2021

Data Modeling of Cognitive Structure in Physiotherapy Students Learning Gross Anatomy

William Allan Besselink
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

William Allan Besselink

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Wellesley Foshay, Committee Chairperson, Education Faculty
Dr. David Perry, Committee Member, Education Faculty
Dr. Ioan Ionas, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

Data Modeling of Cognitive Structure in Physiotherapy Students Learning Gross Anatomy

by

William Allan Besselink

DPT, College of St. Scholastica, 2016

BSc.PT, Queen’s University, 1988

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2021

Abstract

Cognitive structures that promote deep learning of gross anatomy are integral to musculoskeletal physiotherapy practice yet poorly understood. This quantitative, criterion-related validation study addressed two data modeling strategies (multidimensional scaling and Pathfinder networks) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. The study was grounded in the Adaptive Control of Thought-Rational theory of cognition. The research questions addressed the agreement (reliability, accuracy, and association) between student and expert cognitive structures and included the derived quantitative parameters as predictor variables in multiple regression to examine potential relationships with unit grades. An online survey of paired comparisons of 20 anatomical concepts relevant to musculoskeletal clinical practice generated the raw data used in the data modeling strategies for cognitive structure mapping. Convenience sampling was used to recruit 31 physiotherapy students, four course instructors, and three domain experts who completed the online survey. The results indicated moderate to high effect sizes regarding the agreement between student and expert. Six predictor variables accounted for 68.9% of the variance in unit grade indicating a large effect size. Preliminary evidence of concurrent and predictive validity was reported. Positive social change is reflected in this innovative use of data modeling strategies to represent cognitive structure and potentially enhance competency-based education critical to effective musculoskeletal physiotherapy practice.

Data Modeling of Cognitive Structure in Physiotherapy Students Learning Gross Anatomy

by

William Allan Besselink

DPT, College of St. Scholastica, 2016

BSc.PT, Queen’s University, 1988

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2021

Dedication

“Education is the most powerful weapon which you can use to change the world” (Nelson Mandela).

This dissertation is dedicated to the educators who see something special in us that we have yet to comprehend, who bring out the best in us when we don’t know it exists, who inspire us to explore our outer and inner worlds, and who give us tangible examples of what it is like to change the world, one person at a time.

Several exemplary individuals have provided those sparks of inspiration, realization, innovation, and actualization for me over the years: John Ballachey, David Huether, Norie Spence, Dave Ross, Robin McKenzie, Grant Watson, Barbara Melzer, and Barbara Sanders. I can only aspire to lead others the way these individuals have inspired me.

Acknowledgments

"What lies behind us and what lies before us are tiny matters compared to what lies within us" (Ralph Waldo Emerson).

There is no dress rehearsal for the journey that is a PhD. We bring to it the sum of our life experiences—that which lies behind us—with will, determination, and grace. However, we must also stand on the shoulders of giants to see what lies before us and, more importantly, within us. I extend many thanks

To Marshall Burt, for always pushing me to think outside the box,

To Pete Wilde, for planting the PhD seed during one of our many discussions, and

To my dissertation committee—my chair, Dr. Foshay, for his wisdom, war stories, and uncanny ability to challenge me on so many levels at just the right time; Dr. Perry, Dr. Lacy, and Dr. Ionas, for their committee support and guidance; and Dr. Toledo and Dr. Harland for helping to keep the wheels on the PhD cart along the way.

Finally, I want to thank all my students past and present. Robin McKenzie once said that "everything I learned, I learned from my patients." To this, I would respectfully add, "and my students." This dissertation stems from my teaching experiences with you. You inspire me daily, and for that, I am forever grateful.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Physiotherapy (physical therapy) has a clear and evolving role as a primary care provider for musculoskeletal (MSK) conditions (Ojha et al., 2020). The clinical diagnostic accuracy of physiotherapists is equivalent to that of orthopedic surgeons in the context of MSK conditions; as such, it requires a mastery of gross anatomy and its relationship to movement and function (Barrett & Liebman, 2020; Moore et al., 2005). The foundation of physiotherapy education is firmly rooted in anatomical knowledge, an essential aspect of entry-level training leading to clinical reasoning and diagnostic thinking (Timmerberg et al., 2019). However, there is a growing trend among physiotherapy, medical, nursing, and chiropractic students indicating the decline of anatomical knowledge retention (Dayal et al., 2017; Hołda et al., 2019; Narnaware & Neumeier, 2020). This decline becomes crucial for global health policy when 17% of the world's population is affected by musculoskeletal conditions that require the care of a skilled primary MSK provider such as a physiotherapist (Briggs et al., 2020; James et al., 2018).

Knowledge retention and transfer are critical learning outcomes that depend on the student's cognitive structure gained through deep (meaningful) learning (Ausubel, 1963; Mayer, 2002b). Factors related to instructional design, the instructor's pedagogical content knowledge, and the student's self-regulated learning strategies can either enhance or inhibit cognitive structure development (Mayer, 2009; Neumann et al., 2019; van Lankveld et al., 2019). These factors may limit the retention and transfer of anatomical knowledge for subsequent clinical courses and, over the long term, negatively impact clinical practice (Mayer, 2002b; Montpetit-Tourangeau et al., 2017).

Gross anatomy is a foundational course that provides unique challenges to physiotherapy students. Anecdotal experience indicated broad variations in prior knowledge, misconceptions, and an emphasis on rote memorization, features indicative of poor knowledge organization (D'Antoni et al., 2019). Physiotherapy students are often overwhelmed by the volume of the material and lack confidence in understanding how to learn anatomy (Choi-Lundberg et al., 2017). These factors subsequently increase stress and cognitive load, diminishing the student's ability to use prior knowledge for future encoding (Vogel et al., 2018). Much attention has focused on teaching approaches and instruction. However, little research exists on how physiotherapy students learn gross anatomy in a way that develops a cognitive structure associated with meaningful learning and potentially enhances long-term retention, competency, and effective transfer to a clinical context as a primary care provider (Choi-Lundberg et al., 2017; D'Antoni et al., 2019).

The current study addressed two data modeling strategies (multidimensional scaling [MDS] and Pathfinder networks [PFN]) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. Chapter 1 provides an overview of the study and includes the background, problem, purpose, research questions, theoretical framework, nature, definitions, assumptions, scope and delimitations, limitations, and significance. The background provides key relevant literature related to the topic. The problem, purpose, and research questions provide the basis for the investigation. The theoretical framework identifies the foundations for the study that informed the chosen methodology. The scope of the study includes definitions, assumptions, scope and delimitations, and limitations. Finally, the significance of the study provides implications for positive social change.

Background

Deep learning, also known as meaningful learning, is grounded in cognitive science. The primary foundation for deep learning is prior knowledge, and new knowledge builds upon this foundation (Ausubel, 1963; Mayer, 2002a). Cognitive architectures such as Adaptive Control of Thought-Rational, or ACT-R (J. R. Anderson, 1996, 2007), have been designed to align computational and neuroscientific constructs for a functional understanding of deep learning and the cognitive mechanisms underlying knowledge acquisition, encoding, retention, and retrieval. ACT-R utilizes two abstraction levels (symbolic and subsymbolic) to represent these cognitive mechanisms, one of which is cognitive structure.

Cognitive structures are the individual's mental representation of what they know (content) and how they know it (organization); cognitive structures contain facts, personalized meaning, perceptions, and misconceptions. Prior knowledge forms the basis for cognitive structures, which are continually undergoing revision and updating as knowledge and learning progress (J. R. Anderson, 1996; Noushad & Khurshid, 2019). Effective encoding of knowledge into well-organized and relevant cognitive structures free of misconceptions is a goal of learning and instruction. However, the challenge is that the direct measurement of cognitive structures remains elusive; indirect methods are required that demand both reliability and validity in their use (Gisick et al., 2018; Ifenthaler et al., 2011; Moon et al., 2018). Although many studies have addressed cognitive structure and its representation, a broad range of methodological issues has limited the generalizability of findings regarding a preferred representational approach or its reproducibility across domains.

Self-directed learning places a higher demand on the physiotherapy student in developing effective cognitive structures (van Lankveld et al., 2019). The learning process begins with the student's approach to learning, either surface (also known as rote or meaningless) or deep (meaningful; Marton & Säljö, 1976). Deep learning strategies create more developed cognitive structures, enhancing learning outcomes and the transfer of learning to higher order thinking (Krathwohl, 2002; Smith, Stokholm, et al., 2017). Cognitive structures can differentiate students from experts and can be used to establish cognitive performance while monitoring educational progress (Moon et al., 2018). The development of cognitive structure that is optimized for clinical decision making is the goal of any health professions curricula; for the physiotherapy student, this begins with the study of gross anatomy.

The importance of gross anatomy in the education of primary MSK care providers cannot be overstated. Research in medical education has indicated that medical gross anatomy performance is correlated with the United States Medical Licensing Examination Step 1 (Peterson & Tucker, 2005). Peterson and Tucker (2005) reported that this correlation (r = 0.577) is greater than other traditional predictors in use such as the biology section of the MCAT (r = 0.482), science grade point average (GPA; r = 0.213), and undergraduate GPA (*r* = 0.189). Physiotherapy students show a similar trend. Traditional predictors of the first-time pass rate on the National Physical Therapy Exam (NPTE) include preprofessional, undergraduate, first year, and final professional GPA (Bayliss et al., 2017; Cook et al., 2015; S. H. Hayes et al., 1997; Kume et al., 2019; Meiners & Rush, 2017; Roman & Buman, 2019; Wolden et al., 2020). Much like the findings of medical education, gross anatomy grade in physiotherapy students contributed to 49% (younger, traditional students) and 35% (older, nontraditional students) of the variance in final professional program GPA (S. H. Hayes et al., 1997). Wolden et al. (2020) reported that student clinical performance scores had a weak and not statistically significant relationship with first-time NPTE pass rate. However, first- and third-year student GPA, of which gross anatomy is a contributor, had a strong and significant relationship (Wolden et al., 2020). Gross anatomy education plays an integral role in primary MSK provider entry-level training and the first-time pass rate on the NPTE.

Gross anatomy education has remained relatively static over several decades; however, recent developments have included the use of problem-based learning and computer-assisted instruction (Wilson, Brown, et al., 2019; Wilson, Miller, et al., 2018). A critical review by Estai and Bunt (2016), followed by systematic reviews by Losco et al. (2017) and Wilson, Brown, et al. (2019), indicated that gross anatomy teaching methods and instructional strategies attain similar learning outcomes. Wilson, Brown, et al. (2019) noted that there is a need to examine the impact of anatomy pedagogies and learning strategies on the acquisition and long-term retention of anatomical knowledge, especially in the context of student-centered learning. Hulme et al. (2020) noted the importance of tracking changes in anatomical knowledge within the curricula and measuring retention as cognitive levels change. Learner-specific cognitive factors may be significant contributors to the problem of anatomical knowledge retention. However, cognitive learning theories that promote deep learning are often poorly integrated into gross anatomy curricula (Agra et al., 2019; Choi-Lundberg et al., 2017; Smith, Finn, & Border, 2017). For example, a literature review revealed only two studies in gross anatomy and neuroanatomy that included mind mapping to promote knowledge representation and deep learning (Anand et al., 2018; Deshatty & Mokashi, 2013). No studies were identified that focused on students' cognitive structure in the broad context of gross anatomy education, how students' cognitive structure compared to experts (both domain specific and physiotherapy centric), and how students' cognitive structure changed over time. The research was also limited on medical, chiropractic, or physiotherapy students learning gross anatomy to become future primary MSK care providers.

Cognitive structures serve as a construct for knowledge organization, an essential element in learning gross anatomy and a domain that demands cognitive skills such as visualization, spatial ability, and the use of consistent terminology and taxonomy (Amin & Iqbal, 2019; Castro-Alonso & Atit, 2019; Clarkson & Whipple, 2018; Langlois et al., 2020). The student's learning approach can have diverse effects on cognitive structures and subsequent learning outcomes (Marton & Säljö, 1976). Cognitive mapping may represent the multidimensional frames of reference inherent to cognitive structure (Bottini & Doeller, 2020). Data modeling strategies such as MDS and PFN have been used to represent cognitive structures in a broad range of domains (Azzarello, 2007; Balloo et al., 2016; Casas-García & Luengo-González, 2012; Connor et al., 2004; Curtis & Davis, 2003; DiCerbo, 2007; Jaworska & Chupetlovska-Anastasova, 2009; McGaghie, McCrimmon, et al., 2000; McGaghie, McCrimmon, & Thompson, 1998; Neiles et al., 2016; Stevenson et al., 2016; Veríssimo et al., 2017; H. D. White, 2003). MDS provides a visual spatial representation (with associated quantitative measures), whereas a PFN provides a visual network representation (with associated quantitative measures). Although both MDS and PFN are promising approaches to the visual and quantitative representation of cognitive structure, neither have been used to represent cognitive structures in gross anatomy or physiotherapy education.

There was a gap in the research on how physiotherapy students learn gross anatomy, specifically the cognitive structures that promote deep learning in the gross anatomy domain to fulfill their role as primary MSK care providers. Although limited research on cognitive structure was found in other health care professions such as nursing (Alfayoumi, 2019) and medicine (Nicoara, Szamoskovi, et al., 2020), I did not find studies in the physiotherapy domain beyond two concept mapping studies (see Zipp & Maher, 2013; Zipp et al., 2015). The problem was that deep learning of gross anatomy by physiotherapy students is poorly understood based on cognitive structure development. The current study addressed a primary component in this process: the representation of cognitive structure in physiotherapy students learning gross anatomy. The use of cognitive structure mapping via MDS and PFN is a promising and innovative approach to the representation of cognitive structures that merited further research in the context of gross anatomy education for physiotherapy students. Quantitative representation of cognitive structure may provide valuable insight into learning and instruction strategies that enhance the development of well-organized cognitive structures, promote deep learning, and serve as an assessment of learning leading to enhanced retention and transfer (see Leppink, 2020).

Problem Statement

The problem was that the mapping of cognitive structures of physiotherapy students learning gross anatomy is poorly understood. A mastery of gross anatomy is imperative for a primary care provider, given their role in musculoskeletal care (Barrett & Liebman, 2020). Anatomical knowledge retention, an important learning outcome, diminishes over time, creating a need to understand the mechanisms involved in surface (rote/meaningless) and deep (meaningful) learning (Dayal et al., 2017; Hołda et al., 2019; Narnaware & Neumeier, 2020). Learning outcomes in gross anatomy courses do not appear to vary with instructional strategy or learning style (Aslaksen & Lorås, 2019; Husmann & O'Loughlin, 2019; Losco et al., 2017; O'Mahony et al., 2016; L. J. White et al., 2018; Wilson, Brown, et al., 2019). However, gross anatomy grades account for a large percentage of the final professional GPA variance, a predictor of first-time pass rate on the NPTE (S. H. Hayes et al., 1997). Although cognitive structures cannot be measured directly, data modeling strategies such as MDS and PFN have been used as an indirect method to represent cognitive structures in various domains. However, few research studies have addressed the use of these strategies in the health care professions, limited primarily to nursing (Azzarello, 2007) and psychology (Jaworska &

Chupetlovska-Anastasova, 2009); none have focused on the domains of musculoskeletal care or primary care providers such as orthopedic surgeons, chiropractors, or physiotherapists. Finally, methodological challenges in previous studies have limited the generalizability and reproducibility of results. Given the research gap, the goal was to understand the visual and quantitative representation of cognitive structure in the gross anatomy domain, exemplified by MDS and PFN strategies, and to validate the possible meaning of these quantitative measures in the context of entry-level physiotherapy education. Data modeling strategies may serve as a promising and innovative approach to the visual and quantitative representation of the cognitive structures in this domain, fostering deep and meaningful learning.

Purpose of the Study

The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. The study was initially conceived to focus on student cognitive structure in two contexts: how it changes over time and how it compares to two criterion standards (expert cognitive structure and academic performance). However, due to extenuating circumstances, this purpose was revised to reflect one context: comparing student cognitive structure to two criterion standards. The first part of this exploratory study addressed the potential relationships and agreement between student cognitive structure and expert cognitive structure (criterion standard one). There were no independent or dependent variables because an independent variable was not manipulated to examine a change in the dependent variable.

For the second part of this exploratory study, the dependent variable (criterion standard two) was the unit grade. The independent variables were MDS- and PFN-derived quantitative measures and the level of agreement between student and expert cognitive structures. MDS-derived measures included dimensionality, stress-1, Tucker's coefficient of coherence (TCC), R^2 (the variance accounted for by the model), and Euclidean semantic distances. PFN-derived measures included links, degree, eccentricity, coherence, similarity (with another network), and graph-theoretic semantic distances. Covariates addressed prior knowledge and included admission GPA and admission anatomy GPA. Moderating variables that were considered included instructor and mode of program delivery (residential or flexible). Age and gender were used for poststratification to ensure that the sample represented the target population. Validation of measures provided an essential foundation for improving gross anatomy learning, instruction, and assessment to enhance retention, transfer, and competency for physiotherapy students. Insights gained from this study may provide a unique perspective on cognitive structures and serve as an innovation in gross anatomy and physiotherapy education.

Research Questions

The research questions (RQs) addressed in this exploratory study were framed within the context of physiotherapy students enrolled in a first semester foundational gross anatomy course. In this study, cognitive structure (student and expert) was represented by the following measures: MDS dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances, and PFN links, degree, eccentricity, coherence, similarity, and graph-theoretic semantic distances. Prior knowledge was represented by two measures: admission GPA and admission anatomy GPA. The unit grade was measured by a weighted average of written and practical exam grades. The study was guided by the following RQs:

RQ1: Is there a meaningful change over time in the quantitative representation of student cognitive structure?

RQ2: Is there a relationship between student cognitive structure and expert cognitive structure while controlling for prior knowledge?

RQ3: Is there a relationship between student cognitive structure and unit grade while controlling for prior knowledge?

Given the RQs' exploratory nature, hypotheses for each RQ were not appropriate or indicated.

Theoretical Framework for the Study

The theoretical framework integrated cognitive science and data modeling strategies used for dimensionality reduction and data visualization. There has been a growing interest in integrating cognitive science with computational strategies borne of graph theory and network analysis (Siew, 2020). Cognitive science theory includes the ACT-R cognitive architecture (J. R. Anderson, 1996, 2007) as a coherent foundation for cognitive learning theory. Two data modeling strategies were considered within the context of the current study: MDS (Kruskal & Wish, 1978) and PFN (Schvaneveldt, 1990). The integration of cognitive science and data modeling may play an important role in the gross anatomy domain, specifically within the broader scope of physiotherapy education.

The ACT-R model of cognition is a cognitive architecture that provides a foundation for cognitive learning theory (J. R. Anderson, 2007; Ausubel, 1963; Mayer, 2009). This cognitive architecture proposes both symbolic and subsymbolic structures; the former as knowledge chunks stored as declarative knowledge, and the latter as production rules stored as procedural knowledge (J. R. Anderson, 1996). Cognitive units, an integral element of declarative memory, were proposed by J. R. Anderson (1980) as a precursor to what is now considered cognitive structure. Surface learning (also known as rote or meaningless learning) is the rote memorization of (symbolic) information with little coherent integration into an existing cognitive structure, whereas deep (meaningful) learning has clear relationships and associations between concepts that are integrated within a well-organized cognitive structure (Ausubel, 1963; Marton & Säljö, 1976; Mayer, 2009). Expert cognitive structures are typically consistent with those gained through deep learning and serve as a gold standard for student cognitive structure comparisons.

Recent developments in understanding the hippocampus, and its analogous functional representations within the ACT-R computational framework, may serve a role in better understanding cognitive structures (Burgess, 2014; O'Keefe & Nadel, 1978; Spiers, 2020; Theves et al., 2019). Cognitive maps reflect a growing understanding of a multilayered representation of knowledge organization consisting of two frames of reference: a high-dimensional space reflecting semantic spaces and a low-dimensional space reflecting semantic networks (Bellmund et al., 2018; Bottini & Doeller, 2020; Gärdenfors, 2017). The use of data modeling strategies such as MDS (Kruskal & Wish, 1978) and PFN (Schvaneveldt, 1990) may reflect the operationalization of these frames of reference: MDS as a potential visual and quantitative representation of the semantic space, and PFN as a potential visual and quantitative representation of the semantic network. These data modeling strategies provided an indirect yet explicit cognitive structure mapping, which was examined in the gross anatomy domain and physiotherapy students.

Nature of the Study

The nature of the study was initially conceived as a quasi-experimental, longitudinal, criterion-related validation study. It consisted of a one-group pretest-posttest design with pretest and posttest measures of proximity data (see A. D. Harris et al., 2006). However, extenuating circumstances precluded the option of the study being longitudinal in nature. A quasi-experimental design was appropriate for a criterion-related validation design because it minimized selection effects while providing external validity inherent to the design itself (see Burkholder et al., 2016). This study design aligned with the purpose and research questions by providing a foundation to examine the potential use of two data modeling strategies to quantitatively represent the cognitive structures of physiotherapy students learning gross anatomy. This exploratory study addressed the potential relationships and agreement between student cognitive structure and expert cognitive structure (criterion standard one). There were no independent or dependent variables because an independent variable was not manipulated to examine a change in the dependent variable. For the second part of this exploratory study, the dependent variable (criterion standard two) was the unit grade. The independent variables were MDS- and PFN-derived quantitative measures and the level of agreement between student and expert cognitive structures. MDS-derived measures included dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances. PFN-derived measures included links, degree, eccentricity, coherence, similarity (with another network), and graph-theoretic semantic distances. The covariate was prior knowledge (admission GPA and admission anatomy GPA). Moderating variables included instructor and mode of program delivery (residential or flexible). Physiotherapists in clinical practice were provided with a list of concept items derived from the gross anatomy course text *Clinically Oriented Anatomy* (Moore et al., 2018), the *Terminologia Anatomica* (FIPAT, 2019), and the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018) anatomical taxonomy. The physiotherapists were asked to rank the relevance of the concept items to clinical practice. A final list of 20 concept items was used for pairwise comparisons. Student and expert cognitive structure were derived from proximity/similarity data based on pairwise comparison of concept items. Expert cognitive structures served as the primary criterion standard. Experts were drawn from both physiotherapy centric instructors (Doctor of Physical Therapy [DPT] but not PhD in Anatomy) and anatomy centric instructors (PhD in Anatomy but not DPT or clinical doctorate). The sample consisted of volunteer student participants from the DPT program currently enrolled for the first time in the first-trimester gross anatomy course. Between-group analysis was used to examine the agreement with and relationships between student cognitive structures and criterion variables (expert cognitive structure and unit grade) while controlling for prior knowledge.

Although an examination of the current literature on the use of MDS and PFN revealed an absence of specific power calculations, a preliminary a priori power analysis was performed via G*Power with an alpha of 0.05 and power of 0.80 (see Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). Several sample sizes were generated based on the statistical analysis and both medium and large effect sizes for comparison. For paired sample *t* tests, the sample size was 34 (moderate effect size of 0.50) or 15 (large effect size of 0.80). For correlational analysis, the sample size was 84 (moderate effect size of 0.30) or 29 (large effect size of 0.50). For multiple regression, the sample size was 77 (moderate effect size of 0.15) or 36 (large effect size of 0.35). Previous studies provided a foundation upon which to view sample size in the context of a priori calculations. Seminal studies indicated a range of sample sizes from 35 to 71 participants (Acton et al., 1994; Egli, Streule, & Lage, 2008; Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016; Trumpower, Sharara, & Goldsmith, 2010).

Definitions

Agreement: Agreement between raters or measurements is reflected in three statistical measures: reliability, accuracy, and association. Reliability is calculated as interrater reliability via Krippendorff's alpha (A. F. Hayes & Krippendorff, 2007; Krippendorff, 2004). Accuracy is calculated as RMSD or root mean square deviation (Kopp-Schneider & Hielscher, 2019; Looney, 2018). Association is calculated as the strength of linear association via Pearson's product-moment correlation (Kopp-Schneider

& Hielscher, 2019; Looney, 2018). Agreement is used in the context of proximity data, MDS Euclidean semantic distances, and PFN graph-theoretic semantic distances.

Cognition: “The mental action or process of acquiring knowledge and understanding through thought, experience, and the senses” (Lexico, n.d., US dictionary).

Cognitive architecture: “A specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind” (J. R. Anderson, 2007, p. 7).

Cognitive map: “A schematic-like mental representation of the relationships between entities in the world including places, events, people, or even concepts” (Arzy & Schacter, 2019, p. 9).

Cognitive structure: “A hypothetical construct referring to the organization of the relationships of concepts in long-term memory” (Shavelson, 1972, pp. 226–227). This definition is aligned with the cognitive unit described by J. R. Anderson (1980, abstract) as “sets of elements that are stored in long-term memory in a single encoding act and which are retrieved from long-term memory in a single retrieval. By this definition, concepts in a semantic network are generally considered cognitive units”.

Cognitive structure mapping (cognitive mapping): The proposed representation of cognitive structure reflected in a cognitive map defined via two data modeling strategies (MDS and PFN) and their derived quantitative parameters and data visualization. A well-organized cognitive structure reflective of an expert has a greater number of links and stronger associations between them, greater coherence within the PFN model, greater goodness-of-fit, shorter semantic distances, and a higher percentage of variance attributed to the MDS model (Egli, Streule, & Lage, 2008; Schvaneveldt et al., 1985). Higher similarity (PFN) would indicate greater similarity with another cognitive structure, with a change and increase in similarity indicating the cognitive structure is more expert-like (Goldsmith et al., 1991).

Deep (meaningful) learning (construct): Meaningful learning involves the integration of new knowledge with an existing cognitive structure. It is “an iterative process in which learners must continue to refine, rectify, rearrange, and reorganize the content and structure of their knowledge so that their cognitive structure can be improved” (Wei & Yue, 2017, p. 5).

Deep (meaningful) learning (context): “Meaningful learning is distinguished by good transfer performance as well as good retention performance” (Mayer, 2009, p. 20).

MDS-derived measures: The quantitative measures representing both configuration properties and Euclidean semantic distances (Borg & Groenen, 2005). Configuration properties include dimensionality, stress-1 (goodness-of-fit, which represents the coherence of the model), TCC, and R^2. Euclidean semantic distances represent the degree of association of concepts within the spatial model. These measures are also reflected in a spatial visual representation of the proximity data.

PFN-derived measures: The quantitative measures representing both network properties and graph-theoretic semantic distances (Schvaneveldt, 1990). Network properties include links, degree (the number of links attached to each node), eccentricity (the maximum number of links between a node and all other nodes in a network), coherence (the degree to which an individual’s cognitive structure has internal links and associations), and similarity (the degree to which an individual's cognitive structure aligns with the cognitive structure of another individual or group of individuals). Graph-theoretic semantic distances represent the degree of association of concepts within the network model. These measures are also reflected in a network visual representation of the proximity data.

Prior knowledge: The knowledge that the student has before the initiation of the foundational gross anatomy course and is reflected in admission GPA (the student's grade point average on admission to the DPT program) and admission anatomy GPA (the student's grade point average of prerequisite anatomy and physiology courses on admission to the DPT program).

Unit grade: The content module's unit grade is a calculated weighted grade consisting of a multiple-choice exam and a practical, identification-based exam. The unit grade is weighted 50:40 (written 55.56%, practical 44.44.%).

Assumptions

The primary assumption was that cognitive structures exist as a construct grounded in cognitive science. Although this construct has been used extensively in the cognitive science literature, there is a clear lack of consistency in terminology and description. Associated with this assumption was that cognitive structures could be represented visually and quantitatively. The current cognitive science literature indicated that there are no direct representation methods; indirect methods such as natural language and graphical approaches are assumed to represent what they are thought to represent (Ifenthaler & Pirnay-Dummer, 2014). The third assumption was that computational strategies such as MDS and PFN utilizing semantic distances could serve as visual and quantitative representations of cognitive structure. The fourth assumption was that the proximity ratings reflect the perceptual construct; although this is the gold standard in the literature, it is an indirect method of assessing the perceptual nature of concepts and their individualized meaning. Although several assumptions related to cognitive structures and their representation exist, there is an extensive research basis to support indirect methods to represent this construct. These assumptions were necessary to utilize an innovative approach to representing the gross anatomy domain's cognitive structure in physiotherapy students.

Scope and Delimitations

The mapping of cognitive structures of physiotherapy students learning gross anatomy was poorly understood. This study was limited to a narrow subpopulation and context within physiotherapy education: the first-trimester physiotherapy student learning gross anatomy. The research design focused on one content module and its organization (cognitive structure) by the student, with pretest measures reflective of prior knowledge and posttest measures reflective of potential learning and changes over time. This narrow focus impacted internal validity by defining the potential causal inferences that could be made. The study was not intended to address the content within the content module; student generation of content could be assessed via free word association or creating and assessing individual concept maps. However, concept mapping as a representation of cognitive structure would entail a more comprehensive study to do so effectively. Unit grades depended on the validity of written (multiple-choice questions) and practical (identification and free association responses) examinations, limiting their utility as a criterion standard. The results may be generalizable to other physiotherapy students and to the use of data modeling strategies to examine cognitive structures in gross anatomy throughout the physiotherapy curriculum.

Limitations

There were several potential limitations to the study. The study's power and effect sizes were limited by the cohort size and recruitment of participants. Convenience sampling (also known as volunteer response sampling or nonprobability sampling) was used. Although this sampling strategy provided some inherent threats to internal and external validity, these threats were limited via several methodological considerations. The study addressed domain-specific effects, but these may be confused with the domain-general learning that occurs over time within the program. History and maturation bias were potential internal validity issues given the nature of the pretest and posttest measures. Although the examination of cognitive structures throughout a content module may reflect a real-world scenario, it is unknown how much time is required to make significant changes in those cognitive structures. Construct validity of the item list used for proximity/similarity ratings was critical. Instructors and pedagogical content knowledge of instructors varied; however, this was examined as a moderating variable. The quantitative representation of cognitive structure appeared to have construct validity based on research from various disciplines, including psychiatry (Egli, Streule, & Lage, 2008) and nursing (Azzarello, 2007). However, the construct has historically been poorly defined, with significant disparities between the construct definitions proposed by Piaget

(1926), Craik (1943), Quillian (1966), Shavelson (1972), J. R. Anderson (1980), and Jonassen et al. (1993). This disparity necessitated a clear definition and operationalization in the current study, which are described in Chapter 3. Finally, there was little research on the test-retest reliability of cognitive structures or the use of MDS or PFN in this context. Although test-retest reliability will become an essential consideration for practical applications, it was not within the scope of this criterion-related validation design.

Significance

The results of this study may help to fill the gap in understanding how physiotherapy students learn gross anatomy. Insights gained from this study may begin to foster the development of learning and instruction strategies that assist physiotherapy students in developing the cognitive structures necessary for anatomical knowledge acquisition, retention, retrieval, and near transfer to fulfill their role as primary MSK care providers. The potential implications of validated cognitive structure mapping strategies include improvements in the formative assessments of learning used in the context of gross anatomy education for physiotherapy students. However, these potential implications also extend to lifelong learning strategies in clinical practice. Positive social change may occur because of a better understanding of how physiotherapy students develop cognitive structures that promote deep learning in gross anatomy. This may provide the potential for both vertical integration within the physiotherapy curriculum and long-term transfer to clinical practice.

Summary

This study may contribute to the gross anatomy and health sciences literature by providing an enhanced understanding of how physiotherapy students learn gross anatomy. The theoretical framework provided a foundation grounded in cognitive science and emerging concepts. A criterion-related validation design was used to examine the mapping of cognitive structures via data modeling strategies. Assumptions and limitations to the study were clearly defined. Chapter 2 includes a review of the literature related to cognitive learning theory, gross anatomy education, and the data modeling strategies under consideration.]

Chapter 2: Literature Review

The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. The problem was that the mapping of cognitive structures of physiotherapy students learning gross anatomy is poorly understood. Gross anatomy is an integral component of entry-level physiotherapy education, and its importance cannot be overstated. Anatomical knowledge provides the foundation for diagnostic thinking and clinical practice (Timmerberg et al., 2019). A pilot study by Bayliss et al. (2017) indicated that gross anatomy in the first semester of the institution's DPT program was a predictor of success on the NPTE. Gross anatomy grade is also a predictor of final GPA in the physiotherapy program, with the first semester GPA is a predictor of first-time pass rate on the NPTE (S. H. Hayes et al., 1997; Wolden et al., 2020). However, there is a growing trend of decreasing anatomical knowledge in physiotherapy students and the broader scope of health professions (Dayal et al., 2017; Hołda et al., 2019; Narnaware & Neumeier, 2020).

Gross anatomy education, including student-centered learning strategies, has been extensively studied with little difference in learning outcomes reported. Traditional approaches to anatomical knowledge retention have focused on instructional design and teaching methods in gross anatomy. However, I found little research on how to reinforce deep and meaningful learning and promote retention and transfer. This may not purely be a function of the instructional strategy or teaching approach employed; it may also be a function of the learner profile, including how the learner acquires, encodes, and organizes knowledge. Paas and van Merriënboer (2020) noted the importance of learners' strategies in managing cognitive load to promote learning. Cognitive structures are an essential component of the learner profile. Several domains, including mathematics, anesthesiology, accounting, computer science, pulmonary physiology, chemistry, neuroscience education, author cocitation, psychology, research methods, nursing, psychiatry, chronic obstructive pulmonary disease and asthma, and vaccine education, have addressed the cognitive structures of students as they evolve toward those of experts during the learning process (Azzarello, 2007; Balloo et al., 2016; Casas-García & Luengo-González, 2012; Connor et al., 2004; Curtis & Davis, 2003; DiCerbo, 2007; Jaworska & Chupetlovska-Anastasova, 2009; McGaghie, McCrimmon, et al., 2000; McGaghie, McCrimmon, & Thompson, 1998; Neiles et al., 2016; Stevenson et al., 2016; Veríssimo et al., 2017; H. D. White, 2003). However, there was a gap in understanding how physiotherapy students learn gross anatomy, specifically the cognitive structures that promote deep learning in the gross anatomy domain.

Chapter 2 provides a review of the current state of the research that was relevant to the current study. Literature search strategies and topics are described. The theoretical foundations for the study, grounded in cognitive science, are discussed. A theoretical framework in the context of gross anatomy education is presented. Finally, operational constructs related to the study (multivariate analysis and representation) are reviewed to establish the relevance and application of measurement tools related to the study's methodology. The chapter concludes with a summary of the literature review and a transition to Chapter 3.

Literature Search Strategy

I conducted a comprehensive literature search using several online search strategies to attain saturation on topics relevant to the theoretical foundations, methodological approach, and variables used in the study. Walden University's Thoreau search tool was used to search databases related to educational research: Academic Search Complete, APAPsych, Education Source, ERIC, and SAGE Journals. Relevant topics were also searched within allied health and medical research databases: CINAHL Plus, ScienceDirect, and MEDLINE. This focused the search on relevant educational topics in the health professions and, specifically, physiotherapy and gross anatomy domains. Google Scholar was used as a supplemental search tool to establish articles' perceived importance and relevance based on the "cited by" functionality. Citation chaining was implemented both within journal articles and Google Scholar. This strategy deepened the pool of items for consideration based on seminal research or author citations.

A broad range of topics was considered, given the diversity of the constructs utilized within the study. Searches focused on the following key terms: *cognitive architecture*, *ACT-R*, *cognitive learning theory*, *cognitive structure*, *knowledge structure*, *schema*, *mental models*, *cognitive mapping*, *concept mapping*, *mind mapping*, *gross anatomy education*, *Pathfinder associative networks*, and *multidimensional scaling*. Key terms were searched independently and in combination with many references duplicated between databases and searches. The primary inclusion criteria were English language, full-text, peer-reviewed journal articles, books, and book chapters published after January

1, 2016, to encompass a 5-year search window. However, the research considered seminal to the study's theoretical foundations was not subject to this 5-year inclusion criteria. Because there was little research related to multivariate analyses specific to gross anatomy and physiotherapy education, the search scope was expanded to other allied health and medical professions and to all other domains. The Ulrichsweb Global Serials Directory was used to verify the quality and credibility of publications.

Due to the broad scope of the domains integral to the study, searches resulted in a review of several hundred article abstracts, full-text articles, books, and book chapters. After an initial scan of abstracts and a review of the methodology for potentially relevant studies, 327 relevant studies remained. The studies considered for inclusion in this literature review were grouped based on these topics: *cognitive science* (100 citations not inclusive of seminal research), *multivariate analysis strategies* including *multidimensional scaling* and *Pathfinder networks* (44 citations), and *gross anatomy education* (183 citations).

Theoretical Foundation

The acquisition, retention, retrieval, and transfer of knowledge are essential components of learning and instruction, making the cognitive mechanisms underlying these components critical to success. Cognition, or "the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses" (Lexico, n.d., US dictionary), has been the source of philosophical and scientific discussion. Approaches to this discussion have varied from neuroanatomical constructs integrating form and function within the central nervous system to symbolic and subsymbolic representations that address cognition from a computational approach (J. R. Anderson, 2007; Borst & Anderson, 2017; Camina & Güell, 2017). The theoretical foundation of the current study focused on the integration of cognitive science (including cognitive architecture and cognitive structures) and data modeling strategies (spatial and network) to represent cognitive structures in the context of the gross anatomy domain of knowledge in physiotherapy students. This theoretical integration provided the basis for cognitive learning theory relevant to the study of gross anatomy by physiotherapy students.

Cognitive Architecture

The quest for a unified theory of cognition has prompted the development of several cognitive architectures such as ACT-R (J. R. Anderson, 1996) and State Operator and Result (Soar; Laird, 2012). A cognitive architecture's functional goal is to understand better the cognitive mechanisms underlying cognitive functions such as knowledge acquisition, memory encoding and retrieval, and skill acquisition from a computational or connectionist perspective (Laird et al., 2017). J. R. Anderson (2007) defined cognitive architecture as "a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind," and added that "function of the mind can be roughly interpreted as referring to human cognition in all its complexity" (p. 7). Cognitive architectures are categorized based on their knowledge processing pattern: symbolic, emergent, or hybrid (Kotseruba & Tsotsos, 2020; Ye et al., 2018). Both ACT-R and Soar are considered hybrid architectures with symbolic and subsymbolic components; however, they originated in diverse domains: ACT-R in cognitive science and Soar in artificial intelligence (Kotseruba & Tsotsos, 2020; Laird et al., 2017; Ye et al., 2018). A cognitive architecture attempts to provide a unified conceptual framework of abstract representations and computational processes that can be used to understand cognitive function and to predict human behaviors consistent with seemingly diverse cognitive mechanisms (Kotseruba & Tsotsos, 2020).

The ACT-R model of cognition began as the ACT theory (J. R. Anderson, 1976) and evolved through several iterations, making it one of the leading coherent frameworks for cognitive science and cognitive learning theory (Laird et al., 2017; Ritter et al., 2019). As a hybrid cognitive architecture, the ACT-R model utilizes two abstraction levels: symbolic and subsymbolic (J. R. Anderson, 2007). The symbolic level addresses how the brain encodes knowledge, whereas the subsymbolic level addresses how knowledge is made available via retrieval (J. R. Anderson, 2007). Although ACT-R is used to model the mechanisms underlying many conscious cognitive functions and behaviors, it is not considered a comprehensive behavioral or social theory, nor was it intended as such. In developing ACT-R, J. R. Anderson’s (2007) goal was to develop a tool that could effectively link the brain with “functional cognition” (p. 8).

The cognitive architecture of ACT-R is represented by eight modules, each with an associated buffer: visual, aural, vocal, manual, imaginal, intentional, procedural, and declarative (J. R. Anderson, 2007). Four of these modules for the perceptual-motor system (visual, aural, vocal, and manual) interact with the external world (J. R. Anderson et al., 1997). The declarative module is the home of facts and information. The procedural module serves as the central production system and can only interact with the information currently residing in the buffers (J. R. Anderson et al., 2004; Laird et al., 2017; Ritter et al., 2019). The intentional module, also known as the control or goals module, serves to process goals by maintaining the intention of the problem in question (J. R. Anderson, 2007). The imaginal module, also known as the problem module, is focused on attention and the mental representation of the problem (J. R. Anderson, 2007). Buffers are associated with encoding and retrieval to and from the declarative module, as well as matching and execution within the procedural module (J. R. Anderson et al., 2008). The modular organization of the ACT-R model provides a computational yet functional perspective on cognition.

Complex systems (such as those necessary for J. R. Anderson's "functional cognition") are often composed of heterarchies serving a global function and hierarchies serving a local function (Cumming, 2016). Bechtel (2019) described cognition as a heterarchy of cognitive control mechanisms that improve the system's efficiency. The ACT-R model is aligned with these evolving perspectives of complex systems because it entails both parallel and serial processing in understanding the potential limiters to cognitive function. Parallel processing occurs within modules, with multiple operations occurring simultaneously, and between modules, with multiple modules working at the same time (J. R. Anderson, 2007). Serial processing is slower than parallel processing and is one of the most significant bottlenecks to cognition (J. R. Anderson et al., 2004). Within each module, there is a buffer limitation as each buffer accepts and processes one chunk of information at a time (J. R. Anderson, 2007). Between modules, the limiter is the potential dependence of one module awaiting data from another module (J. R.

Anderson, 2007). Parallel and serial processing, essential elements of both heterarchical and hierarchial organization, play a role in cognitive learning theory (Bechtel, 2019; Cumming, 2016).

Extensive research on higher level cognitive processes indicated a growing alignment of empirical data derived from functional magnetic resonance imaging (fMRI) with the modular constructs of ACT-R (J. R. Anderson et al., 2008). Borst and Anderson (2017) found that the modules of ACT-R (vocal, manual, visual, aural, imaginal, intentional, procedural, declarative, and associated buffers) mapped to corresponding brain regions via fMRI. The perceptual-motor system is mapped to the motor cortex (vocal and manual modules), visual cortex (visual module), and auditory cortex (aural module). The imaginal module is associated with the posterior parietal cortex. The intentional module is associated with the anterior cingulate cortex. The procedural module is mapped to the basal ganglia, thalamus, amygdala, and cerebellum. Buffers associated with the modules, most notably the retrieval buffer, are associated with the prefrontal cortex. Finally, the declarative module is mapped to the hippocampus and medial temporal cortex (J. R. Anderson, 2007; J. R. Anderson et al., 2008; Borst & Anderson, 2017; Eichenbaum, 2017; Stocco, 2018). These mappings create a potential direct link between neuroanatomical structure and cognitive function.

A primary goal of the ACT-R cognitive architecture is to understand better the cognitive mechanisms underlying cognitive functions. One of these cognitive mechanisms is the role of memory in cognitive functions such as knowledge acquisition, retention, and retrieval. Traditional descriptions of memory (Camina & Güell, 2017;

Kotseruba & Tsotsos, 2020) focus on duration (short- and long-term) and type (declarative and procedural). Atkinson and Shiffrin (1968) proposed several modalities of memory, including sensory memory leading to short-term memory and ending in long-term memory. Short-term memory includes working memory, whereas long-term memory consists of both declarative and procedural memory (Atkinson & Shiffrin, 1968). Although ACT-R does not represent these specific modalities per the Atkinson and Shiffrin definitions, it does provide mechanisms aligned with them.

Sensory memory, also known as sensory registers, consists of iconic (visual), echoic (auditory), and haptic (touch) perceptions, among others yet to be fully described (Camina & Güell, 2017). Sensory memory is analogous to perceptual information drawn from the vast amount of sensory information available to the individual that registers at any given moment in time (Camina & Güell, 2017). However, the retention of this information lasts for less than 100 milliseconds (Camina & Güell, 2017). If sensory memory is not moved from the sensory registers into short-term memory (more specifically, working memory) and acted upon, it will be lost (Kotseruba & Tsotsos, 2020).

Working memory was considered a part of short-term memory. Baddeley and Hitch (1974) extended the Atkinson and Shiffrin (1968) model to include four functional components of working memory: central executive, episodic buffer, visuospatial sketchpad, and phonological loop. The central executive is analogous to the imaginal module in ACT-R that addresses attentional focus and current mental representation of the problem, and the episodic buffer adds a temporal component to working memory

(Baddeley & Hitch, 1974). The most critical aspect of Baddeley's contribution is the discrete visuospatial and verbal/auditory processing within working memory (Baddeley, 1983, 2010). These elements are analogous to the visual and auditory modules and buffers in ACT-R (J. R. Anderson, 2007; J. R. Anderson et al., 1997). Visuospatial working memory will affect several cognitive skills, including mental rotation and folding, field independence, and general reasoning abilities (Castro-Alonso & Atit, 2019; Keehner, 2011). Baddeley and Hitch's research on working memory paralleled the dual coding theory first described by Paivio (1971, 1986), which consisted of the parallel processing of verbal and nonverbal stimuli into representations that contribute to referential (between-system) and associative (within-system) networks (Paivio, 1971). However, the functional importance of working memory is to serve as the gateway to long-term memory reflected in its role in cognitive processing mechanisms such as knowledge acquisition, decision making, and clinical reasoning (J. R. Anderson, 1996; Chai et al., 2018; Hruska et al., 2016; Ritter et al., 2019). Although ACT-R does not define working memory per the work of Atkinson and Shiffrin (1968) or Baddeley and Hitch (1974), it does provide an analogous functional representation of these constructs via cognitive units, activation, and strength of association (J. R. Anderson, 2007; J. R. Anderson et al., 1996).

Working memory is integral to cognitive function and creates a bottleneck that can limit cognitive capacity and processing (Paas et al., 2004). The primary constraint upon working memory, for both encoding and retrieval, is cognitive load: extrinsic, intrinsic, and germane (Paas et al., 2004). If the cognitive load increases, it can lead to stereotyping, bias, central tendency bias, and fundamental attribution error – all elements will limit the efficacy of knowledge acquisition, learning, and decision-making (Allred et al., 2016). Instructional design, the learner's expertise and prior knowledge, and the inherent complexity of the domain content can create excessive cognitive load (van Merriënboer & Sweller, 2005). However, cognitive load can also be beneficial when cognitive resources are utilized that encourage efficient encoding to and retrieval from long-term memory and subsequently enhance schema construction (Mayer, 2009). Cognitive load, and managing it effectively, thus becomes a significant challenge to working memory.

Long-term memory provides vast storage that is persistent over time. Knowledge acquisition in long-term memory is represented in ACT-R via two modules: declarative memory and procedural memory (J. R. Anderson, 2007). Declarative knowledge, also known as conscious and explicit knowledge, comprises symbolic knowledge chunks representing facts, events, and associations stored in long-term memory (Yee et al., 2017). Declarative knowledge is further subdivided into semantic and episodic knowledge; semantic knowledge is information composed of objects and relationships, whereas episodic knowledge is reflective of past autobiographical experiences (Kotseruba & Tsotsos, 2020; Yee et al., 2017). From an anatomical perspective, the hippocampus has an integral function in spatial and non-spatial episodic and semantic memory organization (Duff et al., 2020; Eichenbaum, 2017). Semantic memory has a high degree of flexibility and is dynamic over time, continuously adapting with the integration of new knowledge, associations, and representations (Duff et al., 2020;

Klooster et al., 2020). There are also individual variations in semantic memory based on the differences in personal experience and meanings associated with them (Yee et al., 2017). In contrast, procedural knowledge, also known as unconscious and implicit knowledge, comprises subsymbolic production rules that utilize knowledge chunks, conditions, and actions stored in declarative knowledge (J. R. Anderson, 2007). The procedural module can only act upon declarative knowledge via knowledge chunks that are shuttled into and out of the declarative (retrieval) buffer as needed (J. R. Anderson, 2007). Experiences influence the content of both declarative and procedural modules (J. R. Anderson, 2007).

Chunking and Activation

Knowledge chunking and activation are integral components of encoding and retrieval in ACT-R and serve as an analogous functional representation of the traditional working memory construct. Knowledge chunks, also known as cognitive units, are symbolic representations of information encoded in the declarative module (J. R. Anderson, 2007; Ritter et al., 2019). Encoding entails a new chunk of knowledge being indexed to a corresponding aspect of prior knowledge based on the context or problem for which it is being encoded (J. R. Anderson, 2007). Each chunk in declarative memory has a base-level activation that indicates how readily available a piece of information is in the declarative module based on the context of the problem being solved (J. R. Anderson, 2007; J. R. Anderson & Matessa, 1997). J. R. Anderson's concept of activation was a refinement of the spreading activation theory initially conceived by Quillian in reference to a computer program (Collins & Loftus, 1975; Quillian, 1962). The more frequently a knowledge chunk is encountered, the more likely it will be retrieved in the future, and thus the higher the level of activation. Buffers can only hold one chunk with memory limitations defined by those chunks in declarative memory with sufficient activation (J. R. Anderson et al., 1996). Activation can increase via repetition or increasing the links between and within cognitive structures (J. R. Anderson & Schunn, 2000). This phenomenon is called the "practice effect," in which any given memory or knowledge chunk can be associated with either a few or many other knowledge chunks. The "fan effect" occurs when retrieval time is affected because of the higher number of associated knowledge chunks based on what J. R. Anderson (2007) called "associative interference." However, the fan effect diminishes when facts are well-organized into cognitive structures, a key element of deep learning.

Cognitive Learning Theory

A robust cognitive architecture such as ACT-R provides a foundation for cognitive learning theory that is consistent with schema theory (Piaget, 1926), assimilation theory (Ausubel, 1963), and the learner's self-directed strategies and approach to learning (Marton & Säljö, 1976). The schema theory of Piaget (1926) described two processes that occur during learning: accommodation (an adaptation of existing knowledge) and assimilation (formation of new knowledge). Ausubel (1963) proposed that new knowledge builds upon prior knowledge, thereby revising and refining its cognitive structure. These theories have further evolved into a differentiation between deep (meaningful) learning and surface (rote or meaningless) learning (Marton & Säljö, 1976). Deep learning involves developing relationships and meaning for the content in question within a well-organized cognitive structure; in contrast, surface learning refers to the memorization of discrete facts within a poorly organized cognitive structure with a goal of factual retrieval (Marton & Säljö, 1976; Mayer, 2002b). Students can experience changes in their approach to learning over time and often transition from a surface learning approach to more strategic and deep learning approaches (McDonald et al., 2017).

The modern-day technological evolution of the work of Piaget, Ausubel, and Marton and Säljö is the cognitive theory of multimedia learning (Mayer, 2002a, 2009). Mayer's work unifies cognitive load theory, active learning, working memory, and the dual coding theory of Paivio (1986). Working memory constraints become critical in meaningful learning, demanding the coordinated and effective use of visuospatial and textual input via dual coding (Mayer, 2009). The primary goals of multimedia instruction are to enhance dual coding (visuospatial and auditory/text) while optimizing cognitive load in the process (Mayer, 2002a, 2009).

Mayer (2009) explicitly defined three learning outcomes: no learning (poor retention, poor transfer), rote learning (good retention, poor transfer), and meaningful/deep learning (good retention, good transfer). Transfer of learning indicates that knowledge can be applied to a new learning scenario; near transfer reflects an activity similar to the context in which the knowledge was encoded, whereas far transfer occurs when the two learning contexts or activities are dissimilar (Mayer, 2009; Montpetit-Tourangeau et al., 2017). Within the context of physiotherapy practice, clinical reasoning and diagnostic thinking require near transfer (Montpetit-Tourangeau et al., 2017). It is believed that deep learning builds upon prior knowledge and experiences, creating more developed cognitive structures that subsequently enhance both retention and transfer of learning to higher-order thinking (Krathwohl, 2002; Smith, Stockholm, et al., 2017).

In the context of cognitive learning theory, the ACT-R model is consistent with the previously noted theories of Piaget (1926), Ausubel (1963), Marton and Säljö (1976), and Mayer (2009). Several mechanisms in both declarative and procedural modules are responsible for learning (J. R. Anderson, 2007). Learning occurs via creating new knowledge chunks in the declarative module (building upon prior knowledge) or creating new production rules in the procedural module via proceduralization, composition, generalization, and analogy (Whitehill, 2013). The strengthening of activation in existing chunks will also produce learning (J. R. Anderson, 2007). Surface learning is represented by the passive learning of symbolic structures in declarative memory; in contrast, deep learning is characterized by linking knowledge chunks via procedural memory with active cognitive structure development as prior knowledge is revised and updated (Whitehill, 2013). J. R. Anderson and Schunn (2000) noted a differentiation in cognition depending upon the goal of learning – be that long-term competency or short-term retrieval of knowledge – that paralleled the work of Marton and Säljö (1976) in the context of student approaches to learning.

