Reconsidering the Role of Synchronous Feedback in Learning Diagnostic Skills: Identifying the Impact of the Instructor

RECONSIDERING THE ROLE OF SYNCHRONOUS FEEDBACK IN LEARNING DIAGNOSTIC SKILLS: IDENTIFYING THE IMPACT OF THE INSTRUCTOR

By SAMUEL JARMAN, BA(HONS)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science McMaster University

MASTER OF SCIENCE (2021) McMaster University
(Health Science Education) Hamilton, ON, CANADA

TITLE: Reconsidering the Role of Synchronous Feedback in Learning Diagnostic Skills: Identifying the Impact of the Instructor

AUTHOR: Samuel Jarman, BA(Hons)
(York University)

SUPERVISOR: Sandra Monteiro, PhD

SUPERVISORY COMMITTEE: Lesley Martin, MD, MSc
Matthew Sibbald, MD, PhD

NUMBER OF PAGES: viii, 50

Lay Abstract

Synchronous feedback has the ability to aid learning. It was hypothesized in this thesis that synchronous feedback that was supportive in nature would improve learning and transfer for learning the skill of visual and auditory cue identification in osteopathic diagnostic procedures. All participants received the same initial learning material, the same instructions for the practice phase of the experiment, and the same videos of a previously learned diagnostic procedure that they identified visual and diagnostic cues from. During the practice phase the three groups were the Supportive Feedback with Specific Content (SC), Supportive Feedback (S), and No Feedback (NF). The differences between groups were evident between diagnostic videos subsequent to the collection of answers for identified cues. The material differences were the delivery of supportive comments regardless of accuracy of answers (SC and S groups), the delivery of specific feedback when accurate cues were identified but placed in the wrong category (SC group only), or the absence of any commentary (NF group). All groups identified cues at similar levels such that the supportiveness of feedback showed no impact on performance. There was a notable difference between groups in relation to the commission of categorization errors where the SC group made less categorization errors with the S group and NF group committing errors at similar rates. The primary benefit of synchronous feedback in this experiment is that the instructor is able to identify errors and provide insight for correction.

Abstract

Introduction: This thesis hypothesized that synchronous feedback which is supportive in nature would have a positive impact on the learning and transfer of the skills of visual and auditory cue identification in osteopathic diagnostic procedures. It was believed that the positive impact of supportive feedback would be evident through accurate identification of both visual and auditory cues. The categories of cues were visually identifiable asymmetrical motion, visual or auditory expressions of pain, and visually identifiable cues of a possible soft tissue tear or motor nerve issue.

Methods: All participants received the same video-based learning resource which was optimized for content (what/how) and cognition (why) followed by the same instructions for the practice phase. During the practice phase all participants were shown a video of a previously learned diagnostic procedure which contained visual and auditory cues. Between videos, participants were all asked the same questions in the same order. In the Supportive Feedback with Specific Content (SC) group participants would receive supportive comments regardless of accuracy of answers and, if they had identified an appropriate physical phenomenon but ascribed it to an incorrect category, they would receive feedback to correct the categorization error. The Supportive Feedback (S) group would receive supportive comments regardless of accuracy of answers but no feedback in relation to categorization errors. The No Feedback (NF) group would receive no supportive comments or feedback in relation to categorization errors. Responses were coded as accurate detection of cues, or categorization errors (detection of cues that were not there, or incorrect categorization of cues).

Results: All groups performed similarly with respect to accurate identification of auditory and visual cues such that there was no identifiable impact in relation to group condition during both the practice and transfer phases. The SC group did commit less categorization errors (11.43%) when compared to the S (28.21%) and NF (31.43%) groups.

Conclusion: The experimental findings supported the hypothesis that supportive feedback enhanced learning outcomes. While not demonstrated through accuracy of cue identification, this was demonstrated through a reduction in cue categorization errors. An additional hypothesis generated from the results of this thesis is that educational designs that allow for the commission of errors by learners followed by correction in the form of direct feedback or group lecture may predict faster attainment of expertise as noted in the reduction of errors.

Dedication

Learning seems to centrally require the ability to identify an error. Thank you to those who have helped me identify my own errors and lack of specific knowledge with or without intentional malice. To the one who told me they knew I would succeed even though they had to leave, thank you. To my daughter, Alice, hopefully one day you read this.

Contents

Chapter 1: Introduction

1.1 Osteopathic Diagnostics

Within the profession of osteopathy, there is a concept that osteopathic practitioners have "seeing and listening hands" (Esteves & Spence, 2014). The concept of "seeing and listening hands" is not supported by research which observes practitioners performing diagnostic procedures. The research observing osteopathic practitioners performing diagnostic procedures consistently notes the use of multiple senses in the diagnostic process (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014). Further research focusing on accurate identification of lumbar spinous processes by osteopathic practitioners is seen to not exceed 78% in experienced practitioners (Snider et al., 2011). It has further been noted that accurate identification of anatomical landmarks is most highly related to experience with palpation (Lavazza et al., 2017). Anatomical variations such as the presence of a lumbar rib and levels of obesity consistently reduce accurate identification of lumbar spinous processes (Snider et al., 2011). Osteopathic practitioners have been observed to perform have variable diagnostic approaches between patients with the same presenting issue (Dinnar et al., 1980). Interobserver reliability is often questionable but is amenable to consensus training (Degenhardt et al., 2005, 2010). The categories where interobserver reliability is most amenable to improvement are tissue texture changes and tenderness (Degenhardt et al., 2005, 2010).

1.2 Diagnostic Teaching Practices

With the observation that osteopathic practitioners utilize different procedural approaches to varying patients with the same presenting issue there is value in considering available information as to how the procedures are taught. The primary information as to the guidance for teaching osteopathic diagnostic procedures is available in textbooks. The common thread within these textbooks is that there is a broad list of procedural steps (Chila, 2010; DiGiovanna et al., 2005; Sammut & Searle-Barnes, 1998). The lists of procedural steps do not present a suggested framework for how or why to choose particular tests so it is generally up to the individual practitioner to make those choices in the moment. The lack of a suggested framework may relate to the issues with interobserver reliability without consensus training (Degenhardt et al., 2005, 2010) as well as the inconsistent approach of individual practitioners with different patients with the same presenting complaint (Dinnar et al., 1980). There is work describing clinical education in a singular Australian context with the note that there is a general paucity of literature describing the phenomenon (Vaughan et al., 2014). There is also work which qualitatively observes the diagnostic reasoning process of osteopathic practitioners while not providing a framework for teaching those skills (Thomson et al., 2014). The author's experience, while not definitive, would echo the general paucity of suggested frameworks for diagnostic choices as well as inconsistent diagnostic practices both within and between practitioners. Along with the aforementioned phenomena there is a lack of suggested frameworks for teaching diagnostic skills aside from long lists of procedures allows for the exploration of literature to identify frameworks which may be useful for the osteopathic context.

1.3 <u>Areas of Consideration</u>

With a general lack of consistent teaching framework there is value in considering phenomenon that would have broad application to teaching in the framework of osteopathic manual therapy. Osteopathic manual therapy includes the need for cue identification in diagnostic procedures to guide treatment which is commonly done with multiple senses (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014), the need to apply motor skills in concert with the use of multiple senses, and the need to be able to apply the aforementioned skills to patients of various presentations. In light of these needs it will be suggested that a framework for teaching and learning activities in osteopathic manual therapy should consider transfer literature for the variation of patient presentation and clinical scenarios as well as motor learning literature to guide general practices for instructors. Without a general framework it is possible that instructors are missing opportunities to effectively guide teaching and learning activities.

1.4 Identification of Cues

So far, the available evidence surrounding osteopathic diagnostic procedures seems to broadly suggest the following:

1. The identification of specific anatomical structures with palpation is not as accurate as may be desired (Snider et al., 2011)
2. Inter-rater reliability with palpation is not highly reliable but may improve with consensus training (Degenhardt et al., 2005, 2010)
3. It is common for osteopathic students and professionals to use a multi-sensory integration approach in diagnostic procedures (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014)

Considering the work presented earlier with respect to inter-rater reliability we are able to identify that the highest levels occur when using palpation to identify tenderness and tissue texture changes (Degenhardt et al., 2005, 2010). The signals of asymmetrical motion may be argued to be most reliable when identified visually and cues of tenderness from the patient may be best identified visually and audibly. Stated another way it is easy to see when things don't move evenly while it is also easy to see and/or hear when a patient is experiencing pain. In light of what seems to be an emergent theme of observing osteopathic diagnostic procedures and identifying when there is a cue from the patient of some form of asymmetrical motion or pain, it may be useful to focus on actively training visual and auditory cue identification during diagnostic procedures as they seem to be the most reliable senses.

