UNIVERSITY OF CALGARY

Design, Development and Evaluation of the Standardized Orthopedic Assessment Tool with Athletic Therapy Staff and Students

by

Mark Richard Lafave

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF KINESIOLOGY

CALGARY, ALBERTA

SEPTEMBER, 2007

Your file Votre référence
ISBN: 978-0-494-34154-4
Our file Notre référence
ISBN: 978-0-494-34154-4

Table of Contents

List of Tables

List of Figures

Abstract

Assessment of clinical competence in medical and paramedical professions is challenging. Teaching those same constructs can be equally challenging. The ultimate goal is to educate future health professionals to be competent. Student assessment is a critical component of this goal. Athletic therapists perform orthopedic assessments on musculoskeletal injuries in order to devise a plan to best rehabilitate the injury. As such orthopedic assessment is an important clinical competency in the profession. The purpose of this study was to develop an assessment tool (the Standardized Orthopedic Assessment Tool or SOAT) that could be employed in performance-based, practical examinations while concomitantly employing the tool to help teach the orthopedic assessment clinical competence.

This study was a multiphase project consisting of content validation (phase I), initial reliability testing (phase II) and integration of the SOAT into various athletic therapy programs in Canada (phase III). Once the SOAT was content validated, reliability testing for the knee, shoulder and ankle region established the initial reliability (internal consistency) for each (α = .83, .82, .91, respectively). The third phase consisted of treatment groups being randomly assigned to three interventions that involved the SOAT and one comparison group. The University of Winnipeg (group 1) was randomly assigned as the comparison group. Interventions for the treatment groups three and four included complete exposure to the e-SOAT (a web-based version of the SOAT) for all educational stakeholders (i.e., students, faculty and clinical supervisors). The primary difference between groups three and four was that group 3 was only given access to the e-

SOAT via desktop computer while group 4 had access to both desktop computer and personal digital assistants (PDAs). Only the instructors were exposed to the SOAT with group 2 and they were not permitted to share the SOAT explicitly with students or clinical supervisors.

All students who participated in the study (n=58) were tested using the SOAT by the primary investigator (who acted as the standardized patient) and local examiners/raters. Knee and shoulder regions were the focus for the final phase of research. Both demonstrated good reliability: $\alpha = .93$ & .90, respectively. A nested design, generalizability coefficient was calculated on group 2 (Concordia University) with positive results: $Ep^2_{(r)} = .73$; $Ep^2_{(s)} = .89$. An analysis of covariance (ANCOVA) revealed a significant difference between group 3 and groups one and two indicating the SOAT had a positive effect on final exam scores [$F(2, 110.4) = 28.6$, $p = .01$.] The covariate of "total number of post-secondary courses" did not have a significant influence on the results of the ANCOVA. Group 4 did not participate in the final phase testing because there were no student volunteers. A Predictive Learning Assessment Model (PLAM) is introduced in light of the research questions and offered as one explanation for the results of this study.

Acknowledgements

I would like to thank Dr. Larry Katz for his help and support throughout my PhD experience. I would also like to thank my committee members, Dr. Norman Vaughn and Dr. Claudio Violato. You all provided wonderful insight and guidance throughout the whole process.

I would like to thank my colleague, Dale Butterwick for his encouragement, leadership and mentorship throughout my PhD and perhaps most importantly, prior to entering the PhD.

I would like acknowledge the University of Winnipeg, Manitoba and Concordia University for their assistance in this study.

I would like to acknowledge my employer, Mount Royal College, for their flexibility and support throughout my educational experience. I was supported with financial resources, and time to complete my studies. In particular, Dean Brenda Hendrickson supported my quest throughout.

Finally, and most importantly of all, I would like to extend my sincerest appreciation to my wife and children for their support. There were many Saturdays and Sundays that were lost throughout my experience and they were always understanding and patient beyond their years (Taylor, Hailey and Caitlin). To my wife, Lynne, I owe you so much. Your support and encouragement during the tough times were critical. I will always remember that Sunday night when I asked you the question "is this really worth it?" And your response was, "you started it … you'll finish it!" Most wives may not be able to say that with such conviction, but because you had done it under what was seemingly tougher circumstances, you inspired me!! I love you!

CHAPTER ONE: INTRODUCTION

Capitalizing on the intimate connection between teaching, learning and assessment can assist achieving educational outcomes and goals (Driscoll, 2005). There have been many studies that attempt to measure the impact of teaching on learning medical-related skills (Fox, Dacre, & Mclure, 2001; Bowen & Irby, 2002; Spencer, 2003; Pratt, Arseneau, & Collins, 2001; Jaques, 2003). It appears there are more questions than answers in this regard. Currently, there is a strong movement away from teacher-centred systems to student-centred systems (Biggs, 1996; Biggs, 1999). However, it is critical that medical and paramedical professions continue to meet curriculum standards set out by professional accrediting bodies such as the Canadian Medical Association or the Canadian Athletic Therapists' Association so the public is protected (Cavanaugh, 1991). As a result, it is challenging to have both student-centred approaches to learning while also meeting the needs of an external accrediting body. The current study will introduce a reframed model, originally described by Earl (2003) to demonstrate how assessment can be employed in multiple ways to guide teaching, learning and assessment, thus, meeting needs of all stakeholders.

Teaching clinical competence with orthopedic assessments skills can be challenging due to the practical nature of these competencies and the way in which these skills have been traditionally taught (Coady, Kay, & Walker, 2003; Kay & Walker, 1998; Bowen et al., 2002). Typically, curricular design has included theoretical concepts in a traditional classroom setting followed by a laboratory session and finally by clinical internship/placements to reinforce those same skills. Teaching and learning throughout a

clinical internship is difficult to guide since learning is often opportunistic and unstructured (Bowen et al., 2002). Lack of structure in this educational setting coupled with the fact that supervisors/preceptors are often far removed from the original curriculum or curricular goals add to the challenge. Many studies have shown there is a limited relationship between the total hours in those placements (quantity) and success in learning the clinical competence construct (Chatenay et al., 1996; Daelmans et al., 2004). However, feedback and quality of supervision have been shown to be good predictors of student success (Chatenay et al., 1996). Structured and guided experiences in practical environments may be the best method of teaching those clinical skills.