A glaring omission in the scientific literature on cognitive load and working memory provides a clear gap. Cognitive learning theory, exemplified by Mayer's cognitive theory of multimedia learning, is built upon the cognitive load theory. In a 2020 review of cognitive load theory and the learning of complex tasks, Paas and van Merriënboer noted the critical elements in managing cognitive load: learning task characteristics, available schemas in long-term memory, the learner, and the learning environment. However, in their review, they addressed learning tasks, the learner, and the learning environment – with no further mention of the schemas in long-term memory and how this can impact cognitive load. This is a clear example of the gap that persists in cognitive learning literature. Although research has focused on instructional design (the learning task), the learner (collaboration, motivation, learning styles), and the learning environment (split attention, stress, instructor pedagogical content knowledge, negative emotions), little is focused on the efficient development of cognitive structures within long-term memory.

Cognitive Structure

Long-term memory is the home of cognitive structures – a term that is traditionally synonymous with a broad range of poorly-defined terms and constructs, including structural knowledge (Jonassen et al., 1993), cognitive units (J. R. Anderson, 1980), semantic networks (Quillian, 1966), schemata (Piaget, 1926), mental models (Craik, 1943), and cognitive structure (Shavelson, 1972). The historical origins of cognitive structures lie in the schema theory proposed by Piaget (1926). Although all share similar themes and represent similar cognitive constructs, the ambiguity in terminology makes consistency and clarity in research and application difficult. For this review, the term "cognitive structure" will reflect the operational definition proposed by Shavelson (1972, p. 226-227): "a hypothetical construct referring to the organization of the relationships of concepts in long-term memory." This term will be used to provide some consistency to this construct in the following discussion.

Several common features of cognitive structures emerge in the literature. A cognitive structure represents knowledge specific to both the individual and the domain that has an internal organization (Liu et al., 2019). This representation is based on the individual's declarative knowledge (including semantic and episodic memory), perceptions, and experiences. Cognitive structures are domain-dependent and aligned with cognitive tasks for that domain (J. R. Anderson & Schunn, 2000). However, there is not one exclusive mental representation in any given domain, though mental representations between individuals may share similar concepts as associations. Jonassen et al. (1993) noted that declarative knowledge is composed of content knowledge (what you know) and structural knowledge (how you organize it), with the term "structural knowledge" often being used interchangeably with "cognitive structure." Contextually relevant information, often consisting of both text and images, is an important element of cognitive structure and defines how it is encoded for future retrieval (Gilboa & Marlatte, 2017; Richter et al., 2019; Ziembowicz, 2017). Cognitive structures are sensitive to chronological order, hierarchical organization, cross-connectivity, and context (Ghosh & Gilboa, 2014). Working memory is critical to the development of cognitive structures; however, it is readily diverted from this task with increases in cognitive load, limiting the learner's ability to attain meaningful learning (Paas et al., 2004). As cognitive structures improve, the overall cognitive load decreases (Wirzberger et al., 2018). However, though all of these phenomena have been observed and documented within the context of knowledge representation, how these cognitive structures develop remains a neurological and cognitive mystery, and their operationalization remains elusive (Ifenthaler et al., 2011; Ziembowicz, 2017).

The foundation for cognitive structures is prior knowledge, making it a critical element in cognitive processing (van Kesteren & Meeter, 2020). Incorporating new knowledge entails indexing it to prior knowledge; this also predicts future behavior (J. R. Anderson, 2007; van Kesteren & Meeter, 2020). Cognitive structures continually undergo revision and updating as knowledge and learning progress (J. R. Anderson, 1996; Noushad & Khurshid, 2019; Zulu et al., 2018). Castro and Siew (2020) proposed that although an understanding of cognitive structure is important, the cognitive structure's transformation with learning is equally important. Flexibility and adaptability are essential for new knowledge and evolving knowledge organization (Ghosh & Gilboa, 2014). However, prior knowledge can also lead to misconceptions that become a part of the cognitive structure that is subsequently difficult to "unlearn" unless the cognitive structure changes (Ziembowicz, 2017). Cognitive structures can develop that promote bias and create false memories; if they are well-established in long-term memory, they can also strengthen misconceptions and be highly resistant to change (van Kesteren & Meeter, 2020). Once again, the knowledge context is critical in encoding knowledge and retrieval based on the problem being solved.

The ACT-R model provides a functional framework for representing cognitive structures based on the chunking of information in the declarative module. This was founded in J. R. Anderson's early work that described cognitive units consisting of concepts (vertices or nodes), propositions (edges or links), and schemata (chunks or clusters) as an abstract representation within the computational framework of the ACT-R model (J. R. Anderson, 1980, 1996). As the retrieval buffer contents are limited to one chunk of information, it is essential to develop chunks containing a higher degree of knowledge or associated data bundled within the cognitive unit (J. R. Anderson, 2007). Retrieval from long-term memory is a hallmark of retention, but effective encoding of knowledge into well-organized and relevant cognitive structures free of misconceptions is a prerequisite.

The premise underlying the declarative module's chunking mechanism provides a degree of implicit structural organization in declarative memory based on the mechanisms of base-level and associative activation proposed by the ACT-R model. Jonassen et al. (2005) envisioned structural knowledge as a bridge between declarative and procedural knowledge. However, these authors did not clearly define the mechanisms underlying structural knowledge, and this conceptualization may be redundant based on the premise of activation within the ACT-R model. Activation involves not only the degree of usefulness in the past (base-level) but also the relevance to the current problem (associative) based on attentional weight (number of sources of activation) and strength of associations with other facts retained in declarative memory (J. R. Anderson, 2007). In this way, declarative memory and knowledge chunks have an implicit structure and organization based on prior value and strength of associations in solving the current problem.

Expertise

Cognitive structures and knowledge organization play a significant role in differentiating novice and expert. As noted by Jonassen et al. (1993), declarative knowledge is a function of both content knowledge and knowledge organization. Expertise and the development of clinical competency is more than just acquiring more knowledge; how you know it is essential (Persky & Robinson, 2017). Experts have two significant differences compared with novices: a high quantity of domain knowledge, and a high quality of internal structure and organization of the domain knowledge (Gardner et al., 2019; Siew, 2020). Improving the structural organization of knowledge through refined cognitive structures has also been shown to improve knowledge transfer success, an essential element in diagnostic thinking and clinical reasoning (Kubsch et al., 2020; Salkowski & Russ, 2018).

Cognitive task analysis in any domain reveals several cognitive factors that differentiate experts from novices, including mental models, perceptual skills, sense of typicality, routines, and declarative knowledge, many of which depend on cognitive structures (Crandall & Hoffman, 2013). However, although research in diagnostic thinking often focuses on experts' cognitive task analysis, this may be problematic without understanding the cognitive structures underlying them. For example, Sullivan et al. (2014) reported that experts would omit 71% of clinical knowledge steps, 51% of action steps, and 73% of decision steps when describing a procedure to learners. Developing expertise – or becoming a competent professional – demands an understanding of what and how they know and an awareness of the learning processes that allow one to become an expert by transforming their cognitive structure eventually (Castro & Siew, 2020; Jung et al., 2016). Experts will tend to emphasize the use of inductive reasoning and pattern recognition. However, experts encounter premature closure and cognitive bias more frequently (Norman et al., 2017). In contrast, novices will use deductive reasoning as their primary strategy based on an overall lack of pattern recognition based on their lack of experience (Norman et al., 2017). They lack the representational skills and competence of experts (Kozma, 2020). Cognitive and metacognitive skills alone account for 22% of the variance in physiotherapy students' clinical reasoning skills (Elvén et al., 2019). These factors may prompt the novice to examine their cognitive structures more fully during clinical reasoning and diagnostic thinking (Shin, 2019).

In the context of ACT-R, experts will display several essential characteristics. They will have enhanced declarative knowledge based on the chunking of information in the declarative module and enhanced activation levels and strength of association. Experts will also have an improved ability to retrieve the knowledge faster and more efficiently due to improved matching via the procedural module with the declarative module. Hruska et al. (2016) noted that novices utilize working memory more so than experts based on the increased activation of the prefrontal cortex on fMRI, the site of the retrieval buffer in ACT-R (J. R. Anderson, 2007). Errors in experts' diagnostic thinking may be related to automation; as knowledge chunks evolve, there will often be a removal of certain details as the information is assimilated (J. R. Anderson, 2007). Experts may also fail to utilize the fan effect characterized by cognitive bias or premature closure in their diagnostic thinking (Norman et al., 2017).

Representation

The representation of a cognitive structure parallels the elements of cognitive task analysis: knowledge elicitation, knowledge representation, and data analysis (Crandall & Hoffman, 2013). Although cognitive structures exist as symbolic mental representations, they remain hypothetical constructs, as noted in the operational definition proposed by Shavelson (1972). This makes a direct measurement of cognitive structures elusive, with indirect methods limited by a lack of psychometric properties such as reliability and validity. Several approaches have been used to indirectly measure cognitive structure dating back to Preece (1976). Ifenthaler et al. (2011) proposed two indirect methods for knowledge elicitation: natural language and graphical. Natural language methods include verbal reporting, think-aloud, free word association, controlled word association, pairwise comparisons, structure formation, and eye-tracking (Ifenthaler et al., 2011; Tsai & Huang, 2002; van Gog et al., 2009). Graphical methods include tree construction, flow maps, concept maps, causal diagrams, DEEP, and Pathfinder analysis (Ifenthaler et al., 2011; Tsai & Huang, 2002; van Gog et al., 2009). Natural language methods are limited by the individual's linguistic skills and fluency in any relevant domain taxonomy or ontology (Clarkson & Whipple, 2018). In contrast, graphical methods are difficult to compare and depend upon the evaluator's interpretation, which may not align with the meanings implied by the picture's creator. Cognitive structures can contain several different contextually relevant modalities, such as text and images. This aligns with the cognitive theory of multimedia learning (Mayer, 2009) and the dual processing theory of Paivio (1971). Cognitive structure representation may benefit from an integration of these two components.

Concept mapping, one of a group of visual mapping strategies, has been proposed as an indirect method of knowledge elicitation and representation for cognitive task analysis and cognitive structures (Crandall & Hoffman, 2013; Davies, 2011; Ifenthaler et al., 2011). In a systematic review by Buitrago and Chiappe (2019), concept mapping was the most widely used knowledge representation approach. This method provides a clear example of the potential to visualize an individual's cognitive structure to foster deep learning. The underlying premise of concept mapping stems from the use of "advance organizers" by Ausubel (1963) to visually represent the development of cognitive structures while building new knowledge on prior knowledge. Novak and Gowin (1984) aligned their concept mapping theory with Ausubel's assimilation theory and Marton & Säljö's approaches to learning to advance the "advance organizer" premise proposed by Ausubel.

Concept mapping has been used in learning, instruction, and assessment across a wide variety of domains. A systematic review by Stevenson et al. (2017) examined the extensive research on concept mapping and found consistently favorable learning outcomes. Concept mapping has been used to promote the development of cognitive structures, knowledge visualization and retention, critical thinking, clinical reasoning, near transfer, and meaningful learning while decreasing the cognitive load of the learner (Abd El-Hay et al., 2018; Bressington et al., 2018; Machado & Carvalho, 2020;

Montpetit-Tourangeau et al., 2017; Schroeder et al., 2018; Si et al., 2019; Wang et al., 2018; Yue et al., 2017; Zulu et al., 2018). Kinchin et al. (2019) noted that concept maps could be differentiated based on their topology (spoke, chain, and network) and that they could be used to represent different types of knowledge (novice, theoretical, practical, and professional). Radwan et al. (2018), in a study of final year medical students, reported a statistically significant correlation between concept mapping scores and clinical reasoning scores as assessed by the Script Concordance Test. Concept mapping could also serve in the natural progression from learning and instruction to structural assessment (Hartmeyer et al., 2018). Concept mapping has extensive research support, making it highly relevant to the discussion of cognitive structures.

Several challenges exist in the effective implementation of concept mapping as a formative and summative assessment. Psychometric properties such as reliability and validity may limit their effective real-world use as a quantitative measurement (Siew, 2020). There is a need for reliable and valid rubrics, graders familiar with the rubric, and the time required to grade each student concept map (Novak & Gowin, 1984). Concept mapping is a cognitive skill that requires training and repetition over time to develop. The visual representation provides rich data in visuospatial and textual references and includes personal meaning, which may be difficult to assess. Concept mapping may offer a tangible, paper- or digital-based visualization of a student's cognitive structure; however, it demands the student's ability to translate a perceived mental representation to an overt visualization. Many may be challenged to do so. Although concept mapping may provide valuable lessons regarding the importance and relevance of cognitive structures to learning and instruction, it may not comprehensively view the cognitive structure's multidimensional nature.

Cognitive Mapping

Cognitive structures suffer from a high degree of ambiguity in both construct description and representation. Good operational definitions are often lacking as they are often used to describe both domain-general and domain-specific applications. Descriptions of these constructs vary from cognitive spaces to conceptual spaces and from semantic networks to cognitive maps (Bellmund et al., 2018; Gärdenfors, 2004, 2017; Lieto et al., 2017). The term "cognitive mapping" was initially proposed by Edward Tolman (1948) in the context of spatial mapping within the hippocampus, potentially merging cognitive structures with neuroanatomy. Tolman envisioned cognitive mapping involving both spatial and non-spatial components. This concept was extended by O'Keefe and Nadel (1978) in the seminal work *The Hippocampus as a Cognitive Map*. The importance of grid-like cells in non-spatial conceptual knowledge, much as Tolman had originally proposed, was reported by Constantinescu et al. (2016). Spiers (2020) noted that the perspective of this "universal cognitive map" was enhanced by the Nobel Prize work of O'Keefe, Moser, and Moser in identifying grid and place cells in the hippocampus and their role in memory (Burgess, 2014).

Tolman's "cognitive mapping" appears to be the construct definition best aligned with cognitive architecture, cognitive structure, and neuroanatomy. Arzy and Schacter (2019, p. 9) provide an operational definition of the cognitive map as "a schematic-like mental representation of the relationships between entities in the world including places, events, people, or even concepts." Behrens et al. (2018) noted that cognitive maps provide a framework for knowledge organization. The premise of a multidimensional cognitive map is on the cutting edge of research focusing on the role of the hippocampus (Theves et al., 2019). This provides the theoretical, computational, and neuroanatomical basis for cognitive structures, both spatially and non-spatially, grounded in complex systems' heterarchical and hierarchical organization (Bechtel, 2019; Bottini & Doeller, 2020; Cumming, 2016; Zemla & Austerweil, 2018).

Cognitive mapping may serve as a more comprehensive representation of the cognitive structure. It reflects both content (in the form of concepts) and structure (in terms of memory organization) that serves as a frame of reference for the individual. Gärdenfors (2004) provided a foundation for this frame of reference that the individual perceptually determines, with the meaning being specific to the individual and not universal. Complex systems, exemplified by the brain and its cognitive mechanisms, can be represented by the system's features, similarity, and connectivity (Comin et al., 2016). Features are measurements used to describe a node or concept, be they intrinsic or induced, and often reflect its spatial position (Comin et al., 2016). Similarity reflects the relatedness between two nodes based on features or correlation (Comin et al., 2016). Connectivity defines the system's network representation and its associated topology (Comin et al., 2016). Cognitive mapping within this context reflects factual and structural knowledge, the individual's perceptions of the topic or domain, and the characteristics of a complex system reflected in the individual (Comin et al., 2016; Egli, Streule, & Lage, 2008). Bottini and Doeller (2020) proposed two interrelated frames of reference for the individual: high-dimensional and low-dimensional spaces. These frames of reference are aligned with the complex systems described by Comin et al. (2016) and provide a foundation for cognitive mapping and its representation.

High-dimensional spaces represent concepts in several dimensions like semantic or conceptual spaces. They are self-centered and relevant to the individual while being egocentric and dependent on the individual's perspective and frame of reference (Bottini & Doeller, 2020). These high-dimensional spaces are believed to be developed within the parietal cortex, are grounded in the sensorimotor experiences of the individual, and may include perceptual, functional, and abstract dimensions. Gärdenfors (1996, 2004, 2017) envisioned these conceptual spaces as having a spatial, geometric, or topological representation based on quality dimensions and perceived similarities. Conceptual spaces may serve as multidimensional frameworks of knowledge hierarchies aligned with high-dimensional frames of reference (Bellmund et al., 2018; Gärdenfors, 2004, 2017).

Low-dimensional spaces represent concepts in 1 or 2 dimensions that can include spatial and non-spatial knowledge. These spaces are like semantic networks in that they are world-centered and factual while being allocentric and independent of the point of view and frame of reference (Bottini & Doeller, 2020). Much of the current literature on cognitive structures refers to some degree of network representation; J. R. Anderson developed the ACT-R cognitive architecture as a computational framework with associative features (J. R. Anderson, 2007; Paivio, 1986). Many neuroanatomical constructs lend themselves to representation as a network, with network analysis used extensively to study brain dynamics (Bhuvaneshwari & Kavitha, 2017; Stam &

Reijneveld, 2007) and connectivity patterns within the brain. This may prove beneficial in representing cognitive structures (Farahani et al., 2019). The structure of a semantic network is also similar to the nodes and links found in concept mapping.

One of the potential failures of cognitive structure representation is the limited dimensionality of the representation. Vukić et al. (2020) described a “multidimensional knowledge network” having a multilayered representation. Bottini and Doeller (2020) noted that individuals might have cognitive maps that reflect two frames of reference, with individuals navigating between both frames of reference to adequately represent their cognitive structure. This entails an individual operating within a global heterarchy of concepts that contains local hierarchical networks (Bechtel, 2019; Cumming, 2016). This multidimensional representation may require a computational framework; this provides a potential role of multivariate techniques to provide data-driven representations of the cognitive structure.

Cognitive Mapping and Data Modeling

Multivariate analysis may serve as a means of quantifying an individual’s cognitive map. Several strategies exist for quantitative cognitive modeling, including cognitive architectures, graphic models, complex systems, and networks (Shiffrin, 2010; Siew, 2020). High-dimensional (egocentric) and low-dimensional (allocentric) representations (frames of reference), as proposed by Bottini and Doeller (2020), combined with the components of complex systems (features, similarity, and connectivity), may provide a comprehensive and quantitative representation of an individual’s cognitive structures.

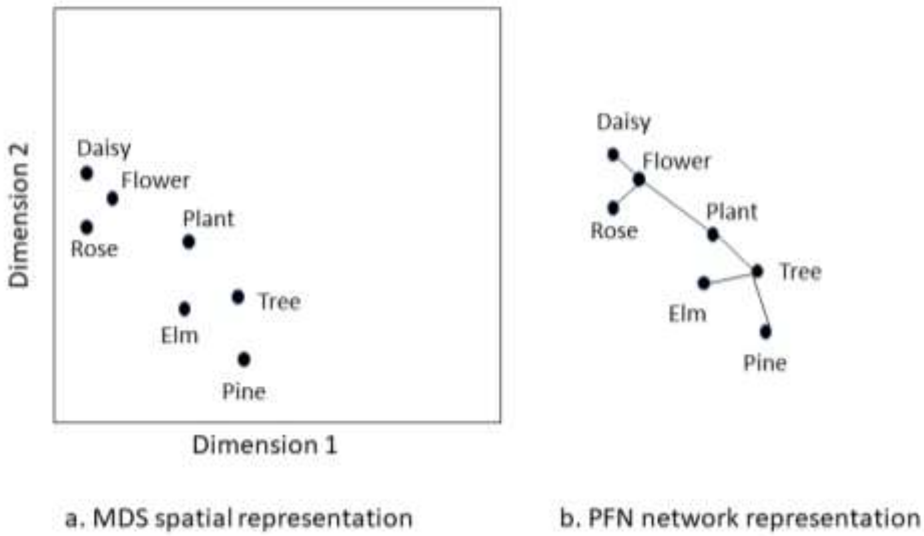
Jonassen et al. (1993) and, more recently, Dozortsev et al. (2017) noted that structural knowledge could be represented empirically by spatial/dimensional and network methods. Dimensional approaches transform the cognitive structure and reveal spatial relationships and clusters of concepts while maintaining semantic distances (Jonassen et al., 1993). Network approaches extract concepts and associative relationships as a part of a semantic network (Jonassen et al., 1993). Although dimensional and network approaches are visuospatial and computational, they can also use a natural language strategy such as pairwise comparisons for semantic similarity or dissimilarity to provide the raw data necessary. These representations, or structural models as described by Schvaneveldt (1990), can potentially be addressed via two multivariate dimensionality reduction techniques grounded in psychometrics that utilize proximity scaling algorithms: MDS and PFN. Scaling algorithms may serve as an effective means of knowledge representation as they can be empirically derived to capture the structure and organization of knowledge (Cooke et al., 1986). Strategies that implement quantitative approaches to analysis are well-aligned with the demands of cognitive task analysis and assessing the structural organization of knowledge structures (Siew et al., 2019).

The key concept consistent between these strategies is perceived semantic similarity and distance (Chen, 1997). The content of the knowledge is predefined with an a priori item list defined by content experts or that which the curriculum deems necessary for acquisition, retention, and retrieval clinically (Gisick et al., 2018). Semantic similarity or proximity data derived from pairwise comparisons can be utilized to represent the underlying structure of the data in two ways: a global overview that examines the feature space using MDS and a local structural view that examines the similarity and connectivity of the network using PFN (Chen, 1997; Goldsmith et al., 1991). MDS, also known as Principal Coordinates Analysis, examines proximity data in terms of pairwise distances in the context of a spatial representation of concepts (Buja et al., 2008; Jonassen et al., 1993). PFN examines the same proximity data in terms of pairwise associations in the context of a network representation of nodes and links (Buja et al., 2008; Jonassen et al., 1993). The visualization of a network's vertices and edges is akin to a concept map, and its analysis is grounded in network science (Newman, 2018). Both MDS and PFN are used to reveal the underlying structure of proximity data based on an individual's perception of the similarity of paired items representing domain concepts (Chen, 1997). Figure 1 provides an example of both MDS and PFN representations.

Although the empirical evidence of cognitive mapping is limited to recent advances in fMRI and neuroanatomical connectivity research, the premise for the functional representation of multiple frames of reference and implicit structural knowledge was noted by Goldsmith et al. (1991). In this seminal study, the authors compared MDS and PFN in a group of 40 college students enrolled in a psychological research techniques course. Their conclusions proposed that MDS may provide more significant insights into the global structure of knowledge, whereas PFN may provide greater insights into the local knowledge structure. Subsequent research by Gillan et al. (1992), Gonzalvo et al. (1994), and Bonebright et al. (2005) supported the work of Goldsmith et al. (1991).

Figure 1

Multidimensional Scaling and Pathfinder Network Representations



Researchers in a diverse range of domains – searching for a computational strategy for the structured assessment of knowledge and cognitive structures – have implemented MDS and PFN analysis. These domains include mathematics (Casas-García & Luengo-González, 2012; Veríssimo et al., 2017), anaesthesiology (Connor et al., 2004), accounting (Curtis & Davis, 2003), computer science (DiCerbo, 2007), pulmonary physiology (McGaghie, McCrimmon, et al., 2000; McGaghie, McCrimmon, & Thompson, 1998), chemistry (Neiles et al., 2016), neuroscience education (Stevenson et al., 2016), author cocitation (H. D. White, 2003), psychology (Jaworska & Chupetlovska-Anastasova, 2009), research methods (Balloo et al., 2016), and nursing (Azzarello, 2007). MDS and PFN have also been used to examine representations within a patient population, including psychiatry (Egli, Riedel, et al., 2009), chronic obstructive pulmonary disease and asthma (Insel et al., 2005), and vaccine education (Amith, Cohen, et al., 2020; Amith, Cunningham, et al., 2017). These studies support the use of quantitative analyses to model structural changes in cognitive structure over time, during learning, or compared to a referent such as an expert (Siew et al., 2019). They can also be used as formative assessments, skill acquisition, and student feedback to address potential misconceptions (Day et al., 2001; Trumpower, Filiz, & Sarwar, 2014; Trumpower, Sharara, & Goldsmith, 2010). However, MDS and PFN are often used independently. The seminal work of Goldsmith et al. (1991) has not been replicated in gross anatomy or physiotherapy domains, making it a potentially innovative approach to address the gap in the research.

In summary, multivariate quantitative analysis and data visualization are well-aligned with the potential quantitative representation of cognitive structures. These mental representations are believed to be based on the individual's perception, prior knowledge, and learning strategies. Although MDS and PFN have been proposed as potential strategies for quantitative representation, an important caveat exists, they do not address the processes necessary to create the structures nor the specific neurological mechanisms involved. In this regard, cognitive mapping may provide insight into the individual's cognitive structures within a computational framework and offer the potential to be used in the structural assessment of knowledge. One approach may not be better or worse than the other to represent cognitive structure; they may represent different layers or strata of the cognitive structure. The ability to examine an individual's cognitive structure then provides a means to compare to other cognitive, educational, and neural measures of importance (Siew et al., 2019).

Gross Anatomy Education

The study of gross anatomy is an integral component of all health professions' curricula, including physiotherapy. However, the retention of anatomical knowledge in many health professions programs such as medical, nursing, chiropractic, and physiotherapy, is poor (Dayal et al., 2017; Hołda et al., 2019; Narnaware & Neumeier, 2020). This is not a new development in health professionals' education; this concern extends back to the research of Prince et al. (2005) and Bergman et al. (2008). Poor retention subsequently contributes to poor near and far transfer (Persky & Murphy, 2019). Gross anatomy education involves a significant volume of content. A broad range of cognitive skills are necessary for the study of anatomy: visualization, spatial ability (visuospatial), consistent terminology and taxonomy (verbal/auditory), and knowledge organization (Amin & Iqbal, 2019; Castro-Alonso & Atit, 2019; Clarkson & Whipple, 2018; Keehner, 2011; Langlois et al., 2020; Lufler et al., 2012). D'Antoni et al. (2019) reported that clinical anatomy students often utilize surface (rote) learning strategies that emphasize the lowest levels of the revised Bloom's taxonomy (Krathwohl, 2002). Several authors have examined retrieval practice and found it a valuable addition to anatomical knowledge retention (S. J. Anderson et al., 2018; D'Antoni et al., 2019; Dobson, Linderholm, & Perez, 2018; Dobson, Perez, & Linderholm, 2017). However, retrieval remains dependent upon encoding; the best retrieval practices will reveal misconceptions and poor anatomical knowledge if the encoding is poorly structured or organized.

Traditional teaching methods have emphasized lectures and cadaveric dissection. Gross anatomy curricula have been impacted by a decrease in curricular time devoted to anatomy and basic sciences education and thus has been increasingly focused on student-centered pedagogy, self-directed learning, and computer-assisted instruction (Amin & Iqbal, 2019; Guimarães et al., 2017; Hulme et al., 2020; van Lankveld et al., 2019; Wilson, Brown, et al., 2019). Cognitive learning theories that promote deep learning are often poorly integrated into gross anatomy and physiotherapy curricula (Agra et al., 2019; Choi-Lundberg et al., 2017; Smith, Finn, & Border, 2017). The application of cognitive learning theories in these domains often focuses solely on instructional design or retrieval practices (D'Antoni et al., 2019; Delgado et al., 2018; Dobson, Linderholm, & Perez, 2018; Dobson, Perez, & Linderholm, 2017; Mukhalalati & Taylor, 2019). Gross anatomy teaching methods and instructional strategies have been extensively reviewed and found to attain similar learning outcomes (Estai & Bunt, 2016; Losco et al., 2017; Wilson, Brown, et al., 2019). As an expert, the instructor provides content knowledge and pedagogical content knowledge to establish what needs to be taught and how it needs to be prepared to attain the curriculum's specific learning outcomes. This is presumably within the context and needs of a physiotherapy student learning gross anatomy. However, L. J. White et al. (2018) noted that the form of content delivery did not affect student outcomes in gross anatomy. Husmann and O'Loughlin (2019) indicated no correlation between learning style and final grade in an undergraduate anatomy course; O'Mahony et al. (2016) found similar results in medical students studying anatomy. Aslaksen and Lorås (2019) reported that working memory performance did not improve by matching instruction to learning style. The equivocal learning outcomes in gross anatomy education suggest that learner-specific cognitive factors may be significant contributors to the problem of anatomical knowledge retention.

Little literature exists related to the cognitive structures necessary for a physiotherapy student nor a gross anatomy student. Many assumptions typically link content knowledge (anatomical knowledge, both declarative and procedural) and diagnostic thinking based on expert opinion and clinical experience. However, cognitive architecture and the organization and indexing of long-term memory are necessary for understanding both learning and instruction (J. R. Anderson, 2007). There should be alignment between cognitive structures within the content domain and how to develop these structures via specific learning and instructional strategies (D’Antoni et al., 2019; Dobson, Perez, & Linderholm, 2017). For example, although concept mapping as a visual representation of a cognitive structure appears in the research within several health professions, including nursing (Alfayoumi, 2019; Jaafarpour et al., 2016; Mohammadi et al., 2019; Si et al., 2019) and medicine (Daley et al., 2016; Nicoara, Szabo, et al., 2018; Nicoara, Szamoskovi, et al., 2020), little research exists in physiotherapy (Zipp & Maher, 2013; Zipp et al., 2015) with just two research studies utilizing mind mapping in gross and neuroanatomy (Anand et al., 2018; Deshatty & Mokashi, 2013). The current literature focuses on learning strategies and cognitive processes but not the student’s cognitive structures necessary for success in the course or beyond (Siew, 2020). This significantly limits the ability of the instructor, the curriculum, and the student in attaining these cognitive structures representative of deep learning that may then serve as valuable in promoting the transfer of learning (Siew, 2020).

The challenge in gross anatomy assessment has been the seeming dichotomy between assessment of learning, assessment as learning, and assessment for learning (Hawe & Dixon, 2017; Leppink, 2020). Assessment of learning is the traditional summative assessment of learning outcomes (Leppink, 2020). Gross anatomy education traditionally utilizes multiple-choice questions or practical examinations to assess learning (Brenner et al., 2015; Choudhury & Freemont, 2017). However, these assessment strategies have either poor validity or have not been tested for validity whatsoever, leaving them as assessments of “meaningless” or surface learning compared to deep learning that promotes near transfer and provides the foundation for competency. Even with the evolution of the digital learning environment, Meyer et al. (2016) found that student performance on gross anatomy practical examinations was unaffected by assessment modality, with no differences between the traditional face-to-face and online variations. Students are often more focused on academic performance based on exam demands and the short-term retention of instructional materials than on developing strategies that promote expertise as a clinician (Choi-Lundberg et al., 2017). However, assessment for learning and assessment as learning are better aligned with the development of deep learning and self-regulated learning skills (Hawe & Dixon, 2017; Kulasegaram & Rangachari, 2018; Leppink, 2020). The challenge is to have assessments aligned with cognitive structure changes based on their importance in deep learning, retention, and expertise.

Gross Anatomy Knowledge for Physiotherapy Students

The ACT-R model of cognition provides a coherent cognitive architecture and computational framework to examine cognition. J. R. Anderson (2007) noted that cognitive function and information processing are consistent with the anatomical structures that allow this processing to be performed effectively. Inherent to the ACT-R model is the concept of chunking information in declarative memory in both a heterarchical and hierarchical organization. Cognitive structures developed by the individual serve as mental representations of content knowledge (what they know) and the organization of knowledge (how they know it) tied to their perceptions and meaning. Changes in cognitive structures over time may indicate that learning has occurred with the potential progression from novice toward more expert cognitive structures. As a learner, merely having the symbolic knowledge isn't sufficient; there must be a deployment of knowledge which demands activation. It is not critical to fully understand the neurological mechanisms underlying the development of cognitive structures as a prerequisite to establish a means of representation of the phenomenon. However, determining the validity of a means of representation becomes critical in utilizing cognitive structures for learning, instruction, and assessment.

The assessment of cognitive structures in physiotherapy students provides a foundation for understanding the deep learning of gross anatomy. However, this gap in the current research has not been considered in traditional approaches to gross anatomy education, with educational research efforts focused on teaching methods and learning styles. Understanding cognitive structures, their development and quantitative representation, may provide insight into the individual's prior knowledge, learning, and organization of anatomical knowledge to enhance retention and transfer. Assessment of gaps in knowledge representation, exemplified by the individual's cognitive structure (what knowledge exists and how it exists), provides a basis for adaptive learning and curricular development that can focus on the underlying strategies and behaviors that will promote their development (Liu et al., 2019).

Literature Review Related to Key Variables and Concepts

Data Modeling and Visualization

Data modeling and visualization can be achieved within a computational framework that includes multivariate analysis. Dimensionality reduction techniques are multivariate analyses that employ scaling algorithms for data visualization (Dzemyda et al., 2013). Data visualization techniques that employ dimensionality reduction may subsequently display hidden structures and organization within both an individual's data (Dzemyda et al., 2013) or a broader "concept landscape" of group data (Muehling, 2017). Data modeling may have the potential to provide a novel and innovative approach to the assessment of and for learning (Morales-Martinez et al., 2017). This aligns well with clinical applications such as the systematic approach proposed by Bonebright et al. (2005). The authors implemented both MDS and PFN to provide a more comprehensive overview of conceptual and perceptual relationships among auditory stimuli. Goldsmith et al. (1991) considered several parameters derived by MDS and PFN as "knowledge indices" based on their correlation between student and expert representations. These indices included the correlations on raw proximity data ($r = 0.61$), MDS distances ($r =$

0.54), PFN distances (r = 0.66), and PFN closeness (also known as neighborhood similarity; r = 0.74). Although the depth of quantitative analysis that can be attained through these approaches may not be directly practical in an educational realm, the theoretical framework has potential value in assessment strategies related to a learner's cognitive structure and, subsequently, cognitive mechanisms and function. It can also provide more significant insights into the current state of development of the student's cognitive structures relative to an expert's cognitive structures and be a potential tool for student feedback. This has the potential to provide a data-driven means of assessment *of* learning, assessment *as* learning, and assessment *for* learning (Leppink, 2020).

Proximity Data

MDS and PFN have been used to provide a broad overview of cognitive structures and serve as the primary methodological constructs for this study. Both MDS and PFN are scaling algorithms that utilize semantic distance and association to develop proximity data representations that provide high predictive utility (Dry & Storms, 2009). Perceived similarity or relatedness between items, keywords, or concepts can be established via pairwise comparisons, a perceptual approach based on Thurstone's law of comparative judgment (Thurstone, 1927). The direct comparison of items to establish perceptual similarity is considered the gold standard for concept organization (Dry & Storms, 2009). Pairwise comparisons have been used repeatedly in psychological research, and their use in educational research is growing (Crompvoets et al., 2020; Heldsinger & Humphry, 2010). Content items are often selected a priori based on the feedback of content experts.

Participants are provided with a paired list of concepts and keywords; for example, if 20 concepts are considered, each participant would have 190 pairwise comparisons. The total number of pairwise items would be 380, with all pairwise items being duplicated, leaving 190 pairwise comparisons. Most studies will use pairwise comparisons with Likert scales ranging from five to seven levels of similarity (ranging from “no similarity” to “identical”). This produces ordinal data, which may limit statistical analysis. However, Wu and Leung (2017) suggest using an 11-point scale may provide greater similarity to interval data and enable statistical analyses oriented to this type of data. Participants do not require any specific training to establish perceptual relatedness other than the primary domain context in which they are working.

This proximity data can then be analyzed by both MDS and PFN scaling algorithms, each producing a different representation. These representations may have some semantic relationship dependent upon the strength of similarity between items; weak similarities between concepts may be influenced by spreading activation and via link associations (De Deyne et al., 2016). Changes in cognitive structure representation may reflect both learning and evolution from novice to more expert cognitive organization levels. Further consideration is now given to both analysis techniques, the measures that are implicit to their potential use as spatial and network representations of cognitive mapping, and their relevance to educational and clinical applications.

Multidimensional Scaling for Spatial Representation

Proximity data analysis via MDS provides a global spatial representation, also known as a perceptual or spatial mapping, consisting of concepts (points) in Euclidean space, much like the high-dimensional frame of reference described by Bottini and Doeller (2020). Dimensional reduction leads to a spatial representation of the data. There is an extensive research history utilizing MDS to examine the cognitive structure in a broad range of practical applications (Balloo et al., 2016; Egli, Riedel, et al., 2009; Egli, Streule, & Lage, 2008; Gillan et al., 1992; Goldsmith et al., 1991; McGaghie, McCrimmon, & Thompson, 1998). Relevant MDS-derived measures and parameters include dimensionality, stress-1 (goodness of fit, which represents the coherence of the model), TCC, R^2, and Euclidean semantic distances. One representative example is that of Egli, Streule, and Lage (2008), in which MDS was used to assess the differences between student and expert psychotherapists in their diagnosis of ICD-10 mental disorders. A total of 26 students participated in the study. As students gained training, their spatial representations (as reflected in their MDS visualizations) became more similar to those of the experts.

Pathfinder Networks for Network Representation

Proximity data analysis via PFN provides a local associative representation consisting of concepts (nodes) and associations (links), much like the low-dimensional frame of reference related by Bottini and Doeller (2020). The representation derived by PFN is in much the same form as a concept map (Meyer & Schvaneveldt, 1976; Schvaneveldt et al., 1988; Schvaneveldt, 1990; Schvaneveldt et al., 1989). Dimensional reduction leads to a network representation of the data. Although PFN has not been in existence as long as MDS, there is still a rich research history utilizing PFN to examine the cognitive structure in a broad range of practical applications (Azzarello, 2007; Curtis

& Davis, 2003; DiCerbo, 2007; Goldsmith et al., 1991; Lyu & Li, 2019; McGaghie, McCrimmon, et al., 2000; Neiles et al., 2016; Stevenson et al., 2016; Trumpower, Filiz, & Sarwar, 2014). Relevant PFN-derived measures and parameters include links, degree (the number of links attached to each node), eccentricity (the maximum number of links between a node and all other nodes in a network), coherence (the degree to which an individual's cognitive structure has internal links and associations), and similarity (the degree to which an individual's cognitive structure aligns with the cognitive structure of another individual or group of individuals). Measures such as coherence (within-subject consistency and reliability of data within the individual network) and similarity (between-subject comparison to a referent structure) can be used to detect change over time. Lyu and Li (2019) noted that engineering students' diagnostic performance improved as their Pathfinder similarity with experts improved. Azzarello (2007) reported a statistically significant relationship between post-course coherence and similarity with mean examination grade in a study of community health nursing students. Neiles et al. (2016) examined PFN in terms of validity and as a measure of assessing cognitive structure change in undergraduate chemistry students. The authors noted that PFN was valid and could be used as a formative assessment for chemistry students. Stevenson et al. (2016) used PFN within an undergraduate neuroscience course. They performed pre- and post-course assessments on 63 students, finding that coherence and similarity improved throughout a course. The authors noted that the post-course assessment had shown improvement and could have promise as an outcome measurement.

Critical Analysis of MDS and PFN

Research using MDS and PFN is relatively abundant in the literature across a broad range of domains, though its prevalence has been diminished over the past decade. Several potential issues exist. A critical review of the literature on MDS and PFN revealed an absence of specific a priori power calculations in any prior studies regardless of the statistical analyses performed. Seminal studies noted previously have a range of sample sizes from 35 to 71 participants (Acton et al., 1994; Egli, Streule, & Lage, 2008; Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016; Trumpower, Sharara, & Goldsmith, 2010). Although the majority of these seminal studies did not report effect sizes as such, further review of the study results revealed large effect sizes based on calculated r^2 and η^2 values (Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016).

There is little research on the test-retest reliability of these data modeling strategies related explicitly to cognitive structures' representation. However, several studies have examined this psychometric property within graph-theoretical networks applied to similar anatomical constructs such as brain networks (Paldino et al., 2017; Welton et al., 2020). Paldino et al. (2017) studied the test-retest reliability of graph theoretical analysis of pediatric patients with epilepsy and found Pearson correlation coefficients ranging from 0.76 to 0.97 with an ICC of 0.74 to 0.96. This indicates good to excellent reliability (Koo & Li, 2016; Liljequist et al., 2019). Welton et al. (2020), in a similar graph theoretical analysis using patients with multiple sclerosis, reported that the ICC was greater than 0.6, a moderate to good reliability (Koo & Li, 2016; Liljequist et al., 2019). Although these studies are based on anatomical constructs, they provide evidence of the potential use of these data modeling strategies to monitor specific cognitive structure changes based on the individual's learning and reorganization of knowledge.

There are several potential methodological concerns with the use of MDS and PFN; however, there does not appear to be any literature that has provided any disclaimer in terms of usage or relevance of these strategies for the intended purpose. Tessmer et al. (1997) provided a clear foundation for future research, noting that structural measures of representation (specifically, MDS and PFN) have predictive validity and can be used to measure changes in learning and differences between experts and novices. However, as exemplified by the commentary of Paas and van Merriënboer (2020), the focus of the cognitive learning theory literature appears to have centered on themes such as multimedia learning, cognitive load, instructional design, and the learning environment with a diminished focus on the efficient development of cognitive structures within long-term memory. Gao et al. (2019) examined the literature related to deep learning in education. The authors noted that although the number of studies related to deep learning theory and strategies has been steadily increasing, those related to evaluation and measurement remain few and consistent over time. It is unknown if this disparity is related to a shift in research agenda, a disparate view of analytical approaches, or a lack of development of the domain over time.

Criterion Standards

One of the primary considerations for a criterion-related validation study is the selection of an appropriate criterion. Content knowledge is often assessed as the criterion standard via multiple-choice questions yet there are implicit issues related to the validity of multiple-choice questions for knowledge assessment. However, this does not examine the structure and organization of the criterion of interest (thus limiting predictive validity), nor does it adequately address issues related to the validity of the multiple-choice questions themselves (thus limiting concurrent validity). Academic grades provide a potential criterion problem. By not representing the full range of differences between students, grading standards can vary significantly and can be highly arbitrary, and the actual meaning of grades in terms of achievement may vary significantly (Borneman, 2012; Hartnett & Willingham, 1980). The use of grades as a criterion may also impact studies aimed at assessing the relationship of learning strategies (Kamath et al., 2018). Grade point average has a long tradition of use as a predictor of success in medical student gross anatomy (Moffatt et al., 1971) and physiotherapy student NPTE success (Bayliss et al., 2017; S. H. Hayes et al., 1997). The primary focus of this study is on the cognitive structure and the internal perceived structural organization and representation of those structures; thus, the criterion selected is that of the expert or instructor. Unit grades will be considered as a secondary criterion.

Summary and Conclusions

Gross anatomy is an integral component of physiotherapy curricula. However, anatomical knowledge retention is poor. Little is known about how physiotherapy students learn gross anatomy. However, cognitive science provides a foundation upon which to understand better the challenges faced by physiotherapy students. The ACT-R model of cognition provides a well-supported theoretical foundation for examining cognition in both domain-general and domain-specific (gross anatomy) cognitive learning contexts. Effective cognitive structures develop through deep and meaningful learning, enhancing retention and near transfer. The emerging cognitive science research indicates the presence and importance of cognitive mapping to represent these cognitive structures.

Gross anatomy education provides unique challenges to the physiotherapy student. Research has focused on learning and instruction strategies, noting minimal differences in learning outcomes. However, self-directed learning strategies may provide a sound framework for examining physiotherapy students engaged in gross anatomy education to develop the skills necessary for effective clinical practice. Anatomical knowledge retention is dependent upon the student's cognitive structures. This underscores the importance of a better understanding of cognitive structures and their quantitative representation as a tool for learning, instruction, and assessment.

Data modeling strategies via scaling algorithms such as MDS and PFN have been used effectively in a broad range of domains to represent cognitive structure. These strategies may provide an innovative approach to the visualization and assessment of gross anatomy cognitive structures grounded in cognitive science. Understanding the cognitive functions essential for success, combined with the backward design of the curriculum, should establish the cognitive structures necessary as the student progresses through the clinical program. These strategies are discussed within the context of the methodology of the current study in Chapter 3.

Chapter 3: Research Method

The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. Chapter 3 provides an overview of the methodology of the study. Major sections include research design and rationale, methodology, data collection strategy, threats to validity, and ethical procedures. The chapter concludes with a summary of the research methods used in this study.

Research Design and Rationale

A quantitative approach was appropriate for this study and the research questions involved. This choice of methodology was consistent with the literature reviewed in Chapter 2 that addressed the proposed use of multivariate analysis and the quantitative representation of cognitive structures (see Acton et al., 1994; Egli, Streule, & Lage, 2008; Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016). The research design was a quasi-experimental, criterion-related validation study using proximity data (see A. D. Harris et al., 2006). A quasi-experimental design was appropriate because selection effects were minimized to better represent real-world scenarios and provide a high degree of external validity (see Bärnighausen et al., 2017).

There were several considerations in the selection of this research design. Researchers on cognitive learning theory have noted the importance of cognitive structures. J. R. Anderson (1996) described the cognitive unit as the precursor to cognitive structure, integrating it as a cognitive mechanism within ACT-R. A review of the literature on gross anatomy education revealed equivocal learning outcomes with various instructional strategies, subsequently providing the rationale to explore self-directed learning strategies that promote cognitive structures development. A quasi-experimental, nonequivalent control group pretest-posttest design was initially conceived as a means of examining the effect of one learning strategy (concept mapping) on the meaningful learning of physiotherapy students enrolled in a gross anatomy course (see Handley et al., 2018). However, several philosophical, epistemological, and methodological concerns were exposed. First, there was little consensus regarding the definition of cognitive structures, how to represent them, and how to measure them. Second, the proposed study design would have been interventional, which provided several methodological constraints such as a lack of good rubrics, the necessity for additional student and instructor training, and the fidelity of implementation. Third, there were many threats to validity, both internal and external. At this point, with several clear limitations methodologically, new avenues were considered.

The study's focus shifted to the emerging cognitive science literature regarding cognitive structures and cognitive mapping and integrating with network science's computational strategies (see Behrens et al., 2018; Bellmund et al., 2018; Siew, 2020; Siew et al., 2019). This literature was well-aligned with the computational framework of ACT-R. Extensive research on MDS and PFN indicated using both data modeling strategies as an indirect means of representing the structure of knowledge in various domains. However, their practical application in the health sciences was limited. Although these computational strategies appear promising and innovative in the gross anatomy and physiotherapy domains, many prior studies were methodologically weak. This necessitated further refinement of the topic, specifically criterion-related validity, and incorporating a research design that would align with this topic. I determined that an innovative and exploratory approach to representing cognitive structure would utilize a criterion-related validation design. The purpose of the study was to investigate the potential use of two data modeling strategies (MDS and PFN) to visually and quantitatively represent cognitive structure in physiotherapy students learning gross anatomy.

The criterion-related validation study would include pairwise comparisons to establish proximity (similarity) data, which could then be used via MDS and PFN to derive spatial and network visual and quantitative representations of cognitive structure, respectively. The study was initially conceived to have student participants complete pretest and posttest pairwise comparisons to examine change over time; however, due to extenuating circumstances, this was revised to focus on one set of pairwise comparisons representative of student cognitive structure. Expert participants also completed one pairwise comparison test. The first part of this exploratory study addressed the potential relationships and agreement between student cognitive structure and expert cognitive structure (criterion standard one). There were no independent or dependent variables because an independent variable was not manipulated to examine a change in the dependent variable. For the second part of this exploratory study, the dependent variable (criterion standard two) was the unit grade. The independent variables were MDS- and PFN-derived quantitative measures and the level of agreement between student and expert cognitive structures. MDS-derived measures included dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances. PFN-derived measures included links, degree, eccentricity, coherence, similarity (with another network), and graph-theoretic semantic distances. Prior knowledge was reflected in admission GPA and admission anatomy GPA and was controlled as a covariate. These measures also reflect academic performance and are factors related to professional program GPA and first-time pass rate on the NPTE (Bayliss et al., 2017; S. H. Hayes et al., 1997; Wolden et al., 2020). The moderating (categorical) variables were the instructor and the program mode of delivery (residential and flex). Although content modules are standardized across program modes of delivery via the Blackboard learning management system, and teaching strategies are often consistent based on the course's lecture and laboratory components, variations can occur. Instructor bias may have had a moderating effect on student cognitive structure's potential changes or the degree of similarity with expert cognitive structure. The program mode of delivery may have had a moderating effect due to potential variations in the degree of synchronous and asynchronous teaching interaction. Both instructor and mode of delivery were potential confounding variables.

The research design was a replication and extension of previous studies using MDS and PFN based on the theoretical justification reported in Chapter 2. Components of the methodology were replicated from several studies, including Goldsmith et al. (1991), Neiles et al. (2016), Stevenson et al. (2016), Egli, Streule, and Lage (2008), and Acton et al. (1994). The concurrent use of MDS and PFN and the basic implementation methodology was exemplified by the Goldsmith et al. study. The knowledge indices described in the Goldsmith et al. study were integrated similarly in the current study, emphasizing the importance of raw proximity data, MDS Euclidean distances, Pathfinder graph-theoretic distances, and Pathfinder coherence, common links, and similarity. The premise of variations in expert cognitive structures and the averaging of raw proximity data to derive the expert cognitive structures was derived from the Acton et al. study. The studies by Neiles et al. and Stevenson et al. were used as examples of the practical implementations of PFN (with undergraduate chemistry and neuroscience students, respectively) and the study of Egli, Streule, and Lage was used as an example of the practical implementation of MDS for examining student–expert differences. Although several methodological challenges were noted in Chapter 2, these were acknowledged to narrow the scope of the current study with these accepted limitations.

I refined the methodology of seminal works and extended the analysis for preliminary use in the gross anatomy content domain. Several refinements were implemented to enhance previous methodological approaches and to build on prior research. An 11-point Likert scale was initially conceived for use in pairwise comparisons to represent proximity data as interval data better and enhance subsequent analysis (see Wu & Leung, 2017). Concept items and functional terms were derived from the course text *Clinically Oriented Anatomy* (Moore et al., 2018), the *Terminologia Anatomica* (FIPAT, 2019) and the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018). The final item list was selected by physiotherapists currently in musculoskeletal clinical practice. This review process increased the external validity and clinical relevance of concepts used for pairwise comparisons. Expert cognitive structures were derived from both course instructors (typically physiotherapy centric) and PhD anatomists using the same procedures and variables used for student cognitive structure. Finally, I controlled for prior knowledge, an integral factor in both the cognitive structure and the target population.

Establishing reliability and validity is critical for any potential learning outcome measure before its use in the field. The reliability of a measure is its stability and internal consistency either within one rater or between raters (Souza et al., 2017). Test-retest reliability is an indicator of the internal validity of the measure. Although the importance of test-retest reliability is evident, establishing intraclass correlations with appropriate power would require its own participant pool (Koo & Li, 2016; Liljequist et al., 2019). Test-retest reliability of experts' cognitive structures would also necessitate the completion of testing within a short period to limit the effects of history and maturation bias (Handley et al., 2018). The validity of a measure is the degree to which it measures what it claims to be measuring (Souza et al., 2017). There is a difference, however, between validity and validation. Sussmann and Robertson (1986) differentiated validity and validation by noting that validation refers to research design, whereas validity is a function of the results attained from the study. Validation studies are critical first steps in the life of an assessment tool. Validity comprises four critical stepwise components: statistical conclusion validity, internal validity, construct validity, and external validity (Sussmann & Robertson, 1986). The current study was a criterion-related validation study in that it was a specific research design used to establish various types of validity through the design of the study.

Several researchers have reported the use of MDS and PFN to represent cognitive structures in a variety of domains (Azzarello, 2007; Balloo et al., 2016; Casas-García & Luengo-González, 2012; Connor et al., 2004; Curtis & Davis, 2003; DiCerbo, 2007; Jaworska & Chupetlovska-Anastasova, 2009; McGaghie, McCrimmon, et al., 2000; McGaghie, McCrimmon, & Thompson, 1998; Neiles et al., 2016; Stevenson et al., 2016; Veríssimo et al., 2017; H. D. White, 2003). These studies appeared to confirm construct validity based on the perceptual representations derived from proximity data and scaling algorithms, with the critical assumption that representations of the cognitive structure are indirect and not direct (Ifenthaler et al., 2011). The study of Neiles et al. (2016) provided a clear example of a relevant validation study. Neiles et al. examined the use of PFN with undergraduate chemistry students by evaluating four types of validity: content, construct, criterion-related, and concurrent. Content validity ensures that the content domain is adequately addressed. Construct validity assesses whether the construct in question is being measured. Criterion-related validity consists of predictive and concurrent validity and establishes whether the assessment or measure predicts future performance or behavior on the criterion of interest (Fink, 2010). The current study paralleled the types of validity assessed by Neiles et al. Content validity was addressed via selecting key terms from the text that adequately represented the domain concepts in question. Construct validity addressed student cognitive structure in comparison to an expert cognitive structure. Criterion-related validity included two criterion standards: expert cognitive structure as a primary criterion (concurrent validity) and unit grade as a secondary criterion (predictive validity).