1.5 Transfer of Learning

Transfer of learning may be defined as retrieving knowledge or skills learned in one setting so they may be used in a different or dissimilar setting (Grierson et al., 2019). Transfer of learning is often separated between near and far transfer. Noted differentiation between near and far transfer suggests that near transfer occurs between very similar contexts while far transfer occurs between superficially dissimilar contexts (Grierson et al., 2019; K. Kulasegaram et al., 2015; K. M. Kulasegaram et al., 2017; Perkins & Salomon, 1999). Related to near and far transfer are the concepts of superficial and deep features. Superficial features map well to near transfer while deep features map well to far transfer (Grierson et al., 2019). There seems to be agreement that superficial features may be seen as literal objects, concepts, or entities that are explicitly described in a problem statement or situation (Chi & VanLehn, 2012). There does not seem to be as much agreement with respect to deep features, however, one description suggests they may be seen as those features that may not be directly perceived such as problem-solving skills/rules/schemas, causal plots in stories, or mental models formed by students in learning (Chi & VanLehn, 2012).

With respect to osteopathic diagnostic procedures there is an opportunity to discuss the presence of near and far transfer. Near transfer may be considered the procedure itself in that the order a diagnostic procedure is performed in will be the same. Near transfer may also occur when an osteopathic diagnostic procedure is performed in varying locations such that the procedure is the same, the treatment table may be the same, but the room may be different. The presence of far transfer in the application of osteopathic diagnostic procedures may be suggested to occur in the difficult to perceive differences between patient presentations. Far transfer may also be present in the application of an osteopathic diagnostic procedure as the need to problem solve for the unique way the procedure needs be performed with an individual patient.

When considering the challenges previously outlined with respect to palpatory accuracy and the seemingly common use of multi-sensory integration in osteopathic diagnostic procedures it is worth noting how near and far transfer may map on to the skills of visual identification of asymmetrical motion as well as visual and/or auditory identification of pain. The procedural aspects of an osteopathic diagnostic procedure may map most accurately to near transfer. The identification of asymmetrical motion and pain from a patient may also best map to near transfer as the cues may present differently between patients however, they are likely easy to identify.

1.6 Teaching for Transfer

Explicitly identifying the deep structures of the procedure, concept, or problem being learned facilitates far transfer (Chi & VanLehn, 2012; Grierson et al., 2019; K. M. Kulasegaram et al., 2017; G. Norman, 2009; Perkins & Salomon, 1999; Sandberg & Barnard, 1997; Smith & Colby, 2007). When explicit learning of deep structures occurs early it seems to be more effective for far transfer (Chi & VanLehn, 2012; K. M. Kulasegaram et al., 2017; G. Norman, 2009). With respect to early explicit learning of deep structures it seems that having two or more examples presented which contain the deep structure seems to improve the capacity for far transfer (Catrambone & Holyoak, 1989; Day & Goldstone, 2012; Gick & Holyoak, 1983; K. M. Kulasegaram et al., 2017; G. Norman, 2009; Quilici & Mayer, 1996). Broadly it seems that, when aiming to optimize for far transfer, explicitly identifying the deep structures of the procedure, concept, or problem being learned should be done early and with a minimum of two examples.

Investigation has been done with respect to teaching procedural skills in a fashion that integrates both the how (procedural knowledge) and the why (conceptual knowledge) (J. Cheung et al., 2017; J. J. H. Cheung et al., 2019). Combining procedural and conceptual instruction with respect to lumbar punctures was performed and compared to other instructional approaches followed by transfer and retention tests which showed that the integrated instruction did not provide much benefit for procedural knowledge but it did improve transfer (J. Cheung et al., 2017; J. J. H. Cheung et al., 2019).

When considering optimizing teaching for transfer with respect to osteopathic diagnostic procedures it may be suggested that there is value in using lessons learned from both integrated (procedural and conceptual knowledge) approaches as well as far transfer approaches. Initial teaching would be well suited to integrate procedural and conceptual knowledge while also presenting a minimum of two explicit examples of the deeper structures of the procedure. When considering the visual identification of asymmetrical motion as well as the visual and/or auditory identification of pain in a patient there may be benefit from applying the integration of procedural and conceptual knowledge as well as explicitly identifying the application of visual and auditory attention during diagnostic procedures.

1.7 Measuring Transfer

Transfer of learning has classically been measured with sequestered problem solving (SPS) tasks that aim to assess if a learner is able to directly replicate or apply the knowledge or skill being learned (Bransford & Schwartz, 1999). In SPS style assessments of transfer, learners are often tasked with replicating or directly applying the knowledge or skill being investigated unaided (Mylopoulos & Woods, 2014). It has been argued that the implicit theoretical underpinning of SPS assessments is characterized by a belief that students should be able to directly apply previous knowledge or skills to a novel situation and this has been termed the Direct Application Theory of Transfer (DA) (Bransford & Schwartz, 1999). In applying the DA/SPS approach to investigating transfer it is notable that transfer is more successful in experts (Bransford & Schwartz, 1999) and this may relate to experts having more experience over time solving novel problems (Day & Goldstone, 2012). An alternate perspective has emerged in Preparation for Future Learning (PFL) which aims to identify signs of learning as minor improvements in speed and quality of learning in knowledge rich environments (Bransford & Schwartz, 1999). An argument has been made that a PFL approach to identifying transfer may identify learners that have the willingness to use strategies and skills to support future learning (Mylopoulos et al., 2016). A PFL paradigm may be more appropriate to identify evidence of learning in novices as it does not look for accurate application or replication when compared to a DA/SPS paradigm.

PFL has been measured with a double transfer design which may be very onerous (Mylopoulos et al., 2016). The application of a double transfer design begins with two separate learning treatments, moves to a standard transfer problem, exposes both initial groups to a new common learning resource, and finishes with a second transfer problem to compare the performance of the two groups to assess the impact of the initial learning treatments (Mylopoulos et al., 2016). In relation to the labor involved with applying a double transfer design there has been research centered around identifying how learners interface with a common learning resource as a way to measure PFL (Mylopoulos et al., 2016). The underlying assumption of assessing how learners interact with a common learning resource is that future learning relies on the will to use skills and strategies to learn in the future (Mylopoulos et al., 2016). It may be generally suggested that PFL centers around identifying behavior that allows a learner to improve speed or quality of learning in knowledge rich environments (Bransford & Schwartz, 1999). In order to assess PFL, clearly defining a specific behavior that shows improved speed or quality of learning in a novel situation allows for identification of those behaviors as opposed to accurate replication of knowledge or a skill.

Overall, measuring transfer seems to fall in the two broad categories of SPS and PFL. SPS seems to identify accurate replication of knowledge or a skill while PFL seems to identify improvements in speed and quality of learning (Bransford & Schwartz, 1999). Comparatively, successful transfer in an SPS assessment will be accurate replication while successful transfer in a PFL assessment will be the engagement of some skill set that may not result in an accurate performance so much as it will show signs of using a skill set to approach an accurate performance. When looking at learning early in an educational continuum or in a novel situation later in an educational continuum a PFL assessment is more likely to identify an appropriate measure of transfer that must be clearly defined. In the case of cue identification in osteopathic manual therapy diagnostic procedures it may be argued that near transfer may be the most appropriate paradigm as it is ideal to measure accurate use of the skills of cue identification in dissimilar situations.

1.8 Transfer Literature in Osteopathic Manual Procedures

There is a paucity of literature available that specifically mentions transfer of learning in the osteopathic profession. The papers previously described with respect to consensus training (Degenhardt et al., 2005, 2010) and how varying training interventions impact palpatory reliability (Lavazza et al., 2017) are not claimed as investigations of transfer and only peripherally examine behaviors of how educational intervention impacts the outcomes of skill application. Macfarlane and Cornall, 2019 present a qualitative examination of unidentified written learner responses describing advice or strategies they would use to transfer learning from a course to practice. The study collected self-generated responses of preferences of learners. The value generated from this investigation may be argued to be in identifying the beliefs of learners with respect to transfer and not an examination of behaviors that support transfer or learning outcomes.

1.9 Motor Learning

In the context of learning a hands-on skill, such as the vast majority of osteopathic manual therapy procedures, there is value in considering research surrounding motor learning. Further, for practitioners aiming to guide patients to create active movement for either diagnostic or treatment purposes, there is value in utilizing effective strategies to teach patients a novel movement.