In the current study, a Standardized Orthopedic Assessment Tool (SOAT) that can be employed in both formative and summative settings is proposed as the solution to a structured learning and assessment model. A tool only earns the title "standardized" once it has successfully passed a number of developmental steps that ensure it measured the underlying construct. The primary construct central to this study is an orthopedic assessment clinical competence. Although, there have been studies that evaluated the impact of teaching and learning orthopedic assessment skills in medicine (Coady et al., 2003; Kay et al., 1998), there have been no studies that have evaluated the impact of employing an assessment tool to measure its impact on learning orthopedic assessment skills in medicine or athletic therapy. The steps associated with tool development (i.e., building evidence of the construct) as well as implementation of the SOAT into various curricula in Canada, underpin the research questions for this thesis.

Statement of the Problem

It may be possible to guide students through the expected competencies if a structure is provided to them, the instructor(s) and the clinical supervisor(s). In other words, if learning goals are explicitly stated and shared with all stakeholders, better learning could result (Mager, 1997; Gagne, 1985). The current study proposed that the e-SOAT (electronic version of the Standardized Orthopedic Assessment Tool) be used to provide the structure to learning the orthopedic assessment competency for athletic therapists in Canada. Care must be taken to ensure the SOAT is a valid and reliable instrument prior to integration to an educational setting. As a result, the first research question/challenge(s) was to create a valid and reliable assessment tool that measures orthopedic assessment skills in athletic therapy students. Validation is a multi-step process that often requires a number of studies to establish the underlying construct (Violato, Marini, & McDougall, 1998). As a result, content validation is the first phase of research followed by reliability testing.

Once evidence for a tool's validity has been established, the next phase of research needs to measure the impact of using that tool formatively and summatively on learning the underlying construct that the tool measures. It is proposed that explicit exposure to all stakeholders (i.e., students, clinical supervisors and instructors) with a valid and reliable assessment tool will have a positive impact on learning when compared with implicit exposure (i.e., instructor is the only stakeholder with access) or no exposure for any stakeholders. In this regard, the e-SOAT is proposed as a tool that can be employed within a predictive learning assessment model (PLAM) context for orthopedic

assessment skills. Some evidence for this model (the PLAM) is provided within this dissertation while other evidence can be found within existing literature.

CHAPTER TWO: LITERATURE REVIEW

The dissertation consisted of three separate phases that culminated into a large study in the third phase. A review of literature for each section is based on the primary purpose of each phase of research. The phase I literature review will focus on content validation, in general with performance-based, practical examinations since no tool to measure orthopedic assessment skills has been reported in the literature. The phase II literature review will focus on testing reliability with performance-based, practical examinations using the Standardized Orthopedic Assessment Tool (SOAT). The most common method to assess skills in medical and paramedical professions has become the Objective Structured Clinical Exam (OSCE), so particular attention is paid to reliability with OSCEs (Probert, Cahill, McCann, & Ben-Shlomo, 2003). The third phase literature review will consist of studies that have attempted to employ OSCEs as a means of teaching, learning and/or assessment. Few studies have attempted to measure this impact in an experimental/quasi-experimental design, so the literature review will be extrapolated from existing OSCE literature.

Phase I – Content Validation of a Clinical Competency

Measurement of Clinical Competence

Assessing clinical competence is complex and even though it is highly unlikely that any one measurement instrument will capture the full spectrum that would define it (Neufeld & Norman, 1985), the pursuit continues. Some common measures of clinical competence include direct observations, oral examinations, written examinations, global rating scales, medical record reviews, patient management problems, computer

simulations and simulated patients. Written examinations lack the validity to measure a clinical competence construct and other high fidelity measures such as global rating scales, direct observation or oral exams may lack reliability (Harden & Gleeson, 1979). Validity and reliability must work together because they are at the heart of sound measurement principles (Hopkins, 1998).

In order for a measurement tool or instrument to be considered "valid" it needs to go through a number of steps and/or studies to build evidence towards the construct that is being measured (Messick, 1998; Violato et al., 1998). Traditionally, validity has been separated into various components (content, criterion and construct) that collectively provide evidence for the ultimate goal of construct validity (Hopkins, 1998; Violato et al., 1998). Face validity is also considered to be an important first step to building validity (Violato et al., 1998; Krathwohl, 1998). Unfortunately, some OSCE studies reported in the literature fail to go beyond face validity building towards construct validity (Vivekananda-Schmidt, Lewis, & Hassell, 2005; Probert et al., 2003). The specific construct being measured in the first phase of the research was orthopedic clinical competence for undergraduate athletic therapy students.

The Objective Structured Clinical Exam (OSCE)

The objective structured clinical exam (OSCE) was first introduced by Harden et al (1975) in attempt to standardize and objectify evaluation of clinical knowledge, skills and abilities (i.e., clinical competence) in medical education (Harden, Stevenson, Downie, & Wilson, 1975). In a follow up to the original concept, Harden and Gleeson (1979) outlined three key questions that, if answered in the affirmative, could be an indicator of clinical competence:

1) Is it valid?

2) Is it reliable?

3) Is it practical?

Each of these questions will be reviewed in sequence below. Considerable work has been done to address psychometric issues (validity, reliability) for written examinations (Barrows, 1993); Violato, Salami, & Muiznieks, 2002). However, written examinations lack the validity to "show how" to complete tasks and demonstrate abilities in medical education (Barrows, 1993; Violato, Salami, & Muiznieks, 2002; Miller, 1990). Miller (1990) outlined a model (Figure 2.1) which demonstrated higher order assessment that should occur with the medical profession, and arguably all allied health care professionals, including athletic therapy. At the peak of Miller's pyramid, the word "does" is intended to represent a clinical competency construct. Hence this is one of the key psychometric issues identified by Harden and Gleeson (1979): validity. The OSCE is considered the gold standard to measure clinical competence according to some authors because it has the greatest validity (Probert et al., 2003; Kramer et al., 2002; Vivekananda-Schmidt et al., 2005) while others are not convinced (Cox, 2000; Hodges, 2003).