Criterion validity, including predictive and concurrent validity, creates a unique challenge when assessing cognitive structure. In the current study, the primary criterion of interest was the representation and organization of the cognitive structure, not the content or the student's ability to recall specific concepts. For this reason, the content was provided to participants a priori via an item list and pairwise comparisons. Because clear distinctions were noted regarding the organization of knowledge between experts and novices, the criterion or reference standard was the experts' cognitive structure. The use of this criterion becomes increasingly essential as curricula evolve toward competency-based education (Bains & Kaliski, 2019; Lucey et al., 2018). Although the limitations for using academic grades were acknowledged, the unit grade (consisting of both multiple-choice and practical examinations) was used as a secondary criterion for consideration. This measure had been used in most previous studies, making its consideration relevant for direct comparison to prior research, practical relevance to the domain-specific application, and promoting the potential generalizability of the current study's findings.

Methodology

The methodology section addresses the study population, sample, sampling procedures, recruitment and participation procedures, and data collection. Operational definitions and instrumentation are described. A concise plan for data preparation and data analysis to address all research questions is presented. Threats to validity (including internal, external, construct, and statistical conclusion validity) are considered. Finally, ethical considerations and procedures are described.

Population

The DPT is a clinical doctorate that serves as the entry level for clinical practice as a physiotherapist. The DPT program at the institution under consideration consists of two separate modes of delivery: an eight-trimester residential program and a 12-trimester online-based flexible program. The target population consisted of DPT students enrolled in Gross Anatomy I during the first 15-week trimester of the program. There are approximately 320 students each trimester institution-wide (260 residential, 60 flexible). The demographics of this population reflect a graduate student who is 26 years of age on average, with a range of 21 to 46 years. Admission prerequisites include a bachelor’s degree and several course prerequisites such as six semester hours of anatomy and physiology. Admission data revealed that entering students have an average cumulative GPA of 3.2 and an average GRE score of 301.

Two groups of experts were used to derive the expert cognitive structures that served as the primary criterion of interest in the current study. The first group consisted of the six lead course instructors within the institutional DPT programs. These course instructors had a clinical doctorate in physiotherapy. Generally, course instructors do not have a PhD in Anatomy, thereby providing an expert cognitive structure that was physiotherapy centric but not domain specific. The second consisted of three anatomy content experts outside of the institutional DPT program. These experts had a PhD in Anatomy but not a clinical doctorate, thereby providing an expert cognitive structure that was domain specific but not physiotherapy centric. Averages were calculated for the instructor group, anatomist group, and combined group (a weighted average based on the number of expert participants).

A group of 10 physiotherapists currently in clinical practice was recruited to rank order concept items and functional terms that would form the 20-item list used for pairwise comparisons by students and experts. Inclusion criteria were a minimum of 10 years of clinical practice focused on musculoskeletal conditions in an outpatient environment. Exclusion criteria were individuals currently involved in teaching physiotherapy students in an entry-level DPT program.

Sampling and Sampling Procedure

As the research study was a criterion-related validation study, voluntary response (nonprobability) sampling was used. Demographic data such as age, gender, admission GPA, admission anatomy GPA, and GRE scores were used for post-stratification weighting to ensure a sample that is as closely representative of the target population as possible (Battaglia, 2008; Farrokhi & Mahmoudi-Hamidabad, 2012). Inclusion criteria consisted of first trimester DPT program students in either the residential or flex modes of program delivery; all students in all delivery modes were allowed to volunteer to participate. There were two exclusion criteria. The first exclusion criterion was students repeating the course as the previous course exposure may create a confounding variable. The second exclusion criterion was those students whose lead instructor is the primary investigator of the current study. This removed any potential bias and influence over study participants due to direct authority over the participants in question.

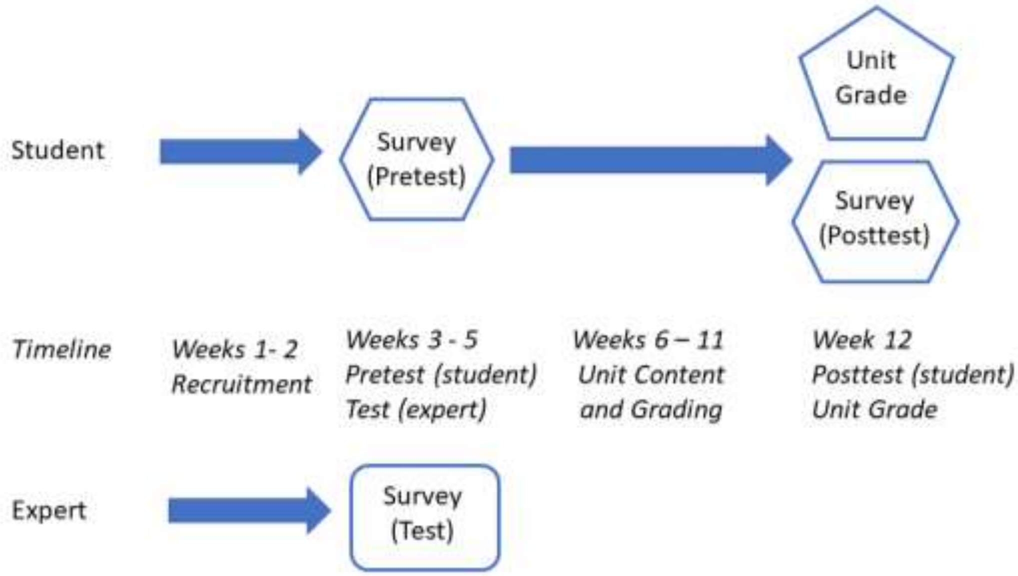
As noted in Chapter 2, a priori power calculations are lacking in most of the seminal studies; however, sample sizes range from 35 to 71 participants (Acton et al., 1994; Egli, Streule, & Lage, 2008; Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016; Trumpower, Sharara, & Goldsmith, 2010). In a review of the studies by Goldsmith et al. (1991), Neiles et al. (2016), and Stevenson et al. (2016), large effect sizes were reported based on calculated r^2 (>0.5) and η^2 (>0.14) values (see Cohen, 1988). Although these studies lacked clearly defined power calculations, a preliminary a priori power analysis via G*Power indicated several appropriate sample sizes (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). To balance Type I and Type II errors, Cohen (1988) suggests using an alpha value of 0.05 and a beta value of 0.20 (with a power of 0.80). All a priori calculations utilized these alpha and beta values. For paired sample *t* tests, the sample size was 34 (moderate effect size of 0.50) or 15 (large effect size of 0.80). For correlational analysis, the sample size was 84 (moderate effect size of 0.30) or 29 (large effect size of 0.50). For multiple regression, the sample size was 77 (moderate effect size of 0.15) or 36 (large effect size of 0.35). Effect sizes tend to be greater in quasi-experimental research designs (compared to randomized clinical trials) and within-group analyses (Bakker et al., 2019).

Procedures for Recruitment, Participation, and Data Collection

The proposed timeline for the study is represented in Figure 2. Student and expert recruitment procedures were consistent, with minor variations in the timeline and demographic information collected. During the first week of the trimester, all groups (student, expert, and physiotherapists in clinical practice) were recruited upon approval from the Walden University Institutional Review Board and the institution offering the DPT program.

Figure 2

Proposed Study Timeline



Prospective student and expert participants were assured that their involvement would be held in strict confidence, their privacy throughout the study was ensured, and that all data associated with the study would be protected and remain anonymous once the data set (demographic data, pretest survey, posttest survey) was complete. During the recruitment process, students and experts were not coerced into participation, and they could choose to opt out at any time without any adverse effect. Anonymous online surveys were a component of the non-coercive recruitment strategy. Participation and non-participation would not impact academic standing in the course, or any subsequent courses offered at the institution, nor would they negatively impact any current or future relationship with the primary investigator. After completing the survey(s), participants received a $10 (appreciation) gift card as a thank you—this encouraged participation in the study.

Ten physiotherapists currently in musculoskeletal clinical practice were recruited upon approval from the Walden University Institutional Review Board and the institution offering the DPT program. These clinicians were provided a list of concept items and functional terms specifically related to the shoulder and glenohumeral joint which was the current student content module. These items were derived from the course text *Clinically Oriented Anatomy* (Moore et al., 2018), the *Terminologia Anatomica* (FIPAT, 2019), and the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018). The *Terminologia Anatomica* (FIPAT, 2019) represents the standardized nomenclature for gross and clinical anatomy (Chmielewski, 2020; Greathouse et al., 2004), whereas the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018) is an ontology of anatomical structures. The list of concept items and functional terms is provided in Appendix A. Physiotherapists were asked to rank order these items according to the level of relevance to clinical practice and musculoskeletal care. Rank order was compiled and averaged, with the top 20 items being used as the final item list for pairwise comparisons.

Students received an explanation of the study in three formats: remotely via teleconference, embedded video within Blackboard, and description posted via announcement in Blackboard (Appendix B). Group teleconferences would be scheduled during the last few minutes of one of the regularly scheduled class sessions, arranged in advance with the lead course instructor(s). Contact information, including the primary investigator’s email address and phone number, was included in all formats. The Blackboard announcement contained a link to volunteer and opt-in for participation in the study; this would generate an email request to the primary investigator for a unique identifier code that would initially be associated with the student ID number. Each participant would receive this unique identifier and link to the online survey via email. Student ID numbers and associated unique identifiers were maintained in an Excel spreadsheet until supplemental data were received from the registrar (admission GPA and admission anatomy GPA) and Blackboard (unit grades). At that time, all student ID numbers were removed from the data set, thereby de-identifying the data set and reducing it to a single unique identifier.

Experts received an explanation of the study in two formats: remotely via teleconference (individual or group), as well as an email containing both a written description and an embedded video. Contact information, including the primary investigator’s email address and phone number, was included in all formats. Each participant would receive a unique identifier and link to the online survey via email. Personal data to be collected from expert participants was the number of years of clinical practice, number of years of teaching anatomy, terminal clinical degree, and terminal academic degree. Expert names and associated unique identifiers were maintained in an Excel spreadsheet. Once all data sets were complete, all expert names were removed from the data set, thereby de-identifying the data set and reducing it to a single unique identifier.

Students and experts that agreed to participate in the study received a link to complete an online survey. The nature of the survey design ensured the fidelity of implementation. This online survey consisted of several components. A description of the study (Appendix B) with all associated risks and benefits, identical to that used during the recruitment procedures, was provided. Informed consent was attained via implied consent; the participant clicked the link to continue with the survey, with a notification that no consent signature was required. This further protected the participant's privacy. A short description of the task's context to be performed, with instructions for completion, was included. All pairwise comparisons followed. It was initially conceived that the students would complete the online survey at two different time intervals: before starting the unit (weeks three through five) and completing the unit (week twelve); however, this was revised so that one survey was completed within the span of the course module. Students received email reminders to limit nonresponse bias and attrition. Experts completed the online survey within the first five weeks of the semester.

The data to be collected for the study was pairwise similarity comparisons. Further details of the pairwise comparisons procedure can be found in the section on instrumentation. A total of 190 pairwise comparisons were collected from all study participants (student and expert). This was projected to take the participant no more than 15 to 20 minutes to complete. Unit grades were collected in week twelve upon completing the content module, written exam, and practical exam. The unit grade was weighted in a means consistent with the course syllabus, such that the written exam accounted for 55.56% of the weighted unit grade.

Data were stored electronically via password-protected Excel spreadsheets and SPSS data files locally (encrypted flash drive) and backed up via cloud storage (Dropbox). Local storage was secured via a flood- and fire-proof safe at the primary investigator's home. Data will be stored for five years per Walden University criteria, at which time the files in question will be deleted. There were not any specific follow-up procedures or debriefing for either student or expert participants to complete. All participants will be provided the opportunity to attend a short institution-wide presentation of the study results at a future date.

Instrumentation

The primary instruments utilized were semantic similarity ratings (via pairwise comparisons) and two independent dimensional reduction scaling algorithms, MDS and PFN, that utilize these ratings (Kruskal & Wish, 1978; Schvaneveldt, 1990). These data modeling strategies address different intents regarding cognitive structure representations. MDS is used to examine potential global (high dimensional) spatial relationships and PFN is used to examine potential local (low-dimensional) network relationships. In this study, cognitive structure (student and expert) was represented by the following measures: MDS dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances, and PFN links, degree, eccentricity, coherence, similarity (with another network), and graph-theoretic semantic distances. Prior knowledge was represented by two measures: admission GPA and admission anatomy GPA. The unit grade was measured by a weighted average of written and practical exam grades.

The cognitive structure for physiotherapy students in the gross anatomy course was initially conceived to be examined in a pre- and posttest fashion to establish if a meaningful change in cognitive mapping occurs over time. However, due to extenuating circumstances, this was revised to entail only one assessment of cognitive structure. Comparisons were made with expert cognitive structures for two subgroups: physiotherapy centric (instructor) and domain specific (Ph.D. anatomist). Comparisons were made between student cognitive structures and unit grades to establish if the two were related.

Similarity ratings were compiled via pairwise comparisons to create a proximity matrix (Roske-Hofstrand & Paap, 1990). An item list was used to generate pairwise comparisons. This item list contained essential concept items and functional terms that were specifically related to the shoulder and glenohumeral joint, the content module in question. Items and terms were derived from the course text *Clinically Oriented Anatomy* (Moore et al., 2018), the *Terminologia Anatomica* (FIPAT, 2019), and the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018), and were based on structure and function as integrating these components is essential to clinical practice. Physiotherapists in musculoskeletal clinical practice rank-ordered the clinical relevance of these terms to establish a final 20-item list. The master list of concept items and functional terms is provided in Appendix A. Pairwise comparisons create a proximity matrix which can then be used for both MDS and PFN calculations.

A review of the relevant literature revealed no specific parameters defined for an optimal number of items used in MDS; however, this has been explored in the PFN research. Schvaneveldt et al. (1985) reported that 15 terms are the lowest number of terms used to generate accurate Pathfinder networks. A greater number of items linearly increases the predictive validity while decreasing the overall variance; however, there is an associated increase in the time necessary to complete the pairwise comparisons (Goldsmith et al., 1991). For example, the use of 20 items would necessitate 190 pairwise comparisons (approximately 16 minutes), whereas increasing to 25 items would necessitate 300 pairwise comparisons (approximately 25 minutes); a 25% increase in terms is reflected in a 50% increase in time per survey. The impact of balancing reliability and efficiency on the study design must be considered. Although a greater number of items may improve the predictive validity, it may also decrease the participant pool due to a greater amount of time necessary to complete the study with pretest and posttest measures (Crompvoets et al., 2020). A total of 20 items were selected, creating 380 pairwise comparisons; with duplicate items removed, 190 pairwise comparisons were collected from all study participants (student and expert). These comparisons were projected to take no more than 15 to 20 minutes per online survey.

Best practices were used in the development of the online survey (Ruel et al., 2015). Initial concerns focused on satisficing behavior such as straightlining of responses based on inattentiveness and optimizing the user experience given the large number of paired comparisons required (Kim et al., 2019; Leiner, 2019). Liu and Cernat (2018) found that straightlining of responses was similar between grid and individual item surveys and that data quality diminished with 9- and 11-column responses. For these reasons, a grid matrix was used with five items and seven scale responses to diminish satisficing behavior (including survey inattentiveness and straightlining) while improving the ease of use (Grady et al., 2019). Pairwise comparisons generated ordinal data; however, Harpe (2015) noted that individual ratings with responses having greater than five categories could be analyzed as continuous data. Both data modeling strategies have options to consider the data as ordinal or interval depending on the analysis of individual data (ordinal) versus aggregated data (interval). Ruel et al. (2015) noted that best practices include using a progress bar and numbering the questions, and these strategies were implemented in the final survey design. All online surveys were designed with mandatory responses for each set of questions, thus preventing missing data. Data were downloaded from the survey website, and data cleaning was performed as the survey data were imported into Excel and SPSS for analysis.

Operationalization of Constructs

The operationalization of variables provides a clear definition of concepts, variables, and indicators. Table 1 represents a summary of relevant construct definitions and their operationalization.

Table 1

Operationalization of Constructs

Construct definition	Construct operationalization
Cognitive structure: "A hypothetical construct referring to the organization of the relationships of concepts in long-term memory." (Shavelson, 1972, p. 226-227)	Pairwise comparisons (raw proximity semantic similarity data) representing perceptual concept organization
Cognitive structure mapping: The representation of cognitive structure reflected in a cognitive map defined via two data modeling strategies (MDS and PFN) and their derived quantitative parameters and data visualization.	MDS spatial representation with MDS-derived quantitative measures (configuration properties including dimensionality, stress-1, R^2, and Euclidean semantic distances) PFN network representation with PFN-derived quantitative measures (network properties including degree, eccentricity, coherence, similarity, and graph-theoretic semantic distances)
Deep/meaningful learning: "Meaningful learning occurs when students build the knowledge and cognitive processes needed for successful problem-solving." (Mayer, 2002a)	Meaningful change in student cognitive structure mapping over time (pretest to posttest) reflected in changes in MDS- and PFN-derived quantitative measures
Prior knowledge: "All knowledge learners have when entering a learning environment that is potentially relevant for acquiring new knowledge" (Biemans & Simons, 1996)	Admission GPA Admission anatomy GPA Pretest student cognitive structure
Criterion: "Human expertise can be defined as displayed behavior within a specialized domain and related domain in the form of consistently demonstrated actions of an individual that are both optimally efficient in their execution and effective in their results" (Herling, 2000).	Primary: expert cognitive structure (Schvaneveldt et al., 1985) Secondary: Unit grade

Paired comparisons data were collected via similarity ratings from 1 to 7 based on the degree of relatedness and similarity between the items. This was based on the Goldsmith et al. (1991) study. As there was not a defined “identical” value and the ratings were perceptual, Kruskal and Wish (1978) reported a need to convert similarity ratings to dissimilarity ratings by subtracting each value from a defined constant; the constant was defined as a value of eight as seven indicated “most similar” but not “identical.”

Cognitive Structure Mapping

Cognitive structure was represented visually and quantitatively via MDS and PFN measures derived from proximity data via pairwise comparisons, the gold standard for the perceptual representation of concept organization (Dry & Storms, 2009). Pairwise comparisons served as the raw data used for both MDS and PFN calculations of key variables. These include MDS-derived measures (dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances) and PFN-derived measures (links, degree, eccentricity, coherence, similarity, and graph-theoretic semantic distances). These values created both a visual and quantitative representation of the proximity data, which served as the cognitive structure mapping.

Multidimensional Scaling. MDS was initially developed by Kruskal and Wish (1978). Data modeling was performed using SPSS software to create a spatial representation of the proximity matrix data. Metric MDS was to be used as the 11-point Likert scale data would serve as interval data (Kruskal & Wish, 1978; Wu & Leung, 2017). This representation reflected the semantic space and consisted of the Euclidean semantic distance between concepts derived from the proximity data. MDS represents the proximity data via multiple dimensions that best represent the semantic distances of the model's proximity data.

Giguère (2006) provided a decision table for the selection of scaling models within MDS. This was based on the number of matrices used, the assumption of perceptual and cognitive differences, measurement conditionality, and data level. Classical MDS (CMDS, also known as the Identity scaling model in SPSS) via the PROXSCAL algorithm is appropriate with one matrix of continuous or ordinal data (for example, an individual's survey responses). This produces one group configuration or stimulus space with Euclidean distances. Replicated MDS (RMDS) uses the same algorithm as CMDS but with multiple matrices to generate one stimulus space. However, RMDS provides a more robust solution as the use of multiple matrices provides increased data to generate the solution and accounting for the difference in how people use the response scale (Davison & Sireci, 2000; Hout et al., 2013). Both CMDS and RMDS produce a stimulus space that can be transformed via rotation, reflection, and scaling without losing the relative locations of the items in the stimulus space. Weighted MDS (WMDS, also known as the Weighted Euclidean scaling model or INDSCAL), also via the PROXSCAL algorithm, is used with multiple matrices and the assumption of perceptual and cognitive differences between the matrices. This produces a common configuration or stimulus space with Euclidean distances and individual spaces and dimension weights that are believed to represent differences in cognitive or perceptual factors (Hout et al., 2013). However, the group space is non-transformable in that the axes used for the dimensions are implicit to that specific data set (Davison & Sireci, 2000). The summary provided by Giguère (2006) is consistent with the more recent work of Borg et al. (2018).

Five variables are derived via MDS: dimensionality, stress-1 (goodness of fit), TCC, R^2, and Euclidean semantic distances. Dimensionality represents the number of dimensions that most accurately represent the proximity data; common values would be two dimensions or greater. The greater the number of dimensions, the more difficult it is to represent the data spatially visually. The dimensionality of the expert cognitive structure was used as the standard of comparison for all student cognitive structures.

The assessment of goodness (or badness) of fit is achieved via a multifactorial approach including scree plot, stress-1, R^2, and a Shepard diagram which plots the observed dissimilarities versus fitted distances (Mair et al., 2016). Stress -1 examines the overall goodness-of-fit of the proximity data to the projected MDS configuration based on the optimal dimensionality, with lower values (less than 0.2) indicating a better goodness-of-fit (Borg & Groenen, 2005; Davison & Sireci, 2000; Kruskal, 1964; Mair et al., 2016). Stress-1 is calculated as the square root of normalized raw stress (Borg & Groenen, 2005). Random stress norms have been calculated for various numbers of objects by Sturrock and Rocha (2000). For a 1% chance of 20 objects being randomly arranged ($p<0.01$), then stress-1 values are 0.446 (one dimension), 0.279 (two dimensions), and 0.189 (three dimensions); thus, if stress values are less than this, there is some certainty that objects are not organized randomly (Sturrock & Rocha, 2000). Finally, TCC provides a measure of congruency within the data set. At the same time, R^2

represents the proportion of variance in the proximity data accounted for by the MDS configuration. Good similarity is reflected in TCC values greater than 0.95 with fair similarity between 0.85 and 0.94 (Lorenzo-Seva & ten Berge, 2006). An increasing R^2 value indicates greater coherence within the MDS model. R^2 is visually represented by a plot of transformed proximities and distance. WMDS produces dimensional weights that indicate the individual's preference of the dimensions defined by the MDS configuration. These values were calculated for each set of student and expert proximity data.

Euclidean semantic distances are the perceived distances between concept items, with shorter distances indicating greater association or similarity. Euclidean distances are unaffected by rotation, translation, or reflection (Borg & Groenen, 2005). Agreement and correlation of semantic distances can indicate a degree of association between student and expert cognitive structures. The reliability of MDS is highly dependent upon the reliability of proximity ratings; however, research related to the test-retest reliability of this approach was lacking.

Pathfinder Networks. PFN was initially developed by Schvaneveldt (1990). Data modeling was performed using the Pathfinder software to create a network representation of the proximity data. This representation reflected the semantic network and consisted of the associated links between concepts derived from the proximity data. Network properties define its overall characteristics. These properties include centrality, links, degree, eccentricity, coherence, and graph-theoretic semantic distances. Eccentricity represents the maximum number of links between a node and all other nodes in a network; the center of a network is the node with minimum eccentricity. Degree is the number of links attached to each node. Coherence represents the degree to which the original proximity data correlates with the inferred relationships of the network -the degree of associated links between nodes within the semantic network. A higher level of coherence would indicate a greater number of links and associations between nodes (concept items). Network properties are derived from the PFN solution for each set of proximity data with visual networks created; group averages were also calculated. Graph-theoretic semantic distances are defined as the shortest path between nodes/concept items, with shorter distances indicating greater association or similarity. Agreement and correlation of semantic distances can indicate a degree of association between student and expert cognitive structures.

PFN also provides two unique measures for comparing two networks: closeness/common links and similarity. Closeness indicates the number of links in common with a second network, whereas similarity is the degree that two networks contain the same nodes and links. Identical networks have a value of one, and no shared links have a value of zero. Similarity is calculated by comparing the nodes and links of two data sets to assess the number of shared links. This provided a direct comparison between a student's cognitive structure and an expert cognitive structure. In effect, the greater the closeness and similarity, the more expert-like the network. The reliability of PFN is also highly dependent upon the reliability of proximity ratings.; however, research related to the test-retest reliability of this approach was lacking.

Agreement Analysis

The intent of examining a relationship between ECS and SCS is to understand if the student's cognitive structure "agrees" with that of the expert, to what degree it agrees (if that is possible to derive), the strength of association between them, and whether this level of agreement impacts future academic success. Assessment of the agreement between student and expert occurs throughout physiotherapy education and requires both interrater reliability and agreement while examining the level of competency displayed in the performance of a clinical activity (Liao et al., 2010). Exploring the relationship between student cognitive structure and expert cognitive structure involves comparing their perceptual organization and relationships between concepts and items. Each cognitive structure representation serves as a measurement tool of the perceived organization of anatomical concepts. These elements are necessary for concurrent and predictive criterion-related validity; the former is reflected in the level of agreement between SCS and ECS (criterion standard one), and the latter is represented by the relationship between SCS properties, the agreement between ECS and SCS, and the unit grade (criterion standard two).

There is a lack of consistency in the literature regarding how to assess the agreement or relationship between raters and the construct validity of measures used to do so. Reliability and agreement are often used interchangeably; however, they represent different constructs and have poor operational definitions (Hernaez, 2015; ten Hove et al., 2018). This has led to the inappropriate use of various statistical analyses (Aggarwal & Ranganathan, 2016, 2017; Ranganathan et al., 2017). Reliability and agreement research often focuses on continuous data; however, ordinal data can be analyzed as continuous data via robust parametric statistical analyses within specific contexts (Harpe, 2015; Norman, 2010; Rhemtulla et al., 2012). A frequent pitfall occurs with the use of Pearson's correlation coefficient. Raters can have a high strength of association (for example, $r = 1$ would be a perfect association) yet have no agreement on any of the ratings. For example, rater one could have three ratings of 1, 1, and 1 and rater two could have ratings of 3, 3, and 3; although the correlation coefficient $r = 1$, the level of agreement is zero. This phenomenon is common in educational research; for example, previous seminal research by Goldsmith et al. (1991) and Gonzalvo et al. (1994) described the correlation of student and expert proximity data, MDS Euclidean distances, and PFN graph-theoretic distances as well as PFN coherence, common links, and similarity. However, they did not report having met the assumptions deemed necessary for the appropriate use of correlational analysis.

Clear differentiation of these terms and the operationalization of constructs is integral to a focused examination of the relationship between student cognitive structure and expert cognitive structure and establishing concurrent criterion-related validity and selecting appropriate and relevant statistical analyses. Stolarova et al. (2014) and Looney (2018) developed frameworks for agreement analysis that provided the foundation for the quantitative comparison of cognitive structures. Stolarova et al. (2014) defined three methods to address agreement: interrater reliability, interrater agreement, and strength of linear association. Looney (2018) provided a framework for agreement aligned with Barnhart et al. (2007) that focuses on absolute agreement via graphical plot, unscaled, and scaled indices. If neither of the individuals serves as a reference, then agreement tends to reflect reliability; if one of the individuals serves as a reference, then agreement tends to reflect validity (Looney, 2018). The latter becomes highly relevant for RQ2 in this study.

In the context of this study, reliability was the extent to which the raters can consistently discriminate between paired comparisons; agreement was the extent to which different raters assign the same value of perceived relatedness (Chaturvedi & Bajpai, 2015). The relationship between student and expert cognitive structures is represented by reliability (reflected in interrater reliability), accuracy (via unscaled and scaled indices), and the strength of linear association between rater variables (Haghayegh et al., 2020; Looney, 2018; Stolarova et al., 2014). First, interrater reliability between student and expert ratings can be examined via Krippendorff's alpha coefficient, a robust tool used to measure reliability and agreement used for various data types and number of raters that accounts for chance agreements (Krippendorff, 2004). This provides Krippendorff's alpha coefficient with greater flexibility in its use (Shabankhani, 2020; Zapf et al., 2016). It also embraces several other known reliability coefficients such as Spearman's rho, Pearson's intra-class correlation, and the kappa statistic as it is calculated based on the differences between raters (A. F. Hayes & Krippendorff, 2007; Shabankhani, 2020). However, as noted in Chapter 2, test-retest reliability, a specific component of interrater reliability and validity, is beyond the scope of this study as it requires a larger sample size. Second, absolute agreement between raters can be established via graphical plots, unscaled, and scaled indices (Barnhart et al., 2007; Looney, 2018). Unscaled indices represent the level of agreement via the mean of differences and standard deviation, limits of agreement with a 95% confidence interval (derived from the mean of differences), or the root mean squared deviation (Kopp-Schneider & Hielscher, 2019). The mean of differences indicates the average agreement across all measurements. It provides a sign or direction, whereas the absolute mean of differences provides an indicator of each measure without sign or direction. The root mean square deviation provides a measure of the average difference in agreement between raters (accuracy) given the units of the original rating (Barnhart et al., 2007; Looney, 2018). This data can often be displayed effectively in a bivariate plot of raw scores compared to a y=x identity line, histogram, or a Bland-Altman plot (Bland & Altman, 1986, 2003; Giavarina, 2015; Haghayegh et al., 2020; Looney, 2018). Scaled indices include intraclass correlation and Lin's concordance correlation coefficient. However, Krippendorff's alpha coefficient is used in place of the intra-class correlation based on the findings of A. F. Hayes and Krippendorff (2007). Finally, the strength of linear association can be calculated via Pearson's correlation coefficient. Many parametric approaches such as Pearson's correlation coefficient that are thought to require continuous data are robust enough to accommodate both non-normality of data and Likert scale data (Harpe, 2015; Norman, 2010). However, the true value of a correlation coefficient may lay in its derived coefficient of determination which describes the proportion of variance accounted for by the statistical solution. It is interesting to note that many of the reliability coefficients currently in use (Cohen's kappa, weighted kappa, ICC, Pearson's correlation, Spearman's rho, and Kendall's tau-b) produce similar results when used with ordinal data, especially if there are seven or more categories of the ordinal variable (de Raadt et al., 2021).

In this study, the construct of cognitive structure was operationalized as "a hypothetical construct referring to the organization of the relationships of concepts in long-term memory" (Shavelson, 1972, p. 226-227) via pairwise comparisons representing perceptual concept organization. However, understanding the relationship between the variables and the context of the raw data (in terms of normality) requires understanding both the agreement between student and expert and the strength of linear association between the two. This promotes the correct use of operational definitions, the careful examination of assumptions for parametric tests given the context of the data, and appropriate comparisons with the previous research. Three levels of analysis were used. A qualitative visual comparison of both MDS and PFN representations was the first step toward examining the potential relationship between student and expert cognitive structure. A quantitative descriptive comparison of raw proximity data, MDS configuration properties, and PFN network properties provided the second comparison level. Descriptive parameters derived from the MDS configuration included stress-1, TCC, and R^2. Descriptive parameters derived from the PFN network properties included degree, eccentricity, number of links, and coherence. Finally, a quantitative statistical analysis examined the raw proximity data, MDS Euclidean distances, and PFN graph-theoretic distances in terms of interrater reliability (Krippendorff's alpha coefficient), accuracy (root mean squared deviation), and strength of linear association (Pearson's correlation coefficient). These measures served to examine the relationship between student cognitive structure and expert cognitive structure in RQ2. Potential predictor variables for RQ3 were subsequently be derived from these measures and used in conjunction with prior knowledge.

Criterion-Related Validity

Criterion-related validity was explored for two criterion standards: expert cognitive structure and unit grade. Cognitive structure was represented by MDS-derived measures (stress-1, TCC, R^2, Euclidean semantic distances) and PFN-derived measures (links, degree, eccentricity, coherence, similarity, and graph-theoretic semantic distances). The first criterion-related validity with expert cognitive structure is exemplified by agreement between student cognitive structure and expert cognitive structure while controlling for prior knowledge (admission GPA and admission anatomy GPA). The second criterion-related validity with unit grade is exemplified by a relationship between student cognitive structure and the agreement between student and expert and unit grade while controlling for prior knowledge (admission GPA and admission anatomy GPA).

Data Analysis Plan

Data collected within the study were analyzed using the Statistical Package for the Social Sciences (SPSS) and the Pathfinder analysis software publicly available at www.interlinkinc.net (Schvaneveldt, n. d.). Online survey data consisting of pairwise comparisons were entered as a proximity matrix for further analysis in both statistical packages. Both MDS and PFN utilize the same proximity matrix data. Data were stored in Excel spreadsheets as well as SPSS data files.

Data Preparation

Each participant (student and expert) was issued a unique identifier code to complete the online survey. This code number would remain associated with the individual's identification number (student) or name (expert) until all data were collected from all sources (survey, registrar, Blackboard). All data were maintained within an Excel spreadsheet. A complete data set for each student participant consisted of a student number, unique identifier code, survey ratings, unit written exam grade (Blackboard), unit practical exam grade (Blackboard), and demographic data collected from the registrar, which included age, gender, admission GPA, admission anatomy GPA, GRE score, program/campus, and mode of delivery. Demographic data were used for post-stratification weighting to ensure a sample that is as closely representative of the target population as possible (Battaglia, 2008). A complete data set for each expert participant consisted of a unique identifier code, program/campus, mode of delivery (residential or flexible), test ratings, number of years of clinical practice, number of years of anatomy teaching practice, terminal clinical degree, and terminal academic degree. Coding designations are noted in Appendix C. A check of email address and IP address ensured unique survey entries without duplication. Once a data set was complete and had all the required elements, the data were de-identified, and only the unique identifier was used.

Each data set was examined manually for missing data and data entry errors. The primary technique for minimizing missing data is an effective data collection strategy built into the study design (Kang, 2013). Missing data can significantly impact quantitative research, resulting in loss of information and bias, decreased statistical power, and increased standard errors (Dong & Peng, 2013). Missing data can occur during the survey proximity ratings or opt-out or non-completion of the posttest survey. A complete description of the process to be completed during the study is essential, with well-defined steps described for the participant. Effective organization of the survey ensured that a participant must complete each item before moving to the next item to minimize the risk of missing data. The survey design prevented item non-response as much as possible, leaving unit-level non-response – when no information is collected from the participant (Dong & Peng, 2013). Email reminders were sent to participants before the projected date of completion of the survey. If a survey was partially completed, this was considered an opt-out. In the event of missing pairwise proximity data, listwise deletion will be used. Most item non-responses would be related to missing data that is considered missing completely at random (Kang, 2013). In these circumstances, listwise deletion, also known as complete case analysis, is the most common approach to missing data while limiting bias associated with removing cases (Kang, 2013). A complete set of proximity ratings are required to derive a cognitive structure; if the data set is incomplete, then it would be preferable to remove the case entirely. Although this may impact the sample size, all correlations and other statistical analyses will be performed on the same set of participants (Warner, 2013). As all pairwise comparisons are necessary to utilize both MDS and PFN and pairwise ratings are perceptual, using a principled missing data method was not indicated in this research design. An SPSS data file was created for proximity matrix data in both MDS and PFN analyses.

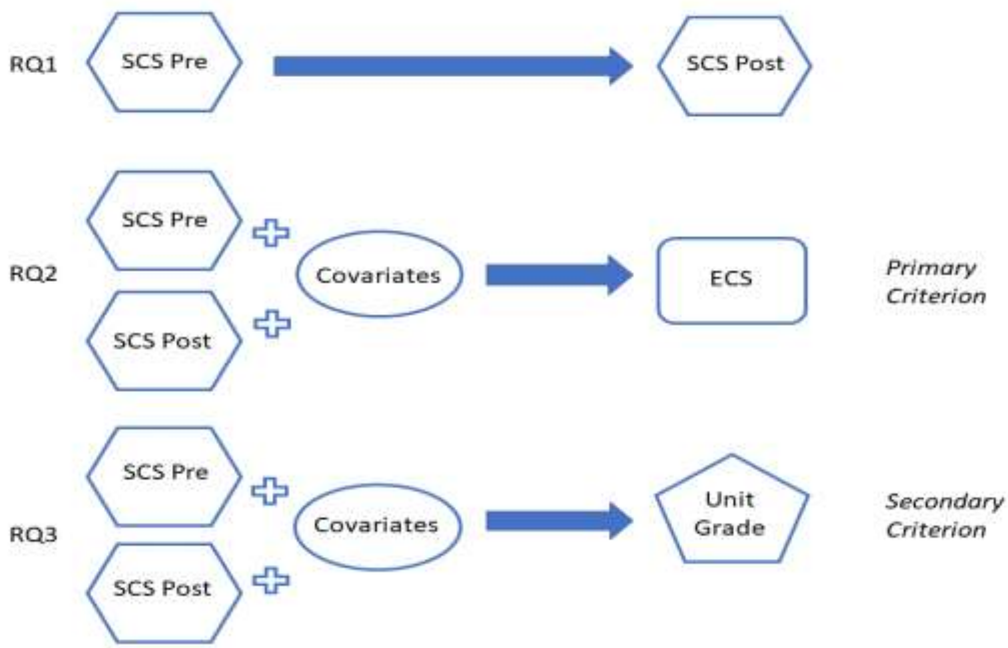
Data screening consisted of proofreading the original data sets and SPSS data file for inconsistencies. Descriptive statistics were used to establish a profile of the data. A histogram, scatterplot, and summary statistics provided descriptive data regarding central tendency and the range of scores. This data established the normalcy of distribution and identified potential outliers outside of +/- 3 standard deviations. Assumptions for all statistical tests were reviewed for the integrity of the statistical analysis. For paired *t* tests, assumptions consist of no significant outliers and normal distribution of differences in the dependent variable (Warner, 2013). For Pearson correlation coefficients, assumptions consist of paired continuous data, linear relationship, no significant outliers, and bivariate normality (Warner, 2013). For multiple regression, assumptions consist of a continuous dependent variable with at least two independent variables, independence of observations, linearity, homoscedasticity, no multicollinearity, no significant outliers, and normal distribution of residuals (Warner, 2013). The data were analyzed after these assumptions were met or addressed. Further commentary on normality and statistical tests is noted in Chapter 4.

Data Analysis and Research Questions

The research study addressed several interrelated components. An overview of the research design and variables is visualized in Figure 3, with a summary of the research questions and associated data analysis noted in Table 2. Methodological discrepancies and revisions are discussed in Chapter 4.

Figure 3

Research Questions and Variables



Note. SCS Pre = student cognitive structure (pretest); SCS Post = student cognitive structure (posttest); ECS = expert cognitive structure.

Both expert and student cognitive structures were reflected in MDS- and PFN-derived measures (MDS dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances, and PFN links, degree, eccentricity, coherence, similarity, and graph-theoretic semantic distances). Averages were calculated for both physiotherapy centric and domain centric expert subgroups as well as the combined expert group. Student cognitive structure (reflected in MDS- and PFN-derived measures) was to be examined for within-subject changes over time, making a pretest and posttest comparison relevant to the study. However, methodological discrepancies are noted in Chapter 4. Criterion-related validity (concurrent and predictive) was examined to assess potential relationships between student cognitive structure and its associated variables and both expert cognitive structure, its associated variables, and unit grade. Concurrent validity was reflected in the level of agreement between SCS and ECS (criterion standard one), whereas predictive validity was represented by the relationship between SCS, level of agreement, and the unit grade (criterion standard two).

A visual comparison of both MDS and PFN representations was the first step toward examining the potential relationship between student and expert cognitive structure. A quantitative descriptive comparison of proximity data, MDS configuration properties, and PFN network properties provided the second comparison level. Descriptive parameters derived from the MDS configuration include stress-1, TCC, and R^2. Descriptive parameters derived from the PFN network properties include degree, eccentricity, number of links, and coherence. The dimensionality of the MDS representation was defined by the average expert cognitive structure representations and the dimensionality that established the data's best fit. The same dimensionality was used to analyze all student cognitive structures to establish consistent comparisons based on the criterion standard. Student MDS stress-1 (goodness-of-fit) and R^2 was calculated based on the dimensionality of the expert cognitive structure. Finally, a quantitative statistical analysis examined the proximity data, MDS Euclidean distances, and PFN graph-theoretic distances in terms of reliability (Krippendorff's alpha coefficient), accuracy (root mean squared deviation), and strength of linear association (Pearson's correlation coefficient) between student and expert. Multiple regression was used to examine potential relationships between student cognitive structure, levels of agreement between student and expert (primary criterion), and unit grade (secondary criterion) while controlling for prior knowledge (admission GPA and admission anatomy GPA). Finally, the instructor and mode of delivery were examined as potential moderating variables. These variables were included as potential confounding variables.

The following research questions were initially conceived and considered in this exploratory study within the context of physiotherapy students enrolled in a first semester foundational gross anatomy course. The study was guided by the following RQs:

RQ1: Is there a meaningful change over time in the quantitative representation of student cognitive structure?

RQ2: Is there a relationship between student cognitive structure and expert cognitive structure while controlling for prior knowledge?

RQ3: Is there a relationship between student cognitive structure and unit grade while controlling for prior knowledge?

Given the RQs' exploratory nature, hypotheses for each RQ were not appropriate or indicated. Methodological discrepancies and revisions are discussed in Chapter 4.

Table 2

Research Questions, Variables, and Data Analysis Plan

Construct validity		Variable	Data analysis
RQ1	Is there a meaningful change over time in the quantitative representation of student cognitive structure?	MDS dimensionality, stress, semantic distances, and R^2 PFN semantic distances, coherence, and similarity	Mean differences and relationship between pretest/posttest student cognitive structures and associated variables Scatterplots Paired *t* tests and Cohen's d Pearson correlation (r) and r^2
Criterion validity		**Variable**	**Data analysis**
RQ2	Is there a relationship between student cognitive structure and expert cognitive structure while controlling for prior knowledge?	MDS dimensionality, stress, semantic distances, and R^2 PFN semantic distances, coherence, and similarity	The relationship between student cognitive structure and the criterion variable (expert cognitive structure) is represented by reliability (Krippendorff's alpha coefficient), agreement (RMSD), and strength of linear association (Pearson's correlation coefficient).
RQ3	Is there a relationship between student cognitive structure and unit grade while controlling for prior knowledge?	MDS dimensionality, stress, semantic distances, and R^2 PFN semantic distances, coherence, and similarity Unit grade Admission GPA Admission anatomy GPA	Multiple regression to examine relationships between multiple student cognitive structure predictor variables and the criterion variable (unit grade). Pearson correlation (r), multiple correlation coefficient (R), standardized coefficient (β), adjusted coefficient of determination (ΔR^2), and regression equation

Threats to Validity

Portney and Watkins (2009) presented a straightforward stepwise process for assessing threats to validity. Statistical conclusion validity refers to the appropriate use of statistical procedures. This allows for appropriate and valid conclusions to be drawn from the relationships between the dependent and independent variables. Internal validity addresses confounding factors that might interfere with these relationships. Construct validity refers to the theoretical constructs representing the variables and their interpretation. External validity refers to the generalizability of results beyond the current study. Each element must be addressed in the study's design, with threats to validity limited or addressed with specific procedures to diminish the threats' impact.

Statistical conclusion validity is the use of statistical analyses appropriate for the data and goals of analysis (Matthay & Glymour, 2020; Portney & Watkins, 2009). Threats to statistical conclusion validity are limited by meeting the appropriate statistical analysis assumptions and ensuring adequate power. Type I and II errors are a component of statistical conclusion validity; however, these factors are addressed via an appropriate selection of alpha and beta values in the study design. The error rate can increase as repeated measures increase; however, this will not be a significant threat to validity with two repeated measures. Factors that influence the study's variability are controlled using standardized protocols (throughout the study's design) and the homogeneity of participants within the cohort. The reliability of both MDS and PFN analyses has not been extensively researched and is an acknowledged limitation of the current study.

A study with internal validity has conditions that promote causal inferences, and the results are due to the study's factors and not due to confounding variables (Matthay & Glymour, 2020; Warner, 2013). Internal validity demands control of extraneous variables (Campbell & Stanley, 1963; Warner, 2013). The current study has several threats to internal validity. As the research design has pretest and posttest measurements, there is a risk of uncontrolled factors occurring over time (history) and changes within the individual (maturation). These factors should be minimized as students throughout the institution have similar admission requirements in each cohort and are within a narrow age group. Social interaction between participants may occur, though the participant's adherence to the study instructions will limit this. There is a risk of a testing effect as participants may better understand the context and process of similarity ratings during the posttest conditions. However, the order of pairwise comparisons can be varied for the posttest ratings, thus minimizing the testing effect. Attrition may impact average group scores depending on which participants drop out of the study by not completing the pretest and posttest. This was limited by using email reminders for posttest rating completion. Attrition rates are noted in the study results. The selection of the criterion or gold standard is critical to internal validity, yet the assumption is made that they are, in fact, the best criterion standard. Although the research design of the current study cannot control this factor, these assumptions are based on the best available research literature related to MDS and PFN. As noted previously, instrumentation and the reliability of measurement is an acknowledged limitation of the current study. Measurement error as a function of instrument reliability can lead to statistical regression (Portney & Watkins, 2009). Each of these threats to internal validity is a function of the research design and serves as a limitation to the study.

The most significant threat to internal validity is selection bias. Voluntary response (nonprobability) sampling allows all students in all modes of delivery to participate. However, this can create a self-selection bias. Exclusion criteria assist in limiting confounding variables in the selection process. Including as many students as possible will dimmish the impact of selection bias by increasing the representativeness of the sample. A lack of a control group (and subsequent assignment of participants) limits the effect of selection bias. Although selection bias issues exist, they are often unavoidable and inherent to many educational studies and institutional procedures. Random (probability) sampling, while limiting selection bias, would severely limit the study's sample size and power. Greater power (with selection bias) is a preferred limitation to the current study compared to random sampling (with a significantly underpowered design), enhancing the potential for generalizable effect sizes within the participant pool's constraints. Repeated measures enhance internal validity by providing each participant with their own control (Warner, 2013). As the target population is DPT students within the first semester of the program, regardless of the mode of delivery or campus, post-stratification and weighting can align the sample's demographic characteristics with those of the target population (Battaglia, 2008). Demographic data such as age, gender, admission GPA, admission anatomy GPA, and GRE scores will be used for post-stratification weighting to ensure a sample that is as closely representative of the target population as possible (Battaglia, 2008; Farrokhi & Mahmoudi-Hamidabad, 2012).

Construct validity reflects the construct being measured accurately, representing the construct in question (Matthay & Glymour, 2020). Threats to construct validity involve issues related to the operationalization of constructs (including construct definitions) and experimental biases introduced by the researcher or participant. The development of the item list for pairwise comparisons is critical to having construct validity. The use of several resources, including the *Foundational Model of Anatomy* (Clarkson & Whipple, 2018), will limit this potential threat. Operational definitions of important constructs are clearly reported, and the operationalization of these constructs is outlined. Data modeling (MDS and PFN) to represent cognitive structure is based on previous research that provides a precedent for potential construct validity. However, as noted in Chapter 2, the construct of a cognitive structure appears to have some characteristics that promote representation, albeit indirectly. Experimental bias related to the researcher is limited as there is limited interaction between researcher and participants. Finally, the Hawthorne effect could also play a role as participants may change their behaviors as they know they are being studied. However, this becomes less of an issue as the study does not have an intervention, and interaction with the researcher is limited.

External validity reflects the generalizability of the study results beyond the current study participants and context (Campbell & Stanley, 1963; Matthay & Glymour, 2020; Warner, 2013). If a study has high external validity, results will translate to real-world scenarios with varied contexts and participants. Two essential issues exist in establishing external validity and the generalizability of results: the effect of the specific setting and context and participants being representative of the target population (Portney & Watkins, 2009). The current study has several threats to external validity. Study results will be specifically relevant to the program and institution of the target population. However, the assumption is made that the institution's admissions are consistent with the broader population of students entering other DPT programs. Admissions data such as admission GPA and admission anatomy GPA can be compared to national DPT program averages if available. Reliability and consistency of measurement in terms of grading are emphasized based on institutional guidelines, though this may limit the generalizability of study results to DPT programs at other institutions. Replication of the study in the future with different student groups at various institutions will be necessary to broaden the impact of the study results. Testing reactivity may impact posttest measures; although this is inherent to the proposed repeated measure research design, the effect should be nominal if posttest ratings have varied the order of pairwise comparisons. The greatest challenge to external validity parallels that of internal validity: the representativeness of the sample. Sampling bias is often used synonymously with selection bias. It is reflected in participants that do not represent the general population due to self-selection and voluntary response (nonprobability) sampling. The strategies utilized to limit this threat to external validity are consistent with those previously noted to limit internal validity. Furthermore, clearly describing the conditions of sampling via an audit trail is integral to reporting study results. Attrition rates reflective of nonresponse bias will also be reported to provide greater clarity of any disparities evident in sampling (McCutcheon, 2008).

Ethical Procedures

Ethical procedures are of utmost importance in the methodology of the current study. Participants completed an informed consent before the initial online survey, which outlined all expectations and rights of the participant. Participants were assured that their personal information related to the study would be held in strict confidence. Data collection was both protected and anonymous to ensure the privacy and security of information. Several layers of encrypted data storage were used to ensure data security and integrity. This consisted of password-protected files stored on both local encrypted storage via flash drive and encrypted cloud storage. Data files were accessible to the primary investigator and were password protected. Local storage was secured via a flood- and fire-proof safe at the primary investigator's home. Data will be stored for five years per Walden University criteria, at which time the files in question will be deleted. The student identification number was used to ensure that a complete data set (with associated unique identifier code) was compiled; all data were de-identified at that time.
Participation in the study was voluntary, and participants were free to withdraw at their discretion without adverse effects. A $10.00 gift card appreciation was given to all student and expert participants who completed the study and all physiotherapists who completed the review and rank-ordering of anatomical concept items. This was used to promote participation without coercion. Academic standing was not impacted by a student choosing to participate or not. Students currently being instructed by the primary

investigator were not eligible for participation to limit any undue influence on data collection at the investigator's institution.

Summary

The methods, sampling, data collection, and ethical procedures discussed in Chapter 3 provide a sound methodological platform for examining the research questions. The research design provides the foundation for examining criterion validation and using these strategies for cognitive structure mapping and quantitative representation in physiotherapy students learning gross anatomy. Threats to validity are described, with potential methodological issues addressed. Finally, ethical procedures are outlined to ensure privacy, confidentiality, and data security.

Chapter 4: Results

The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. The study was designed to address three research questions:

RQ1: Is there a meaningful change over time in the quantitative representation of student cognitive structure?

RQ2: Is there a relationship between student cognitive structure and expert cognitive structure while controlling for prior knowledge?

RQ3: Is there a relationship between student cognitive structure and unit grade while controlling for prior knowledge?

Given the research questions' exploratory nature, hypotheses for each research question were not appropriate or indicated.

Chapter 4 provides an overview of the results of the study. Major sections in this chapter include data collection, an overview of the sample participants (physiotherapist, expert, and student), data analysis including MDS and PFN, and a summary of the findings for each research question. Several exploratory analyses are included that provide further context for the research questions. The chapter concludes with a summary and a transition to the Discussion in Chapter 5.

Data Collection

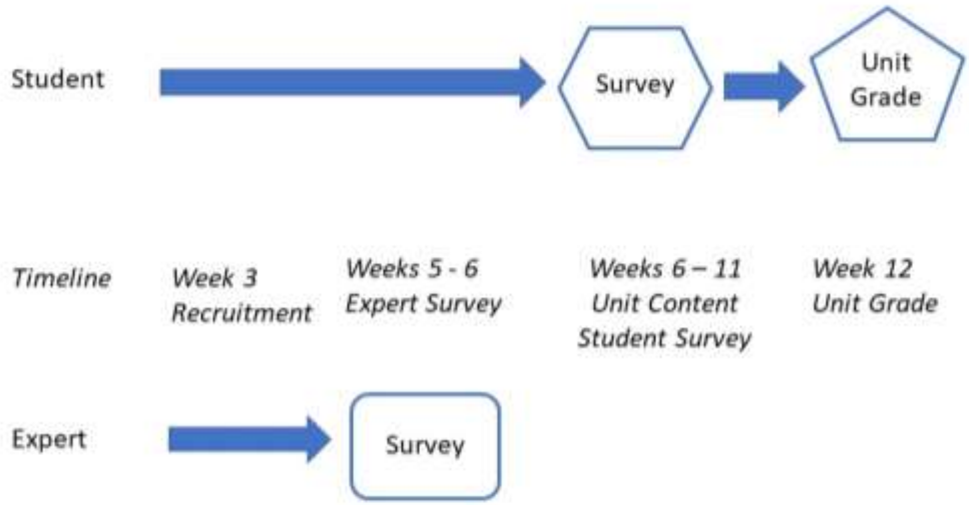
Walden University served as the Institutional Review Board (IRB) of record and the partner organization's IRB entered into an Interagency Authorization Agreement for approval at the institutional level. Once final approval was received from Walden University's IRB and the partner organization (Walden University approval number 12-23-20-0979508) on January 22, 2021, recruitment of subjects was initiated. Data collection began on January 25, 2021 and ended on March 31, 2021.