The OPTIMAL Theory of Motor Learning synthesizes multiple lines of evidence that suggest motivational as well as attentional factors contribute to performance and learning by strengthening the coupling of goals and actions (Wulf & Lewthwaite, 2016). The factors that were noted as components of the OPTIMAL Theory of Motor Learning are enhanced expectancies (EE), autonomy support (AS), and having an external focus of attention (EF) (Wulf & Lewthwaite, 2016). Teaching behaviors that promote enhanced expectancies (EE) that help improve motor learning are noted as positive feedback, positive social comparative feedback suggesting that a learner is doing better than their peers (true or not), as well as commentary suggesting that peers commonly perform well on the learning task currently being undertaken (Lewthwaite et al., 2017). Behaviors that are noted as autonomy supportive (AS) which aid motor learning are allowing the learner to control the delivery of feedback, controlling the use of assistive devices, controlling the extent of practice, controlling the frequency of demonstrations, as well as task irrelevant choices such as the color of a golf ball in a putting task (Lewthwaite et al., 2017). An external focus of attention (EF) is described as instructing a learner to be attentive to the intended movement effect of a task as opposed to an internal focus of attention placed on body movements such that EF is noted to improve motor learning (Lewthwaite et al., 2017). The use varying combinations of EE, AS, and EF are shown to be additive to motor learning performance (Chua & Lewthwaite, 2018; Lewthwaite et al., 2017). In the context of a learner in osteopathic manual therapy the OPTIMAL Theory of Motor Learning may be well employed by an educator to teach them hands-on osteopathic diagnostic and treatment procedures. Further, in the context of a learner or a practitioner attempting to have a patient create an active motion for diagnostic or treatment procedures, using the concept of EF may be useful in effectively guiding the patient to perform the movement. Additionally, the concepts that seem to show positive outcomes in research relating to the OPTIMAL Theory of Motor Learning may guide how an instructor will effectively behave during practice sessions as each of the subcomponents (AS, EE, and EF) are all descriptors of how the instructor may direct, cue, or provide feedback to the learner during practice.

With the general observation that the current teaching practices for osteopathic diagnostic procedures do not seem well defined it is of value to note that the OPTIMAL Theory of Motor Learning may be a good general guide for this process. As already noted, the OPTIMAL Theory of Motor Learning provides guidance for provision of learning choices (AS), provision of instruction for attention (EF), and provision of feedback (EE). EE seems to be the feature of the OPTIMAL Theory of Motor Learning that has the strongest practical use for general behavior when teaching osteopathic students as it includes positive social comparative feedback as well as guiding feedback behavior which may be an easily missed opportunity with no available framework. With no general framework for feedback, it is possible that instructors may both miss opportunities to provide supportive and effective feedback or provide feedback in an ineffective manner. The present evidence that positive feedback improves performance in motor learning tasks (Lewthwaite et al., 2017) may be considered for investigation in relation to cue identification tasks where the EE concept guides provision of feedback behavior for instructors. AS is also likely to be a construct that is easy to utilize regardless of particular skill set as it may be leveraged with either implicit or explicit choices for learners.

1.10 Research Objective

This thesis aims to examine the impact of instructor feedback style as guided primarily by the concepts of AS and EE with minimal to no guidance from the concept of EF on the ability of novice learners to learn and transfer the skill of identifying motion and pain cues from a practiced situation to a novel situation with the following research question:

- To what degree does a supportive instructor feedback style impact novice learner's ability to accurately learn and transfer the skill of identifying motion and pain cues from a practiced patient active diagnostic procedure to observing a patient passive diagnostic procedure?

1.11 Thesis Overview

This thesis aims to examine the impact of supportive instructor feedback on learning and transfer with respect to identifying motion and pain cues in patient active and patient passive motion diagnostic procedures from the osteopathic perspective. A near transfer paradigm was chosen to measure near transfer in novice learners as they aim to replicate the skills of cue identification in dissimilar diagnostic procedures. The transfer measure is the accuracy of cue identification in the transfer test. The goal was to demonstrate differential accuracy between learners who receive supportive feedback with specific feedback on categorization errors (see section 2.2.6 for explanation of the inclusion of categorization errors), learners who receive supportive feedback only, and learners who receive no feedback. The hypothesis was that supportive feedback would enhance learning as noted through accurate cue identification as well as lead to improved transfer. The rationale for utilizing supportive feedback is generated by the observations of the OPTIMAL Theory of Motor Learning which shows positive outcomes with the application of the concept of EE which occurs in the form of positive social comparative feedback as well as positive experiences (Lewthwaite et al., 2017). The null hypothesis would be evidenced by no differences between groups in either accurate cue identification or categorization errors. The learning environment will be primed for transfer by explicit instruction for the deep structures of visually identifying asymmetric motion as well as identifying visual and audible pain cues. Chapter 2 describes the methods used to investigate the research question. Chapter 3 describes the results of the experiment. The results and their implications for training novice learners to identify appropriate cues in patient active and patient passive diagnostic procedures are discussed in chapter 4.

Chapter 2: Methods

2.1 Overview and Rationale

The primary goal of this work is to assess the extent to which supportive instructor feedback impacts the ability of a learner to learn and transfer the skill of cue visual and auditory cues from a learned osteopathic diagnostic procedure to a novel procedure. In order to assess this phenomenon an experiment was devised to randomly group leaners in to practice phases where instructor feedback was supportive with content, supportive without content, or no feedback was provided subsequent to initial learning from a pre-recorded video. The available literature on osteopathic diagnostic procedures suggests that palpatory accuracy is higher in experts than novices (Lavazza et al., 2017; Snider et al., 2011) while still not reaching levels of accuracy that could be argued to be desirable. In light of the issues with palpatory accuracy as well as the observation that it is common for osteopathic manual practitioners to utilize vision and audition alongside palpation in diagnostic procedures (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014) it was decided to exclude training on palpation and focus on identification of visual and auditory cues: motion asymmetry, lack of motion due to a possible motor nerve issue or possible soft tissue tear, as well as visual and auditory pain cues.

2.2 Methods

2.2.1 Participants

As it has been observed that osteopathic manual therapists commonly utilize both vision and audition in diagnostic procedures (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014) it may be believed that students currently learning osteopathic manual therapy would not be ideal candidates for this study as they have already learned the skills of visual and auditory cue identification either explicitly or implicitly. Considering that current students of osteopathic manual therapy were not ideal candidates, novice undergraduate students at McMaster University were seen as ideal candidates as they are less likely to have previous exposure to utilizing visual and auditory cue identification in diagnostic procedures. All participants were students enrolled in the Introduction to Psychology course with an average age of 18.9 years.

Sample size calculations may be best informed by a review of normative sample sizes for other similar studies (Norman et al., 2012). Using this strategy, the current study initially chose to have a total of 60 participants with each of the three sub-groups (supportive feedback with content, supportive feedback with no content, and no feedback) having 20 participants. The studies of a similar nature that guided this initial choice were done by Dr. Jeffrey Cheung on integrating content (procedural knowledge) and cognition (conceptual/cognitive knowledge) to improve transfer for lumbar puncture procedures (J. Cheung et al., 2017; J. J. H. Cheung et al., 2019). It will be noted that statistical analysis was performed at the mid-point of the projected total such that it was evident that, based on the trend towards the final finding (discussed in section 3.2) a decision was made to stop data collection at 42 total participants (14 participants in each condition).

2.2.2 Recruitment

Ethics approval was sought and provided from the Hamilton Integrated Review Ethics Board on through an original application. The Southern Ontario Neuroscience Association (SONA) system at McMaster was utilized to reach out to undergraduate students in the psychology department at McMaster. Considering the timeframe this experiment was performed in (January – March, 2021) and the varying restrictions to in-person meetings in the province of Ontario in relation to the COVID-19 pandemic, the experiment was run through Zoom. Zoom was deemed to be an acceptable platform for this experiment as it would allow for transmission of visual and auditory data that the learner could identify as well as having the possibility for incidental identification of phenomena that may be unique to teaching and learning activities in relation to visual and auditory cue identification in osteopathic diagnostic procedures. Interested participants signed up for available slots in the SONA system which provided them the link for the scheduled Zoom meeting. Prior to the beginning of the Zoom meeting the participant was emailed a link to a LimeSurvey so they were able to indicate their consent prior to continuing with the Zoom meeting. No markers of personal identification were collected outside of demographic information in the form of age.

2.2.3 Initial Learning Phase

As a note on the nature of the Zoom platform it will be said that the necessary interface for this experiment is for the learner to be able to see and hear the elements of the initial learning resource as well as the practice videos while also being able to transmit their audio so that the instructor is able to hear their answers. The non-essential element of Zoom is for the learner to transmit video of themselves with their camera. The choice to transmit video of themselves or not was considered an implicit choice for the learners with 62% (N = 26) choosing to transmit their video and the remaining 38% (N=16) choosing to not transmit their video. The rationale for maintaining the transmission of video by the learner as an implicit choice is that there is general evidence that there is a positive outcome when learners are able to exert some control over the practice environment (Wulf & Lewthwaite, 2016). It will be specifically noted that the Zoom platform is not specifically mentioned in the research suggesting control over the practice environment; available research was primarily performed prior to Zoom becoming as ubiquitous as the COVID-19 pandemic has made it.