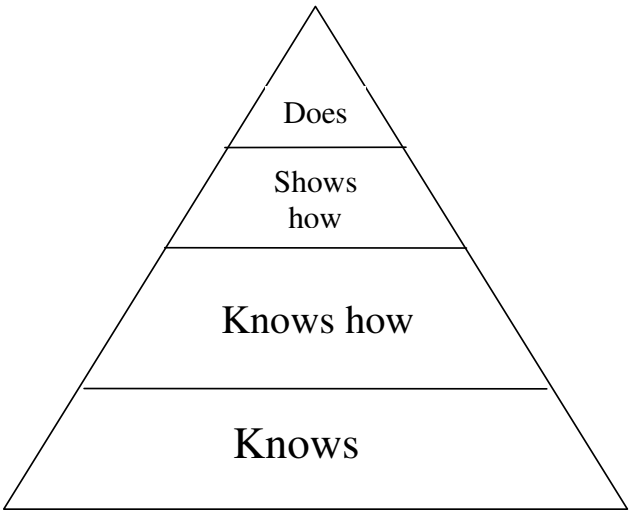


Figure 0.1 Miller's (1990) Pyramid of Competence

OSCE Structure

The physician diagnostic process can employ either a deductive or inductive reasoning approach to solve a problem. Pieces that a physician would put together (deduce) are comprised of components such as a history, a physical examination, identification of patient problems to reach a diagnosis, identification of appropriate investigations, interpretation of investigation results and diagnosis management including patient education to assist them in their diagnosis (Nayer, 1993). The OSCE separates each component (listed above) into the tasks (task analysis) then creates an objective checklist of items that students should follow in order to accomplish a goal. Each of these components represents stations where students read a scenario and carry out the necessary activities (i.e., history, physical exam, interpret findings from diagnostic imaging, etc..) with a standardized patient. The OSCE is comprised of multiple stations

each measuring individual tasks associated with the overall clinical construct. The summation of each station creates a score of the underlying clinical competence construct.

A standardized patient (SP) is a person who is scripted to act like a patient with all the same signs and symptoms a patient would display if they had the diagnosis associated with the specific case/condition. The SP typically rates the student's performance using an objective checklist, but the student may also be evaluated by an expert rater(s) simultaneously. The SP and expert raters are both potentially sources of error in the measurement of clinical competence and this will be discussed in greater detail in the reliability section.

The *objective* part of the exam comes from the fact that all students are graded using the same criteria (the checklist) and the *structured* component of the exam comes from the fact that all students move from station to station completing purposefully chosen components of clinical competence in question. Essentially, the exam standardizes how all students are evaluated to ensure that all students have met "the standard." In professions (such as medicine, athletic therapy, physiotherapy, chiropractic), standards are established to ensure quality control and moreover, so patient safety is upheld (Cavanaugh, 1991). OSCEs have been employed in medical schools as summative exams, but are also used in certification and licensure examinations (LaDuca, 1994; Haladyna, 1994).

Medicine was the first discipline to describe the OSCE (Harden et al., 1975), but many other professions have followed suit. The OSCE is so popular that it was employed in 111 U.S. medical schools in 1994 (Kaufman, Mann, Muijtjens, & van

der Vleuten, 2000). The OSCEs can also be found in dentistry (Brown, Manogue, & Martin, 1999), physiotherapy (Nayer, 1993), chiropractic (Lawson, 2002), optometry (Violato, Marini, & Lee, 2003), massage therapy (Violato et al., 2002), pharmacy (Austin, O'Byrne, Pugsley, & Munoz, 2003) and athletic therapy(Butterwick, Paskevich, Vallevand, & Lafave, 2006).

OSCE Validity and Reliability

The validity of an educational instrument is defined as how well an instrument measures what it is purported to measure (Hopkins, 1998). Reliability provides the means of excavating error associated with measurement (Streiner & Norman, 2003). Validity and reliability work together to build a measurement instrument that assesses an underlying construct, such as clinical competence. The relationship between the two principles associated with sound psychometric measurement will be discussed later. However, each will be covered independently first so as to better appreciate the interaction between them.

Traditionally, validity has been separated into three basic types: content; criterion and construct. Content, criterion and construct validity will be reviewed in sequence and empirical evidence of their application with OSCEs will be provided later in this chapter.

Content Validity

Content validity is the process of constructive alignment (Biggs, 1996) between learning objectives in each of the educational domains (knowledge, psychomotor and affective) and the test items (Bridge, Musial, & Frank, 2003). The constructive alignment process

is also known as blueprint creation or blueprinting (Smee, 2003). Table 2.1 illustrates how tasks on a checklist align with learning objectives and learning goals. Further, Table 2.1 demonstrates how these checklist tasks capture varying levels of expertise for the psychomotor skill domain. When an instructor or examiner wants to test a construct such as clinical competence, the blueprint provides a mechanism to ensure appropriate depth and breadth (i.e., a representative sample) across objectives and domains (Hopkins, 1998). Smee (2003) over-simplified the OSCE development process, particularly as it relates to the scale or checklist development. Alternatively, tremendous thought and planning must go into the development of an OSCE checklist. However, it seems most OSCE research skips sound content validation, disregarding the impact on the tool's validity and reliability.

Table 0.1 Sample Excerpt from a Blueprint for the SOAT

Sub-goals	Learning Objectives	Checklist Tasks (Taxonomy Level)[1]
Active Range of Motion	Students will be able to: communicate clearly to their patient to outline the specific motion they wish a patient to perform. identify the primary tissue that is being tested with active movements during the examination conduct active ROM assessments accurately, safely and efficiently demonstrate appropriate patient positioning and therapist hand positioning throughout the assessment differentiate normal from abnormal function and verbalize deficits when they are present. Assess quality and quantity of movement for each motion listed.	Communicates to patient: The procedures to be undertaken (G) Motion that the patient is to perform (G) Tissue being tested (S) Flexion (M) Extension (M) Adduction (includes ulnar deviation) (M) Abduction (includes radial deviation) (M) External Rotation (includes supination) (M) Internal Rotation includes pronation) (M) Horizontal Abduction (M) Horizontal Adduction (M) Scapular Protraction/Retraction (M) Scapular Elevation/Depression (M) Alter testing procedure when patients exhibit pain, dysfunction or stability (P, A)

Various authors have identified the importance of content validation and outlined the appropriate procedures to complete content validation with OSCEs (Butterwick et al., 2006; Austin et al., 2003; Palarm, Griffiths, & Phillips, 2004). There are a variety of scientific methods employed in each article and these will be reviewed to demonstrate some content validation options available to researchers.

Austin et al. (2003) described the design and development of an OSCE with the intention to certify pharmacologists. Seven stages (Table 2.2) were described in great detail. Their process began in 1997 and it took until 2001 to practically implement the examination.