Methodological Discrepancies

There was one primary methodological discrepancy in data collection compared to the research design that was originally proposed. Due to unforeseen delays in IRB approval, the timing of the study was impacted. These delays prevented the completion of a pretest survey scheduled to be offered in the first 3 weeks of the semester. This necessitated either the delay of the study by 15 weeks or the removal of RQ1. I chose the latter option. The data collection for RQ2 or RQ3 was not adversely affected, and there was no impact on the associated validation components of the study. The finalized timeline is displayed in Figure 4.

Figure 4

Final Study Timeline



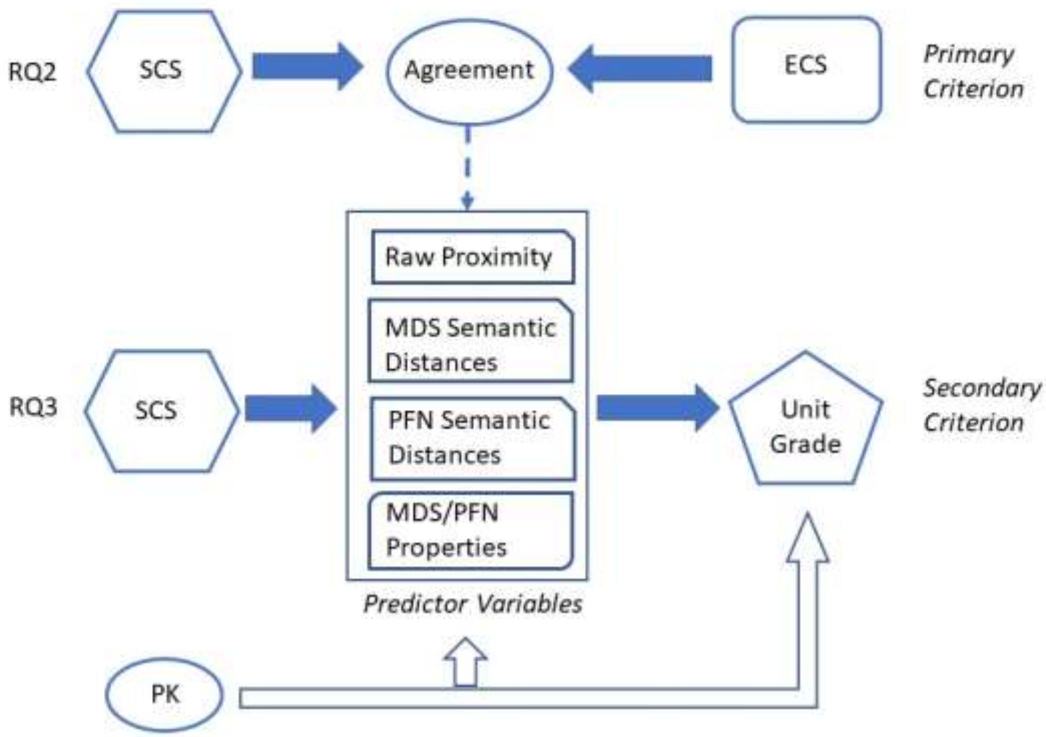
Several smaller discrepancies and refinements in the recruitment process were based on the logistics involved with multiple cohorts in multiple programs. Initially, a comparison based on the mode of delivery (residential versus flexible) was planned. However, as the study data were collected during the COVID-19 pandemic of 2020–2021, all students in all cohorts were in a remote learning environment for the duration of the study. This removed the potential differentiation between residential and flexible modes of delivery. However, comparisons related to the instructor and their specific cohort were maintained. Teleconference arrangements were difficult to make based on the scheduling of classes in multiple time zones and institutional scheduling changes driven by the pandemic. The student recruitment efforts subsequently focused on the

Blackboard course email distribution and the associated video description of the study embedded in the call for participants. Student ID numbers were not used in the initial survey registration process. All prospective survey participants obtained a unique survey token and login directly from the survey website by using their email address. This process simplified registration and decreased the barriers to participation; it also enhanced the confidentiality of the process by removing the initial email request via student ID number.

There were several adjustments in how the study data were analyzed. These are summarized in Figure 5. Moderating variables such as the program mode of delivery were removed because of changes in all modes of delivery necessitated by the 2020–2021 pandemic. With the absence of a pretest survey, the SCS-derived parameters were used as the predictor variables for RQ3. An 11-point Likert scale for the paired comparisons was replaced by a 7-point scale to better align with previous seminal research as well as to improve ease of use by the participant and to be better aligned with best practices in the use of grid formats and online surveys (see Goldsmith et al., 1991; Grady et al., 2019; Liu & Cernat, 2018). Prior knowledge was initially planned as a covariate; however, with the shift in research design given the absence of a pretest, prior knowledge became a predictor variable for RQ3. A weighted average of the expert group participants was not used because it became apparent that subgroups were more relevant to the analysis. Poststratification weighting via demographic data were not deemed necessary due to a smaller-than-expected student sample.

Figure 5

Updated Research Questions and Variables



Note. SCS = student cognitive structure; ECS = expert cognitive structure; MDS-derived parameters = stress-1, TCC, $R^{2;}$ PFN-derived parameters = links, coherence, similarity; agreement = reliability, accuracy, and association; prior knowledge = admission cumulative GPA and admission core sciences GPA.

Correlations and multiple regression were initially proposed as the primary statistical analyses; however, further evaluation of the research questions and data indicated that agreement analysis (reflected in interrater reliability, level of agreement, and strength of linear association between student and expert) were all integral to understanding the relationship between student and expert cognitive structures for RQ2. Multiple linear regression remained the primary statistical approach used for RQ3 with predictor variables derived from RQ2. These variables (derived parameters and agreement analysis) are summarized in Table 3.

Table 3

Summary of Constructs and Variables for RQ2 and RQ3

Construct definition	Construct operationalization	Variable
Cognitive structure: "A hypothetical construct referring to the organization of the relationships of concepts in long-term memory." (Shavelson, 1972, p. 226-227)	Pairwise comparisons (raw proximity semantic similarity data) representing perceptual concept organization	PRX
Cognitive structure mapping: The representation of cognitive structure reflected in a cognitive map defined via two data modeling strategies (MDS and PFN) and their derived quantitative parameters and data visualization.	MDS spatial representation MDS configuration properties MDS Euclidean distances PFN network representation PFN network properties PFN common links/similarity PFN graph-theoretic distances	MDS data visualization MDS stress-1, TCC, R^2 MDS Euclidean distances agreement (α, RMSD, *r*) PFN data visualization PFN links, coherence PFN common links/similarity PFN graph-theoretic distances agreement (α, RMSD, *r*)
Research Question 2	Relationship between SCS and ECS reflected in MDS/PFN properties and agreement analysis	
Research Question 3	Relationship between RQ2 predictor variables, prior knowledge, and unit grade	

Note. PRX = proximity data; TCC = Tucker's coefficient of congruence; R^2 = coefficient of determination; α = Krippendorff's alpha, RMSD = root mean square deviation; *r* = Pearson's correlation coefficient.

Sampling

Physiotherapists, experts (including lead course instructors and domain experts), and first trimester DPT students were recruited for participation in the study via voluntary response (nonprobability) sampling. Recruitment and response rates varied based on the population in question.

Physiotherapists

Thirteen physiotherapists were recruited to participate in the study; an email address was used during registration to generate a unique survey token and login. All data sets were deidentified once the data collection was completed. The response rate was 92.3%, providing a sample of 12 physiotherapists currently in musculoskeletal clinical practice. Descriptive statistics for the physiotherapist sample are displayed in Table 4. The highest clinical degree attained by physiotherapists was the doctorate in PT (n = 6), followed by a bachelor's in PT (n = 4) and a master's in PT (n = 2). All are reflective of the entry-level to practice in the United States and to attain state licensure. The number of years in clinical physiotherapy practice in musculoskeletal care ranged from 11 to 35 years, with a mean of 22 years. The sample represented 264 total years of clinical practice. Three of the physiotherapists (25%) had taught gross anatomy in the past; however, the mean duration of teaching gross anatomy among the three was negligible (1.3 years). These physiotherapists were included in further data analysis.

Experts

The seven Gross Anatomy lead instructors in the partner organization and seven gross anatomy domain experts from outside the partner organization were invited to participate in the study; an email address was used during registration to generate a unique survey token and login. All data sets were deidentified once the data collection was completed. The overall expert response rate was 57.1%, providing a sample of 5 lead instructors (71.4% response rate) and three gross anatomy domain experts (37.5% response rate). Descriptive statistics for the expert sample are displayed in Table 4. Lead instructors were physiotherapy centric in that they were teaching gross anatomy and had clinical degrees in physiotherapy. In contrast, domain experts were domain centric and did not have a clinical degree. Seven of the eight experts had a doctoral degree as their highest academic degree (PhD = 5, DHSc = 1, Sc.D. = 1) with one master's degree reported. There were two primary subgroups: clinical (those having a clinical degree; n = 5) and nonclinical (those not having a clinical degree; n = 3). The clinical group consisted of instructors (n = 4), who were responsible for instructing the cohorts in question, and noninstructors (n = 1). Three lead instructors had a master's in PT and one had a doctorate in PT.

There were two unique cases. The first case was a lead instructor who did not have a cohort represented in the sample population and was an outlier lacking a clinical degree. This subject was subsequently considered in the domain expert subgroup because these experts were domain centric and not physiotherapy centric with clinical degrees. The second case, initially recruited as a domain expert, was the lone participant in this subgroup with a clinical degree. Because this subject's clinical degree was in physiotherapy and the subject had 40 years of clinical experience, they were considered an outlier in the domain expert group. This subject was subsequently considered in the clinical subgroup.

Data were initially analyzed based on total expert group (ECST, *n* = 8) as well as three subgroups: nonclinical (ECSD *n* = 3), clinical (ECSC *n* = 5), and clinical lead instructors (ECSI, *n* = 4). Based on the initial descriptive analysis, it appeared that the physiotherapist sample was consistent with the expert clinical and instructor group.

Table 4

Physiotherapist and Expert Demographic Data

		YTA			YCP		
Group	*n*	*M*	*SD*	Range	*M*	*SD*	Range
Physio	12	0.33	0.65	0-2	22.00	8.30	11-35
ECST	8	22.38	13.56	5-44	18.88	17.10	0-40
ECSC	5	15.00	9.77	5-25	30.20	9.18	22-40
ECSI	4	12.50	9.26	5-25	27.75	8.50	22-40
ECSD	3	34.67	9.50	25-44	0	0	0

Note. YTA = years of teaching anatomy; YCP = years of clinical practice; ECST = ECS total group (*n* = 31); ECSC = ECS clinical subgroup (*n* = 5); ECSI = ECS instructor subgroup (*n* = 4); ECSD = ECS domain expert subgroup (*n* = 3).

DPT Students

Five concurrent cohorts of students totaling 224 students (165 residential program, 59 flexible program) were invited to participate in the study. This sample was much smaller than the expected target population of 320 because the partner organization had one fewer cohort registered for the trimester and fewer students per cohort than projected. The student's email address during registration was used to generate a unique survey token and login. All data sets were deidentified once the data set was complete.

Initial student registration for the online survey was 21.9% (49 students) which was consistent with expectations based on survey research; however, the overall student response rate was 13.9%, providing a sample of 31 students. A total of 18 students (8%) failed to complete the online survey. Seventeen of the 31 students were enrolled in the residential program (54.8% of the sample with a 10.3% response rate), and fourteen students were enrolled in the flexible program (45.2% of the sample with a 23.7% response rate). However, as noted previously, the mode of delivery for both residential and flexible programs shifted to an exclusively remote learning environment during the pandemic.

Demographic data were collected for the five concurrent cohorts. Admission data included age, gender, admission cumulative GPA, admission core science GPA, and GRE. The data representing both the total cohort and study sample are displayed in Table 5.

Table 5

DPT Student Demographic Data

	Total	(*n* = 224)	Sample	(*n* = 31)
Characteristics	*M*	Range	*M*	Range
Age	25.94	20-48	26.35	23-40
Gender (% F:M)	52:48		55:45	
GRE	298.23	280-329	296.68	283-308
Cumulative GPA	3.23	2.48-4.0	3.28	2.63-4.0
Core Sciences GPA	3.30	2.60-4.0	3.37	2.96-4.0

Note. The study sample was representative of the target population, and thus post-stratification was not indicated.

Covariates (as defined by age, gender, program location, and program type) and prior knowledge (as defined by admission cumulative GPA and admission core sciences GPA) were consistent between the cohorts and the study sample, indicating that the study sample was representative of the target population. The use of voluntary response (nonprobability) sampling can often necessitate post-stratification to represent the target population more accurately. However, given the small sample size and consistency between the target population and sampling frame, post-stratification was not needed due to differences in the covariates.

Data Preparation

There were four key components in preparing the data for use in each of the three research questions. First, I screened the survey data and prepared it for further statistical analysis. Second, I assessed the relevant statistical assumptions. Third, I used the study data to procedurally describe the development of cognitive structure (both expert and student) via the data modeling strategies. Finally, two exploratory analyses related to the MDS scaling model (and its selection) were necessary to provide important context for the subsequent selection of models for data analysis.

Several clusters of data were collected, with each serving a specific purpose based on the operationalization of constructs employed in the study. Student demographic data included age, gender, GRE score, program location, and program type (both the study participants and the cohort target population). Prior knowledge was represented by two measures: admission cumulative GPA and admission core science GPA. The students' unit grades served as a criterion standard. Physiotherapist and expert demographic data such as highest academic degree, highest clinical degree, number of years teaching gross anatomy, and number of years in clinical practice in musculoskeletal care were collected. Finally, cognitive structure (student and expert), in this study, was represented by the raw proximity data as well as the primary measures derived from MDS (dimensionality, stress-1, TCC, R^2, and Euclidean semantic distances) and PFN (links, degree, eccentricity, coherence, common links/closeness, similarity, and graph-theoretic semantic distances). Interrater reliability, accuracy, and strength of linear association were calculated from student and expert comparisons (proximity data, MDS Euclidean distances, and PFN graph-theoretic distances).

Preliminary Data Screening

Preliminary data screening was performed. Survey data were downloaded and compiled with admissions data and was screened for missing data, errors, and inconsistencies. However, the design of the online surveys prevented the submission of incomplete surveys. Email and IP addresses confirmed that duplicate responses were not submitted. Data sets were subsequently deidentified and prepared for use in Excel and SPSS.

Statistical Test Assumptions

The data analysis focused on the potential relationship between SCS, ECS, and unit grade via factors such as MDS configuration and PFN network properties as well as agreement analysis for semantic distances (MDS Euclidean and PFN graph-theoretic) that included interrater reliability, level of agreement, and strength of linear association. Relevant factors were then considered as potential predictor variables for multiple linear regression. Each of the associated statistical tests has implicit assumptions for their correct use and application. Reliability, represented by interrater reliability, was assessed via Krippendorff's alpha coefficient (Krippendorff, 2004), which provides the flexibility to use all data types and any number of raters. Krippendorff (2004) noted that the alpha coefficient is an umbrella for other commonly used reliability tests, including Spearman's rho, Pearson's intra-class correlation (ICC), and Cohen's kappa. Accuracy was assessed by the root mean square deviation to establish absolute agreement. Scatterplots, Bland-Altman plots (Bland & Altman, 1986), and histograms displayed the data visually when appropriate. Association, represented by the strength of linear relationship, was calculated via Pearson's correlation coefficient and had four assumptions per Warner (2013): two paired continuous variables (research design), linearity between the variables (noted via scatterplot), no significant outliers (noted via scatterplot or Cook's distance), and bivariate normality (Shapiro-Wilk test, histograms, and normal Q-Q plots). Multiple linear regression has eight assumptions per Warner (2013): continuous dependent variable (research design), two or more independent variables (continuous or categorical, based on the research design), linearity between the variables (noted via scatter plots and partial regression plots), no significant outliers (noted via scatterplot, casewise diagnostics, or Cook's distance), independence of observations (Durbin-Watson test), homoscedasticity (via scatterplot), no multicollinearity (tolerance/VIF values), and residuals are approximately normally distributed (histogram and P-P plot or normal Q-Q plots). Many of the assumptions (for example, the nature of the dependent variable, linearity, no significant outliers, and normality) are shared between statistical analyses. As subgroupings of the data remained the same throughout the study (student, expert, instructor, cohort instructor), the assumptions remained consistent throughout the MDS and PFN analyses that utilize the same raw proximity data.

The issue of normality, a fundamental assumption in both correlational analysis and multiple linear regression, becomes problematic based on the context of the data. Paired comparisons and their associated perceived relatedness are assumed to be normally distributed based on Thurstone's law of comparative judgment (Brown & Peterson, 2009). However, this normality is at the level of each individual comparison only. Multiple paired comparisons produce ordinal data that is inherently not normally distributed across the multiple comparisons, and as such, histograms would not indicate a normal distribution. When comparing proximity data between groups or individuals, assessment of normality may be a function of the individual's perception, knowledge, and experiences regarding their perceived relatedness of that specific paired comparison instead of it being considered an "outlier" in the context of a normal distribution. As a result, removing outliers (in the traditional context) would effectively compromise and remove the representation of certain paired comparisons. Norman (2010) noted that many parametric tests, including Pearson's correlation coefficients, are robust tools and are not adversely impacted by non-normal distributions, especially given the context of the data noted above. Aggregated Likert scale ratings were considered as continuous data for group analyses when appropriate (Harpe, 2015; Norman, 2010). Unless expressly noted otherwise, histograms (mean of differences) and the visual inspection of normal Q-Q plots (Mishra et al., 2019) revealed a normal distribution of all data sets. The Shapiro-Wilk test indicated that the variables were not normally distributed ($p < 0.05$), however, this is expected as each variable represents one independent paired comparison. I included an example of statistical assumptions testing for Pearson's correlation coefficient is provided for group SCS – group ECS comparisons for raw proximity data, MDS Euclidean distances, and PFN graph-theoretic distances. As subsequent analyses (MDS and PFN) were mathematical derivations of proximity data, all Pearson's correlation coefficient assumptions were deemed to have been met if these assumptions were met for the raw proximity data. Assumptions for multiple regression are discussed independently for RQ3.

Cognitive Structure

The intent of examining a relationship between ECS and SCS is to understand better if the student's cognitive structure "agrees" with that of the expert, to what degree it agrees, the strength of association between them, and whether this level of agreement impacts future academic success. In retrospect, after further evaluation of the literature, the use of the term “relationship” was incorrect. The intent was to examine if students align their cognitive structure with experts (concurrent validity) and does the level of agreement relate to predictive validity via a criterion standard reflecting academic performance. In this study, the construct of cognitive structure was operationalized as “a hypothetical construct referring to the organization of the relationships of concepts in long-term memory” (Shavelson, 1972, p. 226-227) via pairwise comparisons representing perceptual concept organization. The issues of construct validity and the appropriate use of statistical analyses were discussed in the operationalization of constructs in Chapter 3.

The raw proximity (similarity) data compiled from the paired comparisons of the survey instrument served as the basis for all statistical analyses with both data modeling strategies. The paired comparisons component of both expert and student surveys was identical to promote the fidelity of implementation. I assumed that as physiotherapy students had completed prerequisite courses in anatomy before program admission, they would clearly understand and be familiar with the items noted. The raw proximity data were converted from similarity to dissimilarity ratings by subtracting each value from a constant value. The defined constant was a value of eight as seven on the relatedness scale indicated “most similar” but not “identical.” The raw proximity data were also analyzed as it represented the direct perceptions of the participant regarding the paired comparisons.

Three levels of agreement analysis were used. First, I presented a visual comparison of student and expert cognitive structures using MDS and PFN data visualizations. Second, I based a quantitative descriptive comparison of student and expert cognitive structures on the MDS configuration properties (stress-1, TCC, and R^2) and PFN network properties (degree, eccentricity, number of links, and coherence). Third, I performed a quantitative statistical analysis of student and expert cognitive structures using the proximity data, MDS Euclidean distances, and PFN graph-theoretic distances in terms of agreement based on reliability (interrater reliability calculated via Krippendorff's alpha coefficient), accuracy (level of agreement based on the root mean squared deviation), and association (the strength of linear relationship calculated via Pearson's correlation coefficient).

Measures derived from these analysis levels served to examine the relationship between student cognitive structure and expert cognitive structure in RQ2. If there was a relationship established between SCS and ECS, then it might be conceived that a higher level of agreement could translate into improved academic performance. Potential predictor variables for RQ3 were subsequently derived from these measures and used in conjunction with prior knowledge.

There are several important considerations regarding the use of raw data to derive or represent cognitive structure. The "meaningful aggregation" (Segalowitz et al., 2016) of group data is a critical consideration for both data modeling strategies as they involve mathematical manipulation of the raw data to generate spatial and network representations. Group data in MDS can be examined via the aggregation of multiple matrices within the MDS solution (see Gonzalvo et al., 1994) or as mean (average) values used to generate the MDS solution (see Goldsmith et al., 1991). Group data in PFN can be examined as mean or median values generated from single or multiple matrices. Segalowitz et al. (2016) noted that the mean data might, in fact, not be truly representative of any of the input data and thus may create an artifact or skewed representation of the data (Janska & Clark, 2010). Mean data are generally vulnerable to outliers in the raw data; however, mean MDS configurations (with multiple matrices) may limit the impact of any one matrix that may be an outlier from the others; it is taken into consideration but allows the MDS configuration to retain representative and reliable results (Janska & Clark, 2010). Previous research has used average/mean values to represent expert cognitive structure via MDS and PFN. For this research study, both approaches were used where appropriate and noted accordingly. I included an exploratory analysis of the potential differences in data aggregation and their impact on the MDS solution.

MDS was used to examine the data at the global/spatial level, whereas PFN was used to examine the data at the local/network level. I calculated the MDS configurations using the Proxscal 1.0 algorithm in SPSS (Busing et al., 1997) and the PFN properties using the Pathfinder Network Java application (Schvaneveldt, n.d.). Both data modeling strategies required specific parameters before analysis: dimensionality in MDS and the q- and r-parameters in PFN. I have described these within the context of the compiled research data.

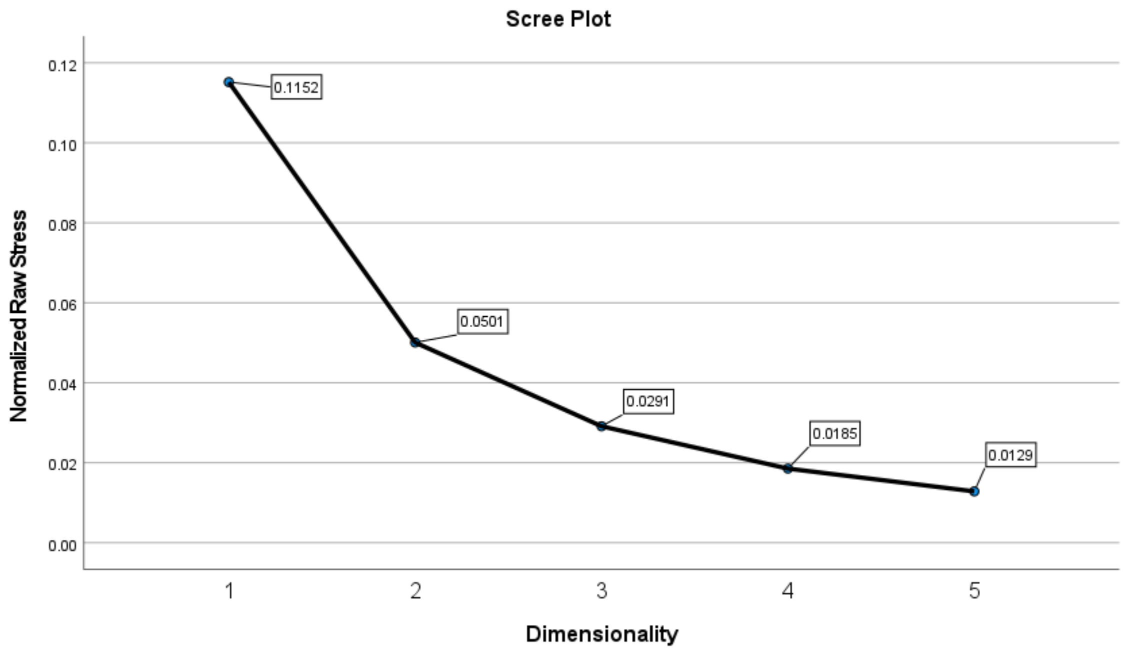
MDS Dimensionality

Best practices for MDS, as proposed by Borg et al. (2018), were used. The authors proposed using the PROXSCAL algorithm with Torgerson scaling for the initial configuration, ordinal proximity transformations, primary approach to ties (untie tied observations), a stress convergence of 0.0000001, minimum stress of 0.0000001, and maximum iterations of 1000. CMDS uses one matrix and generates an MDS solution, whereas RMDS uses multiple matrices to generate an MDS solution based on the aggregation of the data. In this study, RMDS used Identity model scaling to establish the MDS configuration. R^2 in the context of MDS is the coefficient of determination derived from transformed proximities and distances. This value represents the proportion of variance accounted for by the MDS solution.

Dimensionality in MDS is a critical factor in determining how the data are represented spatially. In this study, dimensionality was defined by the expert data and was subsequently applied to the student data to maintain consistency. The initial examination of the eight expert data sets utilized RMDS (CMDS with multiple matrices) with the PROXSCAL algorithm and the Identity scaling model in SPSS. The eight matrices established a common group space to determine the appropriate dimensionality of the solution initially. Several factors are important in determining the appropriate dimensionality of the MDS solution. Davison and Sireci (2000) and Mair et al. (2016) advocated for a multi-factorial approach to goodness of fit that included a scree plot (Figure 6), the residual plot of disparities and transformed proximities, and general interpretability of the solution.

Figure 6

Scree Plot for Assessment of MDS Dimensionality



Note. The scree plot assists in determining dimensionality based on the "elbow" or inflection point of the normalized raw stress values.

The "elbow" in the scree plot appears to be at a dimensionality of two, although this is not distinct; thus, I also considered the R^2 values. Kruskal and Wish (1978) noted that the maximum number of dimensions (D) should be a factor of the number of items/stimuli (I) with $I - 1 \geq 4D$; this established an upper limit of four dimensions to be considered (Table 6).

Table 6

Dimensionality of MDS Solution

Dimension	Stress-1	R^2
1	0.3518	0.627
2	0.2238	0.704
3	0.1708	0.719
4	0.1367	0.721

Kruskal (1964) considered a value of 0.2 as "poor" and 0.1 as "fair." However, although the use of 3 dimensions would improve the stress-1 value, it would not significantly impact R^2 while making the interpretability of the solution more difficult. A two-dimensional solution was selected based on the data in conjunction with the potential interpretability of the findings in the context of the research questions and the potential practicality of use in an educational environment (Davison & Sireci, 2000). This dimensionality was used throughout both expert and student analyses. Decisions regarding the selection of the MDS scaling model (data aggregation to be used and in what context it was used) were made after several preliminary exploratory analyses.

PFN Parameters

PFN can use both individual and multiple matrices (mean or median values) to generate a Pathfinder network which can then be examined and compared to other networks. Two parameters are required to generate a Pathfinder network: the q- and r-parameters. The q-parameter is the number of links in the generated network and is a value between 2 and n-1, where n is the number of nodes (in this study, 20). The r-parameter defines how distances are calculated, using values from 1 to infinity. In the context of this study, to generate a network with ordinal data and the minimum number of links, I set the q-parameter to 19, calculated based on the 20 content items/nodes (20 nodes – 1 = 19). I set the r-parameter to infinity per best practices described by Schvaneveldt (1990), which was consistent with previous seminal research.

Cognitive Structure Procedures

Each sample (ECS, SCS) was examined as a group, and relevant subgroups were identified. Expert subgroups included nonclinical/domain (ECSD, $n = 3$), clinical/instructor (ECSI, $n = 4$), and individual instructor by cohort (ECSIC). Student subgroups were arranged by cohort. I performed within-group and between-group comparisons of cognitive structure, using a similar process for both ECS and SCS using MDS and PFN data modeling strategies to address RQ2. The group data visualizations are displayed. I presented individual SCS in comparison to ECSD, ECSI, and the cohort instructor ECSIC. Agreement analysis was performed between ECS and SCS proximity data, MDS-derived parameters, and PFN-derived parameters. Key MDS- and PFN-derived parameters, interrater reliability, accuracy, and correlations were subsequently used as predictor variables for RQ3 along with prior knowledge variables (admission cumulative GPA, admission core sciences GPA).

Preliminary Exploratory Analysis

Several preliminary exploratory analyses were performed to address relevant statistical issues related to the research questions. These focused on the impact of the MDS scaling model and data aggregation methods given the Likert scale data.

Impact of MDS Scaling Model

RMDS provides an aggregated visual representation of the structure of the stimuli based on multiple matrices. Items have unique coordinates (Euclidean distances) in a configuration; however, their orientation is not fixed and can be transformed via rotation, reflection, and translation. WMDS provides a visual representation of the structure of the stimuli but with two key additions: an individual space and weights and unique coordinates in a fixed orientation of the dimensions/axes. Previous research by Gonzalvo et al. (1994) reported their findings based on the INDSCAL scaling model. Both RMDS and WMDS are known to produce similar group spaces; however, these differences in cognitive structure representation are unknown. To assess potential differences between RMDS and WMDS in assessing group and individual differences, WMDS (multiple matrices, PROXSCAL algorithm, weighted Euclidean scaling model, two dimensions) was used to examine the group spaces for the ECSD, ECSC, and ECSI subgroups. I compared the WMDS results to those attained via RMDS. The results of both RMDS and WMDS are summarized in Appendix D (Table D1).

The results indicated a consistency between group MDS configurations with an overall improvement in R^2 based on the scaling model used. The R^2 values indicated a greater percentage of variance accounted for by the weighted model than the replicated (classical) model across all groups, though these differences were not large. There was also a variation in the orientation of the axes and dimensions. The differentiation between RMDS and WMDS using the same data set has not been reported in previous studies examining cognitive structure. The use of either MDS scaling model provides comparable results consistent with the context of analysis (between-group and within-group), providing support for the context-specific use of both scaling models within the study.

Aggregation Strategy and MDS Configuration

Previous seminal research studies, including Goldsmith et al. (1991) and Gonzalvo et al. (1994), reported their findings based on the mean values of expert groups. As noted in Chapter 3, raw proximity data aggregation has the potential to not fully represent individual data sources within the context of the overall MDS configuration. CMDS, using one matrix of mean values instead of the multiple matrices of raw values used by RMDS, was used to compare the mean ECSD, mean ECSC, and mean ECSI groups. These are summarized in Appendix D (Table D2).

The use of median values may be a more statistically accurate derivation from the initial ordinal data. However, both the mean and median values produce results that overestimated all values compared to the aggregated data derived from multiple matrices. This differentiation between RMDS (multiple matrices of raw data) and CMDS (mean values within one matrix) has not been reported in previous studies of cognitive structure. This is an important consideration during the analysis of expert and student cognitive structures to provide relevant context. Aggregate data (RMDS) was subsequently used for the MDS analyses as appropriate. It is of note that PFN generates mean values for links and distances in determining aggregate data for analysis.

Survey Instrument Development

The online survey consisted of paired comparisons of 20 items (anatomical concepts and structures) related to the shoulder complex. Physiotherapists with ten or more years of musculoskeletal clinical practice defined the items used in the online survey. The physiotherapist participants were responsible for rank-ordering the 40 items representing anatomical concepts and structures, with one being the most relevant to musculoskeletal practice and 40 being the least relevant (Appendix A). The survey data were compiled by ranking the sum of individual rank values for each item. The 20-item list that the physiotherapists perceived to be the most relevant to musculoskeletal physiotherapist practice would then form the instrument used for paired comparisons to define the cognitive structure of both experts and students. The top 20 items based on ranking are summarized in Appendix E.

The online survey containing paired comparisons was designed as a series of grids such that five paired comparisons were displayed per grid per page. The grid format was used to promote the speed of completing the survey, as it contained a total of 190 paired comparisons. Chapter 3 noted the specifics of the survey instrument development.

Interrater reliability amongst the 12 physiotherapist raters is reported in Table 7. Krippendorff’s alpha was used to examine multiple raters with ordinal data. Interrater reliability is noted for three item groups after rank ordering: 40 items (total), top 20 items, and top 10 items. The results indicated poor interrater reliability across the 12 physiotherapist participants and all groups of rank-ordered items. It is important to note that all 12 physiotherapist raters had ten or more years of experience in musculoskeletal clinical practice.

Table 7

Physiotherapist Interrater Agreement

	α	95% CI
40 items	.36	.33 – .39
20 items	.33	.27 – .38
10 items	.32	.25 – .39

Note. Sample of 12 physiotherapists.

As noted in Chapter 3, the small sample size limited the conclusions. A greater sample size is necessary to further examine the impact of the interrater agreement in this research context and was not within the scope of the current study.

Research Question 1

RQ1: Is there a meaningful change over time in the quantitative representation of student cognitive structure?

Due to the unforeseen issues related to the timing of IRB approval, a pretest survey could not be completed by students before the start of the unit module in week 6. In order to complete a pretest data collection as initially planned, the study would have been delayed a minimum of 15 weeks. The second option was to remove RQ1, eliminating the need for a longitudinal pretest-posttest design and the associated data analysis. The premise of “meaningful learning” (reflected in the pretest and posttest measures) was supplemental to better understanding criterion validity through the exploratory nature of RQ2 and RQ3. I chose to remove RQ1 based upon this rationale.

Upon further consultation with and approval by the dissertation committee, RQ1 was removed from the scope of the current study.

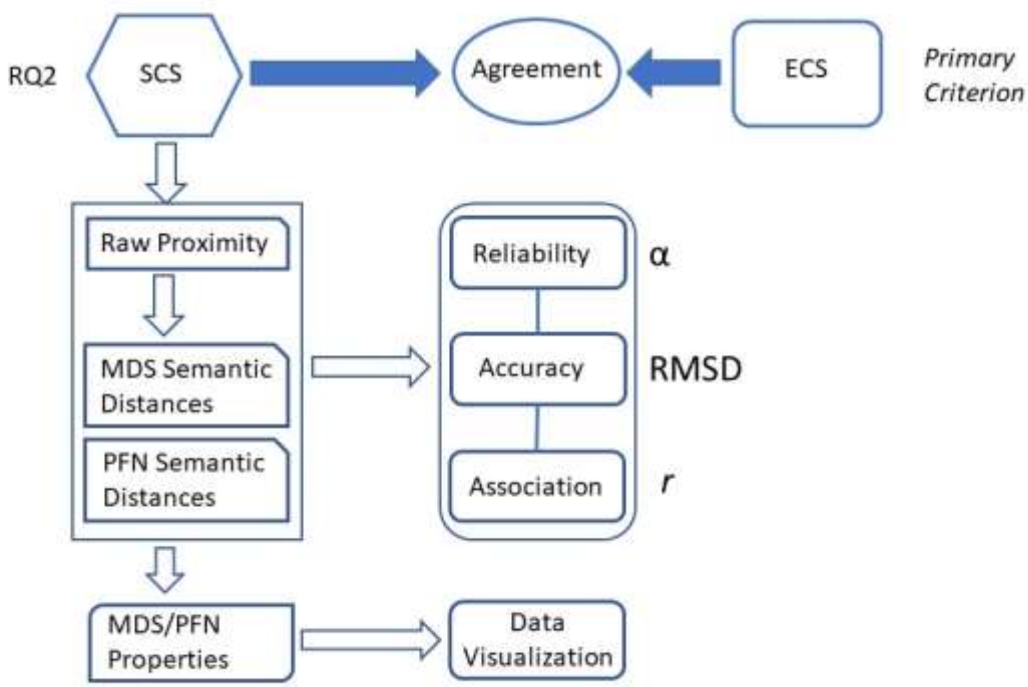
Research Question 2

RQ2: Is there a relationship between student cognitive structure and expert cognitive structure while controlling for prior knowledge?

Figure 7 provides a detailed overview of the approach used to examine RQ2.

Figure 7

RQ2 Overview



Expert cognitive structure (ECS) served as the criterion standard used for RQ2. I used the two data modeling strategies to derive spatial (MDS) and network (PFN) representations from the proximity data. The data visualizations and the derived statistical parameters representing student cognitive structure (SCS) were compared to those of the expert cognitive structure (ECS). This included an agreement analysis of proximity data and both Euclidean and graph-theoretic semantic distances examining reliability (Krippendorf's alpha coefficient with 95% confidence intervals), accuracy (root mean square deviation with bivariate and Bland-Altman plots used where appropriate to display the data visually), and association (Pearson's correlation coefficient). I completed these analyses for three levels of comparisons: group SCS (n = 31) with group ECS (n = 4), individual SCS with group ECS, and individual SCS with cohort instructor ECS. As the primary goal of RQ2 was to establish potential relationships and agreement between ECS and SCS visually and quantitatively (via derived parameters and agreement analysis of MDS and PFN semantic distances), controlling for prior knowledge was not an appropriate inclusion in this research question. However, prior knowledge was considered in the context of RQ3.

Proximity Data

The survey instrument generated 190 paired comparisons of proximity data for each participant; these data were then converted to dissimilarity data (subtracting from a constant value) and subsequently used for both MDS and PFN data modeling strategies that follow.

Group SCS and Group ECS

Results for the agreement analysis between group SCS and group ECS raw proximity data are summarized in Table 8.

Table 8

Group SCS–ECS Agreement: Proximity Data

	Reliability	Accuracy	Association
SCS	α [95%CI]	*RMSD* (units)	*r*
ECSI	.75 [.70,.79]	0.9	0.82**
ECSD	.59 [.58,.66]	1.1	0.66**

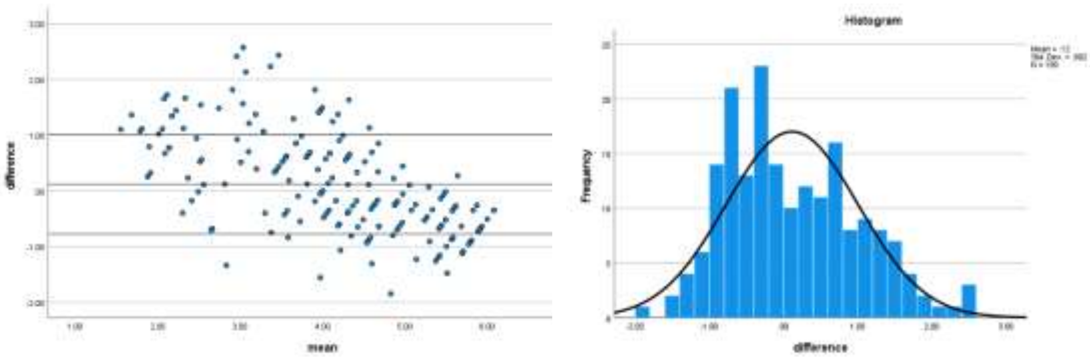
Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, r = Pearson's correlation coefficient. SCS (n = 31), ECSI (n = 4), ECSD (n = 3).

* $p < 0.05$, ** $p < 0.01$

Interrater reliability between student and instructor (α = 0.75) was improved compared to student and domain expert (α = 0.59). In comparison, there was good interrater reliability between ECSI and ECSD expert subgroups (α = 0.64). Accuracy (via RMSD) between SCS and ECSI indicated that SCS proximity values for paired comparisons were within a +/- 0.9 points range on the perceived relatedness rating scale. There were greater differences between SCS and ECSD for paired comparisons on the perceived relatedness rating scale. The data are represented visually by the Bland-Altman plot and histogram of differences (Figures 8 and 9). Students tended to over-rate items with low expert ratings (more dissimilar than experts) and under-rate those with higher expert ratings (less dissimilar than experts).

Figure 8

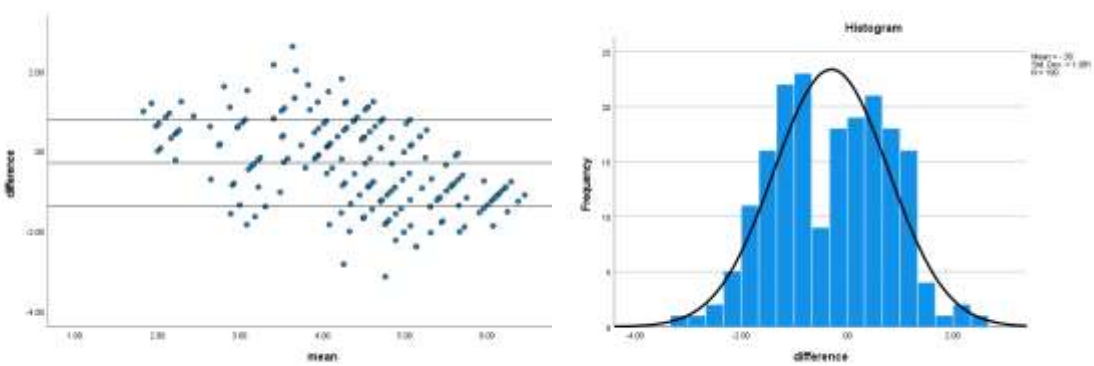
SCS and ECSI Differences: Proximity Data



Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Figure 9

SCS and ECSD Differences: Proximity Data

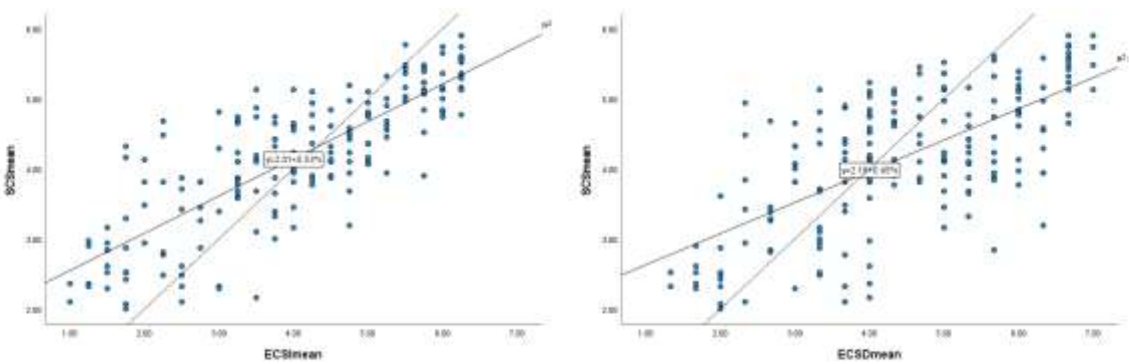


Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Scatterplots (Figure 10) indicated a positive linear relationship between SCS and both ECSI and ECSD. The Shapiro-Wilk test indicated that the variables were not normally distributed ($p < 0.05$), however, this would be expected as each pair of values represents one independent paired comparison. Pearson's correlation test is robust and not adversely impacted by non-normal distributions, especially given the context mentioned above.

Figure 10

Scatterplots of SCS, ECSI, and ECSD: Proximity Data



a. SCS and ECSI

b. SCS and ECSD

Note. Dotted lines indicate line y = x in which student and expert would be in full agreement in terms of perceived relatedness and dissimilarity.

There was a statistically significant strong positive correlation between groups, $r(188) = 0.66 – 0.82$, $p < 0.01$. This indicated a large strength of linear association. SCS had a higher correlation with ECSI than it did with ECSD, which may reflect the importance of a clinical degree in teaching gross anatomy to physiotherapy students.

Based on the results of the subgroup analysis and their practical application educationally, the instructor (ECSI) subgroup was defined as the primary criterion standard for all further SCS comparisons.

Individual SCS and ECS

I compared each individual SCS to both group ECSI and individual ECSIC (ECS for their cohort instructor). Results for the agreement analysis between individual SCS and ECSI proximity data are summarized in Table 9.

Table 9

Individual SCS–ECS Agreement: Proximity Data

	Reliability	Accuracy	Association
SCS	α (*SD*)	*RMSD* (units)	*r*
ECSI	.37 (.20)	1.8	.46**
Range	-.03 - .66	1.2 – 2.4	.12 - .79**
ECSIC	.29 (.23)	2.3 (0.4)	.40
Range	-.25-.61	1.6-3.3	.10-.62**

Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, r = Pearson's correlation coefficient. SCS (n = 31), ECSI (n = 4), ECSD (n = 3), ECSIC individual cohort instructor

* $p < 0.05$, ** $p < 0.01$

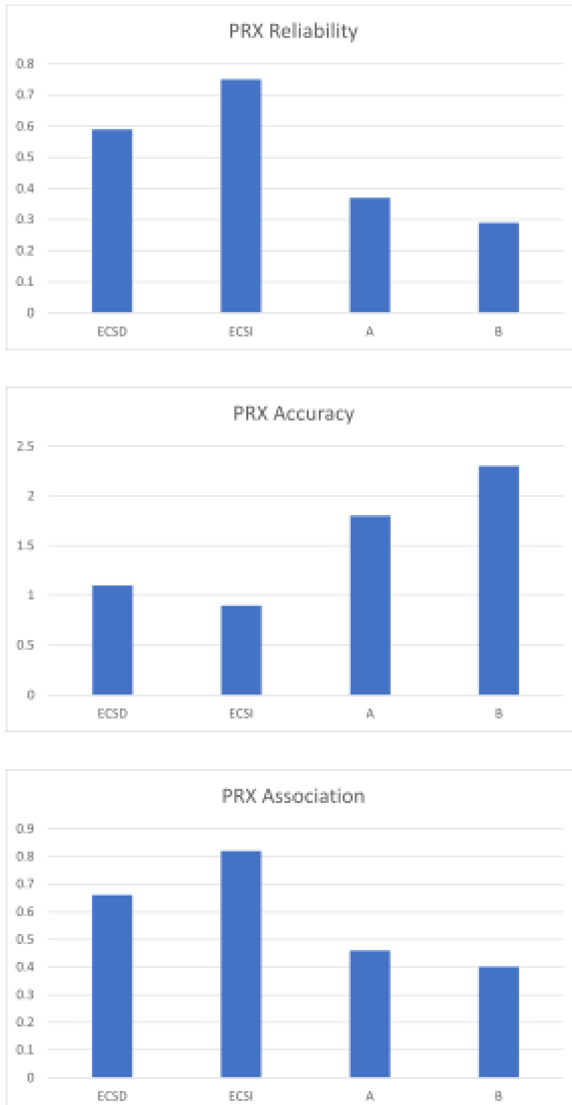
The mean interrater reliability between all individual students and group ECS (α = 0.37) indicated a fair level of interrater reliability. This value decreased when comparing individual SCS and cohort instructor ECSIC (α = 0.29). It is notable that only four students (13%) had an overall increase in interrater reliability with their cohort instructor than with the group ECS. There is a high degree of variability in accuracy among students with mean ratings having a range of +/- 1.8 points on the perceived relatedness rating scale for any given paired comparison. This may reflect perceptual differences or chance agreement. However, nine students (36%) displayed an overall improved level of agreement with the cohort instructor ECSIC compared to the group ECSI. There is also a high degree of variability in association (ECSI: $r(188) = 0.12 - 0.79$, $p < 0.01$; ECSIC: $r(188) = 0.10 – 0.62$).

Proximity Data: Overview

A summary of the agreement analysis across all comparison levels (group SCS and group ECS, individual SCS and group ECSI, and individual SCS and individual ECSIC) is presented in Figures 11 and 12. SCS raw proximity data was aligned with instructors more so than with domain experts. On a more granular level, students did not appear to display a consistently higher relationship with their specific instructor than the group ECSI with a trend toward disparity between cohorts. However, this observation was limited by the small sample size.

Figure 11

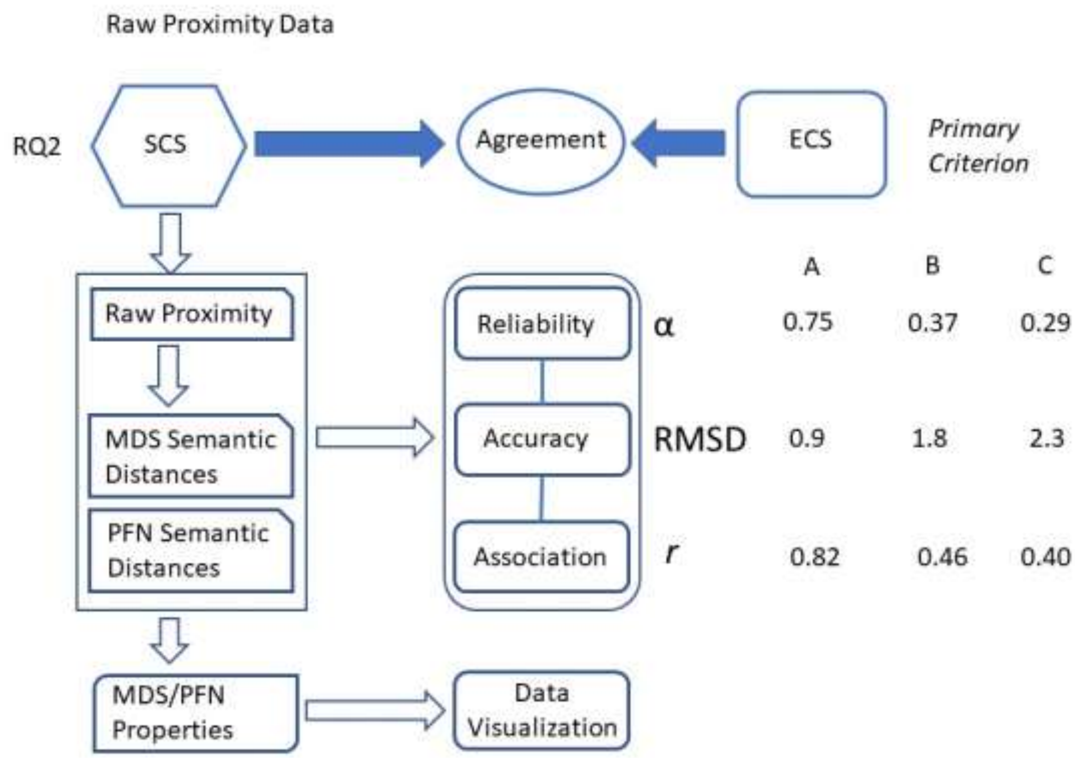
Summary of Agreement Analysis: Proximity Data



Note. Each bar represents a specific level of comparison. ECSD and ECSI are compared to group SCS. A = individual SCS and group ECSI, B = individual SCS and individual cohort instructor ECSIC. At the level of the cohort instructor, accuracy and association increased in 36% of students.

Figure 12

RQ2 Summary: Proximity Data



Note. A = Group SCS–Group ECSI; B = Individual SCS–Group ECSI; C = Individual SCS–Individual ECSIC

Data Modeling: Multidimensional Scaling

I used the proximity data to generate MDS configurations and all derived parameters (stress-1, TCC, R^2, and Euclidean semantic distances). I made comparisons between groups in terms of MDS configuration, MDS-derived parameters, and agreement analysis (reliability, accuracy, and association).

Group SCS and Group ECS

The initial analysis considered the total expert data set (ECST, n = 8) with relevant subgroupings subsequently examined to see if there were changes in the goodness of fit of the MDS configuration. I performed all analyses using RMDS, PROXSCAL algorithm, and Identity model scaling. Refer to Figure 13 for a visual representation of the MDS configurations and Table 10 for all relevant derived parameters.

