Frameowrk | Validation of Assessment | Assessment Criteria | Levels (Number) | Rating Scale | Rubric Type Analytic vs. Holistic | Rubric Type General vs. Task Specific (only one skill) | Feedback Use: Formative, Summative, Both |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cuenca-Martínez, 2023 | PT | Lumbar Manipulation | Randomized comparision | Motor learning | Not reported | 5 criteria: patient position, therapist position, sequence, accuracy, correct application of parameters | 3 | zero-one-two | Analytic | General | Both |
| Gradl-Dietsch, 2016 | Medicine | 4 TJM: CS, T, L and SIJ | Randomized comparision | Peyton teaching method | Not reported | 6 criteria with checklist items for each: Diagnosis, Technique, Position of Patient, Hand Placement, Test Traction [wk pre-thrust?], Manipulation | 2 | correct, incorrect | Analytic | Checklist, Specific | Summative |
| Luedtke, 2023 | PT | 3 Mobilizations: C, L, Knee | Randomized comparision | Motor learning | Face validity, Inter-rater reliability on 56 subjects | 5 criteria: Positioning of patient, Positioning of Therapist, Hand Placement, Direction of Movement, Correct Gradeof Movement | 2 | 0= Did not show, 1= Showed required behavior | Analytic | General | Summative |
| Manton, 2023 | PT | Lumbar Manipulation | Randomized comparision | Motor learning | Reported unpublished, Intra-rater Reliability | 5 Overall General Criteria: Positioning of Patient, Student Hand/Body Position, Localizing Lever Arm, Direction & Rate of Force Application each with unique items; 23 individual checklist items total | 2-3 | Scoring 2-3 scale on each specific to observed behavior | Analytic | Specific | Summative |
| McCallister, 2023 | PT | MT Skills | Comarision non-randomized | Motor learning, Cognitive | Not reported | 4 Criteria: Positions subject and joint, places hands correctly, force perpendicular to joint, grades of mvt and rate | 5 | Max-Mod-Min-VC-Independent | Analytic | General | Both |
| Pérez-Guillén, 2022 | PT | MT Skills, Peer feedback | Survey Cross Sectional | Social Constructivist | Not reported | 5 criteria: Patient position, PT position, Execution procedure, Effect, Clinical Reasoning | 4 | Novel, Competent, Proficient, Expert | Analytic | General | Formative |
| Rossettini, 2017 | PT | C1-C2 Passive R Rotation | Randomized comparision | Peyton teaching method | Face validity, Inter-rater reliability on 10 students | 12 item checklist criteria | 3 | 3 levels specific to the criteria to be scored | Analytic | Checklist, Specific | Both |
| Seals, 2016 | Osteopathy | Still Technique C-T spine & FPR for C-T | Comparision of prior cohort | Motor learning, Cognitive load | Not reported | 5 criteria: Diagnosis, Setup, Contact of tissues, Application of principles, Retest | 4 | 0-not performed/poor, 1-needs improvement, 2-competent, 3-outstanding | Analytic | General | Formative |
| Washmuth, 2020 | PT | 6 TJM's | Randomized comparision | Motor learning | Face validity | 7 criteria: exposure of area, patient positioning, PT position-Body, PT position-Hands, direction of force application, force intensity and speed | 4 | 0=prevents desired effect, 1=Student, 2=Entry level graduate, 3=Seasoned | Analytic | General | Both |
| | Non OMPT Interventions Below | | | | | | | | | | |
| Sattelmayer, 2020 | PT | Vestibular Rehab & Floor to Stand Transfer | Randomized comparison | Motor learning | Content, Inter-rater reliability, structural validity | 6 overall criteria: Preparation, Knowledge & decision making, Safety, Communication, Procedure Execution, Comfort each w/ 3-7 sumcomponents | 5 | 0=Very Poor, 1=Poor, 2=Acceptable, 3=Good, 4=Very Good | Analytic | General | Both |
| Garcia-Ros, 2024 | PT | Neuro PT | Randomized comparison | Motor learning | Content, Internal structure | 5 Criteria: PT position, Patient position, Verbal facilitation, Holds, Execution | 4 | 0= inadequate, 1=needs improvement, 2=adequate, 3 advanced | Analytic | General | Formative |

Note: Abbreviations used are Cervical (C), Facilitated positional release (FPR), Lumbar (L), Physical therapy (PT), Manual Therapy (MT), Right (R), Spine (S), Thoracic (T), Thrust Joint Manipulation (TJM), Verbal Cues (VC)