[1] Psychomotor taxonomy: P=perception; S=set; G=guided response; M=mechanism; C=complex overt response; A=adaptation; O=organization

The development process was quite extensive since it also included reliability testing (Austin et al., 2003). When multiple stages include content validation and reliability testing, the instrument's construct validity is developed concomitantly. Building construct validity will be covered later in this chapter. Steps one, two, three and four are the steps that should be completed in order to achieve content validation while the remaining steps contribute to building construct validity (Table 2.2).

Table 0.2 Seven steps outlined in Austin et al.'s (2003) description of an OSCE development for pharmacologists

Step	**Description**
1	Blueprint development for OSCE pilot
2	Conduct station **development** workshops with practising pharmacists from a variety of practice settings
3	Conduct station **review** workshops with practising pharmacists from a variety of practice settings
4	Develop assessment instruments including answer sheets, scoring rubrics and methodologies, and feedback reports for candidates
5	Field test the OSCE
6	Undertake standard setting workshop
7	Collate and analyze data from the OSCE

A qualitative method called action research was employed to develop an OSCE for radiologists (Palarm et al., 2004). In this approach, researchers were able to draw on "several methodological tools to collect, analyze, and present data in a cyclical fashion" (p. 80). The researchers acted as facilitators over a six month period consisting of six "cycles" involving three lecturers and 33 students. Data consisted of the researchers' reflective diaries, invigilator's notes and observations during examinations, student evaluation forms, digital photographs, meeting minutes with academic colleagues, semi-structured interviews with staff and students, external examiner reports, and email

communications with visiting lecturers. The final OSCE evolved as the researchers gathered more information throughout the research cycles. The final OSCE was a culmination of all data gathered including what the researchers learned throughout all six research cycles.

The psychometric legitimacy of the process for content validation is questionable, particularly because the authors did not report reliability findings (Palarm et al., 2004). In all the other studies (Butterwick et al., 2006; Austin et al., 2003) reliability findings were either reported or forthcoming in future research. In order for a reader to evaluate the overall validity of an assessment tool, it is imperative that the methods employed for validity and reliability testing be shared explicitly. If this is not done, a reader can not truly evaluate the psychometric soundness of the tool.

A measurement instrument that could be used to assess the technical skill of athletic taping (a significant skill which is part of athletic therapy curricula) was developed using a well structured model (Butterwick et al., 2006). In this method, a five step process was introduced (Figure 2.2). In the first step, a tool that was used for summative assessment purposes in a class was adopted which established its face validity. Next, the tool was distributed to local experts and to national experts in the third step. In the third step, a modified Ebel procedure (Ebel & Frisbie, 1986) was employed whereby expert raters judge each task on a marking sheet on three levels of difficulty (hard, medium and easy) and three levels of importance (essential, important and not as important). Items that did not achieve 80% consensus in the initial evaluation were targeted for a face to face discussion and consensus building process in step four. Once again, 80% consensus was targeted in order to keep tasks/items on the list. If items did

not achieve 80%, they were eliminated from the instrument. It could be argued that the 80% consensus is not, by matter of fact, true consensus or unanimity. As a result, without full acceptance of the content by all experts, poorer reliability could result. In the final step, tasks/items were weighted using a guideline in Figure 2.3.

The Ebel and modified Ebel procedures was first described by Ebel (Ebel et al., 1986) and subsequently by Cantor (1989), respectively. The procedure was designed to create a minimal passing score or criterion referenced scores. There are numerous criterion referencing methods for both written and practical (performance-based examinations). Cantor concluded the Ebel procedure was the most appropriate method to create a minimal passing score for performance-based examinations (Cantor, 1989). Two other criterion referenced testing methods were compared to the Ebel procedure for United States naval technical skills: contrasting groups method; and the borderline group methods. Moreover, Cantor stated the Ebel method was the most effective "process for establishing performance standards which are reliable indicators of minimum competency and job readiness" (Cantor, 1989, p. 719). As a result of this study, the SOAT content validation method of choice was the Ebel procedure.