At the beginning of the Zoom meeting, the instructor greeted the learner with their video on and asked if the learner had any questions about the consent process. Subsequent to providing consent and demographic information (age) the instructor turned their own video off, played the initial learning video through the Zoom platform while turning their camera off. The initial learning video was slightly over twenty-six minutes in length. The initial learning video covered the following topics:

1. Definition of osteopathic manual therapeutics
2. Common features of osteopathic diagnostic procedures as observed in research settings (Dinnar et al., 1980)
3. The common use of multiple senses to identify cues in osteopathic diagnostic procedures (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014)
4. Challenges with palpatory accuracy and inter-rater reliability (Degenhardt et al., 2005, 2010; Snider et al., 2011)
5. Visual cues for motion dysfunction (asymmetrical motion between paired structures and complete lack of patient active motion as a possible sign of motor nerve damage or a soft tissue tear)
6. Visual and auditory cues for pain (facial expressions of pain as visual as well as grunts/groans or more direct verbal expressions of pain)
7. Over-arching concepts of how a practitioner may build a useful procedure for osteopathic diagnostics
8. Description of a specific patient active osteopathic diagnostic procedure which will subsequently be demonstrated to the learner

The explanation of each of the above points was delivered in a manner that highlighted the deep features of the concepts. Deep features may be considered to be those features that may not be directly perceived such as problem-solving skills/rules/schemas, causal plots in stories, or mental models formed by students in learning (Chi & VanLehn, 2012). In the case of the initial learning video in this experiment the deep features for visual and auditory cue identification are the cues of asymmetrical motion in paired structures (visual motion), complete lack of motion as possible motor nerve damage or soft tissue tears (visual motion), facial expressions of pain (visual pain), as well as verbal expressions of pain. All of the noted deep features will have different presentations in different patients however they all fall under those stable classifications. The rationale for explicitly teaching those deep features is that there is evidence that explicit identification of deep features in early learning aids far transfer (Chi & VanLehn, 2012; K. M. Kulasegaram et al., 2017; Norman, 2009).

The final portion of the initial learning video demonstrated the specific patient active diagnostic procedure. It will be noted that the term patient active specifically means the patient moves themselves and this is in contrast to patient passive procedures where the practitioner moves the patient. During the demonstration of the specific patient active diagnostic procedure two examples of each form of visual and auditory cue were explicitly demonstrated. The explicit demonstration of a deep feature a minimum of two times has been shown to improve far transfer (Catrambone & Holyoak, 1989; Day & Goldstone, 2012; Gick & Holyoak, 1983; K. M. Kulasegaram et al., 2017; Norman, 2009; Quilici & Mayer, 1996). Upon finishing the second video the initial learning phase was complete.

2.2.4 Practice/Skill Acquisition Phase

The second phase of the experiment, the practice/skill acquisition phase, was performed subsequent to cessation of the initial learning phase. At this point the instructor turned their video back on and informed the learner that they would play a series of videos which displayed the patient active diagnostic procedure and that, in the videos, there would be some combination of the visual and auditory cues they had learned about in the initial learning phase. Each video contained cues engineered to depict the topics covered in training. The instructor informed the learner that they would play a video only once and, at the end of one of the practice videos, the learner would be asked to identify the cues that were present in the practice video. The questions that were to be asked at the end of each practice video were on screen and visible for the learners at this point and the instructor informed them that these were both the questions and the order in which they would be asked subsequent to each video. In order to attempt to be clear in what was said to the learners in relation to each question, the questions will be noted in series along with the description of what was stated to all learners, regardless of condition, when describing the question:

1. **Did you see a motion asymmetry and, if so, where was it (don't worry about technical anatomy terms)?** The instructor clearly informed learners that they did not need to worry about specific anatomical terms in their communication of the location of visually identifiable motion cues. The learners were asked to identify the side of the patient's body and the location as best they were able (ie right ankle/right foot, left shoulder/left arm). The learners were told that their left was the patient's right and, if the instructor needed to clarify right and left, they would ask "your left or the patient's left".

2. **Did you identify pain cues and, if so, did you see it or hear it?** Minimal clarification was provided in relation to this question as the possible answers were constrained well enough within the category
3. **Did you identify cues of a motor nerve issue/soft tissue tear and, if so, where was it (don't worry about technical anatomy terms)?** In relation to the instructions for this question, the instructor made the clarification that cues for this category would include obvious motion asymmetry however, the differential quality was not partial motion but a complete lack of expected motion.

Regardless of condition (supportive feedback with content, supportive feedback without content, no feedback), all learners were informed that the instructor would write their responses on paper, by hand. As the instructor's camera was on during the practice video, they purposefully looked away from the diagnostic procedure to avoid the learner tracking the eyes of the instructor for insight in to cue identification. The instructor performed the following actions between practice videos dependent on condition:

1. Supportive feedback with content – provide a general positive comment such as "good job", "awesome", "good work" OR a positive comment with specific feedback if the learner identified a physical cue but put it in the wrong category such as "remember the difference between asymmetrical motion and a possible motor nerve issue/soft tissue tear is that there is NO movement in a possible motor nerve issue/soft tissue tear". This feedback centered around the deep features relating to categories of cues as presented in section 2.2.3
2. Supportive feedback without content – provide a positive comment that does not refer to the specifics of the skill such as "good job", "awesome", "well done"

3. No feedback – provide no commentary in between practice videos

Subsequent to each practice video, the same slide would appear on the screen for all conditions as a prompt to identify the visual and/or auditory cues they identified. The instructor asked the questions on the slide in series and wrote the learners answers by hand on the data recording sheet. The instructor was directed to record the answers as they were stated for subsequent analysis. Accuracy was determined by appropriate mode of cue identification (visual or auditory), appropriate identification of the side of the body (right or left), and general proximity of term used to actual location of the body. Answers that clearly identified an accurate cue with improper ascription of cue category (i.e. noting a motion asymmetry while ascribing it improperly as a sign of motor nerve/soft tissue tear) were noted as an accurate cue identification with the additional recording of a categorization error. There were engineered cues of all type put in to all videos (practice and transfer) and participants identified non-engineered cues that were present and verified through review of the video. All cues that were identified accurately in each video were scored for all participants such that there were scores for all possible cues as well as engineered cues. To further clarify, all cues accurately noted were used to make a scoring matrix for all participants such that all identifiable cues were measurable, engineered cues were measurable, and categorization errors were measurable.

2.2.5 Transfer Test

Subsequent to the practice/skill acquisition phase where twenty total practice videos were viewed and responded to with varying combinations of visual and auditory cues, learners were shown a video of a patient passive diagnostic procedure that they had no previous exposure to. In the patient passive diagnostic procedure there was a patient on the table and a practitioner performing an organized set of passive motion tests to the patient. Visual cues of motion dysfunction and pain were presented as were audible cues of pain. Subsequent to watching the video of the novel patient passive diagnostic procedure learners were prompted by the same slide they had seen between all practice videos. At this point the instructor verbally requested that the learner identify the cues they saw to the best of their ability. As with the practice videos the instructor wrote the answers down by hand on the data recording sheet as they were stated.

It will be noted here that this transfer test has elements that may map well to far transfer however it functions primarily as a near transfer test. In order to describe the rationale for this claim there will be explanation in relation to a taxonomy of transfer. There is a taxonomy that has been suggested for classifying transfer (Barnett & Ceci, 2002) with the most relevant aspects of that taxonomy being the learned skill, the performance change, the memory demands, and the physical context. In this experimental setting the learned skills will be visual identification of asymmetrical motion, visual identification of pain cues, and auditory identification of pain cues. The performance change will be assessed in terms of accurate identification and categorization of cues. The memory demands will be for the learner to recognize the need to identify cues in the novel transfer situation and to execute the skills learned and practiced prior to do so. The timing of the test, directly after the practice phase, likely reduces the memory demands as there is no temporal dissociation. The physical context is where the novelty will exist such that the learners will practice the skill in a patient active diagnostic procedure while the transfer test presents a novel patient passive diagnostic procedure. The variation in physical context (patient active to patient passive) is where the claim of far transfer primarily lives. The superficial features change quite a bit between a patient active diagnostic procedure and a patient passive diagnostic procedure with the introduction of a practitioner who is now moving the patient as well as the introduction of a treatment table and a change in patient position (standing in active while supine in passive).

The claim that this transfer test primarily functions as a near transfer test is made on the results (see section 3.1.3) as well as the timing of the test being directly after the practice phase making the memory demands for recalling and applying the skill low.

2.2.6 Pilot Study

A pilot study was performed to test the delivery of the experiment. The pilot study identified that the learners were able to identify physical cues accurately however would place the cues in the wrong category. With this observation it was decided that the specific content for feedback in the supportive feedback with specific content group would center on this categorization error. This choice was further supported by observing that learners would accurately identify physical cues or miss them entirely with a generally high level of accurate cue identification.

2.3 Data Collection and Analysis

2.3.1 Data Collection

As noted in section 2.2.4, between each video the instructor would ask the series of questions on the slide in the same order and the learner's answers were recorded by hand on data recording sheets (see Appendix A). Answers were recorded in the manner that participants provided them.