Table 10

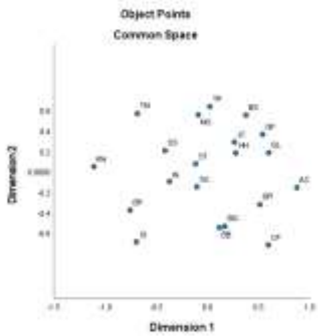
Group RMDS Configuration Properties

Group	Stress-1	TCC	R^2
ECST	.224	.98	.70
ECSD	.222	.98	.73
ECSI	.196	.98	.78
SCS	.265	.96	.53

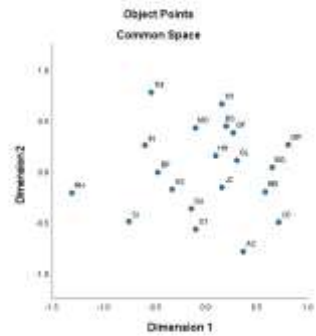
Note: RMDS with multiple matrices, PROXSCAL algorithm, Identity scaling model, two dimensions.

Figure 13

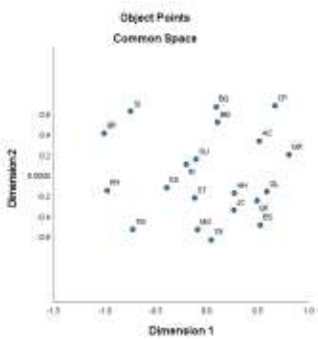
Group CMDS Data Visualizations



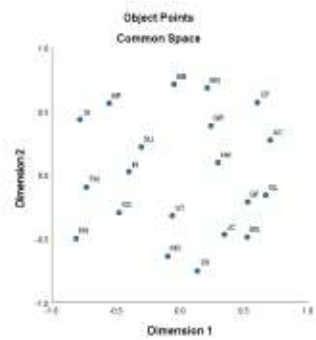
a



b



c



d

Note. All CMDS configurations can be reflected, rotated, and translated without a change in Euclidean distances between items in the configuration space. a = ECST, b = ECSD, c = ECSI, d = SCST. The common space for SCS has a greater spatial range overall.

MDS configurations based on raw proximity data have the risk of stress-related to a random error within the data; in other words, raw proximities may be a function of random choice instead of perceptual differentiation. Sturrock and Rocha (2000) calculated random stress norms noting that a 1% chance of random arrangement for 20 objects in two dimensions would have a stress of 0.279. This indicates that at a level of $p < 0.01$, the null hypothesis is rejected, and the data arrangement is not considered random. In all MDS configurations reported, reported stress values were well below this random stress norm, indicating that the data were not random and reflected perceptual differences. A stress-1 value of 0.2238 indicated a poor overall fit of the configuration for the total expert group but several other factors were considered. TCC was high, indicating good congruence, and $R^2 = 0.70$, indicating a moderate fit with 70% of the variance accounted for by the configuration. The R^2 values indicated that there was a difference between subgroups. A higher R^2 was noted for the ECSC (clinical) and ECSI (the instructor subgroup within the clinical group) MDS configurations. The derived MDS configuration accounted for a greater percentage of the variance within the data indicating a greater internal consistency or coherence of the MDS configuration (McGaghie, McCrimmon, et al., 2000). This suggests that the MDS configuration of domain experts without a clinical degree may differ from the MDS configuration of those with clinical degrees teaching anatomy within the DPT curriculum.

I derived the Euclidean distances from the individual MDS configurations. MDS configurations may vary via reflection, rotation, and translation (using CMDS), but Euclidean distances remain consistent within the context of the scaling model used.

However, the interpretability of MDS distances in the study context was purely referential and contextual; there was not a defined minimal interpretable difference of importance. As the raw proximity data were used within similar subgroups, Pearson's correlation coefficient assumptions were considered to have been met.

Results for the agreement analysis between group SCS and group ECS MDS Euclidean distances are summarized in Table 11.

Table 11

Group SCS–ECS Agreement: MDS Euclidean Distances

	Reliability	Accuracy	Association
SCS	α [95%CI]	*RMSD* (units)	*r*
ECSI	.83 [.77,.87]	0.231	.83**
ECSD	.53 [.45,.60]	0.384	.53**

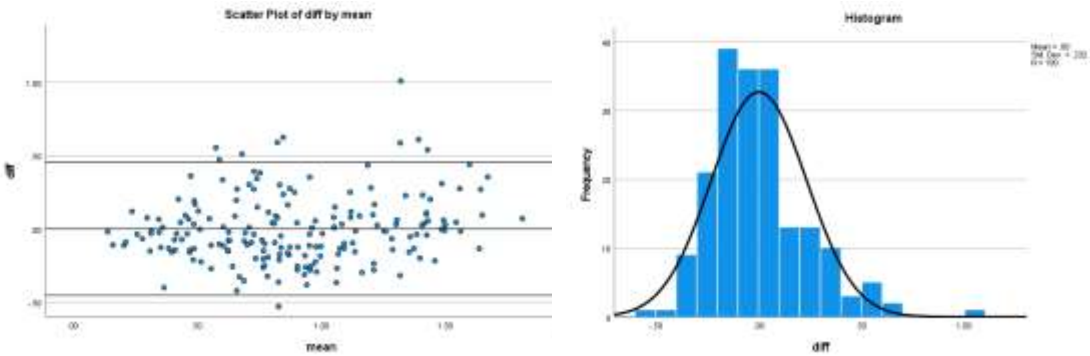
Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, r = Pearson's correlation coefficient. SCS (n = 31), ECSI (n = 4), ECSD (n = 3)

* $p < 0.05$, ** $p < 0.01$

Interrater reliability between group SCS and ECSI (α = .83) was again greater than that between group SCS and ECSD (α = .53). These values indicated a greater disparity between SCS and ECSD MDS distances. There was moderate interrater reliability between ECSI and ECSD (α = .63). The data are represented visually by the Bland-Altman plot and histogram of differences for SCS and ECSI (Figure 14) and SCS and ECSD (Figure 15).

Figure 14

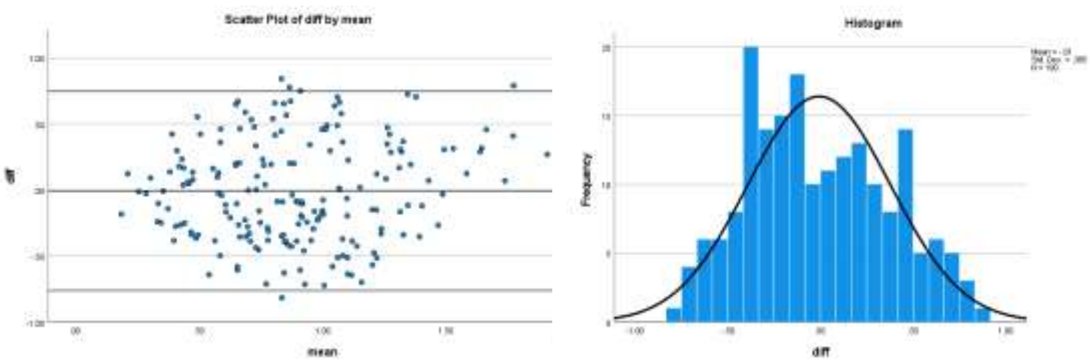
SCS and ECSI Differences: MDS Euclidean Distances



Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Figure 15

SCS and ECSD Differences: MDS Euclidean Distances

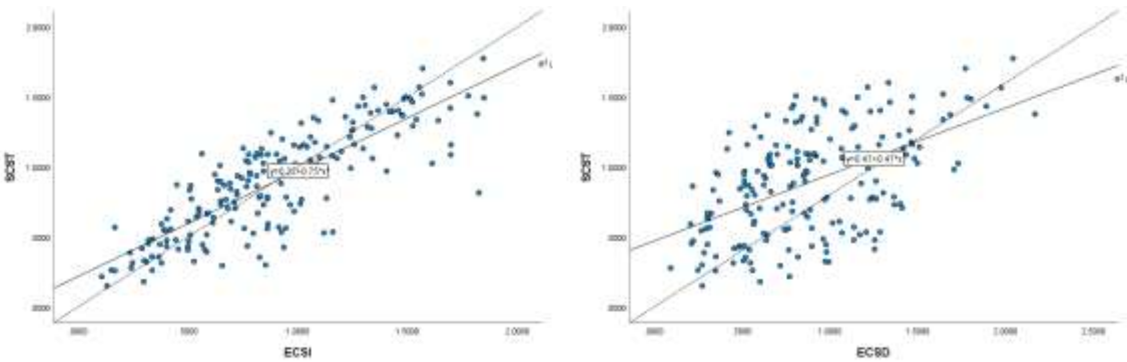


Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Scatterplots (Figure 16) indicated a positive linear relationship between SCS and both ECSI and ECSD. The Shapiro-Wilk test indicated that the variables were not normally distributed ($p < 0.05$), however, this would be expected as each pair of values represents one independent paired comparison. Pearson's correlation test is robust and not adversely impacted by non-normal distributions, especially given the context mentioned above.

Figure 16

Scatterplots of SCS, ECSI, and ECSD: MDS Euclidean Distances



a. SCS and ECSI   b. SCS and ECSD

Note. Dotted lines indicate line y = x in which student and expert would be in full agreement in terms of perceived relatedness and dissimilarity.

There was a statistically significant, strong positive correlation between groups, $r(188) = .53 – .83$, $p < 0.01$). This indicated a large strength of association between groups. SCS had a higher correlation with ECSI than it did with ECSD; once again, this difference may reflect the importance of a clinical degree in teaching gross anatomy in a clinical context.

Individual SCS and ECS

I calculated the individual MDS configurations to derive stress-1, TCC, and R^2 for each participant. These are summarized as mean values in Table 12. Individual SCS displayed a fair goodness of fit and high R^2, reflecting an internal consistency within the individual. I have included all individual MDS data visualizations in Appendix F.

Table 12

Individual RMDS Configuration Properties

	Stress-1	TCC	R^2
Group	*M* (*SD*)	*M* (*SD*)	*M*
SCS [a]	.170 (.033)	.98 (.01)	.82
Range	.103-.243	.97-.99	.61-.94
ECSD [a]	.167 (.019)	.99 (.003)	.84
Range	.145-.180	.98-.99	.82-.89
ECSI [a]	.122 (.039)	.99	.92
Range	.064-.145	.990-.998	.89-.98

Note: Further exploratory analysis regarding RMDS and WMDS appears in a later section.

[a] RMDS with multiple matrices, PROXSCAL algorithm, Identity scaling model, two dimensions.

I compared each individual SCS to both group ECSI and individual ECSIC (ECS for their cohort instructor). I have summarized the results of the agreement analysis between individual SCS and ECS MDS Euclidean distances in Table 13.

Table 13

Individual SCS–ECS Agreement: MDS Euclidean Distances

	Reliability	Accuracy	Association
SCS	α [95%CI]	*RMSD (units)*	*r*
ECSI	.49	0.405	.49
Lower	.01 [-.17,.18]	0.270	.00
Upper	.68 [.61, .75]	0.556	.79**
ECSIC	.41 (.17)	0.456 (0.065)	.41 (.17)
range	.42-.70	.329-.608	.04-.70**

Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, r = Pearson's correlation coefficient. SCS (n = 31), ECSI (n = 4), ECSD (n = 3).

* $p < 0.05$, ** $p < 0.01$

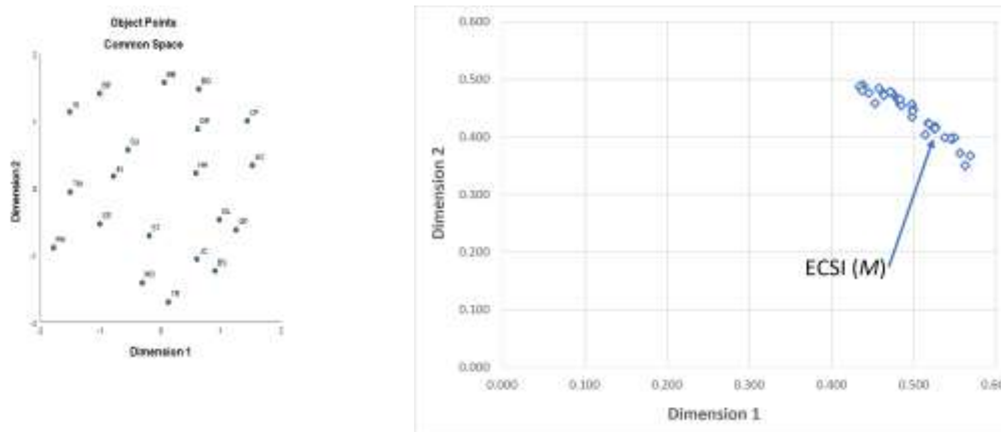
The mean Krippendorff's alpha coefficient of 0.49 indicated a moderate level of interrater reliability between individual students and ECSI. Although the interrater reliability decreased between students and cohort instructors, all four cohorts increased interrater reliability (α = 0.58 to 0.64). These results indicated that there is a trend in improved interrater reliability within an instructor's cohort. Accuracy, within the context of the MDS Euclidean distances, increased with cohort instructors as nine students (29%) displayed an overall decrease in RMSD with the instructor ECS compared to the group ECSI. There is broad variability across student participants. The strength of association also showed a slight overall decrease as compared to the previous group ECS comparison.

A weighted Euclidean scaling model (WMDS) was used to derive common and individual spaces using 32 subjects: 31 students and one instructor. The instructor data represented the mean aggregate values of the four course instructors (ECSI). The WMDS configuration generated individual dimensional weights within the context of the group MDS configuration. MDS-derived parameters for the group data (common space) included stress-1, TCC, and R^2 values. A WMDS configuration of two dimensions had a stress-1 value of 0.2609, indicating a poor fit; this was confirmed by the R^2 of 0.558. In comparison, the cohort-by-cohort MDS configurations had a range of stress-1 values between 0.2150 and 0.2656, indicating a poor fit, but the R^2 values ranged from 0.543 to 0.711, indicating a moderate fit.

The visual representation of the common space and the subject space and dimensional weights for all subjects and instructor (mean) are displayed in Figure 17. Dimension one weights ranged from 0.4350 – 0.5680 (M = 0.496, SD = 0.040) and dimension two weights ranged from 0.350 – 0.489 (M = 0.441, SD = 0.040). The mean instructor dimensional weights were 0.526 and 0.415. It appears that perceptual differences between individuals can be represented spatially. However, the potential importance or relevance of dimensional weights is unknown.

Figure 17

WMDS Group Space and Dimensional Weights



a. Group space   b. Dimensional weights

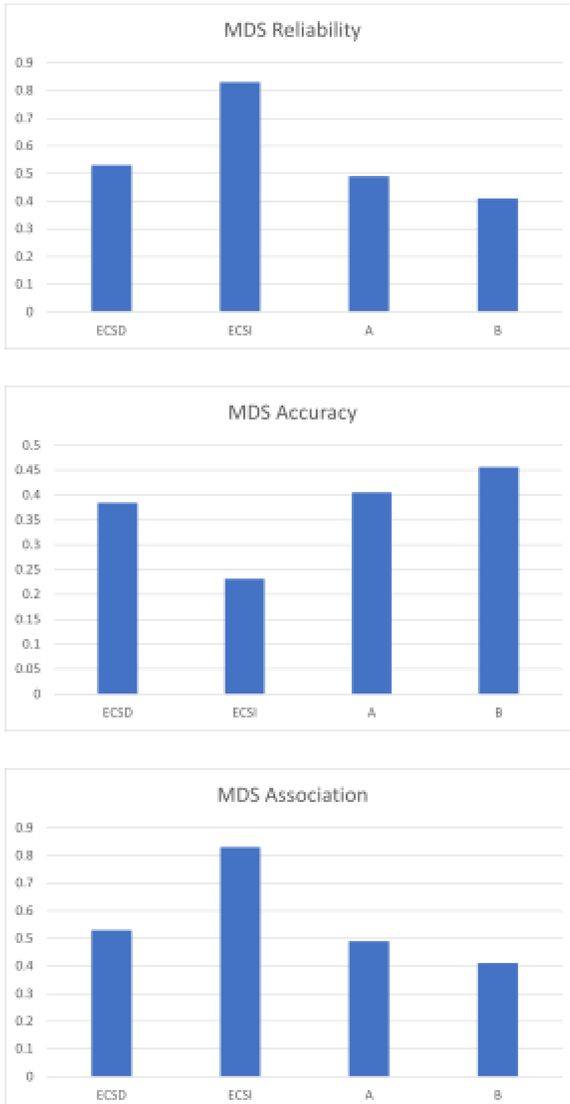
Note. WMDS = Weighted MDS; ECSI (*M*) refers to the mean instructor values.

MDS Overview

A summary of the agreement analysis across all levels of comparison (group SCS and group ECS, individual SCS and group ECSI, and individual SCS and cohort instructor ECSIC) is presented in Figures 18 and 19. SCS was aligned with instructors more so than with domain experts. On a more granular level, students appeared to display a greater relationship with their specific cohort instructor than the group ECSI. However, this trend between cohorts was impacted by the small sample size of each cohort.

Figure 18

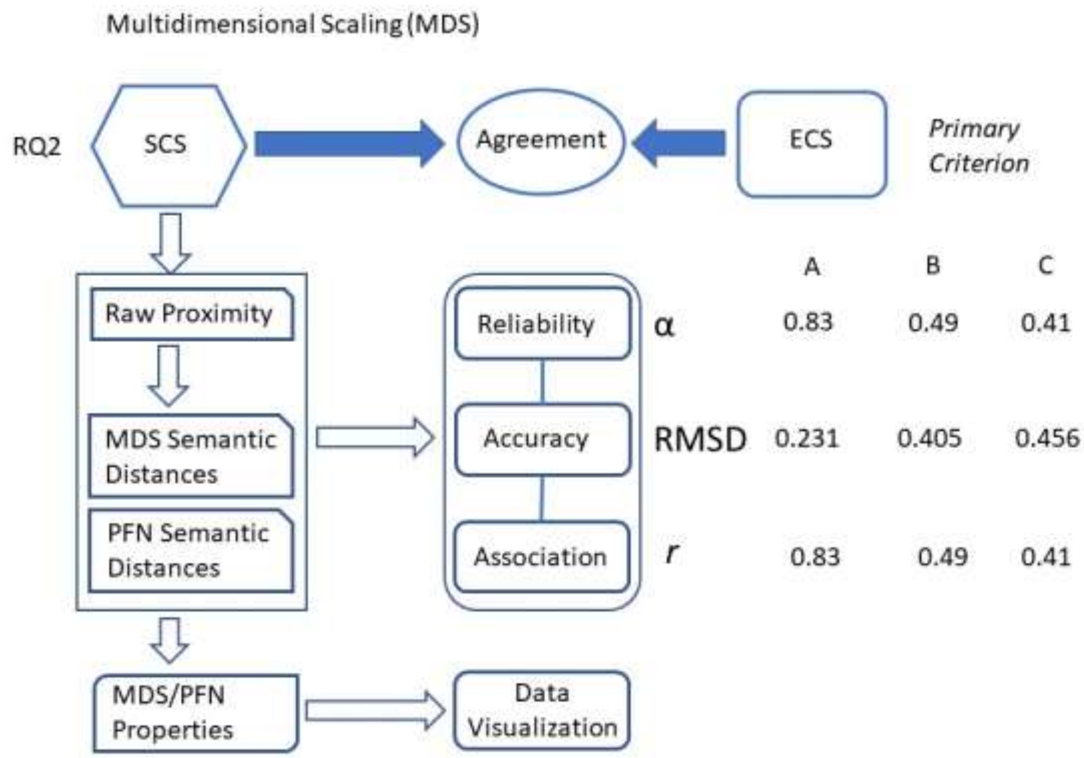
Summary of Agreement Analysis: Multidimensional Scaling



Note. Each bar represents a specific level of comparison. ECSD and ECSI are compared to group SCS. A = individual SCS and group ECSI, B = individual SCS and individual cohort instructor ECSIC. At the level of the cohort instructor, accuracy increased in 29% of students.

Figure 19

RQ2 Summary: Multidimensional Scaling



Note. A = Group SCS–Group ECSI; B = Individual SCS–Group ECSI; C = Individual SCS–Individual ECSIC

Data Modeling: Pathfinder Networks

I used the proximity data to generate PFN representations and all derived parameters. I made comparisons between groups in terms of PFN representation, PFN-derived parameters, and agreement analysis.

Group SCS and Group ECS

The initial analysis used the total expert data set (ECST, *n* = 8) with relevant subgroupings subsequently examined to see if there were differences noted in the network properties. Refer to Figure 20 to visualize the Pathfinder networks and Table 14 for all relevant derived network properties.

Table 14

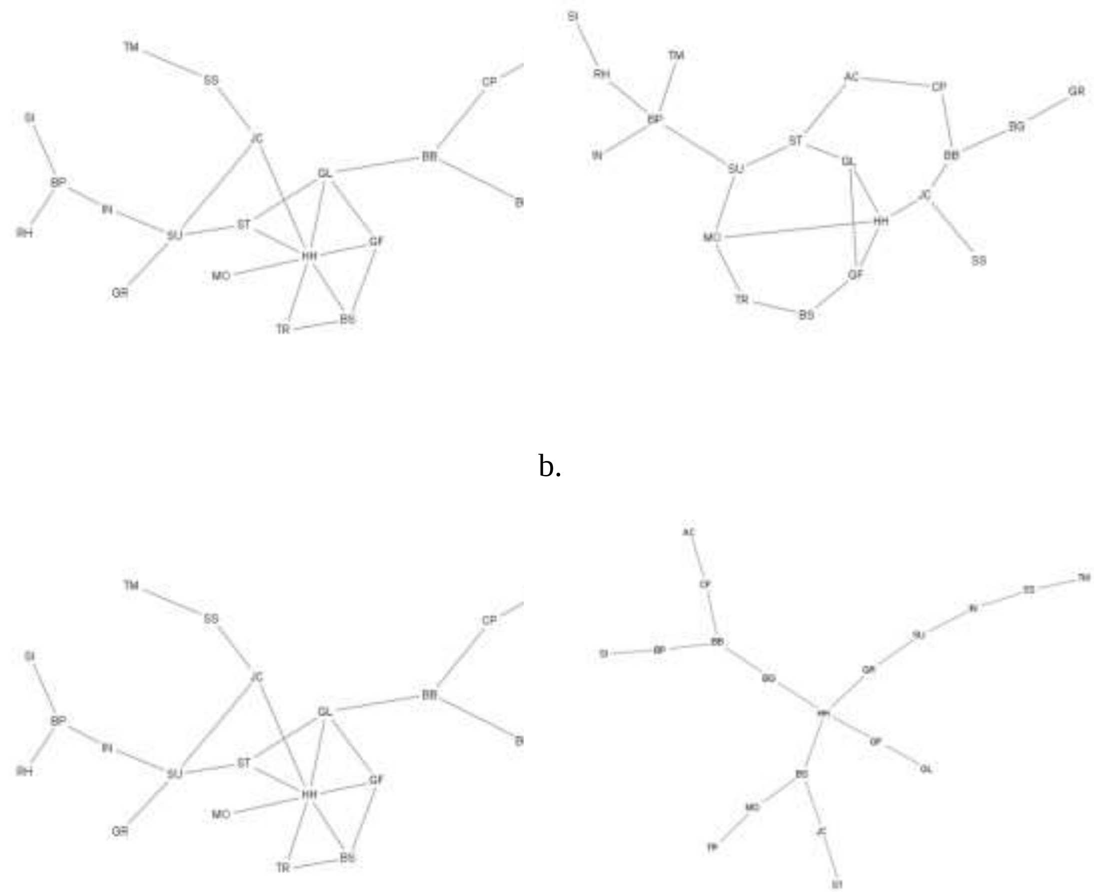
Group Pathfinder Network Properties

Group	Links	Max. Degree	Center	Eccentricity	Coherence
ECST	24	HH	ST	6.2	0.736
ECSD	23	HH-BP	HH-AC-GL-MO-ST	6.6	0.599
ECSI	21	HH	HH-SU	5.5	0.699
SCS	19	HH	GR	8.0	0.796

Note. Maximum degree = item with the greatest number of links to it, eccentricity = the maximum number of links between a node and all other nodes in a network; coherence = the degree to which the original proximity data correlates with the inferred relationships of the network (higher value = greater coherence).

Figure 20

Pathfinder Network Data Visualizations



a. b. c. d.

Note. a = ECST (total), b = ECSD (domain experts), c = ECSI (instructors), d = SCST (student group)

All groups had the same node (HH) with the maximum degree (greatest number of links); however, the center of the network varied from group to group. Students exhibited fewer total links than all the expert groups and displayed greater eccentricity (maximum distance between nodes/items). Coherence is an internal measure that relates to the internal consistency of the data. A high coherence indicates that the original proximities are more consistent with the indirect relationships between the items. Students had similar coherence values compared to experts; however, the data visualizations indicated that their individual networks varied.

Pearson correlation coefficients were calculated for several network properties, including total links, eccentricity, and mean links for each node in the network (corresponding to the 20 items used in the paired comparisons). I presented the results in Table 15. In all three network properties, students have a higher correlation with ECSI than ECSD.

Table 15

Group Correlation of Pathfinder Network Properties

	Degree	Eccentricity	Mean Links
SCS	*r*	*r*	*r*
ECST	0.57**	0.54*	0.59**
ECSD	0.31	-0.27	0.17
ECSI	0.51*	0.51*	0.46*

Note. Degree = the number of links attached to each node; eccentricity = the maximum number of links between a node and all other nodes in a network; mean links = mean number of links per node

Pathfinder networks also produce two other unique derived parameters: common links and similarity. I have presented these parameters in Table 16. SCS has more common links with the instructor ECS and a greater calculated similarity between the Pathfinder networks generated.

Table 16

Group Pathfinder Common Links and Similarity

SCS	Common Links	Common Link %	Similarity
ECSD	6	26.1	0.167
ECSI	10	47.6	0.333

Note. SCS has more common links and greater similarity with ECSI than ECSD.

I derived the graph-theoretic (PFN) distances for each individual network. These were calculated using the r-parameter consistent with the generation of the network (r = infinity). However, the interpretability of PFN distances in the study context is purely referential and contextual; there is not a defined minimal interpretable difference of importance. I compared the PFN distances for each student with those derived from the ECSD and ECSI networks. As proximity data were used within similar subgroups, the assumptions for correlation were assumed to have been met.

Results for the agreement analysis between group SCS and group ECS PFN graph-theoretic distances are summarized in Table 17.

Table 17

Group SCS–ECS Agreement: PFN Graph-Theoretic Distances

	Reliability	Accuracy	Association
SCS	α [95%CI]	*RMSD* (units)	*r*
ECSI	.19 [.02,.36]	2.5	.34**
ECSD	.10 [-.11,.28]	2.7	.15*

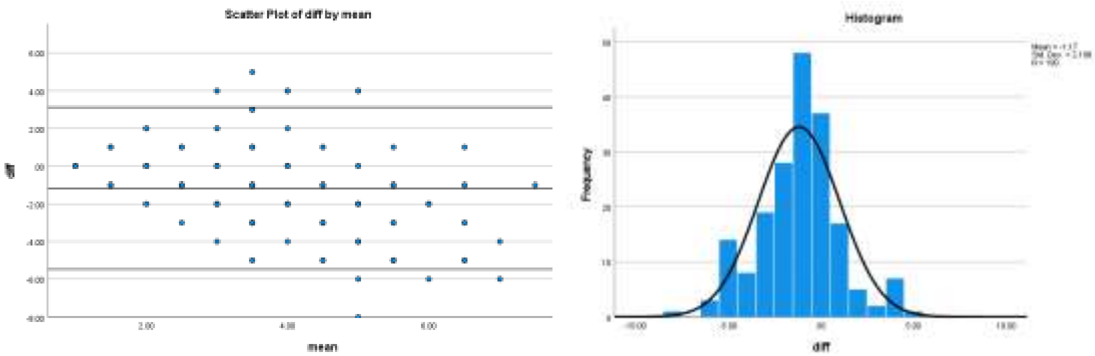
Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, *r* = Pearson's correlation coefficient. SCS (n = 31), ECSI (n = 4), ECSD (n = 3)

* $p < 0.05$, ** $p < 0.01$

There was poor interrater reliability between SCS and ECSI PFN distances (α = 0.19) and between SCS and ECSD PFN distances (α = 0.10). As noted in previous analyses, students had greater interrater reliability with instructors than with domain experts. RMSD served purely as a contextual reference. However, greater differences were noted for ECSD than ECSI. The data are represented visually by the Bland-Altman plot and histogram of differences for SCS and ECSI (Figure 21) and SCS and ECSD (Figure 22).

Figure 21

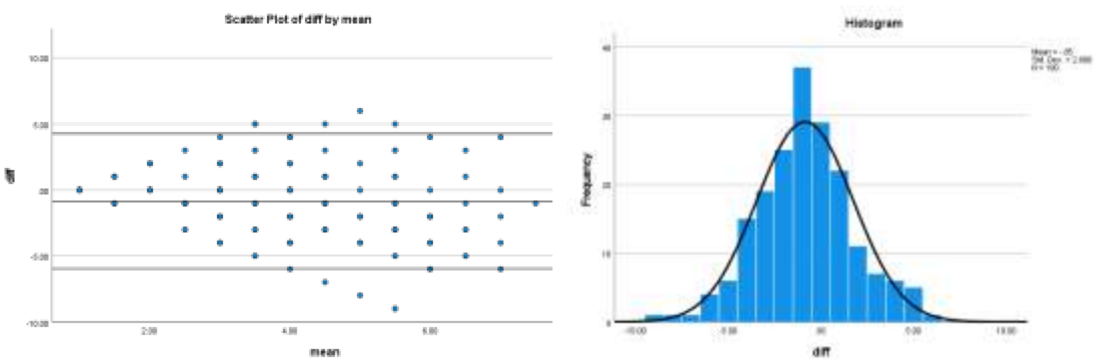
SCS and ECSI Differences: PFN Graph-Theoretic Distances



Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Figure 22

SCS and ECSD Differences: PFN Graph-Theoretic Distances

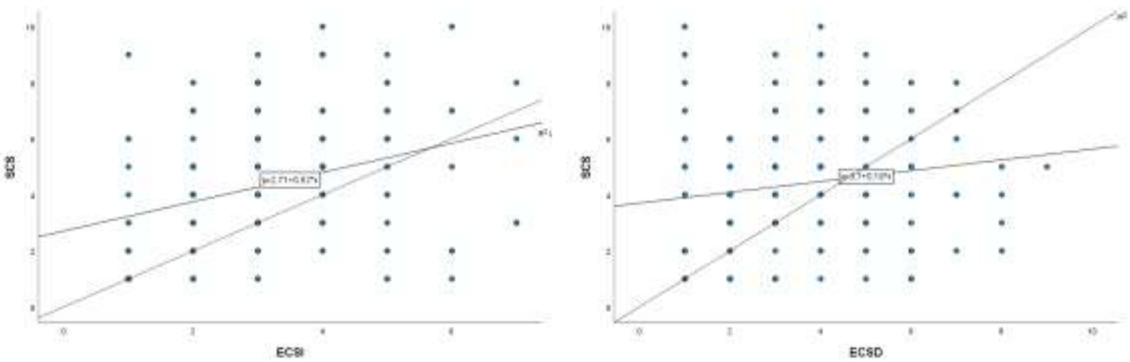


Note. Left – Bland-Altman plot of the mean of differences versus differences. Right – histogram of differences.

Scatterplots (Figure 23) indicated a positive linear relationship between SCS and both ECSI and ECSD. The Shapiro-Wilk test indicated that the variables were not normally distributed ($p < 0.05$), however, this would be expected as each pair of values represents one independent paired comparison. Pearson's correlation test is robust and not adversely impacted by non-normal distributions, especially given the context mentioned above.

Figure 23

Scatterplots of SCS, ECSI, and ECSD: PFN Graph-Theoretic Distances



a. SCS and ECSI

b. SCS and ECSD

Note. Dotted lines indicate line y = x in which student and expert would be in full agreement in terms of perceived related and dissimilarity.

There was a statistically significant small to moderate positive correlation between groups, $r(188) = .15 – .42$, $p < 0.05$. This indicated a moderate strength of linear association. SCS had a higher correlation with ECSI than ECSD; this may again reflect the importance of a clinical degree in teaching gross anatomy in a clinical context.

Individual SCS and ECS

Individual PFN configurations were created to derive network properties such as coherence and the number of links. Pathfinder networks also produce two other unique derived parameters: common links and similarity. PFN provides the ability to make direct comparisons between individuals. These are summarized as mean group values in Table 18. Individual SCS has similar common links, common link percentage, and similarity with ECSI as it does with ECSD. I have included all individual PFN data visualizations in Appendix F.

Table 18

Mean Values of Group Pathfinder Network Properties

	Links	Coherence	Common	Common %	Similarity
Group	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
SCS	41.4 (14.3)	0.430 (0.225)	---	---	---
Range	26-88	0.010-0.765			
ECSD	41.0 (6.0)	0.410 (0.238)	9.6 (3.4)	24.1 (8.3)	.181 (.068)
Range	35-47	0.140-0.588	3-17	10.3-42.1	.061-.356
ECSI	36.8 (14.4)	0.597 (0.033)	9.7 (4.1)	24.0 (8.3)	.189 (.078)
Range	24-52	0.555-0.627	3-19	9.7-40.0	.061-.339

Note. Instructor coherence: C1 = 0.627, C2 = 0.589, C3 = 0.555, C4 = 0.618

Common links, common link percentage, and similarity between the individual student and cohort instructors as a group showed a slight improvement over the group comparison. However, there was a marked improvement when examining individual students by cohort. Common links ranged from 12.4 to 20.6, common link percentage ranged from 27.3 to 54.9, and similarity improved to a range of 0.215 to 0.332.

I compared each individual SCS to both group ECSI and individual ECSIC (ECS for their cohort instructor). Results of the agreement analysis between individual SCS and ECS PFN graph-theoretic distances are summarized in Table 19.

Table 19

Individual SCS–ECS Agreement: PFN Graph-Theoretic Distances

	Reliability	Accuracy	Association
SCS	α [95%CI]	*RMSD* (units)	*r*
ECSI	.12	1.8	.26
Lower	-.31 [-.56, -.07]	1.5	-.18*
Upper	.39 [.25, .51]	2.2	.58**
ECSIC	.16 (.19)	1.6	.24 (.17)
Range	-.22-.52	1.0-2.0	-.12-.52

Note. α = Krippendorff's alpha coefficient, RMSD = root mean square deviation, *r* = Pearson's correlation coefficient.

[a] probability * $p < 0.05$, ** $p < 0.01$

[b] SCS ($n = 31$), ECSI ($n = 4$), ECSD ($n = 3$)

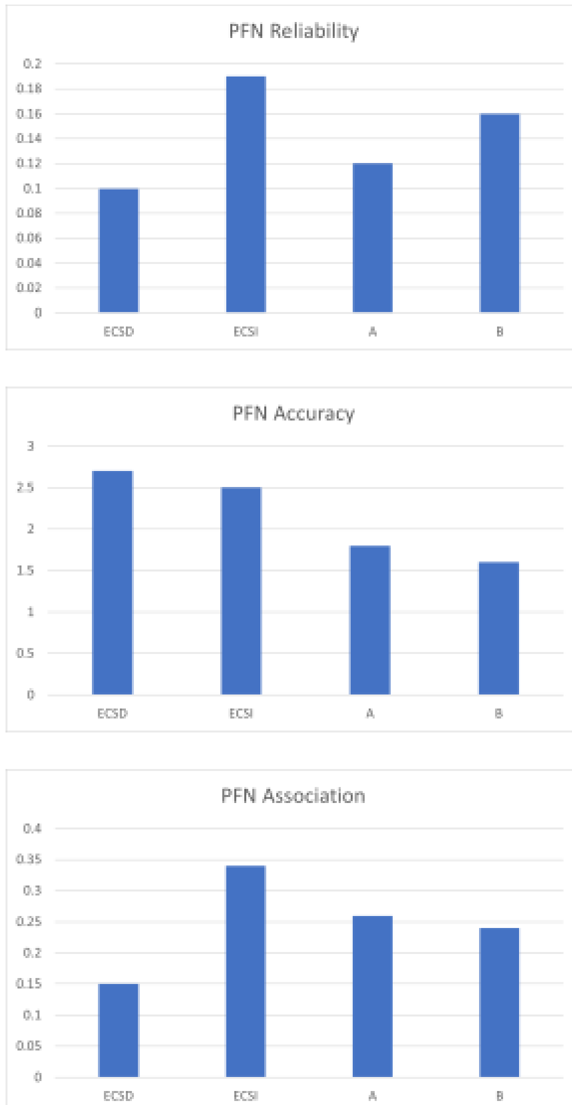
The mean Krippendorff's alpha coefficient (α = 0.12) indicated poor interrater reliability between individual students and ECSI PFN distances; however, the interrater reliability with the cohort instructor ECSIC PFN distances improved (α = 0.16). It is notable that 11 students (44%) had an overall increase in interrater reliability with their cohort instructor than with the group ECS. There was a broad variability amongst students. Accuracy improved with the cohort instructor, with twenty students (80%) displaying an overall decrease in RMSD with the instructor ECSIC compared to the group ECSI. The strength of association between individual SCS and ECSI was consistent with the strength of association between individual SCS and individual ECSIC, though there appeared to be differences between cohorts.

PFN Overview

A summary of the agreement analysis across all comparison levels (group SCS and group ECS, individual SCS and group ECSI, and individual SCS and cohort instructor ECSIC) is presented in Figures 24 and 25. SCS was aligned with instructors more so than with domain experts. On a more granular level, students appeared to display a greater relationship with their specific cohort instructor than the group ECSI. However, this trend between cohorts was impacted by the small sample size of each cohort.

Figure 24

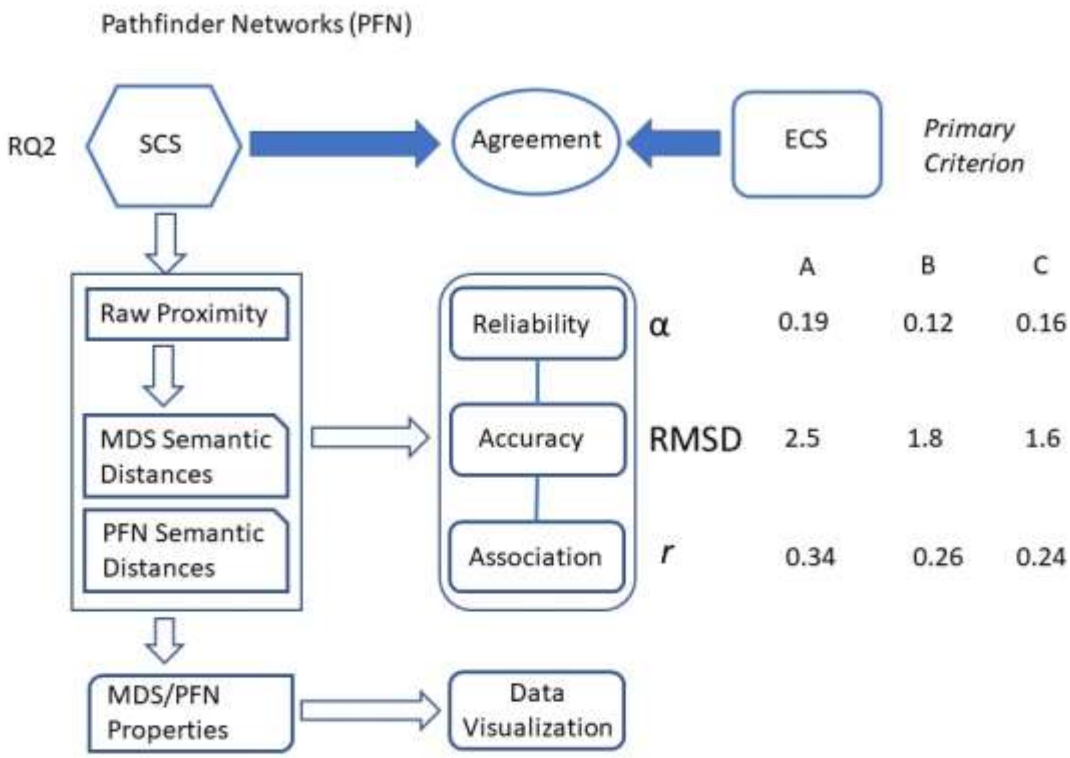
Summary of Agreement Analysis: Pathfinder Networks



Note. Each bar represents a specific level of comparison. ECSD and ECSI are compared to group SCS. A = individual SCS and group ECSI, B = individual SCS and individual cohort instructor ECSIC. At the level of the cohort instructor, reliability increased in 44% of students, and accuracy increased in 80% of students.

Figure 25

RQ2 Summary: Pathfinder Networks



Note. A = Group SCS–Group ECSI; B = Individual SCS–Group ECSI; C = Individual SCS–Individual ECSIC

Summary of Findings

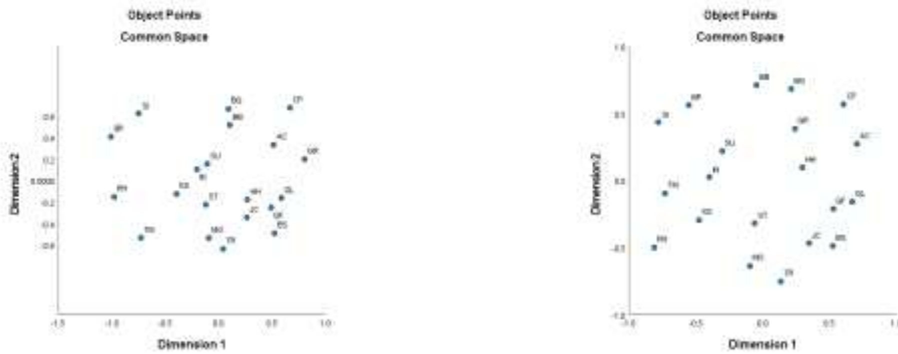
MDS and PFN provided both a qualitative and quantitative representation of anatomical concepts based on the raw proximity (paired comparisons) data. There were qualitative differences in how students and experts perceive the paired comparisons as reflected in the MDS and PFN data visualizations. Quantitative properties of both MDS configuration and PFN network also displayed differences between groups and individuals. Semantic distances (MDS Euclidean distances and PFN graph-theoretic distances) provided a reference for the perceived relationship of anatomical concepts with smaller distances inferring a closer perceptual relationship between items. Agreement analysis based on semantic distances provided a quantitative representation of the degree to which students and experts agree. All these factors were essential in establishing a relationship between SCS and ECS.

The MDS data visualization provided a broad qualitative overview of the spatial relationships of the anatomical concepts with many commonalities between expert and student MDS configurations. Figure 26 displays an example of the potential spatial differentiation between ECSI and SCS. Euclidean distances are unaffected by rotation, reflection, and translation of the CMDS configurations. MDS data visualizations exhibited greater spatial distances in the SCS configuration as compared to the ECSI configuration. However, the specific perceptual meaning of the two dimensions was unclear regarding the organization of anatomical concepts.

The PFN data visualization provided a more specific network relationship of the perceptual organization of anatomical concepts represented in the paired comparisons. Figure 27 displays an example of the potential network differentiation between ECSI and SCS. The PFN data visualization exhibited two critical differences compared to the MDS representation: it provided a direct linking of items and a derived quantitative parameter of direct similarity with another generated network. The SCS coherence values were consistent with those of the ECSI, indicating that both groups had internal coherence though this may not be represented in an identical visual fashion.

Figure 26

MDS Data Visualizations: ECSI and SCS

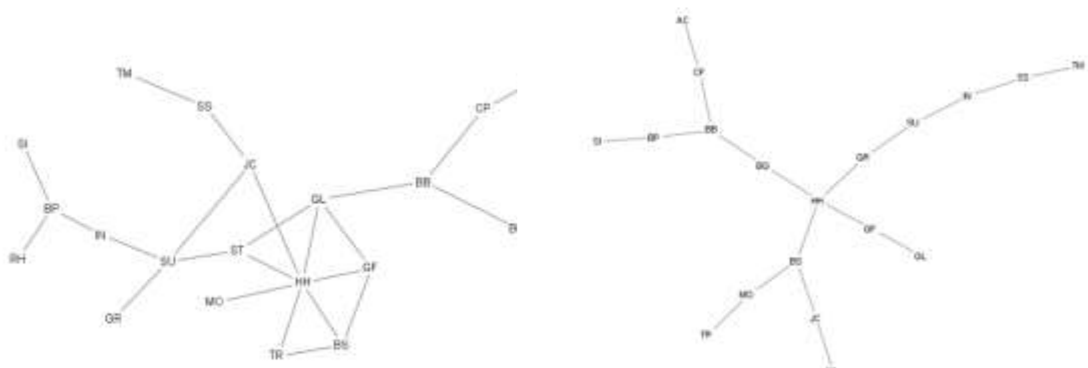


a. b.

Note. a = ECSI, b = SCS

Figure 27

PFN Data Visualizations: ECSI and SCS



a. b.

Note. a = ECSI, b = SCS

Reliability and association may be the two most readily understood measures to indicate a potential relationship between student and expert cognitive structures. Accuracy was quantifiable via the derived parameters of both MDS and PFN; however, RMSD provided a purely contextual reference with smaller distances indicating a closer perceptual relationship. Table 20 summarizes the agreement analysis across all data modeling strategies for both ECSD and ECSI.

Table 20

Summary of Agreement Analysis: Group and Individual

		Group SCS		Individual SCS	
SCS	Measure	ECSD	ECSI	ECSI	ECSIC
PRX	α	.59	.75	.37	.29
	r	.66	.82	.46	.40
MDS distances	α	.53	.83	.49	.41
(Euclidean)	*r*	.53	.83	.49	.41
PFN distances	α	.10	.19	.12	.16
(Graph-theoretic)	*r*	.15	.34	.26	.24
	Common %	26.1	47.6	24.0	44.0
	Similarity	.167	.333	.189	.284

Note. Group SCS is more closely aligned with group ECSI than group ECSD. Individual SCS is generally more closely aligned with group ECSI than with the individual ECSIC; however, alignment with the instructor is higher at the level of common link percentage and network similarity.

There were several important observations regarding the potential relationship(s) between student and expert cognitive structure. The most important finding established with the data analysis was the difference between expert groups and the variation in the agreement between SCS and both ECSI and ECSD. Student cognitive structure had greater reliability, accuracy, and association with instructors than with domain experts. Perceptual differences were also noted within the group of instructors.

There appeared to be some variation between cohorts and their respective instructors reflected in the student's cognitive structure and agreement with the instructor. In terms of proximity data, 36% of students had a greater association with their specific cohort instructor than with ECSI. MDS analysis revealed that 29% of students had a higher degree of accuracy with their specific cohort instructor than ECSI. PFN analysis revealed that 44% of students had greater reliability, with 80% of students having greater accuracy, with their specific cohort instructor than with ECSI.

The data comparing SCS and ECSI indicated large effect sizes for proximity data and MDS Euclidean distances with small effect sizes for PFN graph-theoretic distances. In contrast, the data comparing SCS and ECSD indicated small to moderate effect sizes across all measures. PFN similarity scores displayed improvement with individual comparisons. In summary, the findings provided evidence of concurrent criterion-related validity based on the first criterion standard and the study's operational definitions noted in Chapter 3.

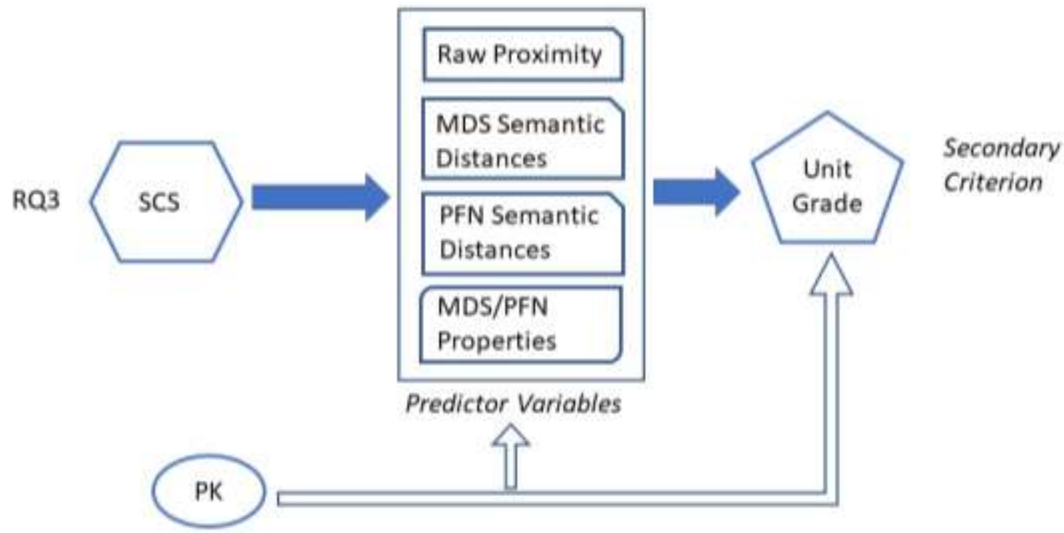
Research Question 3

RQ3: Is there a relationship between student cognitive structure and unit grade while controlling for prior knowledge?

Figure 28 provides an overview of the approach used to examine RQ3.

Figure 28

RQ3 Overview



Note. Agreement between SCS and ECS for each component were derived in RQ2.

Student Unit Grades

Unit grades served as the criterion standard used for RQ3. Exam grades often have poor validity; however, they have served as a standard criterion for comparison with the previous research. Unit grades were calculated based on the weighting prescribed in the Gross Anatomy course syllabus. The written exam grade comprised 55.56% of the unit grade, with the practical grade accounting for 44.44%. Table 21 summarizes the grade data for the total (five cohorts) and sample. The sample frame was representative of the target population, given the small sample size.

Table 21

DPT Student Unit Grades

	Total		Sample	
	M	Range	*M*	Range
n	224		31	
Unit Written Exam	79.9	38 – 100	81.6	56 – 98
Unit Lab Exam	82.6	47 – 100	84.2	58.8 – 100
Unit Grade (weighted)	81.1		82.8 (10.01)	

Note. Standard deviation in parentheses.

Prior Knowledge as a Predictor Variable

The construct of prior knowledge was operationally defined in Chapter 3 as "all knowledge learners have when entering a learning environment that is potentially relevant for acquiring new knowledge" (Biemans & Simons, 1996). The partner institution used several variants of admission GPA and admission anatomy GPA. These included admission cumulative GPA (undergraduate degree) and admission core science GPA (consisting of 2 chemistry courses, two physics courses, two biology courses, and two anatomy/physiology courses). As the institution did not explicitly define admission anatomy GPA, the core science GPA was used. Table 22 summarizes student prior knowledge represented by admission cumulative GPA and admission core sciences GPA (five cohorts). The sample frame was representative of the target population, given the small sample size.

Table 22

DPT Student Prior Knowledge

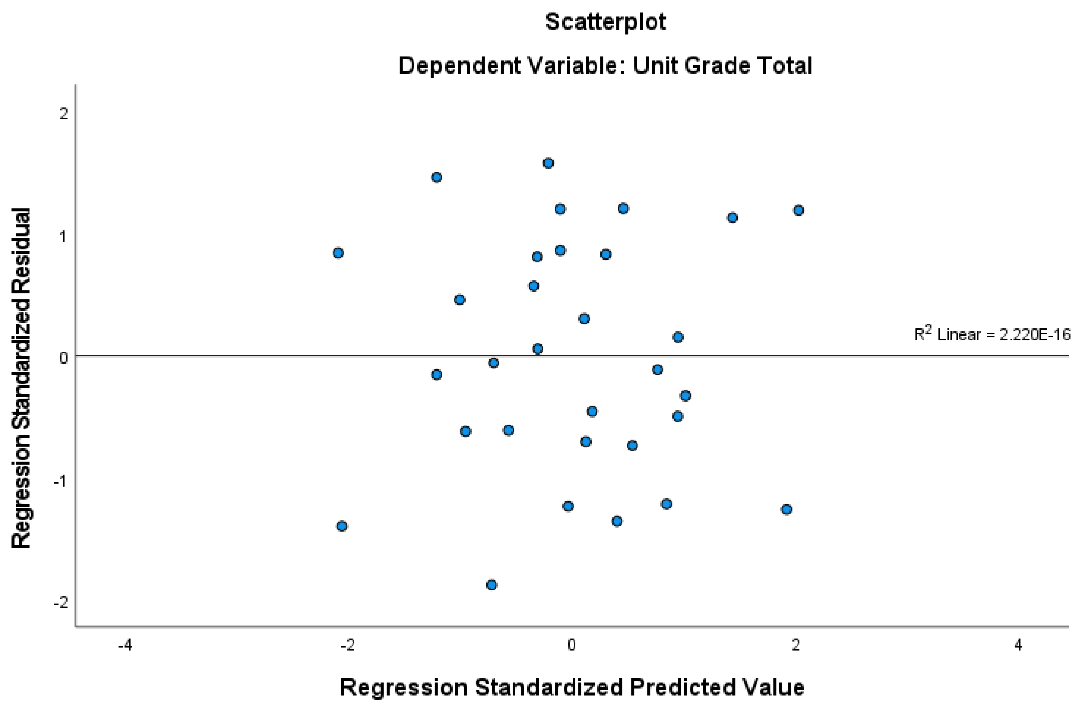
	Total		Sample	
n	224		31	
Cumulative GPA	3.23	2.48-4.0	3.28	2.63-4.0
Core Sciences GPA	3.30	2.60-4.0	3.37	2.96-4.0

Note. The sample was representative of the target population.