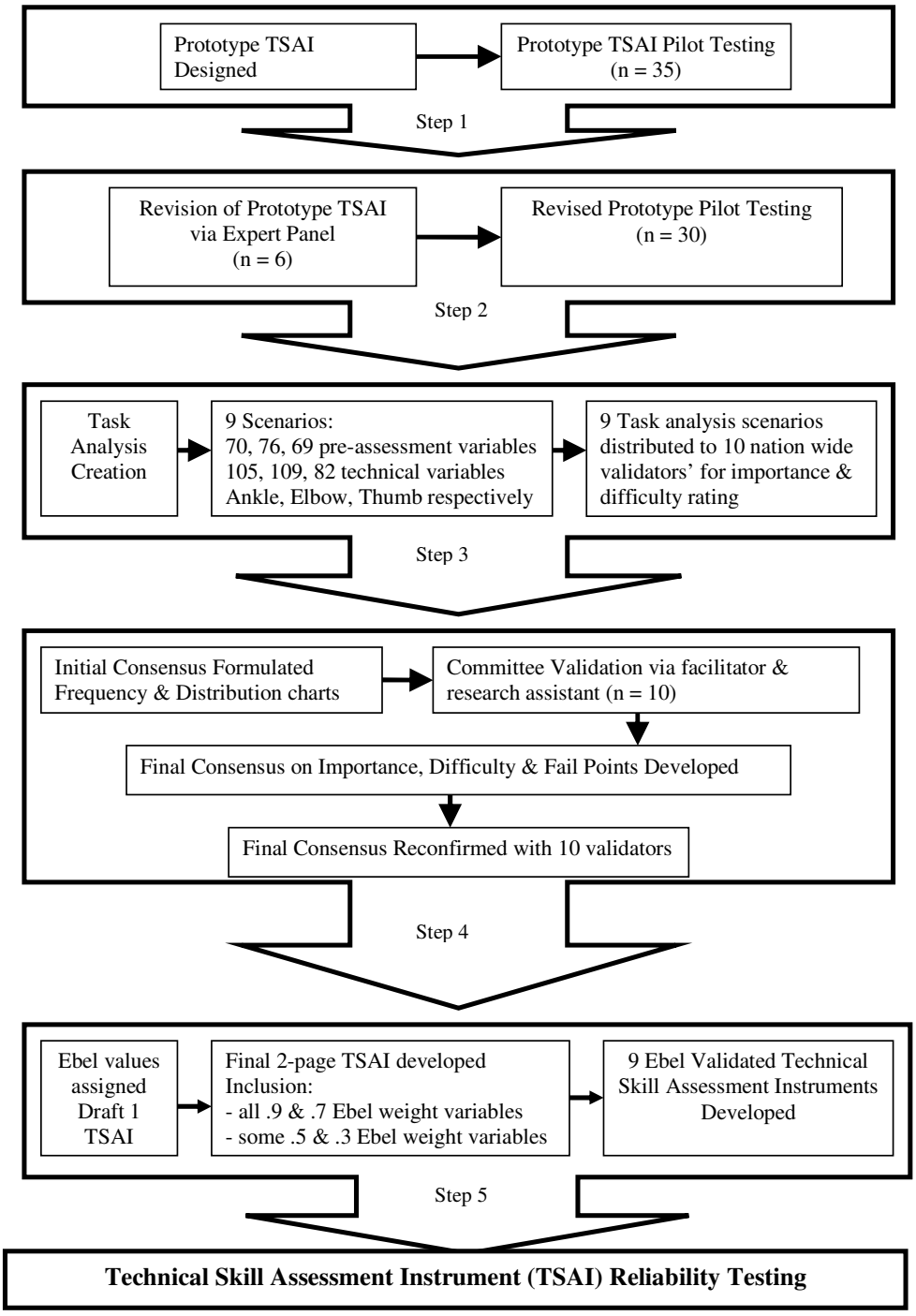


Figure 0.2. Technical Skill Assessment Instrument (TSAI) Methodological Evolution

Figure 0.3. The Ebel Grid with each box numbered and associated weighting for each.

Criterion Validity

Face and content validity are based on sound, rational analysis by experts in the field while criterion validity is based on empirical evidence to prove that a measurement instrument has met a pre-determined criterion (Hopkins, 1998). Criterion validity has been separated into two types: ***predictive*** and ***concurrent***.

Predictive validity refers to how a current performance on an examination may predict future performances on the same examination or a measurement of the same or similar construct (Violato et al., 1998). Typical statistical procedures employed with this type of validity to provide the empirical evidence may include Pearson r correlation

coefficient or stepwise multiple regression analysis. For example, if the OSCE was truly a measure of clinical competence as is claimed by some (Harden, 1990; Harden et al., 1975), and if there was another established measure of clinical competence such as the Physical Achievement Review (PAR), then a study which correlated OSCE scores to the PAR may help prove its validity (Hall et al., 1999). In this hypothetical example, there are a number of "if" statements and the most important relate to whether these two measurements evaluate the same construct. Essentially, if two metrics assess the same or a similar construct then this would be considered "convergent" validity (Streiner et al., 2003). If the two measurements did not correlate highly (i.e., divergent validity), then the two measurements may be measuring two different constructs (Streiner et al., 2003).

Simon et al. (2002) set out to determine the predictive validity of a well established OSCE compared to the United States Medical Licensing Exam (USMLE), Step one scores. In this study, 355 students' scores on the OSCE were correlated with their USMLE scores. Results indicated a moderate correlation (0.41) leading to the conclusion that the two tests were measuring similar constructs, but not identical (Simon, Volkan, Hamann, Duffey, & Fletcher, 2002). Hypothetically, program administrators might be interested in knowing how student performances in both the program OSCEs and the licensing examinations. If the two are strongly correlated, then the curriculum is probably working well. However, if there is poor or moderate correlation, then the program should adjust the curriculum to address the shortcomings. Comparing one measure to a gold standard is a common method employed to determine predictive validity. In the Simon et al. (2002) study, it is not clear that the USMLE is the gold

standard, but it is certainly a standard that must be met in order to practice medicine in the U.S.A.

Concurrent validity differs from predictive validity in that it evaluates the correlation (or relationship) between two measurements that occur simultaneously in contrast to predicting a future performance on the same or similar measurement (Ebel et al., 1986). Van Luijk and Van der Vleuten (1990) studied concurrent validity of OSCE rating scales while concomitantly illustrating one of the pitfalls of the OSCE. In attempt to objectify the evaluation of students, OSCE checklists may trivialize the clinical competence construct thus rewarding thoroughness, but not necessarily competence (Norman, van der Vleuten, & De Graaff, 1991; Van der Vleuten, Norman, & De Graaff, 1991; Van Luijk & Van der Vleuten, 1990; Cox, 2000). To test the theory of whether more global rating scales have the same level of objectivity and reliability as checklists, they concurrently compared the reliability of three scales used to measure student performance in an OSCE (Van Luijk et al., 1990). The examiners used a standard OSCE checklist that was established in a previous study. In addition, the study had experts employ two global rating scales: a general impression rating to judge the overall quality of the performance and a specific rating related to the technical of skill, the proficiency of the skill and the patient approach. Unfortunately, authors did not randomly choose which checklist or scale was employed first, so the checklist may have biased how the global scale was graded. Since the scales were measured simultaneously, it is a good example of concurrent validity. This sparked a number of other studies related to the use of checklists and global rating scales which may impact the instrument reliability. Most OSCEs now develop a rating scale that employs both a detailed checklist and a global

rating on the same form in order to best capture the construct (i.e., validity) while still ensuring good reliability (Reznick et al., 1998; Hodges, Regehr, Hanson, & McNaughton, 1998; Regehr, MacRae, Reznick, & Szalay, 1998).

Construct Validity

Educational research, unlike traditional wet-bench science, involves studying cause and effect factors which are not overtly observable or measurable (Krathwohl, 1998). As a result, researchers are left with observing and interpreting behaviours (Streiner et al., 2003). The underlying factors associated with these behaviours are referred to as traits or constructs (Streiner et al., 2003). Observation of behaviours and measurement instruments which assess certain behaviours permits researchers to make inferences about those behaviours. For example, success in OSCE X implies that a person is clinically competent at Y. Construct validity is definitely the most complex of all types of validity because there is no discretely defined process to establish construct validity in one experiment. There are some studies which claim to establish construct validity, but the construct they refer to is often clearly defined with distinct limitations (Winckel, Reznick, Cohen, & Taylor, 1994; Martin et al., 1997; Reznick, Regehr, Macrae, Martin, & McCulloch, 1997; Ault et al., 2001; Goff, Lentz, Lee, Houmard, & Mandel, 2000; Goff et al., 2001). In all of these studies, each tested the reliability of an objective structured assessment technical skills (OSATS) exam. Further, the researchers were able to differentiate reliability statistics between varying levels of experience with residents who participated in the study, thus allowing them to conclude the OSATS had construct validity. Based on the fact that most authors (Streiner et al., 2003; LaDuca, 1994) believe that construct validity only happens across a number of studies, concluding that a tool

possesses construct validity just because it differentiates between various levels of interns or residents may be overly optimistic. Perhaps this type of validity should be coined construct validity with a small "c" as opposed to construct validity with a big "C?" Application of this analogy to those studies listed above, their tools would measure construct validity with a small "c" since they could detect differences between various levels of residents. Those studies would not conclude that the tools possess construct validity with a big "C" thus stating the tool is a definitive measure of clinical competence in their respective disciplines.