2.3.2 Data Analysis

Data was transcribed from the recording sheets to a spreadsheet. All cues identified by any participant that were proved present by analysis of each video were included as valid cues whether they were the intended engineered cues or not. This created a variable number of cues per video outside of the intended engineered cues against which all participants may be observed. All cues were judged against the appropriate mode of cue identification (visual/audio), the appropriate side of the body (right/left), and general proximity of the cue identified to the appropriate part of the body. Additionally, all cues identified were judged against the appropriate categorization (asymmetry, visual pain, auditory pain, and possible motor nerve issue/soft tissue tear) such that, when a category error was identified, it was scored as such. A category for identified cues that were not present was also recorded. Scores for total cues identified in relation to all possible cues, engineered cues identified, categorization errors, and cues that were identified but not present were calculated for both the practice phase and the transfer test then subdivided by experimental condition (supportive feedback with specific content, supportive feedback, no feedback). Averages for all categories (accurate cue identification, engineered cue identification, categorization errors, and cues identified that were not present) were calculated for analytic comparison. In order to assess for group difference in all of the aforementioned, a one-way ANOVA was performed.

Chapter 3: Results

3.1 Group Averages and One-Way ANOVA

3.1.1 Group Averages Overview

As noted previously, the three groups were: supportive feedback with specific content, supportive feedback, and no feedback. Averages were calculated for cues identified against all available cues, engineered cues identified, categorization errors, and cues identified that were not present for both practice and transfer phases. Along with group averages, a One-Way ANOVA was performed to identify if there were statistically significant differences between group performances.

3.1.2 Group Averages in Practice Phase

During the practice phase it was evident that with both cues identified against all available cues as well as identification of engineered cues was highly similar between groups. Table 1 displays results for these two categories.

	SC Group	S Group	NF Group
All Cues	40.2% (SD 5.0%)	42% (SD 4.7%)	43.1% (SD 5.3%)
Engineered Cues	86.7% (SD 4.7%)	88.8% (SD 4.8%)	87.95% (SD 4.2%)

Table 1. Percent of correct cues detected during practice.

Also, with respect to the engineered cues, all of the groups had fairly high levels of identification as novices. This is perhaps noteworthy in relation to previous identification of issues with palpation where identification lumbar spinous processes does not exceed 78% in experienced practitioners

(Snider et al., 2011). This is not a claim that these tasks are directly equivalent however it is worth considering that novices with no claimed interest in osteopathic manual therapy are able to identify obvious visual cues of asymmetry, visual cues of pain, and auditory cues of pain at a fairly high rate.

In the case of identification of cues against all possible cues there was no notable statistical difference with F (2,39) = 1.19 and p = 0.32. Restricting this analysis to only the engineered cues did not rise to the level of statistical significance with F (2,39) = 0.70 and p = 0.50.

3.1.3 Transfer Averages

During the transfer phase it was again evident that both identification of cues against all possible cues and engineered cues were very similar in all three groups. Table 2 outlines the averages for these two categories at transfer.

	SC Group	S Group	NF Group
All Cues at Transfer	60% (SD 13.6%)	48.6% (SD 15.1%)	58.6% (SD 9.5%)
Engineered Cues at Transfer	92.9% (SD 14.2%)	78.6% (SD 21.1%)	95.2% (SD 12.1%)

Table 2. Percent of correct cues identified at far transfer.

In relation to transfer, these results are conspicuously high in all groups. This will be discussed more fully in section 4.1.3 with a cursory note that this is likely related to the task not seeming difficult overall even for novices.

One-Way ANOVA of identification of cues against all possible cues during transfer shows F (2,39) = 3.24 and p = 0.05. One-Way ANOVA of engineered cues identified during transfer shows F (2,39) = 4.3 and p = 0.02. Examining the group scores (see Table 2) revealed an unexpected pattern. Exploring the scores from individual participants revealed a single outlier in the S group who performed 2 SD below the group mean. Additionally, the confidence intervals for each group had notable overlap such that the range for NF (0.52 to 0.66) overlapped with S (0.42 to 0.56) and SC (0.53 to 0.67). A re-analysis of the transfer scores without the outlier, revealed a different outcome for the detection of all cues (F (2, 38) = 2.2, p = 0.1) and engineered cues (F (2, 38) = 3.1, p = 0.06.

There was only one transfer trial with three engineered cues with a total of five possibly identifiable cues. It is possible that the transfer test design was not sufficiently powered to measure the impact of training.

3.1.4 Categorization Errors

With respect to categorization errors however, there was a difference between groups as displayed in Table 3.

	SC Group	S Group	NF Group
Categorization Errors in Practice	11.43% (SD 7.5%)	28.21% (SD 18.5%)	31.43% (SD 24.1%)
Categorization Errors at Transfer	35.7% (SD 49.7%)	35.7% (SD 49.7%)	50% (SD 51.9%)

Table 3. Percent of categorization errors by training group.

To demonstrate the variation in groups with respect to categorization errors at practice a One-Way ANOVA was again performed. With respect to categorization errors during practice One-Way ANOVA shows F (2,39) = 4.97 and p = 0.01. With respect to categorization errors during transfer One-Way ANOVA shows F (2,39) = 0.37 and p = 0.69. As has already been noted, based on errors of categorization during the practice phase, it is possible to identify the difference between the groups while it is not possible to do so at transfer in relation to a minimal amount of data points.

3.1.5 Cues Not Present

The identification of cues by participants that were not actually present were confirmed by review of the associated video trial. While this was not a common occurrence it was analyzed to ascertain if there was any pattern that was relatable to group condition. Table 4 displays the average occurrence of identification of cues that were not present both in practice and at transfer.

	SC Group	S Group	NF Group
Cues not present in practice	2.15% (SD 3.2%)	1.43% (SD 2.3%)	6.17% (SD 9.6%)
Cues not present in transfer	7.14%	0%	0%

Table 4. Percent of cues described that were not present in practice videos.

While there is the presence of the identification of cues that were not present in each group and there are superficial differences based on the averages at practice, this data does not rise to the level of statistical significance. With respect to the identification of cues that were not present at transfer it will be noted that there was only one individual who identified such a cue and they just happened to be part of the SC group. One-Way ANOVA was performed for identification of cues that were not present during *practice* with F (2,39) = 2.41 and p = 0.10. The p Value for identification of cues that were not present in practice shows that the variation between groups is not statistically significant. One-Way ANOVA was not calculated for identification of cues that were not present in *transfer* as there was only one individual who identified such a cue.

3.2 Summary of Primary Findings

As already noted, the only finding that rose to the level of statistical significance between groups that shows value for further investigation is categorization errors. As already noted, the SC group had a lower incidence of categorization errors (11.43%) then the S group (28.21%) and the NF group (31.43%) with a $p = 0.01$. There is clearly less variance between the S group and the NF group such that they are statistically indistinguishable with $F(1,26) = 0.16$ and $P = 0.69$. This lends credence to the claim that there is a difference with respect to specific behaviors in the SC group condition. The primary difference, as already noted, was feedback with specific content subsequent to categorization errors. This particular behavior, feedback with specific content subsequent to a categorization error, does not seem to do anything to performance with respect to perceptual recognition to visual or auditory cues. With the current experimental setting there was no significant difference between groups with respect to accurate cue identification. The result that is statistically significant was a lower incidence of categorization errors for the SC group, and this may be attributable to specific instructor behavior. The considerations with respect to hypothesis generation will be examined in the discussion presented in Chapter 4.

3.3 Experimental Limitations

Prior to engaging in hypothesis generation, note will be given to limitations within the present experimental protocol. The Covid-19 pandemic led to difficulties with experimental delivery such that the common in-person mode of interaction for synchronous feedback in osteopathic education was not available. There is the possibility that there may be an unknown variable with respect to in-person delivery of this experiment may have shown. Data collection was all performed by a single individual which does allow for consistency in experimental delivery however the lack of variability may confound the findings. The findings are robust enough in supporting the null hypothesis that the single individual collecting data does not seem to have created issues with the findings however it would be of benefit to compare this with multiple individuals collecting data. The nature of the cue identification tasks seemed to be easy enough that all individuals were able to perform well and support the null hypothesis leading to the opportunity to examine whether the findings of this experiment would hold true with subtler cues.

Chapter 4: Discussion

4.1 Interpretation of Results

4.1.1 Primary Finding Showing the Need for Reconsideration of the Initial Hypothesis

As was noted in section 3.2, the primary finding of this experiment is that feedback with specific content subsequent to a categorization error reduces the incidence of further categorization errors. This was not a finding that was predicted. Prior to engaging in this experiment, the belief was that supportive instructor behavior would have some impact on the ability of students to learn the skills of visual and auditory cue identification in a practiced patient active diagnostic procedure as well as transfer those skills to a novel patient passive diagnostic procedure. In this experimental setting there was no discernable difference between groups. Some initial discussion of the results will be undertaken prior to considering evidence which may guide the hypothesis generating aspect of this discussion.