I used multiple regression analysis to determine a prediction model between prior knowledge (represented by the continuous independent variables of cumulative GPA and core sciences GPA) and unit grade (continuous dependent variable). The independence of observations was confirmed via a Durbin-Watson value of 1.754. There was linearity and homoscedasticity assessed by partial regression plots and a plot of residuals against the predicted values (Figure 29). Tolerance values greater than 0.1 confirmed no evidence of multicollinearity. Casewise diagnostics confirmed that there were no outliers greater than +/- 3 standard deviations. The assumption of normality was met based on an examination of the histogram and P-P plots.

Figure 29

Residuals Plot of Unit Grade (Variable: Prior Knowledge)



Note. Linearity and homoscedasticity were confirmed.

The multiple regression model did not predict unit grade, $F(2, 28) = 0.090$, $p = 0.914$, adjusted $R^2 = -0.065$. There was not a statistically significant relationship between the predictor variables (prior knowledge) and unit grade.

I used multiple regression analysis to determine a prediction model between prior knowledge (represented by the continuous independent variables of cumulative GPA and core sciences GPA) and each of the potential predictor variables (including proximity-, MDS-, and PFN-derived parameters as well as variables derived from the agreement analyses between SCS and ECS). All the dependent variables tested were continuous. For each dependent variable tested, all statistical assumptions were met. However, there was not a statistically significant relationship between prior knowledge and any of the values derived from the proximity-, MDS-, or PFN-derived parameters used to represent student cognitive structure, nor the values derived from the agreement analysis between SCS and ECS or any of its subgroups. Prior knowledge was not a statistically significant predictor of MDS Euclidean distances or PFN graph-theoretic distances nor any of the values derived from the agreement analysis, including interrater rater reliability, level of agreement, or strength of linear association between students and experts.

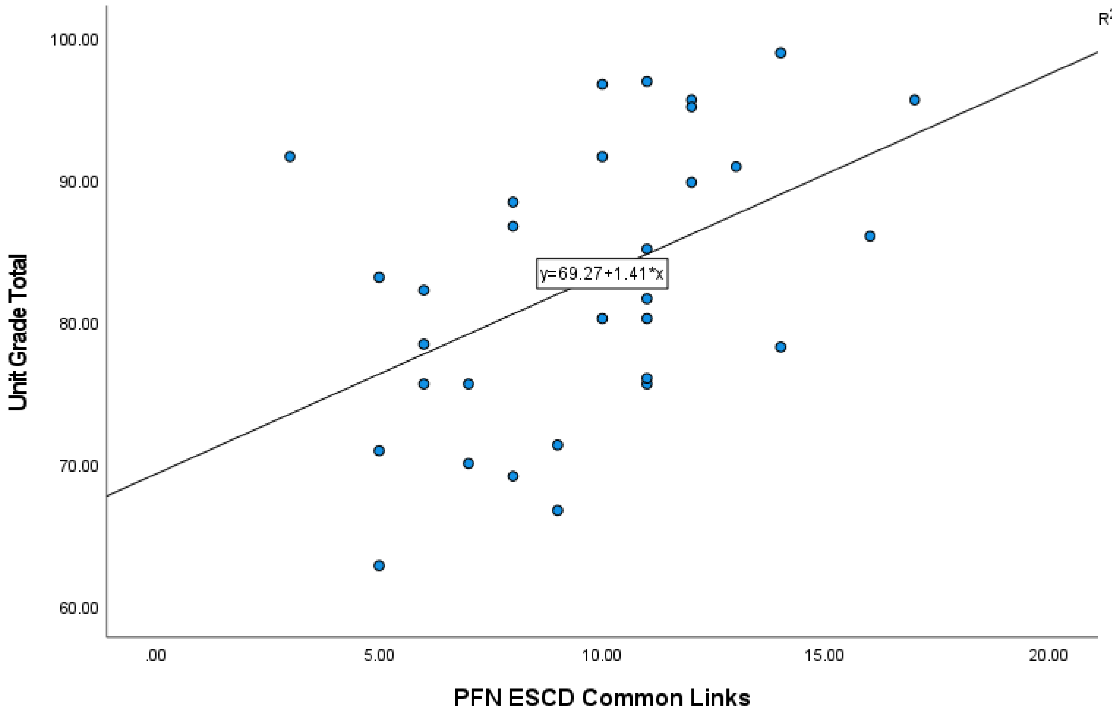
MDS and PFN Predictor Variables

I examined all parameters derived from the proximity data, MDS configurations, PFN networks, and student-expert agreement analyses as potential predictor variables within the context of linear regression. Potential predictor variables included all variables derived directly from the data modeling strategy (for example, MDS stress-1 and PFN number of links) and those derived via direct comparison of SCS and ECS agreement (interrater reliability, level of agreement, and strength of linear association). Refer to Table 3 for an overview of variables derived and subsequently tested.

One predictor variable was noteworthy during the preliminary analysis of individual dependent variables: PFN common links between SCS and ECSD. I performed a linear regression analysis to understand the effect of the agreement between SCS and ECSD in terms of PFN common links and unit grade. A scatterplot of PFN common links against unit grade with a superimposed regression line was plotted (Figure 30). Visual inspection indicated a linear relationship between these variables.

Figure 30

Scatterplot of PFN ECSD Common Links Versus Unit Grade

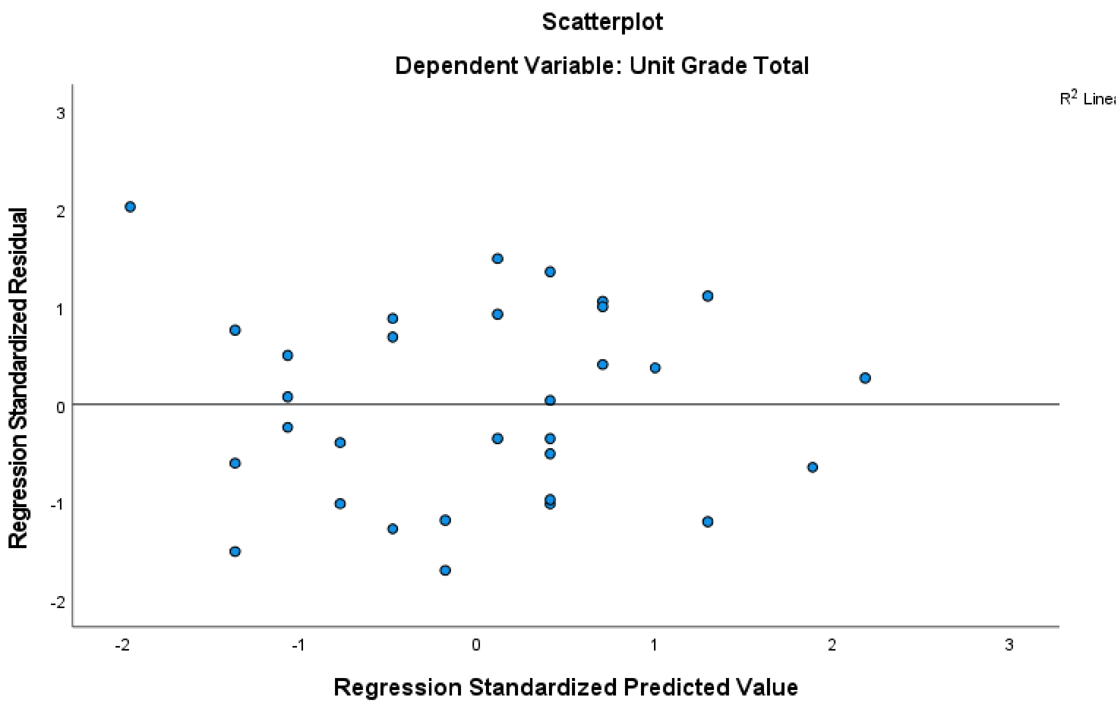


Note. Visual inspection reveals a linear relationship between the variables.

The independence of observations was confirmed by a Durbin-Watson value of 2.069. Casewise diagnostics confirmed that there were no outliers greater than +/- 3 standard deviations. There was linearity and homoscedasticity assessed by partial regression plots and a plot of studentized residuals against the predicted values (Figure 31). The assumption of normality was met based on an examination of the histogram and P-P plots.

Figure 31

Residuals Plot of Unit Grade (Variable: PFN ECSD Common Links)



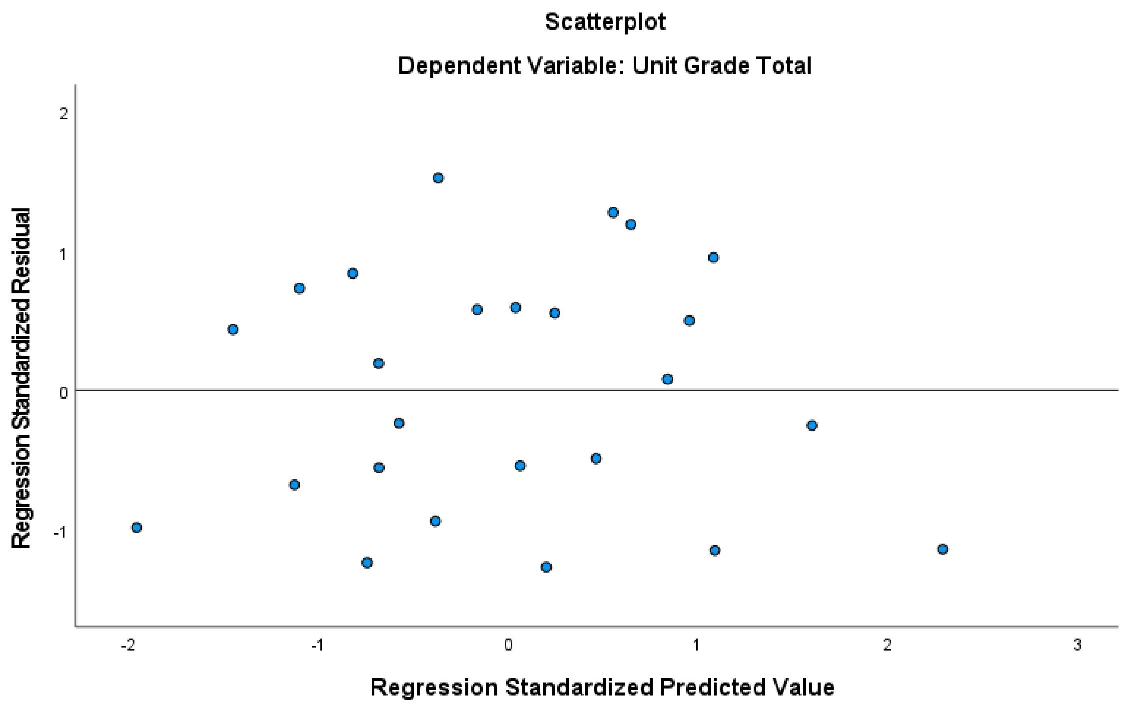
Note. Linearity and homoscedasticity were confirmed.

The level of agreement between SCS and ECSD in terms of PFN common links statistically significantly predicted unit grade, $F(1, 29) = 8.474$, $p = 0.007$, accounting for 22.6% of the variance in unit grade with adjusted $R^2 = 19.9\%$, a medium effect size (see Cohen, 1988). An increase of one common link with ECSD increases the grade by 1.4% (95% CI, 0.42 to 2.40).

Based on the preliminary analysis of all potential independent variables (including the agreement between SCS and ECSD in terms of PFN common links), I used a multiple regression analysis to determine a prediction model between six agreement analysis independent variables and unit grade (continuous dependent variable). One independent variable was derived from the proximity data (PROX correlation with ECSD), one independent variable was derived from the MDS Euclidean distances (CMDS correlation with ECSD), and four independent variables were derived directly from the PFN network properties (PFN common links with ECSD, PFN common links with ECSIC, PFN ECSD similarity, and PFN ECSIC similarity). The independence of observations was confirmed via a Durbin-Watson value of 1.810. There was linearity and homoscedasticity assessed by partial regression plots and a plot of studentized residuals against the predicted values (Figure 32). Tolerance values greater than 0.1 confirmed no evidence of multicollinearity. Casewise diagnostics confirmed that there were no outliers greater than +/- 3 standard deviations. The assumption of normality was met based on an examination of the histogram and P-P plots.

Figure 32

Residuals Plot of Unit Grade (Variables: MDS and PFN)



Note. Linearity and homoscedasticity were confirmed.

The multiple regression model statistically significantly predicted unit grade, $F(6, 18) = 6.645$, $p = 0.001$, adjusted $R^2 = 0.585$. All six variables added statistically significantly to the prediction, $p < 0.05$. Regression coefficients and standard errors can be found in Table 23.

Table 23

Multiple Regression Results for Unit Grade

Unit Grade	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	UL				
Model						.69	.59***
Constant	67.53***	58.06	77.06	4.52			
PRX	34.67*	0.09	69.25	16.46	0.59*		
CMDS	-40.03**	-69.36	-10.69	13.96	-0.80**		
PFN C1	3.54***	1.79	5.29	0.83	1.23***		
PFN C2	-1.89**	-3.05	-0.73	0.55	-1.31**		
PFN S1	-86.49*	-168.88	-4.11	39.22	-0.59*		
PFN S2	119.56**	42.24	196.89	36.81	1.33**		

Note. PRX = Proximity correlation SCS and ECSD; CMDS = MDS Euclidean distance correlation SCS and ECSD; PFN C1 = PFN common links SCS and ECSD; PFN C2 = PFN common links SCS and ECSIC; PFN S1 = PFN similarity SCS and ECSD; PFN S2 = PFN similarity SCS and ECSIC.

B = unstandardized regression coefficient, CI = confidence interval; LL = lower limit; UL = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

*$p < 0.05$, **$p < 0.01$, *** $p < 0.001$

Six predictor variables accounted for 68.9% of the variance in unit grade with adjusted R^2 = 58.5%, a large effect size (see Cohen, 1988). However, direct interpretation of the regression coefficients within the context of the proximity data and data modeling strategies is important. Attaining a one percent increase in unit grade would require several minor changes in any or all of the six parameters. For example, changes of 0.050 in correlation and PFN similarity or a 0.1 change in PFN common links may produce large changes in unit grade. This is a high level of granularity subject to the small sample sizes and requires greater numbers of experts and students to generalize these observed trends.

Summary of Findings

Prior knowledge was not a statistically significant predictor of any of the MDS- or PFN-derived parameters and was not a statistically significant predictor of unit grade. There was a medium effect size in predicting unit grade via PFN common links between SCS and ECSD. Multiple linear regression indicated that there was evidence of predictive criterion validity based on the second criterion standard. In terms of unstandardized beta coefficients, it is important to remember the context of the derived parameter. Small changes in the level of agreement between student and expert can be notable given the context of the derived parameters. However, the minimal interpretable difference for all derived variables was unclear at this point.

Exploratory Analysis

Several exploratory analyses were performed as contextual issues arose during the examination of RQ2 and RQ3. These focused on within-group differences for both instructors and students.

Instructor ECS

An exploratory analysis was performed to examine within-group differences for cohort instructor ECSIC. This was based on the preliminary results of RQ2, indicating that students within a cohort may have a greater agreement with their specific instructor than the group ECS. Proximity data and MDS- and PFN-derived parameters (Euclidean distances, graph-theoretic distances, and associated interrater reliability, RMSD, and Pearson's correlation coefficient) were calculated for each pair of instructors. Table 24 summarizes the ranges of calculated values.

Proximity Data

There was fair to good interrater reliability across all instructor ECSIC ($\alpha = 0.41$, 95% CI 0.37 - 0.45). The RMSD representing the level of agreement between instructor ECSIC indicated that ratings could be up to 2 points different on the seven-point perceived relatedness rating scale. There was a medium to large strength of association between cohort instructors ($r = 0.39 – 0.62$, $p < 0.01$), indicating a degree of internal consistency between instructors regarding their proximity data.

Multidimensional Scaling

There was moderate interrater reliability across all instructor ECSIC ($\alpha = 0.56$, 95% CI 0.51 - 0.60). The RMSD representing the level of agreement between instructor

ECSI MDS distances served primarily as a contextual reference; similar values were present for all instructor comparisons. There was a moderate to strong positive correlation between cohort instructors (r = 0.45 – 0.72, $p < 0.01$), indicating a degree of internal consistency between instructors regarding their MDS Euclidean distances.

I briefly considered the variations of instructor MDS and the preference of dimensions one and two as derived by using a weighted Euclidean scaling model. The WMDS configuration, using data from all four instructors, produced a derived stress-1 = 0.1879, TCC = 0.9822, and R^2 = 0.812. Dimensional weights varied from 0.382 to 0.524 (dimension one) and from 0.432 to 0.555 (dimension two). Experts appeared to utilize dimensions in differing proportions, which may reflect perceptual differences. However, it was unknown if these dimensional differences were relevant or statistically significant.

Table 24

Summary of Agreement Analysis: Cohort Instructor

		Individual Instructor ECSIC	
	Measure	Lower	Upper
PRX	α	.07[a]	.62
	RMSD	2.0	2.2
	r	.39**	.62**
MDS distances	α	.46	.72
(Euclidean)	*RMSD*	.326	.465
	r	.45**	.72**
	Dim 1[b]	.382	.524
	Dim 2[b]	.432	.555
PFN distances	α	.00	.36
(Graph-theoretic)	*RMSD*	1.3	1.8
	r	.20**	.36**
	Common Links	11	30
	Coherence	0.555	0.627
	Similarity	0.250	0.441

Note. Agreement analysis for individual cohort instructors. α = Krippendorff''s alpha; RMSD = root mean square deviation; *r* = Pearson's correlation coefficient.

[a] PRX α range was 0.34 to 0.62 with one outlier.

[b] WMDS (weighted Euclidean scaling model) was used for comparison.

** $p < 0.01$

Pathfinder Networks

There was poor interrater reliability across all instructor ECSIC PFN graph-theoretic distances ($\alpha = 0.19$, 95% CI 0.12 - 0.27). The RMSD representing the level of agreement between instructor ECSI PFN distances served primarily as a contextual reference; similar values were present for all instructor comparisons. There was a small to moderate strength of linear association between cohort instructors ($r = 0.20 – 0.36$, $p < 0.01$), indicating a degree of internal consistency between instructors regarding their PFN graph-theoretic distances. Instructors as a group appeared to have potentially large differences in the number of links generated in their Pathfinder network. However, coherence values were consistent between instructors indicating that the derived measures had internal consistency with the proximity data. These network properties may reflect perceptual differences.

Summary of Findings: Instructor ECS

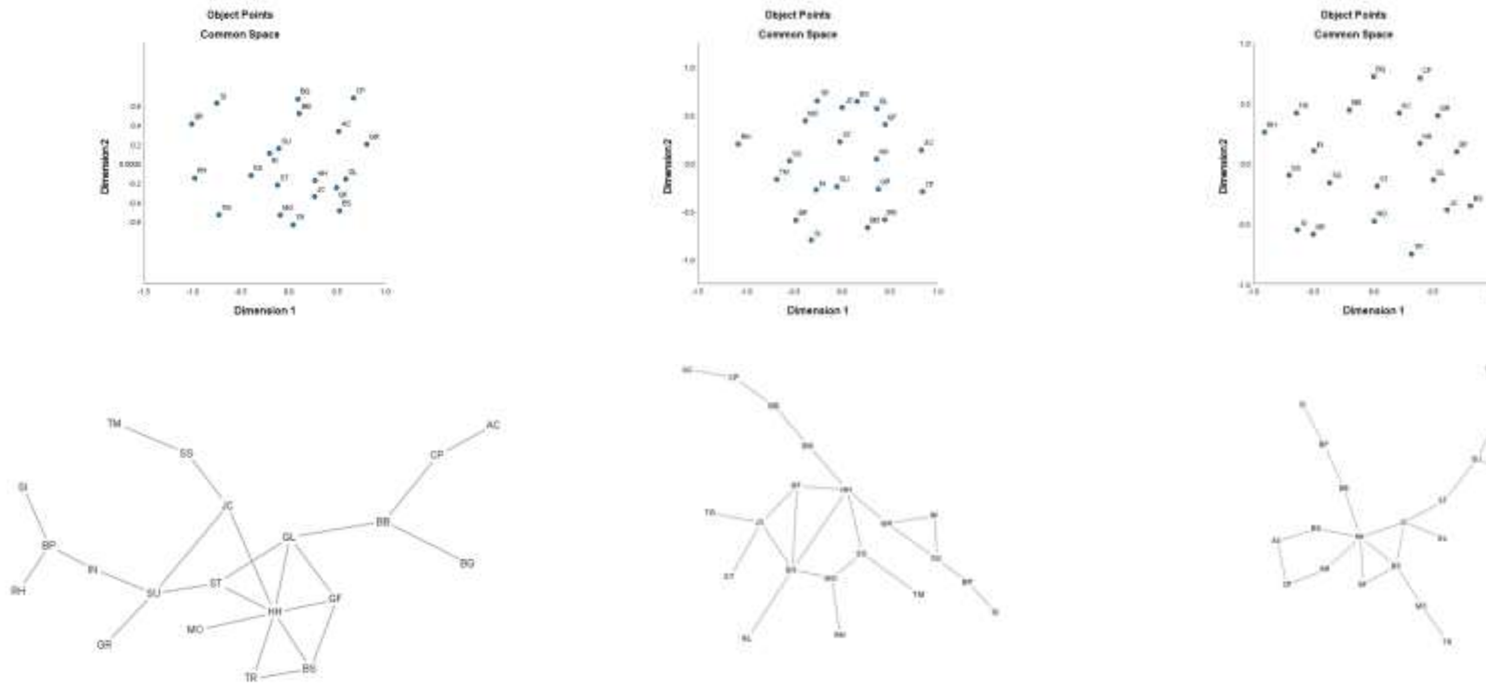
There appeared to be broad variability across instructor ECSIC. However, these were purely observational trends; a larger sample size of instructors may provide less variability with improved statistical power. These variations between instructors provided a rationale for further exploring the relationship between SCS and their specific cohort instructor's ECS.

ECS, SCS, and Academic Performance

I explored a follow-up to RQ3 to compare the visual representations of students having unit grades over 90 and those having unit grades under 75. However, the sample size for the groups (high grades $n = 8$, low grades $n = 6$) provided insufficient power for statistical comparisons. Preliminary observations were limited to qualitative visual assessment. Figure 33 displays the MDS and PFN visual representations of these two student groups with ECSI. The MDS configuration for both groups had similar values for stress-1 (0.254 and 0.244), TCC (0.967 and 0.970), and R^2 (0.58 and 0.61). Both groups also had similar PFN coherence (0.687 and 0.724). These factors indicated that students developed internal coherence regardless of grade. PFN-derived parameters provided the ability for direct comparison between groups. It is of note that students with high grades consistently displayed more common links and higher similarity with ECSI and ECSIC. However, these trends were observational and were not subjected to the scrutiny of statistical analysis given the small sample sizes and inherently low statistical power. These trends suggest the need for further research to define any generalizable conclusions.

Figure 33

ECSI, SCS, and Academic Performance



ECSI | Students with > 90% | Students with < 75%

Note. An observed trend toward greater visual similarity with ECSI was noted in students with higher academic performance.

Summary

The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. Qualitative and quantitative findings were reported based on the proximity data and both MDS and PFN.

MDS and PFN data visualizations produced an initial qualitative overview of commonalities and differences between students and experts. Anatomical concepts occupied similar spatial relationships in MDS but were linked in different ways in PFN. There were also differences noted in the visual representations within the expert subgroups. The first level of quantitative analysis examined the properties derived from MDS configurations and PFN networks. The MDS configurations of experts displayed higher R^2 values with anatomical concepts having closer spatial relationships (decreased semantic distances) than students. The PFN network properties of experts displayed a greater degree of linking anatomical concepts than students.

Agreement analysis (reliability, accuracy, and association) was used to examine the potential relationship between student and expert cognitive structure represented by MDS and PFN data modeling strategies. MDS Euclidean distances and PFN graph-theoretic distances may provide a contextual reference for students based on their agreement with experts. However, the relevance of the measures and the minimally educationally relevant values remain to be studied. The findings of RQ2 indicated that agreement analysis varied between groups, and moderate to large effect sizes were evident in several of the reported measures.

The specific relevance of many derived parameters, either based directly on MDS configuration or PFN network properties or based on semantic distance agreement analysis, was unknown from a practical educational perspective. However, the PFN common links between students and domain experts (ECSD) accounted for 19% of the variance in unit grade. Six of the derived parameters (Proximity correlation between SCS and ECSD, MDS Euclidean distance correlation between SCS and ECSD, PFN common links between SCS and ECSD, PFN common links between SCS and ECSIC, PFN similarity between SCS and ECSD, and PFN similarity between SCS and ECSIC) accounted for 58.5% of the variance in student unit grade, the second criterion standard. Four of the six predictor variables involved comparisons with the domain expert (ECSD), and two involved comparisons with the specific cohort instructor (ECSIC). Four of the predictor variables involved the use of PFN and its derivations.

This study provided preliminary evidence of concurrent and predictive criterion-related validity. Given the operational definitions outlined in Chapter 3, it appeared that these data modeling strategies may provide the potential for a quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy.

Chapter 5: Discussion, Conclusions, and Recommendations

The mapping of cognitive structures of physiotherapy students learning gross anatomy is poorly understood. The purpose of this quantitative study was to explore two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. The nature of the study was a quasi-experimental, criterion-related validation study. The study was conducted to understand better the quantitative representation of cognitive structure in the gross anatomy domain, exemplified by MDS and PFN strategies, and to validate the possible meaning of these quantitative measures in the context of entry-level physiotherapy education.

The study's key findings provided preliminary evidence that MDS and PFN data modeling strategies may serve as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. Cognitive structure mapping was reflected in both the MDS and PFN data visualizations, descriptive properties (MDS configuration and PFN network), and derived parameters such as semantic distances (MDS Euclidean and PFN graph-theoretic). Differences in these quantitative parameters may reflect perceptual differences and level of agreement between student and expert and within expert subgroups. It is unclear whether this represents cognitive structure or some other cognitive, perceptual, or educational construct. Differences were noted between expert subgroups based on the presence or absence of a clinical degree. Preliminary evidence of content, construct, and criterion-related validity (concurrent and predictive) was reported. Six predictor variables derived from the proximity data, MDS configurations, and PFN networks accounted for 68.9% of the variance in unit grade with adjusted R^2 = 58.5%, a large effect size (see Cohen, 1988). The biggest single predictor of the unit grade was the PFN common links between the student and the domain expert. It is unknown whether these factors differentiate academic performance among students. Given the context of the study's operational definitions, there appears to be some potential in using MDS and PFN as a visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy.

Interpretation of the Findings

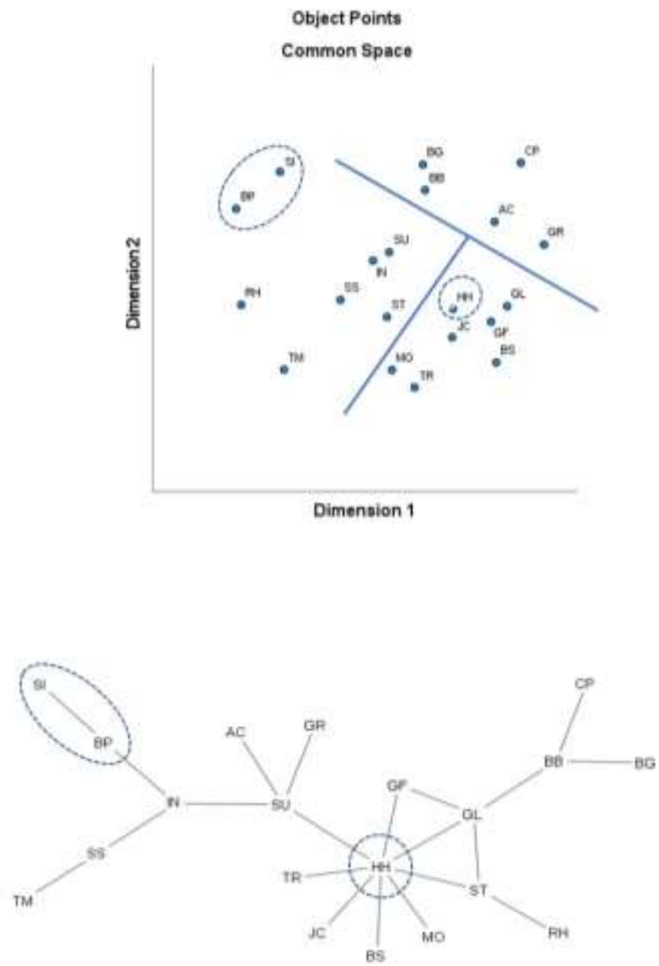
Cognitive structure is a construct rooted in declarative and procedural knowledge developed in long-term memory. J. R. Anderson's (1996, 2007) ACT-R cognitive architecture model served as a foundation for understanding the construct's potential mechanisms. Two of these mechanisms (chunking and activation) provided a context for long-term memory and knowledge acquisition. However, there remains little agreement on a clear operational definition of cognitive structure as a construct, the representation (either directly or indirectly) and validation of the construct, and its practical application and relevance educationally. The current study addressed two data modeling strategies (MDS and PFN) as a potential visual and quantitative representation of the cognitive structures of physiotherapy students learning gross anatomy. Although little research has simultaneously addressed the data visualizations and quantitative representations of MDS and PFN (Branaghan, 1990), the current study provided an extensive and rigorous analysis of the derived parameters of both data modeling strategies.

Data Visualization

The MDS and PFN data visualizations of cognitive structure provided an initial visual representation of differences between students and experts and within expert subgroups. These differences varied in scope and magnitude depending on the data modeling strategy. Branaghan (1990) noted that MDS could represent semantic dimensions underlying a domain and PFN could visualize the direct relationship between items. Figure 34 displays the MDS and PFN data visualizations for the ECSI subgroup. Several potential clusters of items appear to be consistent across MDS solutions once transformed via reflection, rotation, and translation (Figure 34, upper panel). These clusters may reflect the grouping of anatomical constructs and relationships: for example, neurological function (brachial plexus [BP] and segmental innervation [SI]), functional stability including the rotator cuff musculature (stability [ST], supraspinatus [SU], infraspinatus [IN], subscapularis [SS], rhomboids [RH], and teres major [TM]), the joint capsule and articulating surfaces and their role in joint mobility (mobility [MO], triplanar motion [TR], ball and socket [BS], joint capsule [JC], humeral head [HH], glenoid fossa [GF], and glenoid labrum [GL]), and the biceps brachii and its relationship to the shoulder complex (acromion [AC], coracoid process [CP], greater tubercle [GR], biceps brachii [BB], bicipital groove [BG]). The clustering of items may provide a broad overview of how a student is organizing their knowledge, especially in comparison to the organization of the expert or the specific cohort instructor. Visual differences of PFN (Figure 34, lower panel) are readily discernible based on item links within the network and a derived parameter of direct similarity with another generated network. Both data visualizations are consistent with J. R. Anderson's (2007) ACT-R model and the processes of chunking and activation used in knowledge organization. Although differences may be readily apparent in the PFN representations, they are far less so with MDS configurations because they lack a direct assessment of configurational similarity (see Borg & Leutner, 1985). However, MDS representations remain highly consistent and may provide a broad overview of their concept organization, which becomes more granular within the PFN's direct linking.

Figure 34

MDS and PFN Clustering of Anatomical Concepts



Note. The upper panel is MDS data visualization for ECSI. The lower panel is PFN data visualization for ECSI. Dotted lines emphasize clusters of anatomical items and concepts.

Criterion-Related Validity

The results indicated preliminary evidence of content and construct validity as well as concurrent and predictive criterion-related validity. The selection of items by physiotherapists in musculoskeletal clinical practice enhanced content and construct validity because the items reflected relevant competency, expertise, and knowledge organization. Concurrent (criterion-related) validity is the correlation of a measurement with a criterion (Barnhart et al., 2007). In the current study, the level of agreement between SCS and ECSI (criterion standard one) and how this differs from the level of agreement between SCS and ECSD provided preliminary evidence of concurrent validity. Predictive (criterion-related) validity is the correlation of a measurement with a future criterion (Barnhart et al., 2007). In the current study, predictive validity was represented by the predictor variables (the descriptive properties of MDS and PFN as well as the agreement between SCS and ECS) and the unit grade (criterion standard two). Six predictor variables accounted for 68.9% of the variance in unit grade with adjusted R^2 = 58.5%. Of the six predictor variables, one variable was derived from the proximity data (SCS correlation with ECSD), one variable was derived from the MDS Euclidean distances (SCS correlation with ECSD), and four variables were derived from the PFN network properties (SCS common links with ECSD, SCS common links with ECSIC, SCS similarity with ECSD, and SCS similarity with ECSIC). In the context of gross anatomy and physiotherapy education, these results are consistent with the predictive validity reported by Goldsmith et al. (1991) and Johnson et al. (1994).

Agreement Analysis

The study addressed the potential quantitative agreement between student and expert reflected in reliability (interrater reliability), accuracy, and strength of linear association. Assessment of agreement between student and expert occurs throughout physiotherapy education and requires both interrater reliability and agreement in the performance of a clinical activity, making it highly relevant in the assessment and comparison of cognitive structures in the context of physiotherapy education (Liao et al., 2010). Cognitive structure is essential in the development of expertise and, subsequently, clinical performance and diagnostic thinking. However, the use of agreement analysis between student and expert cognitive structures to derive unit grade and academic performance predictors had not been previously reported in the literature. There appeared to be preliminary evidence supporting this approach based on the findings of RQ2 and RQ3 with several derived parameters linked to academic performance. However, most parameters had no direct relationship with unit grade. Results suggest that a rigorous, data-driven approach to the representation of cognitive structure is promising and deserves further consideration.

Internal Consistency

Knowledge acquisition involves the addition of personal meaning above and beyond factual information. Students and experts will develop cognitive structure in long-term memory based on their perceptions, life experiences, emotional meaning, and misconceptions. How the individual organizes their knowledge will vary based on these factors. In the current study, the individual MDS configurations displayed consistently high R^2 values (indicating large effect sizes) and the PFN representations indicated good coherence within both students and experts. Both groups organize their knowledge, inclusive of misconceptions, in a way that provides internal consistency and coherence within their individual cognitive structure mapping. The MDS and PFN data visualizations displayed these individual variations; changes in these representations over time may provide evidence of a cognitive structure that evolves with learning, knowledge organization, or competency development. However, the degree to which these variations are relevant educationally is unclear and requires further examination.

Prior Knowledge

Prior knowledge is an integral component of learning (Ausubel, 1963). Meaningful learning occurs as students scaffold new knowledge upon prior knowledge via J. R. Anderson's (2007) chunking and activation mechanisms. However, the variables used to define prior knowledge in the current study (admission cumulative GPA and admission core sciences GPA) were not predictors of any of the MDS- or PFN-derived parameters, SCS agreement with ECS, or unit grade. However, a large proportion of variance in unit grade could be accounted for by six derived parameters that appear to be otherwise unrelated to prior knowledge. This creates a potential disparity between a construct (prior knowledge) and its current academic representation (GPA). Several researchers, including Bayliss et al. (2017) and S. H. Hayes et al. (1997), have examined the relationship between GPA as a predictor of success on the NPTE. Because prior knowledge was not associated with unit grades and the testing associated with it, there is a question of whether GPA (in any form) is a good metric to use for prior knowledge. Its use as a predictor of success on the NPTE may represent factors unrelated to cognitive structure or physiotherapy curriculum.

Expert Differences

There were differences between expert subgroups (domain expert ECSD, instructor ECSI, and cohort instructor ECSIC) across all group comparisons with proximity data, MDS, and PFN. The agreement between SCS and ECSI appeared to be higher than that between SCS and ECSD. This may provide early evidence of the importance of the presence or absence of a clinical degree in the teaching of gross anatomy to physiotherapy students. Although reliability, accuracy, and association appeared to have moderate to good reliability and medium to large strength of association (with the associated medium to large effect sizes), these values generally diminished at the level of the cohort instructor. However, large numbers of students improved their agreement with the cohort instructor. Reliability improved in 44% of students for PFN graph-theoretic distances. Accuracy improved in 29% of students for MDS Euclidean distances and 80% of students for PFN graph-theoretic distances. Association improved in 36% of students for the proximity data. These results provide evidence of a potential cohort-instructor-specific effect and highlight the importance of the instructor’s academic content knowledge, pedagogical knowledge, and pedagogical content knowledge (see Neumann et al., 2019). However, this may result from either the actual differences between student and instructor or may be a function of the small sample size. The use of data modeling strategies may provide insight into these perceptual differences. This could provide valuable information regarding how students organize their knowledge based on their interaction with a specific instructor to attain learning outcomes.

The biggest single predictor of the unit grade was the student's PFN common links with the domain expert ECSD. Initial exploratory analysis indicated a trend for students with higher grades to be more highly correlated with ECSD than with ECSI. Domain experts may have knowledge that is more representative of nonclinical gross anatomy, which aligns more clearly with the content and context of the course text used to develop unit exams and is not specifically physiotherapy centric. Although students may align more closely with instructors in terms of agreement analysis, what defines their unit grade may be more closely aligned with the domain expert and course text. A gross anatomy course offered in the first trimester of the program may be testing primary anatomical organization that aligns more with the ECSD domain experts (basic anatomical knowledge) than its application (ECSI). This provides a challenge for effective clinically based education that promotes near transfer and clinical competency.

Previous Research

Previous research by Goldsmith et al. (1991), Gonzalvo et al. (1994), Neiles et al. (2016), and Stevenson et al. (2016) found relationships between MDS- and PFN-derived parameters and student grades. Large effect sizes were reported based on calculated r^2 (>0.5) and η2 (>0.14) values (see Cohen, 1988). The moderate to large effect sizes reported in the current study were consistent with these studies. However, many of these studies focused solely on correlation, a statistical test that may not have been used judiciously nor with a clear delineation of operational definitions, level of measurement, and assumptions necessary for their statistical analyses. Agreement is not solely an issue of correlation as raters can be highly correlated with little to no agreement. This makes the conclusions from previous research problematic. The current study provided a more extensive and granular analysis of agreement within expert subgroups.

Implications

Cognitive structure is a fuzzy construct for clinicians and educators to conceptualize due to inconsistencies in operational definitions, the inability to represent it readily, and the lack of a clearly defined relevance and practical application for students, educators, and future clinicians. The exploratory nature of the current study provided a narrow window into the construct of cognitive structure in a small sample of physiotherapists, experts, and DPT students in the context of gross anatomy education. However, the current study has implications on theory development, research methodology, educational practice, and positive social change.

Theory Development

The cognitive architecture proposed by the ACT-R model consists of eight modules – four of which are related to the perceptual-motor system and interact directly with the external world (J. R. Anderson et al., 1997). The remaining four modules (declarative, procedural, intentional, and imaginal) are related to facts, procedures, goals, and a mental representation of the problem. Representations of cognitive structure, such as those proposed via MDS and PFN, provide a snapshot of cognition at that specific moment in time. This may reflect a summation of the function of these four modules at that moment. However, long-term memory and knowledge organization is a dynamic system (Schuelke et al., 2009). J. R. Anderson (2007) proposed that chunking and activation are factors in long-term memory and cognitive structure development. Jonassen et al. (1993) proposed another aspect of long-term memory, that of structural knowledge. If MDS and PFN serve as a potential visual and quantitative representation of cognitive structure, then they may only reflect cognitive structure at that moment in time. This underscores a need to examine test-retest reliability.

Cognitive learning theory is an integral factor in the development of competent physiotherapists. However, this depends on the goal of learning: short-term retrieval of knowledge or long-term competency (J. R. Anderson & Schunn, 2000). The development of cognitive structure parallels the development of epistemic cognition leading to critical thinking, clinical reasoning, diagnostic thinking, near transfer, and academic achievement (Greene & Yu, 2016; Montpetit-Tourangeau et al., 2017). Cognitive structures developed through deep learning enhance retention and transfer of learning to higher-order thinking and decrease cognitive load (Krathwohl, 2002; Smith, Stockholm, et al., 2017). Experts learn to categorize problems based on deeper features and have improved knowledge organization (Fatima, 2020; Schuelke et al., 2009). Critical thinking is necessary for clinical practice specifically and as a 21st-century skill more broadly, but it is also limited by surface (rote) learning. Novices tend to use surface learning strategies to accumulate facts, which generates a higher level of cognitive loading without enhancing cognitive structure development (Fatima, 2020; Schuelke et al., 2009; Zulu et al., 2018). This may promote the development of misconceptions which can severely hamper academic progression and clinical competency. A visual and quantitative representation of cognitive structure may have potential implications in monitoring a student's cognitive structure development and refining their learning strategies.

The development of a reference standard for expert cognitive structure could serve a valuable role in assessment for learning in the context of competency-based medical education (P. Harris et al., 2017). A reference ECS could be derived from experts currently in musculoskeletal clinical practice. It should not just reflect safety and requisite knowledge but also evidence-based practice and clinical efficacy, the standards by which physiotherapists will solve the challenges of global disability. Expert perceptual data could then be integrated with data derived from an outcomes management system to establish a reference standard of expert cognitive structure. Subsequent agreement analysis between student and expert may reflect the student's progression toward that which may ultimately reflect expert clinical practice.

Research Methodology

Data analysis in the current study indicated that two important aspects of methodology need to be considered in the discussion of cognitive structure. First, agreement analysis should consider several measures representing various aspects of agreement, such as reliability, accuracy, and association. Agreement is not solely an issue of correlation as raters can be highly correlated with little to no agreement. Individuals describing their perceptual experiences via psychometric scaling are, in essence, measurement tools of that individual's perceptual experience and state of declarative knowledge and semantic memory. As such, approaches to measurement like those used in assessing the level of agreement between measurement tools in a laboratory environment are indicated.

One of the biggest challenges in research methodology is understanding what semantic (Euclidean and graph-theoretic) distances represent cognitively and perceptually. Smaller semantic distances are equated with concepts that are perceptually closer to each other or having a higher degree of agreement or perceived relatedness. However, how these distances relate to differences in perception and knowledge organization between individuals is unknown. Results from the current study indicated that these measures may reflect contextual perceptual changes, although they should be used cautiously to represent cognitive structure.

Educational Practice

Wainer and Kaye (1974) described several challenges in education related to developmental psychology which are relevant to physiotherapy education to this day:

> A major goal of any course of instruction is the integration of concepts into a cohesive structure. The recall of facts and the ability to define concepts are fairly easy outcomes to assess, but the extent to which students understand interrelationships among the facts and concepts is problematic. Relationships are more difficult to define; there is far less agreement among instructors and among authors as to the meaningful structure of the subject matter; and the instructor is usually ambivalent about whether his students should be acquiring the structure, his structure, or their own structure. (p.591)

Physiotherapy students need to develop good cognitive structure to promote competency. This begins in foundational courses such as gross anatomy. As noted by Wainer and Kaye (1974), there must be meaningful structure. This may be reflected in the content structure (as defined by the course text and aligned with a domain expert), the instructor's structure (aligned with clinical practice or the perceived needs for success on the NPTE), or the student's structure. As noted in the results of this study, it appears that all are relevant in terms of cognitive structure. Using an expert's cognitive structure as a referential structure becomes increasingly essential as curricula evolve toward competency-based education (Bains & Kaliski, 2019; Lucey et al., 2018).

The impact of the instructor in terms of the development of a student's cognitive structure cannot be overstated. The current study found preliminary evidence of student cognitive structure aligning with that of the instructor (ECSI) more so than the domain expert (ECSD), although the best predictor of academic performance was the number of PFN common links with the domain experts. Housner et al. (1993) examined student cognitive structures in relation to those of the instructor and found increasing correspondence throughout the course. This highlights the importance of the instructor's academic content knowledge, pedagogical knowledge, and pedagogical content knowledge (Depaepe et al., 2013; Neumann et al., 2019; Shulman, 1987). However, Gess-Newsome et al. (2019) reported that in terms of pedagogical content knowledge, the instructor's academic content knowledge was the only variable directly correlated with a student's academic performance. Misconceptions are readily developed and hard to revise once chunked into long-term memory. It is important to note that misconceptions can arise as a function of the student's self-directed learning and an instructor's level of understanding or how they present the concept(s) to a student. This becomes increasingly important with foundational courses such as gross anatomy. Student cognitive structure will have its own internal coherence, but it can be favorably (or adversely) impacted by that of the cohort instructor. How a student represents their knowledge compared to an expert (specifically, their cohort instructor) may provide valuable information regarding the assessment of and for learning within the specific and broad educational contexts.

Physiotherapy education straddles two curricular concepts: time-based and competency based. However, the two concepts are often used interchangeably in health care education though most program lengths are fixed. However, the learning curve for both is not the same (Pusic et al., 2015). Much of the focus on assessment in physiotherapy education revolves around knowledge-based multiple-choice exams (aligned with the NPTE) or clinical/practical exams such as Objective Structured Clinical Examinations (OSCE). Students will adapt their learning strategies to what they perceive to be expected of them given the testing environment; this often emphasizes surface learning strategies (Rovers et al., 2019). Course grades composed of knowledge-based multiple-choice exams and practical exams often have limited validity and are used primarily as a marker for success in the educational system. Teaching strategies now implement simulations designed to reflect clinical scenarios. However, effective simulations are dependent upon an understanding of cognitive task analysis, which is derived from the practical application of cognitive structure development. Although these teaching and assessment strategies may be considered to reflect the development of competency, physiotherapy education does not examine the actual cognitive structure of the developing clinician. This cognitive structure drives clinical reasoning and diagnostic thinking and not just the ability to render a successful performance in a simulation or practical exam that signifies perceived proficiency or competency.

Assessments for learning, designed around data modeling strategies such as MDS and PFN, may provide a window into the dynamic development of cognitive structure. The current study may provide some potential insight into how cognitive structure could be used as an assessment strategy to refine a student's learning path. MDS- and PFN-derived parameters appear to display a relationship between expert subgroups and between students and experts and are also predictive of academic grades within the context of this specific gross anatomy course. A subset of these parameters may be a potentially relevant assessment *of* learning, assessment *for* learning, or at least progression toward success on the unit exams based on the predictor variables noted. However, the perceptual and educational relevance to the derived quantitative parameters is unclear. If there is a minimal interpretable difference for parameters such as level of agreement and its impact on learning, then it is unknown at this point. This makes an evaluation of the derived parameters and their differences more theoretical and (presently) less practical as a formative or summative assessment.

Positive Social Change

The current study's findings suggest a potential role in promoting positive social change on many levels – individual, domain specific, and physiotherapy centric. Self-directed learning places a higher demand on the physiotherapy student in developing effective cognitive structures (van Lankveld et al., 2019). Liu et al. (2019) noted the importance of cognitive structure in adaptive learning and the development of learning paths to foster individual learning, emphasizing the importance of the cognitive mechanisms of the individual and not on outdated learning styles. This becomes increasingly important in a domain such as gross anatomy, in which a wide range of teaching and learning strategies have been shown to produce similar outcomes (Estai & Bunt, 2016; Losco et al., 2017; Wilson, Brown, et al., 2019). Cognitive mechanisms associated with adaptive learning may not currently be addressed effectively. Knowledge level and knowledge structure are necessary components of adaptive learning, though the latter is rarely examined as a part of physiotherapy curricula in either a formative or summative fashion. Using data modeling strategies such as MDS and PFN to represent cognitive structure in physiotherapy students learning gross anatomy could provide an individualized self-assessment and reflection aligned with self-directed learning and its role in 21st-century education.

One significant finding of the study which has a potential impact on health professions education is the notable differences between the expert subgroups and the student's alignment with a particular subgroup. There is much debate within the anatomical education community about the lack of Ph.D. programs to train anatomists and the urgency of continuing anatomical education. However, the results of this study present a different picture in that a clinician with anatomical education experience and training may be better suited to provide more context-specific anatomical teaching to other student clinicians that is also aligned with the stated goals of the curriculum and the

National Physical Therapy Exam. This brings into question the importance of a clinical degree in educating those who will eventually be practicing as clinicians. Although the sample size was small and limited the findings' generalizability, it provides a foundation for further research.

The current study highlights the dichotomy between creating competent clinicians and passing the licensure exam to ensure patient safety. Physiotherapists in active clinical practice were surveyed to define relevant anatomical concepts in the clinical realm to enhance content validity. Students have a greater agreement with instructor ECS (becoming more "expert-like") as compared to domain expert ECS, yet the greatest predictor of the unit grade was the PFN common links with domain expert ECS. A domain expert may exhibit knowledge that is better aligned with the course text and subsequently knowledge-based exams. A clinician with ten or more years of experience may represent competency, clinical efficacy, and expertise, but these traits may not necessarily be consistent with the goals of the current physiotherapy curriculum defined by the NPTE. In the end, assessing agreement with clinical experts may not provide a reasonable predictor of academic performance relevant to the NPTE. However, it remains to be seen if the development of cognitive structure that aligns with ECSI does reflect the cognitive structure of physiotherapists with the clinical reasoning and diagnostic thinking reflective of effective musculoskeletal clinical practice.

Several significant assumptions have been made in physiotherapy education regarding teaching, learning, and subsequent diagnostic thinking and clinical performance. Many of these elements are defined by the nature of program accreditation and the NPTE. The primary goal of entry-level professional education, and one of the most important outcome measures in the accreditation of physiotherapy curricula, is the first-time pass rate on the NPTE. Decisions regarding curricular developments and faculty retention are often based on an educational program's need to have a high first-time pass rate. The purpose of the NPTE is to "assess your basic entry-level competence after graduation from an accredited program" (Federation of State Boards of Physical Therapy, 2021b). This is aligned with their stated mission to "protect the public" (Federation of State Boards of Physical Therapy, 2021a). The NPTE is a multiple-choice exam that focuses primarily on the components of clinical practice that ensure patient safety more so than clinical efficacy. The goals of public safety, requisite knowledge, and clinical competency may be aligned, but they are not synonymous; a clinician can provide safe care without having a high level of clinical competency by simply abiding by the words of Hippocrates: "do no harm."

The current study may serve as a catalyst for the reevaluation of physiotherapy curriculum in terms of the role of competency-based education based on the cognitive structure of the developing clinician. Competency-based education requires both formative and summative assessments that promote assessment for learning and reflect either a change in competency or alignment with an expert or evidence-based practices. Cognitive structures are dynamic systems and undergo revisions and adaptations based on the scaffolding of new knowledge and the revision of prior knowledge. The representation of cognitive structure via data modeling strategies may provide an innovative approach to better understanding how the developing clinician perceives the world, how they organize their knowledge, and how it compares to experts in musculoskeletal clinical practice. MDS may provide a spatial overview of the individual’s representation of anatomical concepts and structures. In contrast, PFN may provide a more granular representation of the association and linking of these anatomical concepts and structures. These strategies may provide an important perspective on how students organize their knowledge compared to an expert as early as their foundational courses, including gross anatomy. The dichotomy between assessment, competency, knowledge, and their relationship to the development of cognitive structure reveals a need for fundamental change not only in how gross anatomy is taught to physiotherapy students but how it is assessed and vertically integrated into the scope of competency-based physiotherapy education leading to long-term transfer to clinical practice.

Limitations of the Study

There are several limitations of the study. The COVID-19 pandemic of 2020-2021 created a unique educational experience for students. Two modes of delivery were initially going to be examined: one based on on-campus experiences and one based on remote learning with face-to-face lab experiences. Students will often self-select their mode of delivery based on their approach to self-directed learning and prior experiences with remote learning. However, during the pandemic, all students were forced to partake in a fully remote learning mode of delivery, which may have provided unforeseen learning challenges to those expecting a different delivery mode as the basis for their physiotherapy educational experience. Although the consistency in the mode of delivery across all students created a potential benefit in terms of the consistency of data collection, it may have also inadvertently added a limiter for those students who were not expecting the change in their model of delivery.