It typically takes a number of studies that have proven the measurement instrument has content validity, perhaps has shown criterion validity and certainly has been shown to be reliable before it can be considered to have construct validity (Violato et al., 1998). Construct validity should involve testing a theory about a behaviour as measured by an instrument or scale. All successful validity and reliability studies help to collectively contribute to the instrument or scale's construct validity, thus proving or disproving the underlying theory (Streiner et al., 2003). Violato et al. (1998) have provided a nine step process one should follow to establish construct validity (Table 2.3).

Table 0.3 A nine step approach to achieving construct validity

Step	Approach
1	Identify and describe the meaning of a construct
2	Derive theoretical support for the construct
3	Based on the theory, develop items, tasks or indicators of the construct
4	Develop a theoretical network for the construct that can be empirically established by correlation.
5	Conduct research to obtain the data necessary to investigate the correlations between the variables in the theoretical framework.
6	Design experiments based on the construct theory and correlations to test for causal relationships.
7	Evaluate all of the relevant evidence and revise the theory, construct and measures if necessary.
8	Fine tune the measures of the construct by eliminating items and revising the tasks
9	Return to step 3 and proceed again

Taken from Violato et al., 1998 (pg. 101).

In order to understand the concept of construct validity more practically, an example is provided below. Cohen, Reznick, Taylor, Provan and Rothman (1990) tested a hypothesis that second year surgical residents should be able to perform better than foreign medical graduates who were applying to the pre-internship program in Ontario. If they were able to prove the OSCE was reliable and that the second year residents performed better than the foreign medical graduates, the authors felt they would have proven their theory and established construct validity. There were a number of problems with this study including their stated conclusions. As mentioned previously, it is challenging to establish construct validity in one study. The authors stated, in general, that "OSCEs have been shown to be reliable and valid format for the assessment of clinical skills" (Cohen, Reznick, Taylor, Provan, & Rothman, 1990). In fact, no previous study was done on this particular OSCE item to confirm its content validity. Content validation consisted of meeting 15 minutes prior to the exam to confirm the content and

scoring of the checklist. Circumventing the content validation process merely provides face validity of the exam, but does not provide a logical development process (Violato et al., 2002; Butterwick et al., 2006; Austin et al., 2003; Palarm et al., 2004).

In contrast to the Cohen et al (1990) study, Newble, Hoare and Elmslie (1981) designed a complex, multifaceted study which studied all three types of validity (content, criterion and construct) and the reliability of an OSCE for medical students in Australia. Although sampling details for each component of validity and reliability were sparse, the concepts of validity and reliability seemed to be covered quite well. The study took place over a number of years building both validity and reliability (Newble, Hoare, & Elmslie, 1981). Due to the complex and extensive nature of each of the sub-studies in this paper, one could conclude it had, in fact, shown signs of construct validity.

In summary, the concept of validity is rather complex and intricate. Further, Hodges (2003) put it quite succinctly when he commented that "*it is incorrect to say OSCEs have validity*". Rather, it is important to contextualize where, how and with whom an OSCE is valid which was apparently lacking in the early development of OSCE literature from the 1970's through to the late 1990's. A critical aspect in creating valid measurement instruments involves a well designed research plan which addresses as many of the types of validity as possible, while also including reliability.

Phase II – Reliability of Performance-Based, Practical Examinations

Reliability and OSCEs

Reliability plays an important role in conjunction with validity to create a psychometrically sound measurement instrument like those employed in OSCEs (i.e., its checklist and/or global rating scales) (Violato et al., 1998). Reliability can be defined as the consistency of a measurement over time, across examinees, examiners and conditions (Hopkins, 1998). A measurement instrument can not be valid unless it is also reliable, but an assessment tool can be reliable without being valid (Hopkins, 1998; Violato et al., 1998). For example, one may wish to measure how consistently a student can perform a specific technical skill such as knot tying for surgery or bouncing a basketball. A student may be able to perform that technical skill (knot tying or bouncing a basketball) quite well over time, but if the technical skill has nothing to do with the underlying construct one wishes to measure (i.e., surgical competence or basketball expertise), then the reliability is pointless. If a measurement tool is considered to be reliable and valid, it permits the examiner to make inferences about the people it was designed to measure.

There are many factors that contribute to poor reliability or inconsistency in measures from exam to exam and thus, it is rare that a student would ever perform exactly the same time after time. Classical test theory states that test scores are comprised of two components: a true score and the error associated with the observation (Streiner et al., 2003). Reliability error uses score variance as the indicator of its measurement error (Ebel et al., 1986). There are a number of ways to calculate reliability and this is dependent on both the study design and purpose.

Reliability Statistics

In the past, studies with OSCE reliability statistics have focused on expert rater/judge variance and/or the specific inter-station variance (Wass, van der Vleuten, Shatzer, & Jones, 2001) in contrast to the internal consistency or reliability measurements of written examinations scored over time or across forms (test-retest; parallel forms; split half) (Ebel et al., 1986). Intraclass correlation coefficients have often been used to measure reliability in OSCEs (Ault et al., 2001; Martin et al., 1997; Reznick et al., 1997). There are six types of intraclass correlation coefficient (ICC) that can be used to measure reliability and Cronbach's alpha measurement has often been the statistic of choice for testing inter-rater reliability (Shrout & Fleiss, 1979; Cortina, 1993; Ebel et al., 1986). Inter-rater reliability is only one source of measurement error and that is the limitation of the ICC. Cronbach's alpha reliability coefficient could be considered a special case of a more robust statistical procedure (one facet design) called Generalizability Theory (G-theory) (Ebel et al., 1986; Shrout et al., 1979). G-theory can account for a number of sources of error in its calculation (Brennan, 2000; Streiner et al., 2003; Brennan, 1992). If one is interested in a more global reliability (beyond one facet's contribution to the overall variance), then G-theory is more appropriate (Brennan, 2000; Swanson, Clauser, & Case, 1999) . The benefits of g - theory is that it takes all sources of error that impact a measure's overall reliability and separates them so that one can measure its overall impact on the measure's reliability (and thus, validity). The challenge with performance-based examinations in the medical profession is to find examiners/raters and SPs who can examine large numbers of students to accomplish a fully crossed design.