4.1.2 High Rates of Accuracy with Engineered Cues

The practice and transfer videos were all created with particular cues that were intended to be obvious. In all groups, these visual and auditory cues were detected at high rates during practice (SC = 86.7%, S = 88.8%, NF = 87.95%) and during transfer (SC = 92.9%, S = 78.6%, NF = 95.2%). It would require further experimentation to determine if subtler auditory and visual cues would be detectable by novices at similar rates. Considering the high rates of cue detection there may be value in examining the aforementioned construct of subtler auditory visual cues in novices as well as comparing them against experienced osteopathic manual practitioners to identify if this is a stable phenomenon in either group or one that may be leveraged to benefit diagnostic accuracy. This finding also warrants further investigation considering the levels of accuracy observed in palpation not exceeding 78% for identification of lumbar spinous processes in an aforementioned experimental setting (Snider et al., 2011). There is some possibility that purposeful utilization of visual and auditory cues may aid, support, or improve diagnostic accuracy considering that there is evidence that both students and experienced osteopathic practitioners use multiple senses in diagnostic processes (Dinnar et al., 1980; Esteves et al., 2008; Esteves & Spence, 2014) while the common refrain seems to remain that practitioners have seeing and listening hands (Esteves & Spence, 2014).

4.1.3 High Rates of Transfer Accuracy with Engineered Cues

As was already noted in section 4.1.2, part of the high rates of accuracy in transfer with engineered cues (SC = 92.9%, S = 78.6%, NF = 95.2%) may be that the cues were fairly obvious. There are two likely contributors to the high rates of transfer in these novice participants in the form of low memory demands base on the transfer test being performed immediately after the practice phase as well as the initial learning environment being primed for transfer. It is not currently possible to tell if one of these factors contributed more to the high rates of transfer however it is possible to consider this through future experimentation. A simple experiment would be to have one group perform only the transfer test as well as another group perform the practice phase and the transfer test without the initial learning phase to compare rates of accurate identification of engineered cues against the current rates noted in this experimental setting. If the rates were similar with those proposed groups, then it would be reasonable to say that the task itself was easy enough that neither the initial learning phase or practice phase contributed to the present results.

4.1.4 Lack of Finding in Primary Objective

It is evident that, with respect to the capacity of novices to identify visual and auditory cues in both a patient active and patient passive osteopathic diagnostic procedure, the present differences in instructor behavior do not have a notable impact. This lack of finding may point to features of experimental design or the possibility that the skills assessed are not particularly difficult to learn and apply. As discussed in section 4.1.3, it is currently not possible to pull apart the impact of either the initial learning phase or the practice phase from the possible ease of the skill being learned an applied. Section 4.1.3 presents a possible experiment to pull those factors apart in the form of having a group simply perform the transfer task as well as another group perform the practice phase as well as the transfer phase without the initial learning phase. Performing this would require the removal of the category of possible motor nerve issue or soft tissue tear as it would be aimed at the ease of the perceptual task. Further, considering the high levels of accurate identification of engineered cues in both the practice and transfer phases, section 4.1.2 highlights the possible utility of examining whether or not the present levels of cue identification are a stable phenomenon by testing subtler cues in both novices and experienced osteopathic manual practitioners. If visual and auditory cue identification are skills that are both easy to learn by novices and show high levels of accuracy in both novices and experts, there may be reason to purposefully teach these skills to aid in diagnostic processes in osteopathic manual therapy.

4.1.5 Considerations with Respect to Experimental Finding

As noted in section 3.2, the central finding in this experimental setting is that specific feedback subsequent to a categorization error reduces categorization errors as seen between group conditions (SC = 11.43%, S = 28.21%, NF = 31.43%). When no feedback occurs subsequent to a categorization error, they continue at similar rates regardless of which group a participant was in that did not receive feedback (S = 28.21%, NF = 31.43%). The difference between receiving or not receiving feedback subsequent to the commission of a categorization error is evident such that an argument may be made that the continuation or cessation of the commission of error may relate to the participant not detecting the error. The phenomenon presented with the current finding is most likely explained by through the instructor detecting the error that the learner did not followed by provision of feedback which drives the current result. It is the detection of error by the instructor and the lack of detection of the error by the learner that is central to the hypothesis generation that will occur subsequent to this point. For the purpose of further discussion this phenomenon will be termed external error detection. The rationale for terming this phenomenon external error detection will be described in relation to evidence for the mechanisms of internal error detection in simple decision or motor tasks in section 4.2.

4.2 Evidence for Mechanisms of Error Detection

Considering the finding from the current experiment is driven by the instructor detecting an error and providing feedback it is important to consider situations where the learner themselves detects the error. In the case of simple decision or motor tasks there is evidence of a stable neural signal measured by electroencephalogram (EEG) termed error related negativity (ERN) which occurs subsequent to a task error and is not often at the level of conscious awareness (Rodríguez-Fornells et al., 2002; Yeung et al., 2004). There is also evidence that slightly prior to or in parallel with the ERN there is a neural signal termed lateralized readiness potential (LRP) that is indicative of corrective actions being initiated during simple motor tasks (Rodríguez-Fornells et al., 2002). Both the ERN and LRP at least suggest cognitive systems that run without clear conscious awareness when there is the ability to detect an error in the simple decision or motor tasks in which they have been experimentally demonstrated. In relation to the ERN, there is experimental evidence that within a short time course (200 – 500 ms) there is a stable EEG signal termed error positivity (Pe) which denotes conscious awareness of errors in simple tasks (Boldt & Yeung, 2015). When asking participants about their confidence in decisions coupled with the relationship to Pe measurements it was noted that lower Pe measurements related to higher levels of decision confidence regardless of objective accuracy (Boldt & Yeung, 2015). Taken together, the ERN, the LRP, and the Pe signals point to cognitive processes that seem robust. In the context of the current activity of hypothesis generation these phenomena point to the ability of an individual who commits errors in simple tasks to detect and begin to correct them. In the experimental setting in this work, it may be worth considering that categorization errors are not simple enough tasks that the learner is able to detect them. There is reason to believe that, when errors rise to an unknown level of complexity that there may be value in an external source detecting them on behalf of the individual who is unable to identify them.

4.3 Error Reduction in Medical Professions

The evidence available for reducing errors in medical professions seems to point in two broad directions in the form of interventions aimed at systems or individual practitioners. Some highlights of interventions aimed at the systems in which medical professionals work the following suggestions have been made:

1. Robust system for continuous knowledge improvement in surgical pathology (Nakhleh, 2008)
2. Where multiple pathology methods exist, standardize the use of one method for the lab (Nakhleh, 2008)

3. Standardizing the approach of radiologists to images (Bruno et al., 2015)
4. Computer supported diagnostic systems (Berner & Graber, 2008)
5. Feedback and calibration for practitioners (Berner & Graber, 2008)

With respect to points 2 and 3 it will be noted that standardizing diagnostic approaches has shown increased observer reliability in osteopathic manual therapy with respect to particular measures (Degenhardt et al., 2005, 2010) such that there is at least initial evidence that standardization does improve agreement. Interobserver reliability may not be the absolute gold standard in diagnostics but it does show agreement between practitioners. The concept that agreement may not be the gold standard is that agreement is not always directly related to accuracy. Point 1 is a concept that would likely have wide agreement with the possible caveat that it is both difficult to enact in a system and difficult to quantify success in the frame of the individual or the system. Point 4 has the opportunity to offload memory demands on practitioners in the moment however any mistake the computer supported diagnostic system contains may lead to material issues for both patients and practitioners. The possible issues with computer supported diagnostic systems may be dealt with effectively over time and that is outside the scope of this thesis project. Point 5 is of the most interest to the finding of this experiment such that, at least for novices, feedback processes may carry the benefit of identifying errors that a practitioner does not know they made.