The predominant limitation of the study was the small sample size of expert (domain experts: n = 3; instructors: n = 4) and student (n = 31) samples. A priori power calculations indicated that correlational analysis would require a sample size of 15 (large effect size) to 34 (moderate effect size). In comparison, multiple regression would require a sample size of 36 (large effect size) to 77 (moderate effect size). The correlational analysis had sufficient power though multiple regression was mildly underpowered. However, sample sizes were consistent with previous research that also reported large effect sizes based on calculated r^2 (>0.5) and η^2 (>0.14) values (see Goldsmith et al., 1991; Neiles et al., 2016; Stevenson et al., 2016). Finally, a priori power calculations indicated that paired sample t tests would require a sample of 15 to 34; this made within-group comparisons (for example, between students attaining high grades > 90 and those attaining poor grades < 75) unrealistic and highly under-powered as the former group had eight students and the latter group had six students. The small sample size limited the generalizability of the findings and impacted the power of this component of the statistical analysis. As such, only preliminary observations were reported in the case of student group comparisons of academic performance.

The survey instrument employed in the study was unique as the physiotherapists in musculoskeletal clinical practice defined it. Larger sample size would have provided greater insight into the perceived importance and relevance of anatomical concepts. The small sample size for physiotherapists (n = 12) limited the conclusions drawn regarding any degree of interrater agreement in item selection and rank ordering which ultimately defined the content and construct validity of the survey instrument.

Paired comparisons formed the basis for the raw proximity data that assessed perceptual differences of anatomical concepts. This approach has a long history of use in psychological scaling (see Brown & Peterson, 2009; Thurstone, 1927); however, paired comparisons may also be viewed as a rather rudimentary means of examining perceptual differences for high-level constructs such as cognitive structure. An item list for paired comparisons may have produced artificial representations of the construct or represented some other perceptual or organizational construct. Further research on the psychometric properties of paired comparisons and other strategies for assessing perceptual and organizational differences in the health care professions is indicated.

Criterion standard one, ECS, was like the “knowledge indices” used by Goldsmith et al. (1991). Criterion standard two, unit grade, was used in this study to maintain some degree of consistency with previous research. However, lecture and lab exams have generally not been assessed for validity, making a unit grade based on these assessment tools potentially problematic. Exams may become reflections of an instructor’s academic content knowledge, pedagogical knowledge, and pedagogical content knowledge (Neumann et al., 2019) or a DPT program’s perception of what is valid and necessary as a prerequisite for later courses in the curriculum or preparation for the National Physical Therapy Examination (NPTE). The two criterion standards used in the current study may not fully represent equivalent criteria. There may be a disparity between ECS (a criterion aligned with knowledge organization, competency, and expertise) and unit grade/academic performance (a criterion aligned with knowledge retrieval). However, they do share similar features.

There was a significant challenge due to the lack of consistent operational definitions in the scientific literature regarding agreement and how this is quantified in a statistically sound and consistent fashion. The quantitative representation of cognitive structure could be viewed in terms of assessing any measurement tool; however, this measurement tool is internalized to the rater. Significant discrepancies exist in terms of the description of agreement, the statistical tests used to assess it, and clearly meeting the assumptions of the statistical test used. At the level of student-expert comparisons (reliability, accuracy, and association), there is little consistency in the literature regarding the meaning and practical application. There are few reports (if any) of what would be considered a minimal perceptible or interpretable change in many of the agreement measures, their relationship to the perceptual data or MDS and PFN implementations of the raw proximity data, or the impact on cognitive structure development, meaningful learning, or academic performance.

One of the challenges with MDS is the direct visual comparison of stimulus spaces. Comparison to a reference standard representation is difficult statistically. An MDS configuration will seek to find the best fit amongst multiple matrices; if a student and expert are used concurrently within the analysis, the resultant configuration will be a composite of both. Although individual differences scaling produces a group space and individual spaces, the group space is the best fit for the entire data set. Assessing the degree of configurational similarity (see Borg & Leutner, 1985) mathematically was beyond the scope of the current analysis. However, several authors have proposed doing so via mathematical transformations such as Procrustes rotation (see Borg & Leutner, 1985; Egli, Streule, & Lage, 2008; Peres-Neto & Jackson, 2001; Rosas, 2017). This was beyond the scope of many health professions researchers and beyond the level of practicality for the educator should these strategies be employed in an educational environment.

Finally, direct interpretation of the regression coefficients within the context of the raw proximity data and data modeling strategies is problematic from a practical perspective. Semantic distances, be they MDS Euclidean distances or PFN graph-theoretic distances, are viewed in a purely referential context. For example, as MDS Euclidean distances decrease, items are closer together perceptually; as PFN graph-theoretic distances decrease, there is a more direct and shorter pathway between concepts perceptually. How these distances relate to individual differences in perception, knowledge organization, and learning is unknown.

Recommendations

Several recommendations for further research emerged based on the results of the current study. Foremost of these recommendations is the need for improved operational definitions of cognitive structure regarding what is being represented and how it is being represented. To measure a construct, you must know what you are measuring. Refined definitions can then be used to delineate and differentiate perceptual changes related to the construct and the psychometrics used for measurement and scaling purposes. MDS and PFN provide preliminary evidence of a quantitative, indirect representation of cognitive structure of physiotherapy students learning gross anatomy that is grounded in the operational definitions of the current study. It is unknown if this reflects an indirect representation of cognitive structure or another construct. The neuroanatomical construct of cognitive structure may require similar (but not identical) definitions as compared to those used in a learning context. Further research to examine these issues in the context of neuroscience and education is critical.

The current study results indicated a need to examine cognitive structure further and its practical application in education. Several aspects of cognitive structure could be examined: the test-retest reliability within any given individual, the changes noted within an individual over time and minimally detectable or relevant changes in cognitive structure that indicate key milestones in competency and academic performance. Test-retest reliability would indicate the consistency of the representation and enhance its validity for practical use. Changes in cognitive structure over time (for example, within the duration of a course of study or between admission, graduation, and ten years of clinical practice) could provide evidence of learning and the development of diagnostic thinking. Finally, understanding a minimal interpretable change could provide evidence of learning benchmarks and progression towards academic proficiency and competency.

The psychometrics of rank-ordering by the physiotherapist must be better understood, especially in relation to their own cognitive structure. The current study indicated that physiotherapists had poor interrater reliability in rank ordering anatomical concepts based on clinical relevance. Further examination of a group of physiotherapists in terms of rank-ordering and the subsequent use of paired comparisons to derive their cognitive structure representation that grounds their rankings could provide greater insight into perceived importance and relevance based on an individual's cognitive structure.

The results of the current study provided preliminary evidence of differences between expert subgroups. This was subsequently related to the level of agreement with the student. An expert cognitive structure may have value as a referent cognitive structure, making the expert subgrouping important (Acton et al., 1994). Examining a broad population of experts could provide the basis for comparison between expert subgroups such as clinicians (new graduates and experienced), domain experts, and clinical instructors. Compilation of this expert perceptual data in a database could expand the understanding of cognitive structure and expertise in physiotherapists. However, this assumes that the assessment and representation of cognitive structure have clear operational definitions.

The current study used a survey of 20 anatomical concepts and items. However, some of the paired comparisons may have greater value in predicting the level of agreement or as predictor variables based on higher correlations. Factor analysis could help establish which items and paired comparisons are better predictors, and the survey could then use a smaller number of items. This may lead to developing a gross anatomy concept inventory along the lines of the Force Concept Inventory used in physics education (see Hestenes et al., 1992). A concept inventory could be used as both a gross anatomy course pretest and assessment of and for learning (see Leppink, 2020).

Conclusion

Preliminary evidence indicates that data modeling strategies such as MDS and PFN have potential as a visual and quantitative representation of cognitive structure. Specifically, using these strategies appears to have some value in describing the cognitive structure of physiotherapy students learning gross anatomy compared to experts and highlights the importance of clinical practice instead of just a deeper understanding of the gross anatomy domain. The visual and quantitative representation of cognitive structure via MDS and PFN data modeling is promising in terms of criterion-related validity and as a foundation for further research on agreement analysis between the cognitive structures of students and experts. It is unclear if these representations genuinely reflect cognitive structure or another educational, clinical, or cognitive construct. The significance of changes in these derived parameters over time is unknown. The study’s findings provide critical perspectives on the real-world relevance and practical application of cognitive structure in competency-based education. The development of expertise reflected in the agreement with expert cognitive structure serves as an integral component of the learning process that begins with foundational courses like gross anatomy. The representation of cognitive structure in physiotherapy students learning gross anatomy may serve as a valuable first step in better understanding this process and innovation in true competency-based education of physiotherapists.

References

Abd El-Hay, S. A., El Mezayen, S. E., & Ahmed, R. E. (2018). Effect of concept mapping on problem solving skills, competence in clinical setting and knowledge among undergraduate nursing students. *Journal of Nursing Education and Practice*, *8*(8), 34–46. https://doi.org/10.5430/jnep.v8n8p34

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, *86*(2), 303–311. https://doi.org/10/d7j4h7

Aggarwal, R., & Ranganathan, P. (2016). Common pitfalls in statistical analysis: The use of correlation techniques. *Perspectives in Clinical Research*, *7*(4), 187–190. https://doi.org/10.4103/2229-3485.192046

Aggarwal, R., & Ranganathan, P. (2017). Common pitfalls in statistical analysis: Linear regression analysis. *Perspectives in Clinical Research*, *8*(2), 100–102. https://doi.org/10.4103/2229-3485.179438

Agra, G., Formiga, N. S., Oliveira, P. S. D., Costa, M. M. L., Fernandes, M. D. G. M., & Nóbrega, M. M. L. D. (2019). Analysis of the concept of meaningful learning in light of the Ausubel's theory. *Revista Brasileira de Enfermagem*, *72*(1), 248–255. https://doi.org/10.1590/0034-7167-2017-0691

Alfayoumi, I. (2019). The impact of combining concept-based learning and concept-mapping pedagogies on nursing students' clinical reasoning abilities. *Nurse Education Today*, *72*, 40–46. https://doi.org/10.1016/j.nedt.2018.10.009

Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, *23*(6), 1825–1831. https://doi.org/10/f9hdnv

Amin, A., & Iqbal, J. (2019). Effective ways to learn and retain gross anatomy. *Pakistan Armed Forces Medical Journal*, *69*(3), 708–714.

Amith, M., Cohen, T., Cunningham, R., Savas, L. S., Smith, N., Cuccaro, P., Gabay, E., Boom, J., Schvaneveldt, R., & Tao, C. (2020). Mining HPV vaccine knowledge structures of young adults from Reddit using distributional semantics and Pathfinder networks. *Cancer Control*, *27*(1), 1–16. https://doi.org/10/gg8whm

Amith, M., Cunningham, R., Savas, L. S., Boom, J., Schvaneveldt, R., Tao, C., & Cohen, T. (2017). Using Pathfinder networks to discover alignment between expert and consumer conceptual knowledge from online vaccine content. *Journal of Biomedical Informatics, 74*, 33–45. https://doi.org/10/gc5zd4

Anand, M. K., Singh, O., & Chhabra, P. K. (2018). Learning with concept maps versus learning with classical lecture and demonstration methods in neuroanatomy—A comparison. *National Journal of Clinical Anatomy*, *7*(2), 95–102. https://doi.org/10.1055/s-0040-1701785

Anderson, J. R. (1976). *Language, memory, and thought*. Erlbaum.

Anderson, J. R. (1980). *Concepts, propositions, and schemata: What are the cognitive units?* [Technical report]. Carnegie-Mellon University.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355–365. https://doi.org/10/dfv7rv

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford

University Press. https://doi.org/10.1093/acprof:oso/9780195324259.001.0001

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., Carter, C., Fincham, J., Qin, Y., Ravizza, S., & Rosenberg-Lee, M. (2008). Using fMRI to test models of complex cognition. *Cognitive Science*, *32*(8), 1323–1348. https://doi.org/10/dwgjft

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*(4), 728–748. https://doi.org/10/dfggnn

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, *12*(4), 439–462. https://doi.org/10/bmhrqt

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*(3), 221–256. https://doi.org/10/dc5k2h

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 1–27). Routledge.

Anderson, S. J., Hecker, K. G., Krigolson, O. E., & Jamniczky, H. A. (2018). A reinforcement-based learning paradigm increases anatomical learning and retention—a neuroeducation study. *Frontiers in Human Neuroscience*, *12*, Article 38, 1–10. https://doi.org/10.3389/fnhum.2018.00038

Arzy, S., & Schacter, D. L. (2019). Self-agency and self-ownership in cognitive mapping. *Trends in Cognitive Sciences*, *23*(6), 476–487. https://doi.org/10/gf2tz5

Aslaksen, K., & Lorås, H. (2019). Matching instruction with modality-specific learning style: Effects on immediate recall and working memory performance. *Education Sciences*, *9*(1), 1–11. https://doi.org/10/ggwncq

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*(4), 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Ausubel, D. G. (1963). Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education*, *14*(2), 217–222. https://doi.org/10.1177/002248716301400220

Azzarello, J. (2007). Use of the Pathfinder scaling algorithm to measure students' structural knowledge of community health nursing. *Journal of Nursing Education*, *46*(7), 313–318. https://doi.org/10.3928/01484834-20070701-05

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10/cz4znd

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London*, *302*(1110), 311–324. https://doi.org/10.1098/rstb.1983.0057

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Bains, M., & Kaliski, D. Z. (2019). An anatomy workshop for improving anatomy self-efficacy and competency when transitioning into a problem-based learning, Doctor of Physical Therapy program. *Advances in Physiology Education*, *44*(1), 39–49. https://doi.org/10/ggz7fw

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, *102*(1), 1–8. https://doi.org/10.1007/s10649-019-09908-4

Balloo, K., Pauli, R., & Worrell, M. (2016). Individual differences in psychology undergraduates' development of research methods knowledge and skills. *Procedia - Social and Behavioral Sciences*, *217*, 790–800. https://doi.org/10/gg35pj

Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics*, *17*(4), 529–569. https://doi.org/10.1080/10543400701376480

Bärnighausen, T., Tugwell, P., Røttingen, J. A., Shemilt, I., Rockers, P., Geldsetzer, P., Lavis, J., Grimshaw, J., Daniels, K., Brown, A., Bor, J., Tanner, J., Rashidian, A., Barreto, M., Vollmer, S., & Atun, R. (2017). Quasi-experimental study designs series—paper 4: uses and value. *Journal of Clinical Epidemiology*, *89*, 21–29. https://doi.org/10.1016/j.jclinepi.2017.03.012

Barrett, J., & Liebman, C. (2020). Gaining competence in musculoskeletal care as a primary care provider. *The Journal for Nurse Practitioners*, *16*(1), 44–47. https://doi.org/10/ghgjp9

Battaglia, M. P. (2008). Nonprobability sampling. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 524–527). Sage Publications. https://doi.org/10.4135/9781412963947.n337

Bayliss, J., Thomas, R. M., & Eifert-Mangine, M. (2017). Pilot study: What measures predict first time pass rate on the National Physical Therapy Examination? *International Journal of Allied Health Sciences and Practice*, *15*(4), Article 1, 1–12.

Bechtel, W. (2019). Resituating cognitive mechanisms within heterarchical networks controlling physiology and behavior. *Theory & Psychology*, *29*(5), 620–639. https://doi.org/10/gg8q5s

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10/gfgkz9

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*, 1–11. https://doi.org/10/gfkrrh

Bergman, E. M., Prince, K. J. A. H., Drukker, J., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2008). How much anatomy is enough? *Anatomical Sciences Education*, *1*(4), 184–188. https://doi.org/10.1002/ase.35

Bhuvaneshwari, B., & Kavitha, A. (2017). Investigations on the brain connectivity patterns in progression of Alzheimer's disease using functional MR imaging and graph theoretical measures. *2017 IEEE 16th International Conference on Cognitive Informatics Cognitive Computing*, 151–160. https://doi.org/10/gg3tw5

Biemans, H. J. A., & Simons, P. R.-J. (1996). Contact-2: A computer-assisted instructional strategy for promoting conceptual change. *Instructional Science*, *24*(2), 157–176. https://doi.org/10/d6sx3c

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound in Obstetrics & Gynecology*, *22*(1), 85–93. https://doi.org/10.1002/uog.122

Bonebright, T. L., Miner, N. E., Goldsmith, T. E., & Caudell, T. P. (2005). Data collection and analysis techniques for evaluating the perceptual qualities of auditory stimuli. *ACM Transactions on Applied Perception*, *2*(4), 505–516. https://doi.org/10/dtwf7w

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Borg, I., Groenen, P. J. F., & Mair, P. (2018). *Applied multidimensional scaling and unfolding*. Springer. https://doi.org/10.1007/978-3-319-73471-2

Borg, I., & Leutner, D. (1985). Measuring the similarity of MDS configurations. *Multivariate Behavioral Research*, *20*(3), 325–334. https://doi.org/10.1207/s15327906mbr2003_6

Borneman, M. J. (2012). Criterion problem. In N. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 288–291). SAGE Publications. https://doi.org/10.4135/9781412961288

Borst, J. P., & Anderson, J. R. (2017). A step-by-step tutorial on using the cognitive architecture ACT-R in combination with fMRI data. *Journal of Mathematical Psychology*, *76*, 94–103. https://doi.org/10/f9wx6v

Bottini, R., & Doeller, C. F. (2020). Knowledge across reference frames: Cognitive maps and image spaces. *Trends in Cognitive Sciences*, *24*(8), 606–619. https://doi.org/10/gg3t6f

Branaghan, R. J. (1990). Pathfinder networks and multidimensional spaces: Relative strengths in representing strong associates. In Schvaneveldt, R. W. (Ed.) *Pathfinder associative networks: Studies in knowledge organization* (pp. 111–120). Ablex Publishing.

Brenner, E., Chirculescu, A. R. M., Reblet, C., & Smith, C. (2015). Assessment in anatomy. *European Journal of Anatomy*, *19*(1), 105–124.

Bressington, D. T., Wong, W., Lam, K. K. C., & Chien, W. T. (2018). Concept mapping to promote meaningful learning, help relate theory to practice and improve learning self-efficacy in Asian mental health nursing students: A mixed-methods pilot study. *Nurse Education Today*, *60*, 47–55. https://doi.org/10.1016/j.nedt.2017.09.019

Briggs, A. M., Shiffman, J., Shawar, Y. R., Åkesson, K., Ali, N., & Woolf, A. D. (2020). Global health policy in the 21st century: Challenges and opportunities to arrest the global disability burden from musculoskeletal health conditions. *Best Practice & Research Clinical Rheumatology*, 101549, 1–24. https://doi.org/10/gg5xzm

Brown, T. C., & Peterson, G. L. (2009). *An enquiry into the method of paired comparison: reliability, scaling, and Thurstone's Law of Comparative Judgment*. General Technical Report. RMRS-GTR-216WWW. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. https://doi.org/10.2737/RMRS-GTR-216

Buitrago, M., & Chiappe, A. (2019). Representation of knowledge in digital educational environments: A systematic review of literature. *Australasian Journal of Educational Technology*, *35*(4), 46–62. https://doi.org/10.14742/ajet.4041

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, *17*(2), 444–472. https://doi.org/10/bdwxvj

Burgess, N. (2014). The 2014 Nobel Prize in Physiology or Medicine: A spatial model for cognitive neuroscience. *Neuron*, *84*(6), 1120–1125. https://doi.org/10/gg3xrb

Burkholder, G. J., Cox, K. A., & Crawford, L. M. (2016). *The scholar-practitioner's guide to research design*. Laureate Publishing.

Busing, F. M. T. A., Commandeur, J. J., Heiser, W. J., Bandilla, W., & Faulbaum, F. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *Softstat*, *97*, 67–74.

Camina, E., & Güell, F. (2017). The neuroanatomical, neurophysiological and psychological basis of memory: Current models and their origins. *Frontiers in*

Pharmacology, *8*, Article 438, 1–16. https://doi.org/10/ggr7wz

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company.

Casas-García, L. M., & Luengo-González, R. (2012). The study of the pupil's cognitive structure: The concept of angle. *European Journal of Psychology of Education*, *28*(2), 373–398. https://doi.org/10/ggsw2p

Castro, N., & Siew, C. S. Q. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings of the Royal Society A*, *476*(2238), 1–25. https://doi.org/10/gg274w

Castro-Alonso, J. C., & Atit, K. (2019). Different abilities controlled by visuospatial processing. In J. C. Castro-Alonso (Ed.), *Visuospatial processing for education in health and natural sciences* (pp. 23–51). Springer International Publishing. https://doi.org/10.1007/978-3-030-20969-8_2

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, *9*. Article 401, 1–16. https://doi.org/10/ggx68p

Chaturvedi, H. K., & Bajpai, R. C. (2015). Evaluation of interrater agreement and interrater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, *41*(3), 20–27.

Chen, C. (1997). Tracking latent domain structures: An integration of Pathfinder and latent semantic analysis. *AI & Society*, *11*(1–2), 48–62. https://doi.org/10/ccn2r5

Chmielewski, P. P. (2020). New Terminologia Anatomica highlights the importance of clinical anatomy. *Folia Morphologica*, *79*(1), 15–20. https://doi.org/10/ghgtdr

Choi-Lundberg, D. L., Williams, A.-M. M., & Zimitat, C. (2017). A psychometric evaluation of the anatomy learning experiences questionnaire and correlations with learning outcomes. *Anatomical Sciences Education*, *10*(6), 514–527. https://doi.org/10.1002/ase.1693

Choudhury, B., & Freemont, A. (2017). Assessment of anatomical knowledge: Approaches taken by higher education institutions. *Clinical Anatomy*, *30*(3), 290–299. https://doi.org/10.1002/ca.22835

Clarkson, M., & Whipple, M. (2018). Does the Foundational Model of Anatomy ontology provide a knowledge base for learning and assessment in anatomy education? *Proceedings of the 9th International Conference on Biological Oncology*, *2285*(24), 1–6.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Collins, A., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/11571-006

Comin, C. H., Peron, T. K. D., Silva, F. N., Amancio, D. R., Rodrigues, F. A., & Costa, L. D. F. (2016). Complex systems: Features, similarity and connectivity. *Physics Reports*, *861*, 1–41. https://doi.org/10.1016/j.physrep.2020.03.002

Connor, O., Weinger, M. B., Cooke, N. J., & Slagle, J. (2004). Using psychological scaling techniques to assess clinical expertise in anesthesiology. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 48, No. 15, pp. 1746–1750). SAGE Publications. https://doi.org/10.1037/e577092012-025

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a grid-like code. *Science*, *352*(6292), 1464–1468. https://doi.org/10/f8rt9r

Cook, C., Engelhard, C., Landry, M. D., & McCallum, C. (2015). Modifiable variables in physical therapy education programs associated with first-time and three-year National Physical Therapy Examination pass rates in the United States. *Journal of Educational Evaluation for Health Professions*, *12*(44), 1–8. https://doi.org/10/ghgjfq

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 538–549. https://doi.org/10/c49cx7

Craik, K. J. W. (1943). *The Nature of Explanation*. CUP Archive.

Crandall, B. W., & Hoffman, R. R. (2013). Cognitive task analysis. In J.D. Lee, A. Kirlik, & M. J. Dainoff (Eds.), *The Oxford handbook of cognitive engineering* (pp. 229–239). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199757183.013.0014

Crompvoets, E. A. V., Béguin, A. A., & Sijtsma, K. (2020). Adaptive pairwise comparison for educational measurement. *Journal of Educational and Behavioral Statistics*, *45*(3), 316–338. https://doi.org/10/ggjq9t

Cumming, G. S. (2016). Heterarchies: Reconciling networks and hierarchies. *Trends in Ecology & Evolution*, *31*(8), 622–632. https://doi.org/10/f8w98p

Curtis, M. B., & Davis, M. A. (2003). Assessing knowledge structure in accounting education: An application of Pathfinder associative networks. *Journal of Accounting Education*, *21*(3), 185–195. https://doi.org/10/dj9s75

Daley, B. J., Durning, S. J., & Torre, D. M. (2016). Using concept maps to create meaningful learning in medical education. *MedEdPublish*, *5*, 1–29. https://doi.org/10.15694/mep.2016.000019

D'Antoni, A. V., Mtui, E. P., Loukas, M., Tubbs, R. S., Zipp, G. P., & Dunlosky, J. (2019). An evidence-based approach to learning clinical anatomy: A guide for medical students, educators, and administrators. *Clinical Anatomy*, *32*(1), 156–163. https://doi.org/10.1002/ca.23298

Davies, M. (2011). Concept mapping, mind mapping, and argument mapping: What are the differences and do they matter? *Higher Education*, *62*(3), 279–301. https://doi.org/10.1007/s10734-010-9387-6

Davison, M. L., & Sireci, S. G. (2000). Multidimensional scaling. In Tinsley, H. E., & Brown, S. D. (Eds.). *Handbook of applied multivariate statistics and mathematical modeling* (pp. 323–352). Academic Press. https://doi.org/10.1016/B978-012691360-6/50013-6

Day, E. A., Arthur, W., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, *86*(5), 1022–1033. https://doi.org/10/ctv59w

Dayal, M., Owens, J., Gibson, W., & Strkalj, G. (2017). Anatomical knowledge retention in physiotherapy students: A preliminary assessment. *International Journal of*

Anatomy and Research, *5*(1), 3474–3479. https://doi.org/10.16965/ijar.2016.485

De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2016). Structure at every scale: A semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General*, *145*(9), 1228–1254. https://doi.org/10/f84m54

Delgado, Á. H. D. A., Almeida, J. P. R., Mendes, L. S. B., Oliveira, I. N. D., Ezequiel, O. D. S., Lucchetti, A. L. G., & Lucchetti, G. (2018). Are surface and deep learning approaches associated with study patterns and choices among medical students? A cross-sectional study. *Sao Paulo Medical Journal*, *136*(5), 414–420. https://doi.org/10.1590/1516-3180.2018.0200060818

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001

de Raadt, A., Warrens, M. J., Bosker, R. J., & Kiers, H. A. (2021). A comparison of reliability coefficients for ordinal rating scales. *Journal of Classification*, 1–25. https://doi.org/10.1007/s00357-021-09386-5

Deshatty, D. D., & Mokashi, V. (2013). Mind maps as a learning tool in anatomy. *International Journal of Anatomy and Research*, *1*(2), 100–103.

DiCerbo, K. E. (2007). Knowledge structures of entering computer networking students and their instructors. *Journal of Information Technology Education*, *6*(1), 263–277. https://doi.org/10/ggzv3r

Dobson, J., Linderholm, T., & Perez, J. (2018). Retrieval practice enhances the ability to evaluate complex physiology information. *Medical Education*, *52*(5), 513–525. https://doi.org/10/gg2n25

Dobson, J. L., Perez, J., & Linderholm, T. (2017). Distributed retrieval practice promotes superior recall of anatomy information. *Anatomical Sciences Education*, *10*(4), 339–347. https://doi.org/10/ggfn4v

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 1–17. https://doi.org/10.1186/2193-1801-2-222

Dozortsev, V., Nazin, V., Oboznov, A., & Mironova, A. (2017). Structural knowledge as an evaluation instrument of trainees progress in learning. *2017 IEEE 11th International Conference on Application of Information and Communication Technologies* (pp. 1–5). IEEE. https://doi.org/10/ggz9jz

Dry, M. J., & Storms, G. (2009). Similar but not the same: A comparison of the utility of directly rated and feature-based similarity measures for generating spatial models of conceptual data. *Behavior Research Methods*, *41*(3), 889–900. https://doi.org/10/bvcghn

Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. (2020). Semantic memory and the hippocampus: Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in Human Neuroscience*, *13*, Article 471, 1–17. https://doi.org/10/gg4sz6

Dzemyda, G., Kurasova, O., & Žilinskas, J. (2013). *Multidimensional data visualization* (pp. 113–177). Springer New York. https://doi.org/10.1007/978-1-4419-0236-8_4

Egli, S., Riedel, M., Möller, H.-J., Strauss, A., & Läge, D. (2009). Creating a map of psychiatric patients based on psychopathological symptom profiles. *European Archives of Psychiatry and Clinical Neuroscience*, *259*(3), 164–171. https://doi.org/10/b599n8

Egli, S., Streule, R., & Lage, D. (2008). The structure-based expert model of the mental disorders - a validation study. *Psychopathology*, *41*(5), 286–293. https://doi.org/10/fq73g8

Eichenbaum, H. (2017). Memory: Organization and control. *Annual Review of Psychology*, *68*(1), 19–45. https://doi.org/10/ggtc5b

Elvén, M., Hochwälder, J., Dean, E., & Söderlund, A. (2019). Predictors of clinical reasoning using the Reasoning 4 Change instrument with physical therapist students. *Physical Therapy*, *99*(8), 964–976. https://doi.org/10.1093/ptj/pzz044

Estai, M., & Bunt, S. (2016). Best teaching practices in anatomy education: A critical review. *Annals of Anatomy*, *208*, 151–157. https://doi.org/10.1016/j.aanat.2016.02.010

Farahani, F. V., Karwowski, W., & Lighthall, N. R. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review. *Frontiers in Neuroscience*, *13*, Article 585, 1–27. https://doi.org/10/ggt6n6

Farrokhi, F., & Mahmoudi-Hamidabad, A. (2012). Rethinking convenience sampling: Defining quality criteria. *Theory and Practice in Language Studies*, *2*(4), 784–792. https://doi.org/10.4304/tpls.2.4.784-792

Fatima, F. (2020). Differences in diagnostic reasoning of expert and novice residents: Underlying reasons and suggestions for improvements. *Journal of the Dow University of Health Sciences (JDUHS)*, *14*(1), 38–41. https://doi.org/10.36570/jduhs.2020.1.772

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Federation of State Boards of Physical Therapy. (2021a). *Mission and vision: Promoting safety and competence*. Retrieved from https://www.fsbpt.org/About-Us/Mission-Vision

Federation of State Boards of Physical Therapy. (2021b). *National exam (NPTE)*. Retrieved from https://www.fsbpt.org/Secondary-Pages/Exam-Candidates/National-Exam-NPTE

Fink, A. (2010). *Survey Research Methods* (pp. 152–160). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.00296-7

FIPAT. (2019). *Terminologia anatomica. International anatomical terminology* (3rd ed.). Georg Thieme Verlag.

Gao, Y., Zhao, B., & Qian, M. (2019). On the compare of evaluation of deep learning in education. *2019 IEEE International Conference on Computer Science and Educational Informatization* (pp. 205–208). https://doi.org/10/gg34bt

Gärdenfors, P. (1996). Conceptual spaces as a basis for cognitive semantics. In A. Clark, J. Ezquerro, & J. M. Larrazabal (Eds.), *Philosophy and cognitive science: Categories, consciousness, and reasoning* (pp. 159–180). Springer Netherlands. https://doi.org/10.1007/978-94-015-8731-0_8

Gärdenfors, P. (2004). Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, *2*(2), 9–27.

Gärdenfors, P. (2017). Semantic knowledge, domains of meaning and conceptual spaces. In P. Meusburger, B. Werlen, & L. Suarsana, L. (Eds.), *Knowledge and action* (Vol. 9, pp. 203–220). Springer International Publishing. https://doi.org/10.1007/978-3-319-44588-5

Gardner, G. E., Ellen Lohr, M., Bartos, S., & Reid, J. W. (2019). Comparing individual and group-negotiated knowledge structures in an introductory biology course for majors. *Journal of Biological Education*, *53*(3), 274–287. https://doi.org/10.1080/00219266.2018.1469537

Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. M. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, *41*(7), 944–963. https://doi.org/10.1080/09500693.2016.1265158

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114.

https://doi.org/10/f5s3hr

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Giguère, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. *Tutorials in Quantitative Methods for Psychology*, *2*(1), 26–38. https://doi.org/10.20982/tqmp.02.1.p026

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10/gdz7mr

Gillan, D. J., Breedin, S. D., & Cooke, N. J. (1992). Network and multidimensional representations of the declarative knowledge of human-computer interface design experts. *International Journal of Man-Machine Studies*, *36*(4), 587–615. https://doi.org/10/bkvs5b

Gisick, L. M., Webster, K. L., Keebler, J. R., Lazzara, E. H., Fouquet, S., Fletcher, K., Fagerlund, A., Lew, V., & Chan, R. (2018). Measuring shared mental models in healthcare. *Journal of Patient Safety and Risk Management*, *23*(5), 207–219. https://doi.org/10/ggxf4q

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, *83*(1), 88–96. https://doi.org/10/bxhd5t

Gonzalvo, P., Canas, J. J., & Bajo, M.-T. (1994). Structural representations in knowledge acquisition. *Journal of Educational Psychology*, *86*(4), 601–616. https://doi.org/10/d74kj7

Grady, R. H., Greenspan, R. L., & Liu, M. (2019). What is the best size for matrix-style questions in online surveys? *Social Science Computer Review*, *37*(3), 435–445. https://doi.org/10.1177/0894439318773733

Greathouse, D. G., Halle, J. S., & Dalley, A. F. (2004). Terminologia Anatomica: Revised anatomical terminology. *Journal of Orthopaedic & Sports Physical Therapy*, *34*(7), 363–367. https://doi.org/10/ff6k9b

Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 45–53. https://doi.org/10.1177/2372732215622223

Guimarães, B., Dourado, L., Tsisar, S., Diniz, J. M., Madeira, M. D., & Ferreira, M. A. (2017). Rethinking anatomy: how to overcome challenges of medical education's evolution. *Acta Médica Portuguesa*, *30*(2), 134–140. https://doi.org/10.20344/amp.8404

Haghayegh, S., Kang, H.-A., Khoshnevis, S., Smolensky, M. H., & Diller, K. R. (2020). A comprehensive guideline for Bland–Altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings. *Physiological Measurement*, *41*(5), 055012, 1–21. https://doi.org/10.1088/1361-6579/ab86d6

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health*, *39*(1), 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in*

Pharmacy Teaching and Learning, *7*(6), 836–850. https://doi.org/10.1016/j.cptl.2015.08.001

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association*, *13*(1), 16–23. https://doi.org/10/cmt2pp

Harris, P., Bhanji, F., Topps, M., Ross, S., Lieberman, S., Frank, J. R., Snell, L., & Sherbino, J. (2017). Evolving concepts of assessment in a competency-based world. *Medical Teacher*, *39*(6), 603–608. https://doi.org/10.1080/0142159X.2017.1315071

Hartmeyer, R., Stevenson, M. P., & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 598–619. https://doi.org/10.1080/0969594X.2017.1377685

Hartnett, R. T., & Willingham, W. W. (1980). The criterion problem: What measure of success in graduate education? *Applied Psychological Measurement*, *4*(3), 281–291. https://doi.org/10/b6kgft

Hawe, E., & Dixon, H. (2017). Assessment for learning: A catalyst for student self-regulation. *Assessment & Evaluation in Higher Education*, *42*(8), 1181–1192. https://doi.org/10/gg2qnj

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77–89.

https://doi.org/10.1080/19312450709336664

Hayes, S. H., Fiebert, I. M., Carroll, S. R., & Magill, R. N. (1997). Predictors of academic success in a physical therapy program: Is there a difference between traditional and nontraditional students? *Journal of Physical Therapy Education*, *11*(1), 10–16. https://doi.org/10/gg9rd7

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, *37*(2), 1–19. https://doi.org/10.1007/BF03216919

Herling, R. W. (2000). Operational definitions of expertise and competence. *Advances in Developing Human Resources*, *2*(1), 8–21. https://doi.org/10/cdr7rx

Hernaez, R. (2015). Reliability and agreement studies: A guide for clinical investigators. *Gut*, *64*(7), 1018–1027. https://doi.org/10.1136/gutjnl-2014-308619

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158. https://doi.org/10.1119/1.2343497

Hołda, M. K., Stefura, T., Koziej, M., Skomarovska, O., Jasińska, K. A., Sałabun, W., & Klimek-Piotrowska, W. (2019). Alarming decline in recognition of anatomical structures amongst medical students and physicians. *Annals of Anatomy*, *221*, 48–56. https://doi.org/10.1016/j.aanat.2018.09.004

Housner, L. D., Gomez, R., & Griffey, D. C. (1993). A Pathfinder analysis of pedagogical knowledge structures: A follow-up investigation. *Research Quarterly for Exercise and Sport*, *64*(3), 291–299. https://doi.org/10.1080/02701367.1993.10608813

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. Wiley interdisciplinary reviews. *Cognitive Science*, *4*(1), 93–103. https://doi.org/10.1002/wcs.1203

Hruska, P., Krigolson, O., Coderre, S., McLaughlin, K., Cortese, F., Doig, C., Beran, T., Wright, B., & Hecker, K. G. (2016). Working memory, reasoning, and expertise in medicine-insights into their relationship using functional neuroimaging. *Advances in Health Sciences Education: Theory and Practice*, *21*(5), 935–952. https://doi.org/10.1007/s10459-015-9649-2

Hulme, A. K., Luo, K., & Štrkalj, G. (2020). Musculoskeletal anatomy knowledge retention in the Macquarie University chiropractic program: A cross-sectional study. *Anatomical Sciences Education*, *13*(2), 182–191. https://doi.org/10/gg2n2t

Husmann, P. R., & O'Loughlin, V. D. (2019). Another nail in the coffin for learning styles? Disparities among undergraduate anatomy students' study strategies, class performance, and reported VARK learning styles. *Anatomical Sciences Education*, *12*(1), 6–19. https://doi.org/10.1002/ase.1777

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time. *Instructional Science*, *39*(1), 41–61. https://doi.org/10/dd2h49

Ifenthaler, D., & Pirnay-Dummer, P. (2014). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 289–301). Springer. https://doi.org/10.1007/978-1-4614-3185-5_23

Insel, K. C., Meek, P. M., & Leventhal, H. (2005). Differences in illness representation among pulmonary patients and their providers. *Journal of Health Psychology*, *10*(1), 147–162. https://doi.org/10/czv8xq

Jaafarpour, M., Aazami, S., & Mozafari, M. (2016). Does concept mapping enhance learning outcome of nursing students? *Nurse Education Today*, *36*, 129–132. https://doi.org/10.1016/j.nedt.2015.08.029

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., … Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10/gfhx3v

Janska, A. C., & Clark, R. A. J. (2010). Further exploration of the possibilities and pitfalls of multidimensional scaling as a tool for the evaluation of the quality of synthesized speech. *7th ISCA Workshop on Speech Synthesis*.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, *5*(1), 1–10. https://doi.org/10/gg264n

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1994). Locus of the predictive advantage in Pathfinder-based representations of classroom knowledge. *Journal* *of Educational Psychology*, *86*(4), 617–626. https://doi.org/10.1037/0022-0663.86.4.617

Jonassen, D., Strobel, J., & Gottdenker, J. (2005). Model building for conceptual change. *Interactive Learning Environments*, *13*(1–2), 15–37. https://doi.org/10/cdrv43

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Psychology Press.

Jung, E., Kim, M., & Reigeluth, C. M. (2016). Learning in action: How competent professionals learn. *Performance Improvement Quarterly*, *28*(4), 55–69. https://doi.org/10/gg869r

Kamath, A., Rao, R., Shenoy, P. J., & Ullal, S. D. (2018). Approaches to learning and academic performance in pharmacology among second-year undergraduate medical students. *Scientia Medica*, *28*(4), 32395, 1–6. https://doi.org/10.15448/1980-6108.2018.4.32395

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10/gftzh3

Keehner, M. (2011). Spatial cognition through the keyhole: How studying a real-world domain can inform basic science—and vice versa. *Topics in Cognitive Science*, *3*(4), 632–647. https://doi.org/10/btjcnd

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review*, *37*(2), 214–233. https://doi.org/10.1177/0894439317752406

Kinchin, I. M., Möllits, A., & Reiska, P. (2019). Uncovering types of knowledge in concept maps. *Education Sciences*, *9*(2), 131–145. https://doi.org/10.3390/educsci9020131

Klooster, N. B., Tranel, D., & Duff, M. C. (2020). The hippocampus and semantic memory over time. *Brain and Language*, *201*, 104711, 1–7. https://doi.org/10/gg4s32

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10/b84r

Kopp-Schneider, A., & Hielscher, T. (2019). How to evaluate agreement between quantitative measurements. *Radiotherapy and Oncology*, *141*, 321–326. https://doi.org/10.1016/j.radonc.2019.09.004

Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53*(1), 17–94. https://doi.org/10/ggcmc6

Kozma, R. B. (2020). Use of multiple representations by experts and novices. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (p. 16). Routledge. https://doi.org/10.4324/9780429443961-4

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (No. 11). Sage. https://doi.org/10.4135/9781412985130

Kubsch, M., Touitou, I., Nordine, J., Fortus, D., Neumann, K., & Krajcik, J. (2020). Transferring knowledge in a knowledge-in-use task—investigating the role of knowledge organization. *Education Sciences*, *10*(1), 1–16. https://doi.org/10/gg2z4n

Kulasegaram, K., & Rangachari, P. K. (2018). Beyond "formative": Assessments to enrich student learning. *Advances in Physiology Education*, *42*(1), 5–14. https://doi.org/10/gct8gs

Kume, J., Reddin, V., & Horbacewicz, J. (2019). Predictors of physical therapy academic and NPTE licensure performance. *Health Professions Education*, *5*(3), 185–193. https://doi.org/10/ghgjd3

Laird, J. E. (2012). *The Soar cognitive architecture*. MIT Press. https://doi.org/10.7551/mitpress/7688.001.0001

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13–26. https://doi.org/10/gctv89

Langlois, J., Bellemare, C., Toulouse, J., & Wells, G. A. (2020). Spatial abilities training in anatomy education: A systematic review. *Anatomical Sciences Education*, *13*(1), 71–79. https://doi.org/10/ggxr3x

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248.

Leppink, J. (2020). Learning processes. In J. Leppink (Ed.), *The art of modeling the learning process: Uniting educational research and practice* (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-43082-5_1

Lexico. (n.d.). Cognition. *Lexico.com dictionary*. Retrieved July 31, 2021, from https://www.lexico.com/en/definition/cognition

Liao, S. C., Hunt, E. A., & Chen, W. (2010). Comparison between interrater reliability and interrater agreement in performance assessment. *Annals Academy of Medicine Singapore*, *39*(8), 613–618.

Lieto, A., Chella, A., & Frixione, M. (2017). Conceptual spaces for cognitive architectures: A lingua franca for different levels of representation. *Biologically Inspired Cognitive Architectures*, *19*, 1–9. https://doi.org/10.1016/j.bica.2016.10.005

Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLOS ONE*, *14*(7), e0219854, 1–35. https://doi.org/10/ghb97s

Liu, M., & Cernat, A. (2018). Item-by-item versus matrix questions: A web survey experiment. *Social Science Computer Review*, *36*(6), 690–706.

https://doi.org/10.1177/0894439316674459

Liu, Q., Tong, S., Liu, C., Zhao, H., Chen, E., Ma, H., & Wang, S. (2019). Exploiting cognitive structure for adaptive learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 627–635. https://doi.org/10.1145/3292500.3330922

Looney, M. A. (2018). Assessment of interrater and intermethod agreement in the kinesiology literature. *Measurement in Physical Education & Exercise Science*, *22*(2), 116–128. https://doi.org/10.1080/1091367X.2017.1395742

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Losco, C. D., Grant, W. D., Armson, A., Meyer, A. J., & Walker, B. F. (2017). Effective methods of teaching and learning in anatomy as a basic science: A BEME systematic review: BEME guide no. 44. *Medical Teacher*, *39*(3), 234–243. https://doi.org/10.1080/0142159X.2016.1271944

Lucey, C. R., Thibault, G. E., & ten Cate, O. (2018). Competency-based, time-variable education in the health professions: Crossroads. *Academic Medicine*, *93*, S1–S5. https://doi.org/10.1097/ACM.0000000000002080

Lufler, R. S., Zumwalt, A. C., Romney, C. A., & Hoagland, T. M. (2012). Effect of visual-spatial ability on medical students' performance in a gross anatomy course. *Anatomical Sciences Education*, *5*(1), 3–9. https://doi.org/10.1002/ase.264

Lyu, X., & Li, Z. (2019). Predictors for human performance in information seeking, information integration, and overall process in diagnostic tasks. *International Journal of Human–Computer Interaction*, *35*(19), 1831–1841. https://doi.org/10/gg2b4s

Machado, C. T., & Carvalho, A. A. (2020). Concept mapping: Benefits and challenges in higher education. *The Journal of Continuing Higher Education*, *68*(1), 38–53. https://doi.org/10/gg7gzg

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research*, *51*(6), 772–789.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Matthay, E. C., & Glymour, M. M. (2020). A graphical catalog of threats to validity. *Epidemiology*, *31*(3), 376–384. https://doi.org/10/ggjn2f

Mayer, R. E. (2002a). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71. https://doi.org/10/cjrpt6

Mayer, R. E. (2002b). Rote versus meaningful learning. *Theory into Practice*, *41*(4), 226–232. https://doi.org/10/dqvj6x

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

McCutcheon, A. L. (2008). Sampling bias. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods* (p. 785). Sage Publications. https://doi.org/10.4135/9781412963947.n509

McDonald, F., Reynolds, J., Bixley, A., & Spronken-Smith, R. (2017). Changes in approaches to learning over three years of university undergraduate study. *Teaching and Learning Inquiry*, *5*(2), 65–79. https://doi.org/10.20343/teachlearninqu.5.2.6

McGaghie, W. C., McCrimmon, D. R., Mitchell, G., Thompson, J. A., & Ravitch, M. M. (2000). Quantitative concept mapping in pulmonary physiology: Comparison of student and faculty knowledge structures. *Advances in Physiology Education*, *23*(1), S72–S81. https://doi.org/10/gft668

McGaghie, W. C., McCrimmon, D. R., & Thompson, J. A. (1998). Multidimensional scaling assessment of medical and veterinary student knowledge organization of pulmonary physiology concepts. *Proceedings of the American Educational Research Association*.

Meiners, K. M., & Rush, D. K. (2017). Clinical performance and admission variables as predictors of passage of the National Physical Therapy Examination. *Journal of Allied Health*, *46*(3), 164–170.

Meyer, A. J., Innes, S. I., Stomski, N. J., & Armson, A. J. (2016). Student performance on practical gross anatomy examinations is not affected by assessment modality. *Anatomical Sciences Education*, *9*(2), 111–120. https://doi.org/10.1002/ase.1542

Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, *192*(4234), 27–33. https://doi.org/10/c6b6s6

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac*

Anaesthesia, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Moffatt, D. J., Jacobs, A. W., & Metcalf, W. K. (1971). Predictors of academic performance in gross anatomy. *Academic Medicine*, *46*(6), 945–948. https://doi.org/10.1097/00001888-197106000-00015

Mohammadi, F., Momennasab, M., Rostambeygi, P., Ghaderi, S., & Mousazadeh, S. (2019). The effect of education through conceptual mapping on critical thinking of nursing students. *Journal of the Pakistan Medical Association*, *69*(8), 1094–1098.

Montpetit-Tourangeau, K., Dyer, J.-O., Hudon, A., Windsor, M., Charlin, B., Mamede, S., & van Gog, T. (2017). Fostering clinical reasoning in physiotherapy: Comparing the effects of concept map study and concept map completion after example study in novice and advanced learners. *BMC Medical Education*, *17*(1), 1–23. https://doi.org/10.1186/s12909-017-1076-z

Moon, B., Johnston, C., & Moon, S. (2018). A case for the superiority of concept mapping-based assessments for assessing mental models. *Proceedings of the Eighth International Conference on Concept Mapping*.

Moore, J. H., Goss, D. L., Baxter, R. E., DeBerardino, T. M., Mansfield, L. T., Fellows, D. W., & Taylor, D. C. (2005). Clinical diagnostic accuracy and magnetic resonance imaging of patients referred by physical therapists, orthopaedic surgeons, and nonorthopaedic providers. *Journal of Orthopaedic & Sports Physical Therapy*, *35*(2), 65–71. https://doi.org/10/ghgjcn

Moore, K. L., Dalley, A. F. D., & Agur, A. M. R. (2018). *Clinically oriented anatomy*

(8th ed.). Wolters Kluwer.

Morales-Martinez, G. E., Lopez-Ramirez, E. O., Castro-Campos, C., Villarreal-Treviño, M. G., & Gonzales-Trujillo, C. J. (2017). Cognitive analysis of meaning and acquired mental representations as an alternative measurement method technique to innovate e-assessment. *European Journal of Educational Research*, *6*(4), 455–464. https://doi.org/10.12973/eu-jer.6.4.455

Muehling, A. (2017). Concept landscapes: Aggregating concept maps for analysis. *Journal of Educational Data Mining*, *9*(2), 1–30. https://doi.org/10.5281/zenodo.3554717

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*, 1–10. https://doi.org/10.1177/2382120519840332

Narnaware, Y., & Neumeier, M. (2020). Second-year nursing students' retention of gross anatomical knowledge. *Anatomical Sciences Education*, *13*(2), 230–236. https://doi.org/10/gg2n2s

Neiles, K. Y., Todd, I., & Bunce, D. M. (2016). Establishing the validity of using network analysis software for measuring students' mental storage of chemistry concepts. *Journal of Chemical Education*, *93*(5), 821–831. https://doi.org/10/f8px77

Neumann, K., Kind, V., & Harms, U. (2019). Probing the amalgam: The relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, *41*(7), 847–861.

https://doi.org/10.1080/09500693.2018.1497217

Newman, M. (2018). *Networks*. Oxford University Press. https://doi.org/10.1093/oso/9780198805090.001.0001

Nicoara, S. M., Szabo, B. A., Micu, C. M., & Badea, A. F. (2018). Meta-analysis on the study with concept maps on the medical field. *Transylvanian Journal of Psychology*, *18*(2), 135–166. https://doi.org/10/ggxr38

Nicoara, S. M., Szamoskozi, S.-E., Mitrea, D.-A., & Leucuta, D.-C. (2020). Concept mapping, an effective tool for long-term memorization of anatomy—A quasi-experimental research carried out among 1st year general medicine students. *European Journal of Investigation in Health, Psychology and Education*, *10*(1), 530–543. https://doi.org/10/gg45dm

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The causes of errors in clinical reasoning: Cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine*, *92*(1), 23–30. https://doi.org/10/gfzq3m

Noushad, B., & Khurshid, F. (2019). Facilitating student learning: An instructional design perspective for health professions educators. *Research and Development in Medical Education*, *8*(2), 69–74. https://doi.org/10/ggx7wd

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University

Press. https://doi.org/10.1017/CBO9781139173469

Ojha, H. A., Fritz, J. M., Malitsky, A. L., Wu, J., Weiner, M. G., Brandi, J. A., Rhon, D. I., Mobo, B. H. P., Fleming, K. M., Beidleman, R. R., & Wright, W. G. (2020). Comparison of physical therapy and physician pathways for employees with recent onset musculoskeletal pain: A randomized controlled trial. *PM&R*, 1–10. https://doi.org/10/ghgjcm

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

O'Mahony, S. M., Sbayeh, A., Horgan, M., O'Flynn, S., & O'Tuathaigh, C. M. P. (2016). Association between learning style preferences and anatomy assessment outcomes in graduate-entry and undergraduate medical students. *Anatomical Sciences Education*, *9*(4), 391–399. https://doi.org/10.1002/ase.1600

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10/dp3ggn

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10/gg745f

Paivio, A. (1971). *Imagery and verbal processes*. Psychology Press.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Paldino, M. J., Chu, Z. D., Chapieski, M. L., Golriz, F., & Zhang, W. (2017).

Repeatability of graph theoretical metrics derived from resting-state functional networks in paediatric epilepsy patients. *The British Journal of Radiology*, *90*(1074), 20160656. https://doi.org/10/f977pm

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*(2), 169–178. https://doi.org/10.1007/s004420100720

Persky, A. M., & Murphy, K. (2019). Investigating whether transfer of learning in pharmacy students depends more on knowledge storage or accessibility. *American Journal of Pharmaceutical Education*, *83*(6), 6809, 1274–1281. https://doi.org/10/gg7mnq

Persky, A. M., & Robinson, J. D. (2017). Moving from novice to expertise and its implications for instruction. *American Journal of Pharmaceutical Education*, *81*(9), 72–80. https://doi.org/10/gcp7sw

Peterson, C. A., & Tucker, R. P. (2005). Medical gross anatomy as a predictor of performance on the USMLE Step 1. *The Anatomical Record Part B: The New Anatomist*, *283B*(1), 5–8. https://doi.org/10.1002/ar.b.20054

Piaget, J. (1926). *The thought and language of the child*. Harcourt, Brace, and Company.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice*. Pearson/Prentice Hall.