Scaling Responses in Tool Development: Contribution to Reliability

The choice between binary/dichotomous (i.e., yes, no; done, not done) rating and those more extensive (i.e., outstanding, very good, good, average, below average, poor) is a question that should be carefully considered since its implications can impact the outcome of any measure's overall validity and reliability (Thorndike, 1971). An appropriate scaling response should be tagged to each task based on the characteristics of the tasks/items that are identified through a structured content-validation process. Stevens (1946) provided a theoretical framework or typology to capture levels of data that is still widely accepted and used today: Nominal; Ordinal; Interval; Ratio. For sake of ease, this framework will be used to outline data types and discuss the statistical implications even though there is some controversy surrounding its use (Velleman & Wilkinson, 1993).

Rating scale responses may differ based on characteristics within the tasks/items that are posed to students in a performance based examination. Appendix A contains a copy of the SOAT with the various components of an orthopedic assessment. Within the patient history (in Appendix A), a student may ask a patient's age. An examiner's choice to grade the student performance is relatively simple: either the student asked the question or they did not. Therefore, some tasks more naturally lend themselves to dichotomous data. In the dichotomous example listed above, the rating scale is most appropriately captured in the nominal data definition. Nominal data can be defined by named categories which do not have an ordered response and do not lend themselves to numerical data (Streiner et al., 2003; Stevens, 1946).

Appendix A also provides an example of data which could be considered a continuous variable: the components of the physical examination. If a student passively moves a patient's arm through a range of motion, there is potential for varying levels of expertise with the technical skill necessary to complete this task/item. The psychomotor skills within the physical examination component (listed as Active Range of Motion, Passive Range of Motion, Isometric Resisted Testing, Special Testing and Palpation in Appendix A) are best suited to continuous levels of data collection (Waltz, Strickland, & Lenz, 2005). This second example is most appropriately captured by the ordinal data definition. Ordinal data builds on nominal data in that it can have numerical values assigned, but only for the purpose of ordering and should not be confused with interval data which has equal intervals between the scales or numbers (Streiner et al., 2003). Ratio data is differentiated from interval data in that its scale has a true zero.

Streiner & Norman (2003) outlined three errors that can occur when ordinal data is treated like nominal data or in other words, treating continuous data like it is discrete or dichotomous data:

1) Raters/examiners may have a different concept of what constitutes a positive response, and thus inter-rater error is introduced;

2) Even if raters/examiners have a similar concept of what constitutes a positive response, the limited number of choices reduces reliability;

3) Dichotomizing continuous data results in a loss of efficiency of the overall tool.

Each of these errors will be further explored in the next section.

Continuous Scaling Response Choices

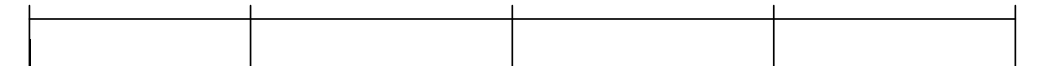
There are many continuous measurement choices, but those can only be explored once the data characteristics (dichotomous level data or continuous data) have been examined and finalized. Common types of continuous scaling responses will be reviewed below, followed by their projected reliability.

1. Numerical Rating Scale - the rater circles or checks the degree to which the characteristic is present (Gronlund, 1981).
Example: To what extend do you feel the examinee performed on XX test?

1.
2.
3.
4.
5.

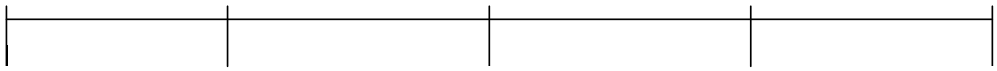
Where 1 = Unsatisfactorily, 2 = Below average, 3 = Average, 4 = Above average, 5 = Outstanding

2. Graphical Rating Scale – place an X or a line where you feel the examinee falls on the horizontal line below (Gronlund, 1981) for a specific behaviour.



Never Seldom Occasionally Frequently Always

3. Descriptive Graphical Rating Scale (also known as an adjectival scale)– similar to the graphical rating scale, but instead of just words along the scale, there are descriptions (Gronlund, 1981)

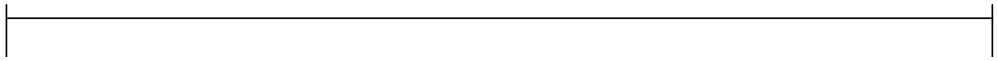


Never participates; quiet and passive

Participates as much as other group members

Participates more than any other group member

4. Visual Analogue Scale – is different from other scales listed above because it has two anchors at either end of the scale with more absolute descriptors (Streiner et al., 2003)



No pain Pain as bad as it can be

5. Juster Scale – this is a slight variation of adjectival scales only different because it adds the variable of probabilities (Streiner et al., 2003)

10	Certain, practically certain	99 in 100 chance
9	Almost sure	9 in 10 chance
8	Very probable	8 in 10 chance
7	Probable	7 in 10 chance
6	Good possibility	6 in 10 chance
5	Fairly good possibility	5 in 10 chance
4	Fair possibility	4 in 10 chance
3	Some possibility	3 in 10 chance
2	Slight possibility	2 in 10 chance
1	Very slight possibility	1 in 10 chance
0	No chance, almost no chance	1 in 100 chance

6. Likert Scale – this is very similar to adjectival scales except Likert scales are typically bipolar and adjectival scales are unipolar. For example, Likert scales describe a level of agreement going from strongly agree to strongly disagree with a neutral descriptor in the middle (Likert, 1952)

Many studies have compared the most reliable type of scale to use when measuring student performance and more specifically, the total number of response categories or descriptions that should be on a scale (McKelvie, 1978). There is little agreement for a gold standard on the number of categories for a scale and some have argued that this is because the optimal number is really a function of the stimulus being measured (Cronbach, 1946). If there are too many rating categories, then the measurement becomes too fine and the level of discrimination between examiners/raters may be lost. Too few rating categories and the accuracy (i.e; reliability) is lost (Streiner et al., 2003). McKelvie (1978) identified five factors that should be considered when deciding on the number of rating scales and their discrimination ability:

1. discriminability scaling;
2. information transmitted;
3. independent duplication judgement;
4. standard error of judgement;
5. reliability.