Turning to the interventions for error reduction aimed at individual practitioners it is possible to put forth a general set of suggestions for these attempted interventions. Broadly speaking, the following suggestions are a representative sample of the interventions aimed at individual practitioners:

1. Cognitive debiasing (Bruno et al., 2015)
2. Checklists and structured reporting (Bruno et al., 2015; Graber, 2009)

3. Metacognitive/reflective strategies (Berner & Graber, 2008; Graber, 2009)
4. Increased expertise (Berner & Graber, 2008; Graber, 2009)

When considering the evidence for these strategies it seems that the results are in no way what would be hoped. With respect to cognitive debiasing strategies there is general evidence that they do not reduce diagnostic errors (Monteiro et al., 2018). There is specific experimental evidence with ECG interpretation that shows cognitive debiasing efforts do not reduce errors (Sibbald et al., 2019). Experimental evidence demonstrates that checklists for knowledge retrieval does not reduce errors for ECG interpretation (Sibbald et al., 2019). It is possible to describe metacognitive/reflective strategies as cognitive forcing strategies which do not reduce errors in a controlled trial (Sherbino et al., 2014). Reflecting on a diagnostic procedure by residents does not show much reduction in errors (Monteiro et al., 2015). At this point, with at least the small amount of evidence cited, it would seem that strategies 1-3 from the above list do not work in the way that would be desired. Strategy 4, increased expertise, does show error reduction as noted in review articles (Monteiro et al., 2018; Norman et al., 2017; Norman & Eva, 2010). In an experimental setting with junior residents higher speed of diagnostic decisions displays higher levels of objective accuracy which is believed to relate to higher levels of knowledge (Sherbino et al., 2012). Broadly speaking, it seems that reducing errors is most amenable to increased levels of knowledge which may suggest that higher levels of specific knowledge allows for understanding of appropriate choices and behaviors as well as enough knowledge to simply not commit identifiable errors. The challenge with increased expertise as a strategy is the time as well as other resources required to enact a strategy that is already occurring.

4.4 Observational Synthesis

The evidence presented so far suggests that there are basic cognitive mechanisms that identify and correct for known errors committed by an individual in simple decision or motor tasks in contrast with poor results related to having individuals committing errors in medical professions identify possible errors. While not directly related fields of inquiry these two observations suggest that if the individuals committing errors were able to identify their errors they would be likely to show corrective behavior. The low levels of error reduction with strategies aimed at individuals in medical professions may be signals of a possible dead end whereby further attempts to implement the same strategies will not yield better results. It seems fair to suggest that increased expertise is likely attained through some mix of both experience as well as directed education. It is the directed education portion of increasing expertise that is likely to be most amenable to purposeful gains in efficiency and effectiveness. The current experimental result points out that the notable benefit of synchronous feedback is the identification of errors that the learner shows no evidence of detecting. It is the meeting point of these observations that will drive the final portion of this discussion.

4.5 Hypothesis Generation

The evidence presented from this experimental setting showing that external error detection by an instructor reduces categorization errors in novices coupled with the low level of efficacy of current error reduction strategies aimed at individual practitioners along with the evidence that increased expertise does reduce errors points in a general direction. The general direction that emerges from the meeting point of these lines of evidence suggests that it is possible that medical errors occur in relation to lack of detection. When errors are detected there seems to be cognitive processes that begin correction quickly in at least simple decision or motor tasks. Novices require increased expertise which is often imparted by some mix of experience and direct educational efforts. Direct educational efforts allow for detection of errors by an expert on behalf of the learner that is currently unable to detect them while also increasing expertise through the experience. Educational design that supports this consideration is presented in the form of PFL literature which shows some benefit from discovery learning followed by guided learning improves performance in novel situations (Mylopoulos et al., 2016). A very particular educational design that further demonstrates this process is productive failure which specifically advocates for discovery learning followed by guided learning (Kapur, 2016; Steenhof et al., 2019). There is a case to be made that the opportunity to contrast generated solutions to novel problems in discovery learning followed by guided learning allows for comparison of failed solutions from the learner to the correct solution provided by an instructor (Bransford & Schwartz, 1999). The benefits of discovery learning followed by guided learning are not noted on initial measures such that learners who are in a discovery learning condition perform similarly to learners in a guided learning condition as measured on acquisition and application tests (Steenhof et al., 2019). The difference between discovery learning followed by guided learning when compared to guided learning alone is noted in novel transfer situations (Steenhof et al., 2019).

The next consideration for this effort in hypothesis generation is the observation that far transfer is far more identifiable in experts (Bransford & Schwartz, 1999) and this seems to relate to experts having more experience identifying abstraction (Day & Goldstone, 2012). Explicit identification of structural features has been shown to improve far transfer (Gick & Holyoak, 1983). Taken together it is notable that far transfer is more identifiable in experts and improved in non-experts when specific and explicit information is provided by an external source. Non-experts perform better in novel situations where errors may be argued to be more likely when provided explicit instruction while experts show higher levels of far transfer as they have more experience in novel situations.

Pulling these threads together leads to the primary hypothesis to be presented from this experimental setting. The primary hypothesis is that training that incorporates opportunity for error detection for novices in cognitively ambiguous circumstances may lead to faster rates of error reduction as well as lead to lower error rates in future cognitively ambiguous circumstances. The rationale to predict that error reduction will be faster in those who receive external error detection in cognitively ambiguous circumstances is supported by this experiment, previous experiments (Gick & Holyoak, 1983), as well as findings from productive failure literature (Steenhof et al., 2019). The primary target of external error detection for efficacy is likely to be novices or at least those earlier in their learning processes as experts already show lower levels of error in medical professions. Based on available evidence it seems as though it would not take a major change in resource allocation to implement a strategy of discovery learning followed by guided learning. There may be many opportunities for both discovery learning followed by guided learning as well as observation of current practice opportunities where instructors identify errors that learners do not know they are making. This approach may also have the opportunity to be implemented where directed at the system level through feedback and calibration such that more experienced practitioners may provide formal feedback and calibration to less experienced practitioners though this may take changes in resource and time allocation. The opportunity to measure this hypothesis is most likely available in current simulation based medical education scenarios where discovery learning followed by guided learning may be compared varying permutations of external error detection in combination with guided and discovery learning. If no difference in groups is identified in error commission in the initial simulation scenario it may then be measured in a novel simulation situation.

The secondary hypothesis from the observations in this experimental setting center on the observation that identification of the engineered visual and auditory cues was quite high. The hypothesis generated from that observation is that purposeful use of visual and auditory diagnostic cues in osteopathic manual therapy may show higher levels of accuracy than palpation currently does. In section 4.1.2 it was already discussed that this would need to be examined with subtler visual and auditory cues as well as comparing novices and experienced practitioners.

4.6 Conclusion

The primary finding in this experimental setting is that external error detection by an instructor reduces categorization errors. This finding was not expected and, as such, was not led to activities of hypothesis generation. The primary hypothesis from this finding in concert with available evidence is the prediction that external error detection for novices in cognitively ambiguous circumstances may lead to faster rates of error reduction as well as lead to lower error rates in future cognitively ambiguous circumstances. The secondary hypothesis is the prediction that that purposeful use of visual and auditory diagnostic cues in osteopathic manual therapy may show higher levels of accuracy than palpation currently does. Broad suggestions for possible experimental investigation are presented in section 4.1.2 as well as section 4.5.

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a Cause of Diagnostic Error in Medicine. *The American Journal of Medicine*, *121*(5, Supplement), S2–S23. https://doi.org/10.1016/j.amjmed.2008.01.001

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A Simple Proposal with Multiple Implications. *Review of Research in Education*, *24*, 61. https://doi.org/10.2307/1167267

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *RadioGraphics*, *35*(6), 1668–1676. https://doi.org/10.1148/rg.2015150023

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147–1156. https://doi.org/10.1037/0278-7393.15.6.1147

Cheung, J. J. H., Kulasegaram, K. M., Woods, N. N., & Brydges, R. (2019). Why Content and Cognition Matter: Integrating Conceptual Knowledge to Support Simulation-Based Procedural Skills Transfer. *Journal of General Internal Medicine*, *34*(6), 969–977. https://doi.org/10.1007/s11606-019-04959-y

Cheung, J., Kulasegaram, K., Woods, N., Moulton, C., Ringsted, C., & Brydges, R. (2017). Knowing How and Knowing Why: Testing the effect of instruction designed for cognitive integration on procedural skills transfer. *Advances in Health Sciences Education*, *23*. https://doi.org/10.1007/s10459-017-9774-1

Chi, M. T. H., & VanLehn, K. A. (2012). Seeing Deep Structure From the Interactions of Surface Features. *Educational Psychologist*, *47*(3), 177–188. https://doi.org/10.1080/00461520.2012.695709

Chila, A. G. (2010). *Foundations of Osteopathic Medicine*. Lippincott Williams & Wilkins.

Chua, L.-K., & Lewthwaite, R. (2018). Onward and upward: Optimizing motor performance. *Human Movement Science, 60*. https://doi.org/10.1016/j.humov.2018.05.006

Day, S. B., & Goldstone, R. L. (2012). The Import of Knowledge Export: Connecting Findings and Theories of Transfer of Learning. *Educational Psychologist*, *47*(3), 153–176. https://doi.org/10.1080/00461520.2012.696438

Degenhardt, B., Johnson, J., Snider, K., & Snider, E. (2010). Maintenance and improvement of interobserver reliability of osteopathic palpatory tests over a 4-month period. *The Journal of the American Osteopathic Association*, *110*, 579–586.

Degenhardt, B., Snider, K., Snider, E., & Johnson, J. (2005). Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: Improvements from consensus training. *The Journal of the American Osteopathic Association*, *105*, 465–473.