Preece, P. E. (1976). Mapping cognitive structure: A comparison of methods. Journal of *Educational Psychology*, *68*(1), 1–8. https://doi.org/10.1037/0022-0663.68.1.1

Prince, K. J. A. H., Scherpbier, A. J. A. A., van Mameren, H., Drukker, J., & van der

Vleuten, C. P. M. (2005). Do students have sufficient knowledge of clinical anatomy? *Medical Education*, *39*(3), 326–332. https://doi.org/10.1111/j.1365-2929.2005.02096.x

Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning curves in health professions education. *Academic Medicine*, *90*(8), 1034–1042. https://doi.org/10.1097/ACM.0000000000000681

Quillian, M. (1966). *Semantic memory*. Bolt Beranek and Newman. https://doi.org/10.21236/AD0641671

Quillian, M. R. (1962). A revised design for an understanding machine. *Mechanical Translation*, *7*(1), 17–29.

Radwan, A., Abdelnasser, A., Elaraby, S., & Talaat, W. (2018). Correlation between concept maps and clinical reasoning for final year medical students at the faculty of medicine -Suez Canal University. *Education in Medicine Journal*, *10*(4), 43–51. https://doi.org/10.21315/eimj2018.10.4.5

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in Clinical Research*, *8*(4), 187–191. https://doi.org/10.4103/picr.PICR_123_17

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Richter, F. R., Bays, P. M., Jeyarathnarajah, P., & Simons, J. S. (2019). Flexible updating of dynamic knowledge structures. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10/gg5jd6

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(3), 1–19. https://doi.org/10/ggtgnk

Roman, G., & Buman, M. P. (2019). Preadmission predictors of graduation success from a physical therapy education program in the United States. *Journal of Educational Evaluation for Health Professions*, *16*, 1–7. https://doi.org/10/ghgjdw

Rosas, S. R. (2017). Multi-map comparison for group concept mapping: An approach for examining conceptual congruence through spatial correspondence. *Quality and Quantity*, *51*(6), 2421–2439. https://doi.org/10.1007/s11135-016-0399-x

Roske-Hofstrand, R. J., & Paap, K. R. (1990). Discriminating between degrees of low or high similarity: implications for scaling techniques using semantic judgments. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 61–74). Ablex Publishing.

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, *14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6

Ruel, E., Wagner, W. E., & Gillespie, B. J. (2015). *The practice of survey research: Theory and applications*. Sage Publications. https://doi.org/10.4135/9781483391700

Salkowski, L. R., & Russ, R. (2018). Cognitive processing differences of experts and novices when correlating anatomy and cross-sectional imaging. *Journal of Medical Imaging*, *5*(3), 1–17. https://doi.org/10.1117/1.JMI.5.3.031411

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: a meta-analysis. *Educational Psychology Review*, *30*(2), 431–455. https://doi.org/10.1007/s10648-017-9403-9

Schuelke, M. J., Day, E. A., McEntire, L. E., Boatman, P. R., Boatman, J. E., Kowollik, V., & Wang, X. (2009). Relating indices of knowledge structure coherence and accuracy to skill-based performance: Is there utility in using a combination of indices? *Journal of Applied Psychology*, *94*(4), 1076–1085. https://doi.org/10.1037/a0015113

Schvaneveldt, R. (n.d.). Pathfinder networks. Retrieved from https://interlinkinc.net/

Schvaneveldt, R., Dearholt, D., & Durso, F. (1988). Graph theoretic foundations of Pathfinder networks. *Computers & Mathematics with Applications*, *15*(4), 337–345. https://doi.org/10.1016/0898-1221(88)90221-0

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization* (pp. xi, 315). Ablex Publishing.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. *Psychology of Learning and Motivation*, *24*, 249–284. https://doi.org/10.1016/S0079-7421(08)60539-3

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & De Maio, J. C. (1985). Measuring the structure of expertise.

International Journal of Man-Machine Studies, *23*(6), 699–728. https://doi.org/10/c4cgc4

Segalowitz, N. S., Doucerain, M. M., Meuter, R. F. I., Zhao, Y., Hocking, J., & Ryder, A. G. (2016). Comprehending adverbs of doubt and certainty in health communication: A multidimensional scaling approach. *Frontiers in Psychology*, *7*, 1–13. https://doi.org/10.3389/fpsyg.2016.00558

Shabankhani, B. (2020). Assessing the interrater reliability for nominal, categorical and ordinal data in medical sciences. *Archives of Pharmacy Practice*, *11*(S4),144–148.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, *63*(3), 225–234. https://doi.org/10/drqjxp

Shiffrin, R. M. (2010). Perspectives on modeling in cognitive science. *Topics in Cognitive Science*, *2*(4), 736–750. https://doi.org/10/dpqfc4

Shin, H. S. (2019). Reasoning processes in clinical reasoning: From the perspective of cognitive psychology. *Korean Journal of Medical Education*, *31*(4), 299–308. https://doi.org/10/ggvqh8

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Si, J., Kong, H.-H., & Lee, S.-H. (2019). Developing clinical reasoning skills through argumentation with the concept map method in medical problem-based learning.

Interdisciplinary Journal of Problem-Based Learning, *13*(1). 1–16. https://doi.org/10.7771/1541-5015.1776

Siew, C. S. Q. (2020). Applications of network science to education research: Quantifying knowledge and the development of expertise through network analysis. *Education Sciences, 10*(4), 101–116. https://doi.org/10/gg3dvb

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, *2019*, Article 2108423, 1–24. https://doi.org/10/gg2734

Smith, C. F., Finn, G. M., & Border, S. (2017). Learning clinical anatomy. *European Journal of Anatomy*, *21*, 269–278.

Smith, C. F., Stokholm, C., Sinha, R., Ponikwer, F., Carter, M., & Birch, M. (2017). Interplays of psychometric abilities on learning gross anatomy. *MedEdPublish*, *6*(2), 1–17. https://doi.org/10.15694/mep.2017.000104

Souza, A. C. D., Alexandre, N. M. C., & Guirardello, E. D. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*, *26*(3), 649–659. https://doi.org/10/gf66kv

Spiers, H. J. (2020). The hippocampal cognitive map: One space or many? *Trends in Cognitive Sciences*, *24*(3), 168–170. https://doi.org/10/gg3t4k

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics*, *1*(1), 1–19. https://doi.org/10/fkchmq

Stevenson, J. L., Shah, S., & Bish, J. P. (2016). Use of structural assessment of knowledge for outcomes assessment in the neuroscience classroom. *Journal of Undergraduate Neuroscience Education*, *15*(1), A38–A43.

Stevenson, M. P., Hartmeyer, R., & Bentsen, P. (2017). Systematically reviewing the potential of concept mapping technologies to promote self-regulated learning in primary and secondary science education. *Educational Research Review*, *21*, 1–16. https://doi.org/10/ghdnwm

Stocco, A. (2018). A biologically plausible action selection system for cognitive architectures: Implications of basal ganglia anatomy for learning and decision-making models. *Cognitive Science*, *42*(2), 457–490. https://doi.org/10/gc733p

Stolarova, M., Wolf, C., Rinker, T., & Brielmann, A. (2014). How to assess and compare interrater reliability, agreement and correlation of ratings: An exemplary analysis of mother-father and parent-teacher expressive vocabulary rating pairs. *Frontiers in Psychology*, *5*, 1–13. https://doi.org/10.3389/fpsyg.2014.00509

Sturrock, K., & Rocha, J. (2000). A multidimensional scaling stress evaluation table. *Field Methods*, *12*(1), 49–60. https://doi.org/10.1177/1525822X0001200104

Sullivan, M. E., Yates, K. A., Inaba, K., Lam, L., & Clark, R. E. (2014). The use of cognitive task analysis to reveal the instructional limitations of experts in the teaching of procedural skills. *Academic Medicine*, *89*(5), 811–816. https://doi.org/10/f54rzc

Sussmann, M., & Robertson, D. U. (1986). The validity of validity: An analysis of validation study designs. *Journal of Applied Psychology*, *71*(3), 461–468. https://doi.org/10/dgvgr8 ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2018). On the usefulness of interrater reliability coefficients. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative psychology: The annual meeting of the psychometric society*, *233*, 67–75. Springer International Publishing. https://doi.org/10.1007/978-3-319-77249-3_6

Tessmer, M., Perrin, B., & Bennett, W. (1997). *Assessing the stability of structural learning measures* [Report]. Air Force Research Laboratory. https://doi.org/10.1037/e443512005-001

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. *Current Biology*, *29*(7), 1226–1231. https://doi.org/10/gg3xqm

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10/b9pn6t

Timmerberg, J. F., Dole, R., Silberman, N., Goffar, S. L., Mathur, D., Miller, A., Murray, L., Pelletier, D., Simpson, M. S., Stolfi, A., Thompson, A., & Utzman, R. (2019). Physical therapist student readiness for entrance into the first full-time clinical experience: A Delphi study. *Physical Therapy*, *99*(2), 131–146. https://doi.org/10.1093/ptj/pzy134

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10/d65zts

Trumpower, D. L., Filiz, M., & Sarwar, G. S. (2014). Assessment for learning using digital knowledge maps. In D. Ifenthaler & R. Hanewald (Eds.), *Digital* *knowledge maps in education* (pp. 221–237). Springer New York. https://doi.org/10.1007/978-1-4614-3178-7_12

Trumpower, D. L., Sharara, H., & Goldsmith, T. E. (2010). Specificity of structural assessment of knowledge. *Journal of Technology, Learning, and Assessment*, *8*(5). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1624

Tsai, C.-C., & Huang, C.-M. (2002). Exploring students’ cognitive structures in learning science: A review of relevant methods. *Journal of Biological Education*, *36*(4), 163–169. https://doi.org/10.1080/00219266.2002.9655827

van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*(2), 325–331. https://doi.org/10.1016/j.chb.2008.12.021

van Kesteren, M. T. R., & Meeter, M. (2020). How to optimize knowledge construction in the brain. *NPJ Science of Learning*, *5*(1), 1–7. https://doi.org/10/gg3f6g

van Lankveld, W., Maas, M., van Wijchen, J., Visser, V., & Staal, J. B. (2019). Self-regulated learning in physical therapy education: A non-randomized experimental study comparing self-directed and instruction-based learning. *BMC Medical Education*, *19*(1), 1–9. https://doi.org/10.1186/s12909-019-1484-3

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Veríssimo, S., Lopes, V. G., Garcia, L. M. C., & González, R. L. (2017). Evaluation of changes in cognitive structures after the learning process in mathematics. *International Journal of Innovation in Science and Mathematics Education*, *25*(2), 17-33.

Vogel, S., Kluen, L. M., Fernández, G., & Schwabe, L. (2018). Stress leads to aberrant hippocampal involvement when processing schema-related information. *Learning and Memory*, *25*(1), 21–30. https://doi.org/10/ggsbmk

Vukić, Đ., Martinčić-Ipšić, S., & Meštrović, A. (2020). Structural analysis of factual, conceptual, procedural, and metacognitive knowledge in a multidimensional knowledge network. *Complexity*, *2020*, 1–17. https://doi.org/10/gg7mn2

Wainer, H., & Kaye, K. (1974). Multidimensional scaling of concept learning in an introductory course. *Journal of Educational Psychology*, *66*(4), 591–598. https://doi.org/10.1037/h0036933

Wang, M., Wu, B., Kirschner, P. A., & Michael Spector, J. (2018). Using cognitive mapping to foster deeper learning with complex problems in a computer-based environment. *Computers in Human Behavior*, *87*, 450–458. https://doi.org/10.1016/j.chb.2018.01.024

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). SAGE.

Wei, W., & Yue, K.-B. (2017). Integrating concept mapping into information systems education for meaningful learning and assessment. *Information Systems Education Journal*, *15*(6), 4–16.

Welton, T., Constantinescu, C. S., Auer, D. P., & Dineen, R. A. (2020). Graph theoretic analysis of brain connectomics in multiple sclerosis: Reliability and relationship with cognition. *Brain Connectivity*, *10*(2), 95–104. https://doi.org/10/gg3tx3

White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science & Technology*, *54*(5), 423–434. https://doi.org/10/ddt5r6

White, L. J., McGowan, H. W., & McDonald, A. C. (2018). The effect of content delivery style on student performance in anatomy: Blended learning and assessment in anatomy. *Anatomical Sciences Education*, *12*(1), 43–51. https://doi.org/10/gg2n28

Whitehill, J. (2013). Understanding ACT-R - an outsider's perspective. *ArXiv:1306.0125*, 1–12. Retrieved from http://arxiv.org/abs/1306.0125

Wilson, A. B., Brown, K. M., Misch, J., Miller, C. H., Klein, B. A., Taylor, M. A., Goodwin, M., Boyle, E. K., Hoppe, C., & Lazarus, M. D. (2019). Breaking with tradition: A scoping meta-analysis analyzing the effects of student-centered learning and computer-aided instruction on student performance in anatomy. *Anatomical Sciences Education*, *12*(1), 61–73. https://doi.org/10.1002/ase.1789

Wilson, A. B., Miller, C. H., Klein, B. A., Taylor, M. A., Goodwin, M., Boyle, E. K., Brown, K., Hoppe, C., & Lazarus, M. (2018). A meta-analysis of anatomy laboratory pedagogies. *Clinical Anatomy*, *31*(1), 122–133. https://doi.org/10.1002/ca.22934

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications*, *3*(1), 1–16. https://doi.org/10/ggx7rg

Wolden, M., Hill, B., & Voorhees, S. (2020). Predicting success for student physical therapists on the National Physical Therapy Examination: Systematic review and meta-analysis. *Physical Therapy*, *100*(1), 73–89. https://doi.org/10/gg9xpm

Wu, H., & Leung, S.-O. (2017). Can Likert scales be treated as interval scales? A simulation study. *Journal of Social Service Research*, *43*(4), 527–532. https://doi.org/10.1080/01488376.2017.1329775

Ye, P., Wang, T., & Wang, F.-Y. (2018). A survey of cognitive architectures in the past 20 years. *IEEE Transactions on Cybernetics*, *48*(12), 3280–3290. https://doi.org/10/gfqhb5

Yee, E., Jones, M., & McRae, K. (2017). Semantic memory. In J. T. Wixted & S. Thompson-Schill (Eds.), *The Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed. Vol. 3, pp. 1–38). Wiley. https://doi.org/10.1002/9781119170174.epcn309

Yue, M., Zhang, M., Zhang, C., & Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, *52*, 87–94. https://doi.org/10.1016/j.nedt.2017.02.018

Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring interrater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, *16*(1), 1–10.

https://doi.org/10.1186/s12874-016-0200-9

Zemla, J. C., & Austerweil, J. L. (2018). Estimating semantic networks of groups and individuals from fluency data. *Computational Brain & Behavior*, *1*(1), 36–58. https://doi.org/10/gfvrws

Ziembowicz, K. (2017). Mental models—their diagnosis and role in knowledge acquisition. *Studia Psychologiczne*, *55*(2), 40–52.

Zipp, G., & Maher, C. (2013). Prevalence of mind mapping as a teaching and learning strategy in physical therapy curricula. *Journal of the Scholarship of Teaching and Learning*, *13*(5), 21–32.

Zipp, G. P., Maher, C., & D'Antoni, A. V. (2015). Mind mapping: Teaching and learning strategy for physical therapy curricula. *Journal of Physical Therapy Education*, *29*(1), 43–48. https://doi.org/10.1097/00001416-201529010-00008

Zulu, E., Haupt, T., & Tramontin, V. (2018). Cognitive loading due to self-directed learning, complex questions and tasks in the zone of proximal development of students. *Problems of Education in the 21st Century*, *76*(6), 864–880. https://doi.org/10.33225/pec/18.76.864

Appendix A: Content Items and Functional Terms

Please rank order these concept items and functional terms in order of anatomical importance and clinical relevance, with 1 = most important/relevant to clinical practice and 40 = least important/relevant to clinical practice.

Item	*Rank*	*Item*	*Rank*
Humeral head		Mobility	
Acromion		Stability	
Bicipital groove		Ball and socket	
Coracoid process		Triplanar Motion	
Glenoid fossa		Subacromial bursa	
Biceps brachii		Levator scapulae	
Triceps brachii		Axillary artery	
Supraspinatus		Circumflex humeral arteries	
Infraspinatus		Suprascapular nerve	
Subscapularis		Lateral pectoral nerve	
Teres minor		Axillary nerve	
Deltoid		Glenoid labrum	
Teres major		Joint capsule	
Latissimus dorsi		Glenohumeral ligaments	
Pectoralis major		Coracoclavicular ligaments	
Pectoralis minor		Coracohumeral ligament	
Coracobrachialis		Transverse scapular ligament	
Rhomboids		Transverse humeral ligament	
Brachial Plexus		Greater tubercle	
Segmental Innervation		Lesser tubercle	

Appendix B: Description of Study for Prospective Participants

Call for Research Study Participants

There is a gap in understanding how physiotherapy students learn gross anatomy, specifically, how they organize concepts to promote learning and retention. As a part of my dissertation research, I am conducting a study regarding the organization of anatomy concepts by first trimester DPT students. This study's findings will enhance understanding of how students learn anatomy, the foundation for all courses within the DPT curriculum.

What Will I Do? Study participants will register and complete an online survey at any time prior to [date removed]. The online survey will take approximately 15 to 20 minutes to complete. The online survey will ask you to compare several pairs of items for similarity/relatedness on a 0 to 10 scale, with 0 being completely dissimilar and 10 being completely similar (identical). As an example, imagine the words "goldfish" and "shark." You might perceive them to have a certain degree of similarity as they are both fish. The next pair of words could be "shark" and "lion," which you might perceive to have a little similarity. Items in the online survey will refer to anatomical concepts, and there is no right or wrong answer - just your perception of their similarity and relatedness. It is not testing your anatomical knowledge.

A link for recruitment to participate in the study is at the end of the announcement. This will generate an email request; simply include your student identification number. In return, you will receive both a unique private identifier code and a link to the online survey. Informed consent will be attained via clicking on a link that acknowledges your understanding before beginning the paired similarity ratings.

Voluntary Nature of the Study: This study is voluntary. Everyone will respect your decision of whether you choose to be a part of the study or not. You will be treated the same at [institution removed], whether you choose to be a part of the study or not. If you decide to join the study now, you can still change your mind later. You may stop at any time.

Payment: Upon completing both surveys, participants will receive a $10 electronic gift card in appreciation of their participation. Participants will submit an email address upon completion of the survey to which the electronic gift card will be sent. Email addresses will not be associated with the online surveys. Study participants will also be provided an opportunity to attend a presentation of the study results upon completing the study.

Privacy: Any information you provide will be kept confidential and anonymous. Surveys will be linked to your unique identifier; no identifying information will be associated with your results. Electronic data will be kept strictly confidential in a fire- and flood-proof safe in my home and encrypted as a private file in my Dropbox account.

Questions: You can ask questions of the researcher by email at [email removed].

Thank you for your consideration in participating in this study.

Appendix C: Data Coding for Participant Data Sets

Student Data Set

Code	*Description*	*Data Source*
SID	Student ID (deidentified once data set compiled)	Student
UI	Unique Identifier	Primary Investigator
Pretest	Pretest ratings (pairwise comparisons)	Student
Posttest	Posttest ratings (pairwise comparisons)	Student
GradeW	Unit grade – written exam	Blackboard
GradeP	Unit grade – practical exam	Blackboard
Age	Age	Registrar
Gender	Gender	Registrar
AdmGPA	Admission Cumulative GPA	Registrar
AdmAGPA	Admission Core Science GPA	Registrar
Campus	Location	Registrar
Mode	Mode of Delivery: residential or flexible	Registrar

Expert Data Set

Code	*Description*	*Source*
UI	Unique Identifier	Primary Investigator
Test	Test ratings (pairwise comparisons)	Expert
YCP	Years of Clinical Practice	Expert
YAT	Years of Anatomy Teaching	Expert
TCD	Terminal Clinical Degree	Expert
TAD	Terminal Academic Degree	Expert
Campus	Location	Expert
Mode	Mode of Delivery: residential or flexible	Expert

N.B. Data sets as described do not include derived MDS and PFN parameters.

Appendix D: Preliminary Exploratory Analysis

Table D1

Impact of MDS Scaling Model

Group	Model	Stress-1	TCC	R^2
ECSD	RMDS	0.222	0.98	0.73
	WMDS	0.217	0.98	0.75
ECSI	RMDS	0.196	0.98	0.78
	WMDS	0.188	0.98	0.81

Note. RMDS with multiple matrices, PROXSCAL algorithm, Identity scaling model, two dimensions; WMDS with multiple matrices, PROXSCAL algorithm, weighted Euclidean scaling model, two dimensions

Table D2

Aggregation Strategy and MDS Configuration

Group	Strategy	Stress-1	TCC	R^2
ECSD	RMDS	0.222	0.98	0.73
	CMDS Mean	0.252	0.97	0.67
	CMDS Median	0.149	0.98	0.89
ECSI	RMDS	0.196	0.98	0.78
	CMDS Mean	0.234	0.97	0.69
	CMDS Median	0.149	0.98	0.89

Note. RMDS with multiple matrices; CMDS mean with one matrix; CMDS median with one matrix. All utilize PROXSCAL algorithm, Identity scaling model, two dimensions.

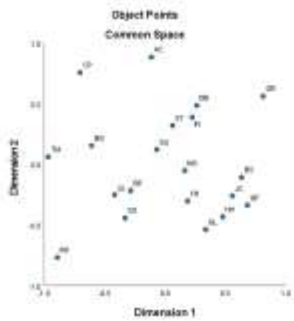
Appendix E: Final Item List

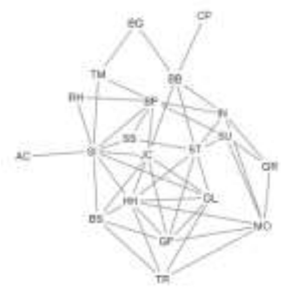
The 20 anatomical structures and concepts noted in bold were used for the paired comparisons survey. Codes associated with these items are noted.

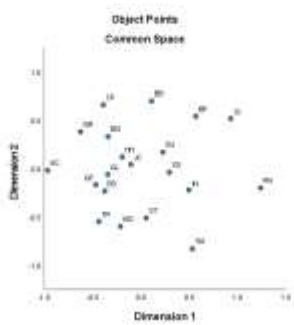
Item	*Code*	*Item*	*Code*
Humeral head	**HH**	Mobility	**MO**
Acromion	**AC**	Stability	**ST**
Bicipital groove	**BG**	Ball and socket	**BS**
Coracoid process	**CP**	Triplanar Motion	**TR**
Glenoid fossa	**GF**	Subacromial bursa	
Biceps brachii	**BB**	Levator scapulae	
Triceps brachii		Axillary artery	
Supraspinatus	**SU**	Circumflex humeral arteries	
Infraspinatus	**IN**	Suprascapular nerve	
Subscapularis	**SS**	Lateral pectoral nerve	
Teres minor		Axillary nerve	
Deltoid		Glenoid labrum	**GL**
Teres major	**TM**	Joint capsule	**JC**
Latissimus dorsi		Glenohumeral ligaments	
Pectoralis major		Coracoclavicular ligaments	
Pectoralis minor		Coracohumeral ligament	
Coracobrachialis		Transverse scapular ligament	
Rhomboids	**RH**	Transverse humeral ligament	
Brachial Plexus	**BP**	Greater tubercle	**GT**
Segmental Innervation	**SI**	Lesser tubercle	

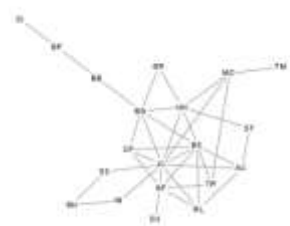
Appendix F: MDS and PFN Data Visualizations

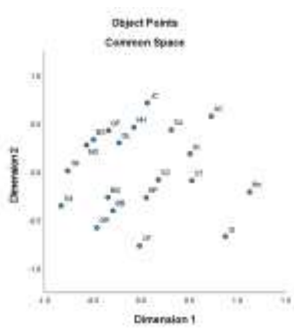
ECSD (Domain Experts; n = 3)

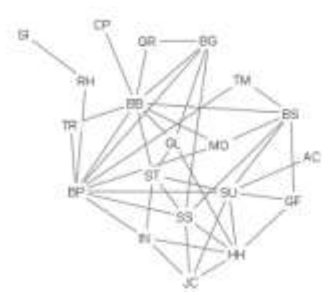




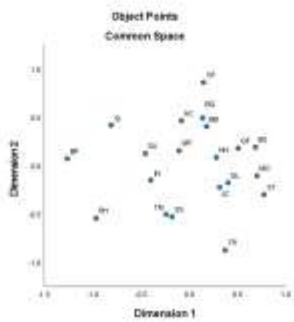


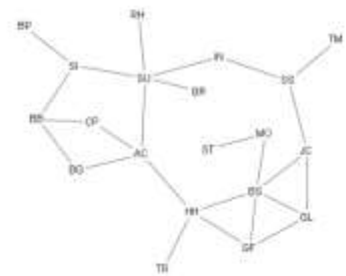


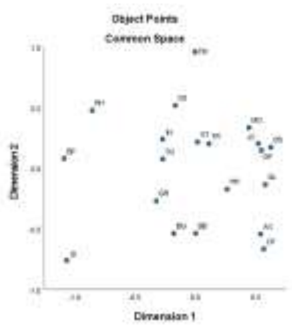


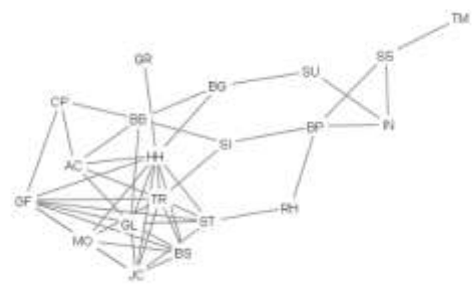


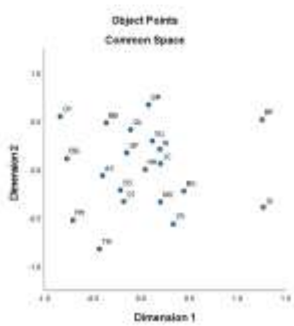
ECSI (Instructors; n = 4)

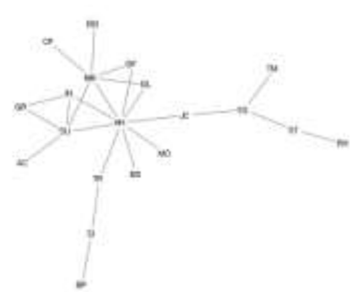


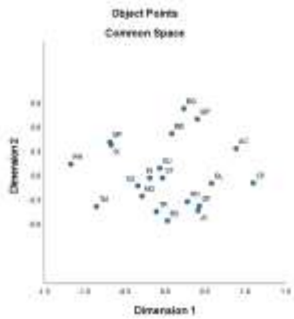


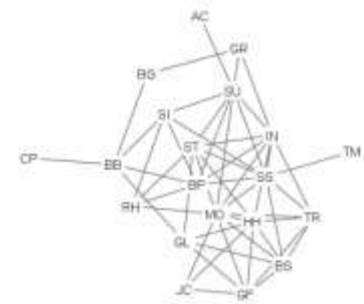




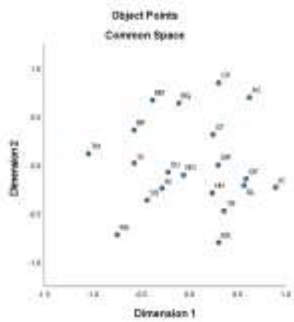


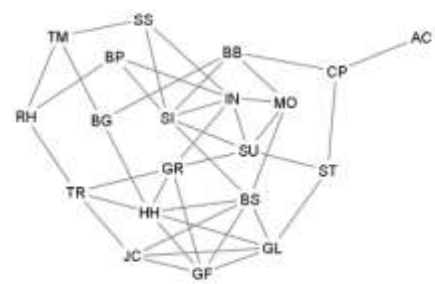


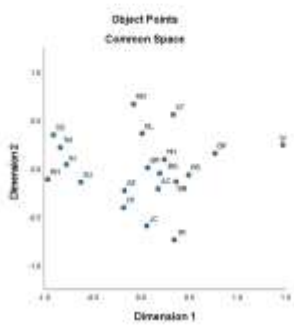


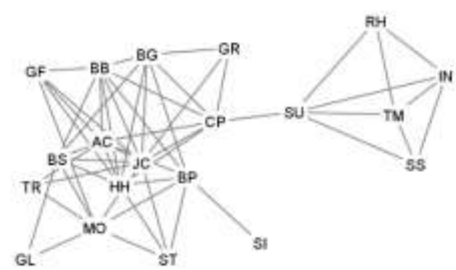


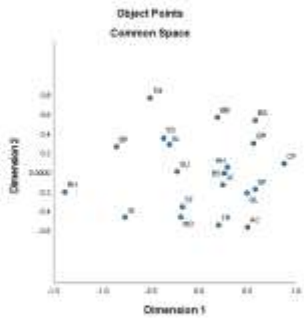
SCS (DPT students; n = 31)

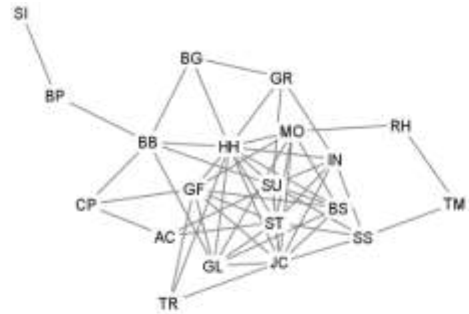


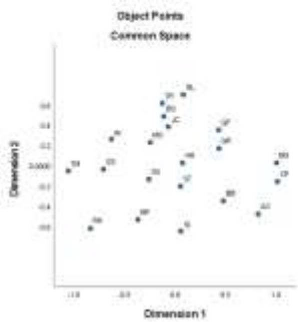


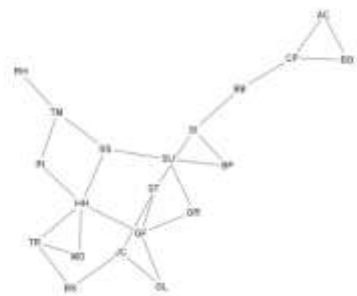


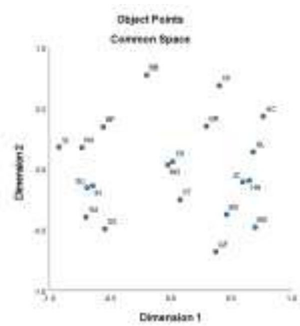


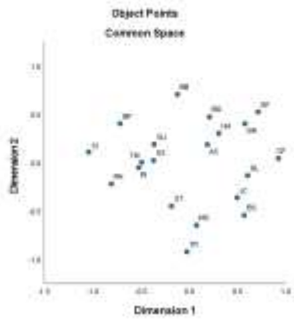

Object Points
Common Space
Dimension 2
Dimension 1

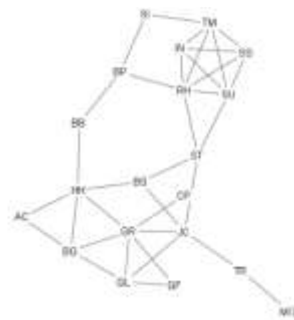

SI
BP
BG
GR
BB
HH
MO
RH
IN
SU
GF
CP
BS
TM
ST
AC
SS
GL
JC
TR

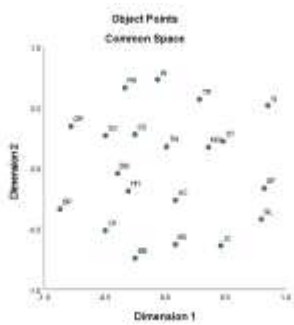

Object Points
Common Space
Dimension 2
Dimension 1

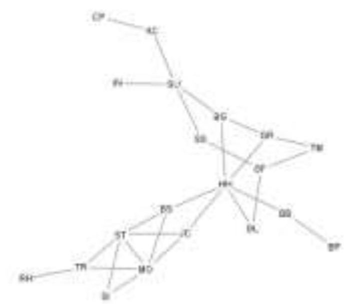


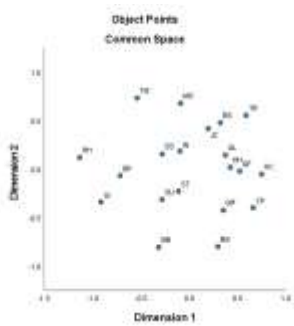

Object Points
Common Space
Dimension 2
Dimension 1

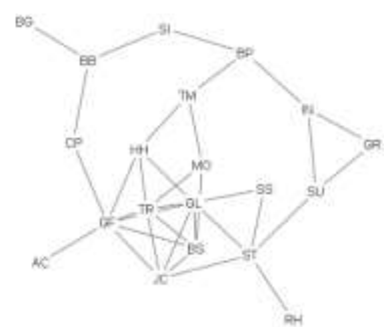

Object Points
Common Space
Dimension 2
Dimension 1

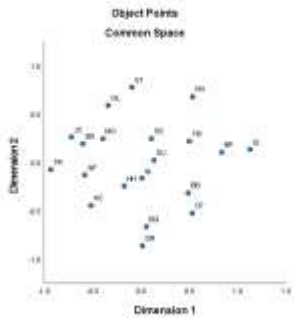



Object Points
Common Space
Dimension 2
Dimension 1

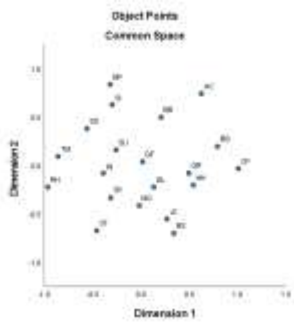


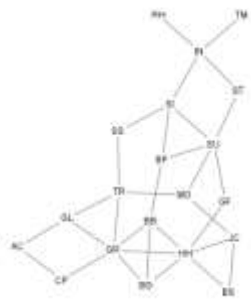

Object Points
Common Space
Dimension 2
Dimension 1

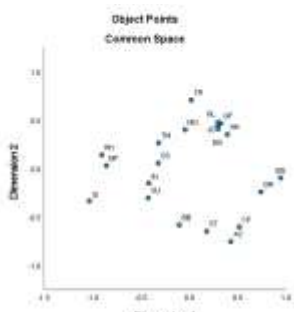

BG
SI
BB
BP
TM
IN
GR
CP
HH
MO
SS
SU
TR
GL
GF
BS
AC
ST
JC
RH

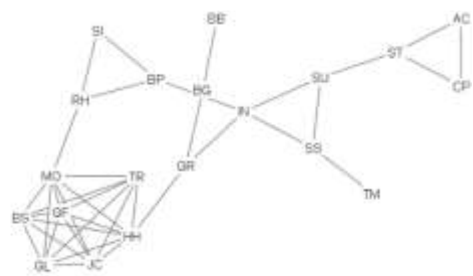

Object Points
Common Space
Dimension 2
Dimension 1

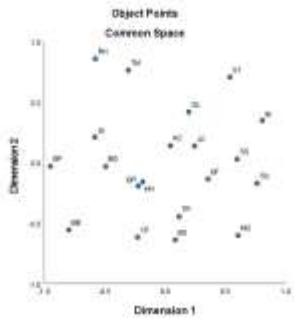

RH
SI
BP
TM
SU
BB
GR
CP
IN
HH
SS
BG
ST
AC
JC
GF
GL
TR
BS
MO

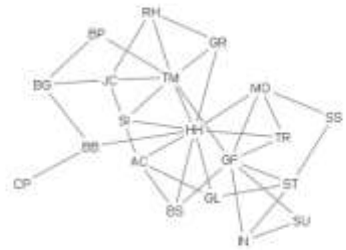

Object Points
Common Space
Dimension 2
Dimension 1

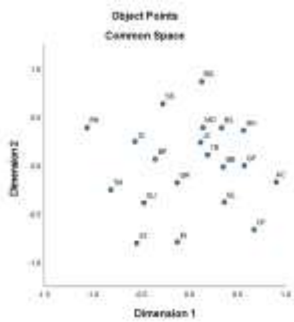


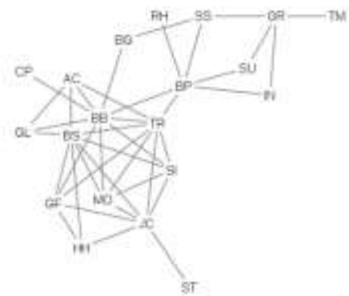

Object Points
Common Space
Dimension 2
Dimension 1

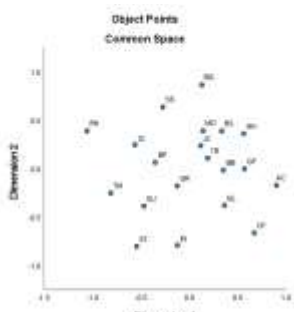

SI
BE
AC
ST
BP
RH
BG
SU
CP
IN
GR
SS
TM
MO
TR
BS
GF
HH
GL
JC

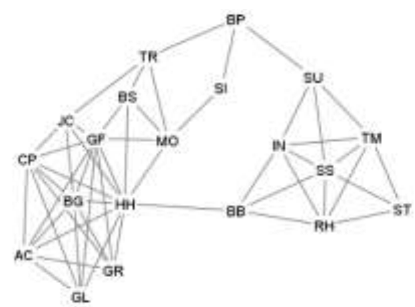

Object Points
Common Space
Dimension 2
Dimension 1

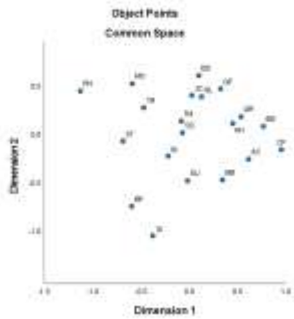

RH
BP
GR
BG
JC
TM
MO
SS
SI
HH
TR
BB
AC
GF
ST
CP
GL
BS
SU
IN

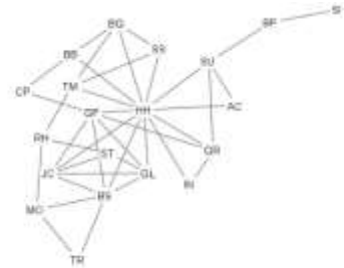

Object Points
Common Space
Dimension 2
Dimension 1

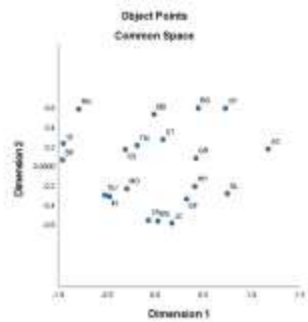

RH
SS
GR
TM
BG
CP
AC
SU
BP
IN
BB
TR
GL
BS
SI
GF
MO
JC
HH
ST

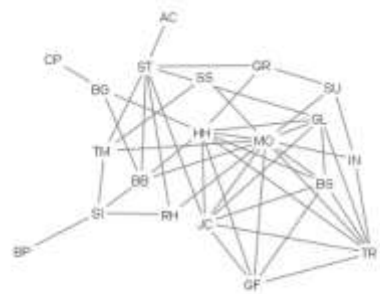

Object Points
Common Space
Dimension 2
Dimension 1

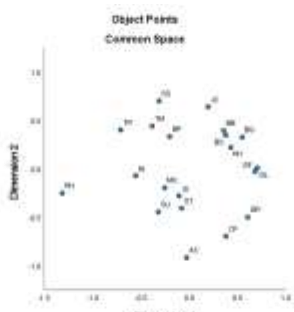

BP
TR
SU
SI
BS
JC
GF
MO
CP
IN
TM
SS
BG
HH
BB
RH
ST
AC
GR
GL

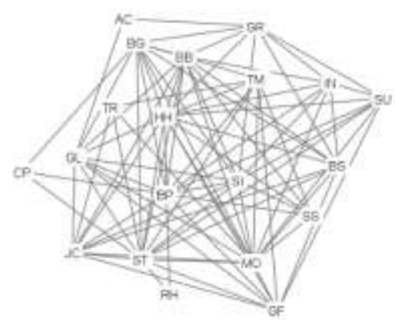

Object Points
Common Space
Dimension 2
Dimension 1

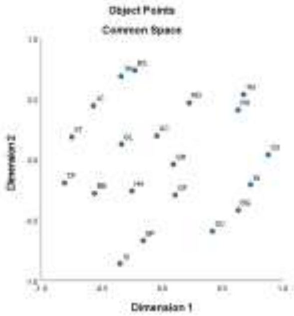

BG
SS
BB
SU
CP
TM
AC
GF
HH
RH
ST
GR
JC
GL
IN
MO
TR

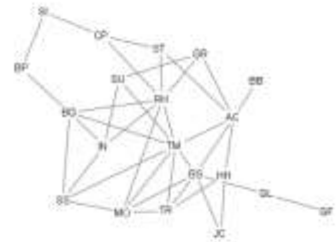

Object Points
Common Space
Dimension 2
Dimension 1

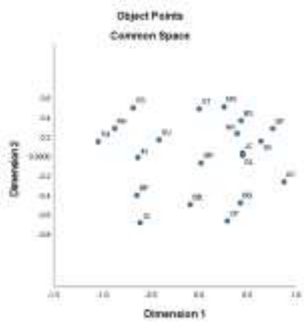

AC
CP
ST
GR
SS
SU
BG
GL
HH
MO
TM
IN
BB
BS
SI
RH
JC
BP
TR
GF

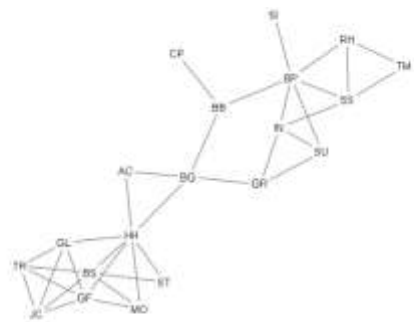

Object Points
Common Space
Dimension 2
Dimension 1

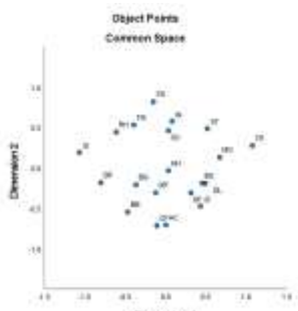

AC
GR
BG
BB
TM
IN
SU
TR
HH
GL
BS
CP
BP
SI
SS
JC
ST
MO
RH
GF

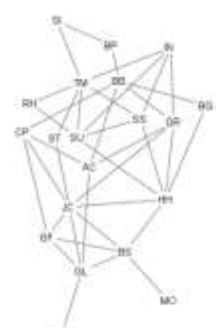

Object Points
Common Space
Dimension 2
Dimension 1

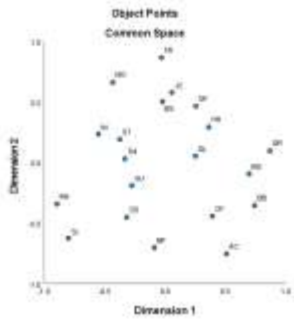

SI
CP
ST
GR
BP
SU
RH
BB
BG
AC
IN
TM
BS
HH
GL
GF
SS
MO
TR
JC

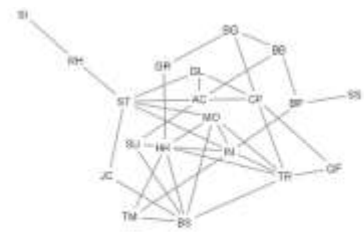

Object Points
Common Space
Dimension 2
Dimension 1

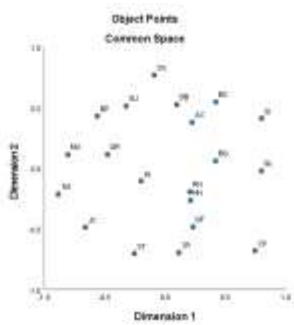

SI
RH
CP
TM
BP
BB
SS
IN
SU
AC
BG
GR
HH
GL
TR
BS
ST
GF
JC
MO

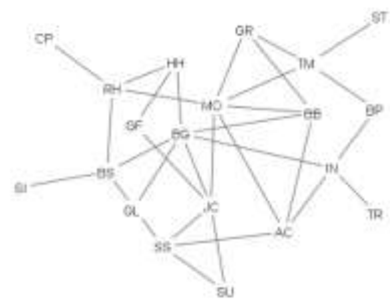

Object Points
Common Space
Dimension 2
Dimension 1

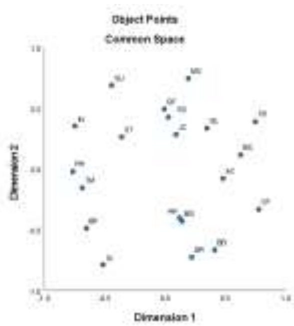

SI
BP
IN
TM
BB
RH
BG
SS
GR
CP
ST
SU
AC
JC
HH
BF
BS
GL
MO
TR

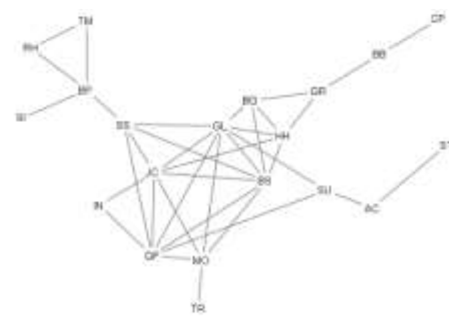

Object Points
Common Space
Dimension 2
Dimension 1

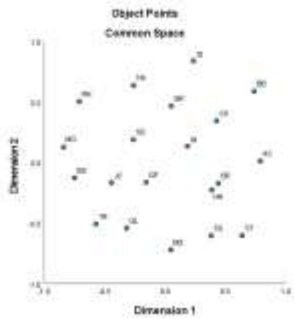

SI
BG
RH
GR
BB
GL
ST
AC
CP
BP
SS
MO
SU
HH
IN
JC
TR
GF
TM
BS

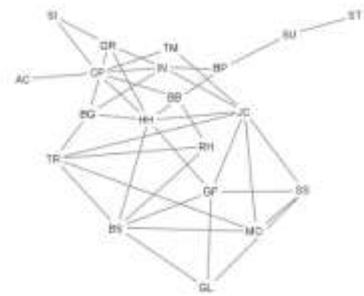

Object Points
Common Space
Dimension 2
Dimension 1

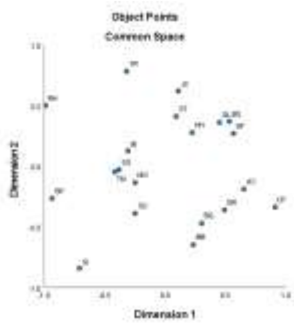

ST
CP
GR
HH
TM
RH
MO
BB
BP
GF
BG
SI
BS
IN
GL
JC
TR
AC
SS
SU

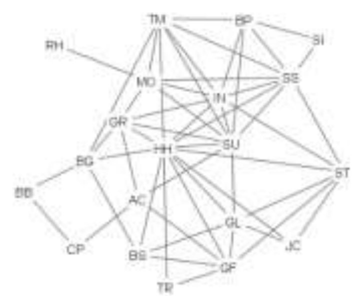

Object Points
Common Space
Dimension 2
Dimension 1

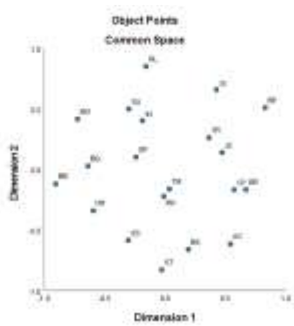

TM
CP
RH
BB
BP
GR
SI
BG
SS
GL
HH
ST
JC
BB
SU
IN
AC
GF
MO
TR

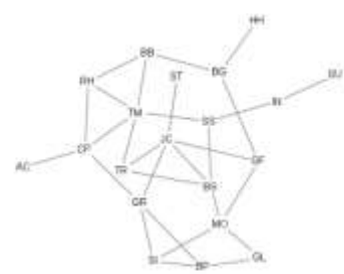

Object Points
Common Space
Dimension 2
Dimension 1

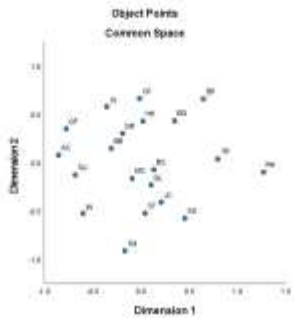

SI
ST
SU
GR
TM
AC
CP
IN
BP
BB
BG
HH
JC
RH
TR
GF
SS
BS
MO
GL

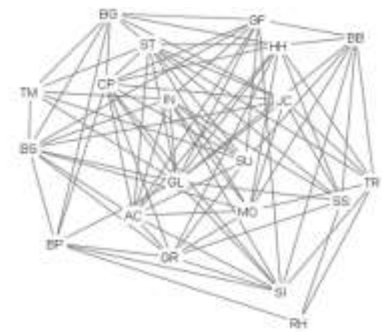

Object Points
Common Space
Dimension 2
Dimension 1

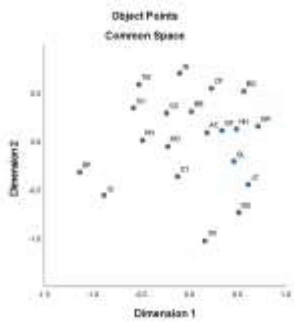

TM
BP
SI
RH
MO
SS
IN
GR
HH
SU
BG
ST
BB
AC
GL
JC
CP
BS
GF
TR

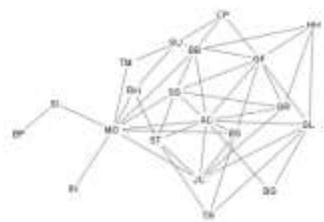

Object Points
Common Space
Dimension 2
Dimension 1

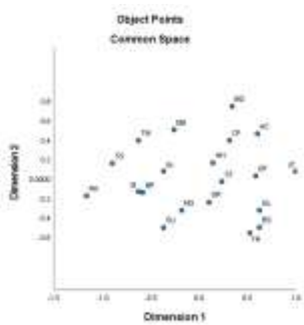

HH
BB
BG
SU
RH
ST
IN
TM
SS
JC
CP
AC
TR
GF
BS
GR
MO
SI
BP
GL

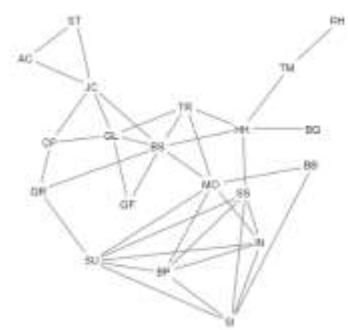

Object Points
Common Space
Dimension 2
Dimension 1

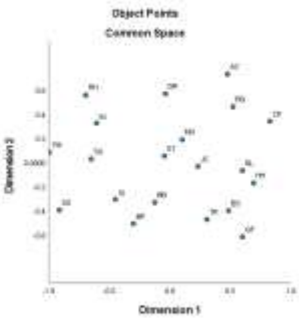

BG
GF
BB
ST
HH
CP
TM
IN
JC
BS
SU
GL
TR
SS
AC
MO
BP
GR
SI
RH

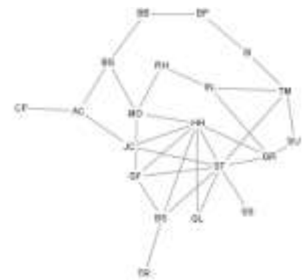

Object Points
Common Space
Dimension 2
Dimension 1

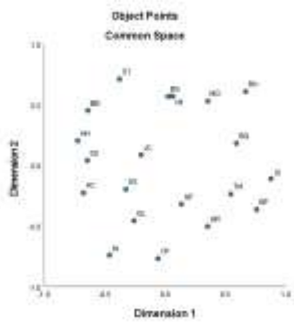



Object Points
Common Space
Dimension 2
Dimension 1




Object Points
Common Space
Dimension 2
Dimension 1




Object Points
Common Space
Dimension 2
Dimension 1