DiGiovanna, E. L., Schiowitz, S., & Dowling, D. J. (2005). *An Osteopathic Approach to Diagnosis and Treatment*. Lippincott Williams & Wilkins.

Dinnar, U., Beal, M. C., Goodridge, J. P., Johnston, W. L., Karni, Z., Mitchell, F. L., Upledger, J. E., & McConnell, D. G. (1980). Classification of diagnostic tests used with osteopathic manipulation. *The Journal of the American Osteopathic Association*, *79*(7), 451–455.

Esteves, J. E., Geake, J., & Spence, C. (2008). Multisensory integration in an osteopathic clinical examination setting. *International Journal of Osteopathic Medicine*, *11*(4), 159. https://doi.org/10.1016/j.ijosm.2008.08.016

Esteves, J. E., & Spence, C. (2014). Developing competence in diagnostic palpation: Perspectives from neuroscience and education. *International Journal of Osteopathic Medicine*, *17*(1), 52–60. https://doi.org/10.1016/j.ijosm.2013.07.001

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Graber, M. L. (2009). Educational strategies to reduce diagnostic error: Can you teach this stuff? *Advances in Health Sciences Education*, *14*(1), 63–69. https://doi.org/10.1007/s10459-009-9178-y

Grierson, L., Norman, G., Monteiro, S., & Sibbald, M. (2019). Chapter 9—Simulation-Based Education and the Challenge of Transfer. In G. Chiniara (Ed.), *Clinical Simulation (Second Edition)* (pp. 115–127). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00009-7

Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*, *51*, 1–11. https://doi.org/10.1080/00461520.2016.1155457

Kulasegaram, K. M., Chaudhary, Z., Woods, N., Dore, K., Neville, A., & Norman, G. (2017). Contexts, concepts and cognition: Principles for the transfer of basic science knowledge. *Medical Education, 51*(2), 184–195. https://doi.org/10.1111/medu.13145

Kulasegaram, K., Min, C., Howey, E., Neville, A., Woods, N., Dore, K., & Norman, G. (2015). The mediating effect of context variation in mixed practice for transfer of basic science. *Advances in Health Sciences Education*, *20*(4), 953–968. https://doi.org/10.1007/s10459-014-9574-9

Lavazza, C., Milano, V., Abenavoli, A., & Maggiani, A. (2017). How type and number of training sessions influence the reliability of palpation. *Journal of Bodywork and Movement Therapies*, *22*. https://doi.org/10.1016/j.jbmt.2017.07.012

Lewthwaite, R., Cardozo, P., & Chiviacowsky, S. (2017). Triple play: Additive contributions of enhanced expectancies, autonomy support, and external attentional focus to motor learning. *Quarterly Journal of Experimental Psychology (2006)*, *71*, 1–22. https://doi.org/10.1080/17470218.2016.1276204

Macfarlane, C., & Cornall, D. (2019). Expanding options: Supporting skills transfer from a post-graduate osteopathy program to clinical practice. *International Journal of Osteopathic Medicine*, *33*, 38–45. https://doi.org/10.1016/j.ijosm.2019.05.002

Monteiro, S. D., Sherbino, J., Patel, A., Mazzetti, I., Norman, G. R., & Howey, E. (2015). Reflecting on Diagnostic Errors: Taking a Second Look is Not Enough. *Journal of General Internal Medicine*, *30*(9), 1270–1274. https://doi.org/10.1007/s11606-015-3369-4

Monteiro, S., Norman, G., & Sherbino, J. (2018). The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. *Journal of Evaluation in Clinical Practice*, *24*. https://doi.org/10.1111/jep.12907

Mylopoulos, M., Brydges, R., Woods, N., Manzone, J., & Schwartz, D. (2016). Preparation for future learning: A missing competency in health professions education? *Medical Education*, *50*, 115–123. https://doi.org/10.1111/medu.12893

Mylopoulos, M., & Woods, N. (2014). Preparing medical students for future learning using basic science instruction. *Medical Education*, *48*, 667–673. https://doi.org/10.1111/medu.12426

Nakhleh, R. E. (2008). Patient Safety and Error Reduction in Surgical Pathology. *Archives of Pathology & Laboratory Medicine*, *132*(2), 181–185.

Norman, G. (2009). Teaching basic science to optimize transfer. *Medical Teacher*, *31*(9), 807–811. https://doi.org/10.1080/01421590903049814

Norman, G., Monteiro, S., & Salama, S. (2012). Sample size calculations: Should the emperor’s clothes be off the peg or made to measure? *BMJ*, *345*. https://doi.org/10.1136/bmj.e5278

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

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/10.1097/ACM.0000000000001421

Perkins, D., & Salomon, G. (1999). *Transfer Of Learning*. *11*.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144–161. https://doi.org/10.1037/0022-0663.88.1.144

Rodríguez-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time Course of Error Detection and Correction in Humans: Neurophysiological Evidence. *Journal of Neuroscience*, *22*(22), 9990–9996. https://doi.org/10.1523/JNEUROSCI.22-22-09990.2002

Sammut, E. A., & Searle-Barnes, P. J. (1998). *Osteopathic Diagnosis*. Nelson Thornes.

Sandberg, J., & Barnard, Y. (1997). Deep learning is difficult. *Instructional Science*, *25*(1), 15–36. JSTOR.

Sherbino, J., Dore, K. L., Wood, T. J., Young, M. E., Gaissmaier, W., Kreuger, S., & Norman, G. R. (2012). The Relationship Between Response Time and Diagnostic Accuracy. *Academic Medicine*, *87*(6), 785–791. https://doi.org/10.1097/ACM.0b013e318253acbd

Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Canadian Journal of Emergency Medicine*, *16*(1), 34–40. https://doi.org/10.2310/8000.2013.130860

Sibbald, M., Sherbino, J., Ilgen, J. S., Zwaan, L., Blissett, S., Monteiro, S., & Norman, G. (2019). Debiasing versus knowledge retrieval checklists to reduce diagnostic error in ECG interpretation. *Advances in Health Sciences Education*, *24*(3), 427–440. https://doi.org/10.1007/s10459-019-09875-8

Smith, T. W., & Colby, S. A. (2007). Teaching for Deep Learning. *The Clearing House*, *80*(5), 205–210. JSTOR.

Snider, K. T., Snider, E. J., Degenhardt, B. F., Johnson, J. C., & Kribs, J. W. (2011). Palpatory accuracy of lumbar spinous processes using multiple bony landmarks. *Journal of Manipulative and Physiological Therapeutics*, *34*(5), 306–313. https://doi.org/10.1016/j.jmpt.2011.04.006

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

Thomson, O. P., Petty, N. J., & Moore, A. P. (2014). Diagnostic reasoning in osteopathy – A qualitative study. *International Journal of Osteopathic Medicine*, *17*(2), 83–93. https://doi.org/10.1016/j.ijosm.2013.08.002

Vaughan, B., MacFarlane, C., & Florentine, P. (2014). Clinical education in the osteopathy program at Victoria University. *International Journal of Osteopathic Medicine*, *17*(3), 199–205. https://doi.org/10.1016/j.ijosm.2013.10.010

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, *23*(5), 1382–1414. https://doi.org/10.3758/s13423-015-0999-9

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Appendix A: Blank Data Collection Sheet

Participant Number:_______________ Participant Age:______

Feedback Condition (please circle): 1. S w/Content 2. S no Content 3. No Feedback

Practice Video #	Participant Identified Cue(s)	Engineered Cue Present
1		Right heel off ground (motion)
2		Pain in face with extension (visual pain)
3		Left knee limited flexion (motion)
4		Grunt on full flexion (auditory pain)
5		Trunk twist left on flexion (motion)
6		Right arm no flexion (nerve/tear)
7		No neck flexion and face (motion and visual pain)
8		Right arm half bent with grunt (motion and auditory pain)
9		Left arm will not flex with long grunt (nerve/tear and auditory pain)
10		Left heel off ground with face and grunt on shoulder flexion (motion, auditory, and visual pain)

11		Trunk twist right with minimal flexion (two motion cues)
12		Grunt on full flexion with left arm better flexion than right (auditory pain and motion)
13		Both knees minimal flexion with no right arm flexion (two motion cues)
14		Trunk twist left on extension with quick yelp (motion and auditory pain)
15		Trunk bend left on flexion and right heel off ground (two motion)
16		Head turn left on flexion and extension (two motion)
17		Long grunt on flexion and quick face on extension (auditory and visual pain)
18		Right elbow minimal flexion and extension (two motion)
19		Right knee limited extension and long grunt (motion and auditory pain)
20		Trunk bend right on extension with face (motion and visual pain)
Transfer Video	Participant Identified Cue(s)	Engineered Cues Present
		Right knee bent, grunt and face on shoulder abduction, neck won’t turn right (motion, auditory/visual pain)