Reliability is the most common measure of discrimination, and hence the focus of this review. Streiner and Norman (2003) suggested that seven response categories seem to be the best choice for a scale. Guilford (1954) advocated for up to 25 response categories, but as Streiner and Norman (2003) pointed out, the differences in reliability after seven are nominal. However, anything less than seven response categories can result in a loss of reliability. The number seven seems to be rooted in an original study which concluded that human cognition and information processing works best with "chunks" of information within that number (Streiner et al., 2003; Symonds, 1924).

Reliability of the various types of continuous scales is also a factor to consider when choosing the appropriate scale. Cook, Heath, Thompson and Thompson (2001) compared the reliability of graphical rating scales to Likert type scales. In this study, web-based tools were used to measure student perception of library service quality with a nine point, radio button Likert scale and a web-based graphical rating scale using a previously established survey tool (Cook, Heath, Thompson, & Thompson, 2001). The benefits of using technology as part of the survey tool is that graphical rating scales are quite laborious to mark where as computer mediated scales can automatically calculate data and drop the results into a spreadsheet format. The results indicated almost identical reliability coefficients of .965 and .960 for the Likert scale radio button and web-based

graphical rating scale, respectively. Thus, use of these types of scales for future marking of performance based examinations may become quite useful and efficient.

Statistical Implications for Levels of Data

Reliability measurement between raters for performance based examinations contributes to a tool's overall validity (Hopkins, 1998). For example, Cronbach's alpha reliability coefficient is commonly employed as an inter-rater reliability measurement (Streiner, 1993). There is a lengthy debate in the literature that has raged for years on the pairing of an appropriate statistical method with a specific level of data, and thus it is difficult to complete a full review of both sides of the argument in the context of this chapter (Streiner et al., 2003). In short, judgements made by examiners on a student performance using a Likert scale or some variation of a graphical rating scale is considered nominal or ordinal level of data (for reasons listed above) (Stevens, 1946; Svensson, 2000; Kerlinger & Lee, 1986). However, since most continuous scales approximate interval level of data, it is commonplace to treat the data as if it were interval level of data, thus permitting parametric statistics to be used (Kerlinger et al., 1986; Streiner et al., 2003). The ultimate test for the appropriateness of using parametric statistics such as Cronbach's alpha reliability coefficient is whether the assessment tool yields high reliability (Streiner et al., 2003). If the reliability is low (below .50), then there is a good chance the scale is not correct or the statistical method used is inappropriate (Streiner et al., 2003). If an experimenter did have low reliability on an assessment tool, they could try to run non-parametric statistics (Svensson, 2000) or perform further analysis such as item analysis or exploratory factor analysis to alter the assessment tool (Streiner et al., 2003).

The OSCE: Checklists versus Rating Scales

The OSCE has evolved considerably since its original description (Harden et al., 1975). Swanson et al. (1999) provided a framework which thoroughly outlined the variables and factors to consider when comparing OSCEs. There are many OSCE models which measure clinical competence thus, making general conclusions about the validity and reliability of OSCEs challenging. However, Harden (1990), in a retrospective analysis of the evolution of the OSCE over a 15 year span, supported the notion that there should be flexibility in the type and format of OSCEs.

Theoretically, the benefit of objective structured clinical exams is that the examiners or SP's have a dichotomous scaled checklist of items that guide the examiner thus removing subjectivity (Harden et al., 1975; Regehr et al., 1998). Neufeld and Norman (1985) outlined two important ways to measure clinical competence: global rating scales and checklists. In Harden et al.'s (1975) original paper, rating student performance, was quite simply a binary (i.e., yes or no) checklist or a three point checklist which included 'carried out satisfactorily, attempted but not satisfactorily and not attempted.' On the surface, the OSCE design with binary responses appears to objectively measure student performance, thus removing the subjectivity introduced by examiners and further increasing the reliability of this the measure (Van der Vleuten et al., 1991). However, checklists have a tendency to trivialize constructs that the researcher or educator intended to measure resulting in a reduction of fidelity and validity (Van Luijk et al., 1990; Van der Vleuten et al., 1991; Norman et al., 1991). As Streiner and Norman (2003) pointed out, dichotomizing continuous data results in a loss of

efficiency of the overall tool. In addition, students may have a tendency to just memorize checklists without a deeper understanding of the underlying construct. These pitfalls have been supported by other experts and deserve further analysis (Reznick et al., 1998; Regehr et al., 1998; Hodges, Regehr, McNaughton, Tiberius, & Hanson, 1999; Van Luijk et al., 1990; Van der Vleuten et al., 1991; Rethans et al., 2002; Barrows, 1993; Dupras & Li, 1995; Hodges et al., 1998).

Reliability Differences Between Global Rating Scales and Checklists

Dichotomous checklists in performance based examinations have been employed due to their objectivity and high inter-rater reliability (Swanson & Stillman, 1990; Reznick et al., 1998). However, many issues have arisen over the years with more use of OSCE-type examinations including fidelity, validity of the measurement, raters' feeling undervalued and with students' superficial understanding of a construct due to memorization of a checklist (Van Luijk et al., 1990; Hodges, 2003). In their study comparing simultaneous use of checklists (using dichotomous measures) and global rating scales (using a 10 point Likert scale), Luijk and Van der Vleuten (1990) found that inter-rater reliability for checklists were significantly better than global rating scales. However, they also calculated a generalizability coefficient and found these differences were averaged out over stations when different raters are used at each station. Further, the global rating scales and checklists rank-order students in a similar way thus making them conclude the two were equally objective. They concluded there was some validity to their initial concerns about exclusive use of checklists in these examinations and that the problem should be studied further (Van Luijk et al., 1990).

Following up on findings in the Van Luijk et al. (1990) study, a surgical education research group at the University of Toronto began to investigate the use of tools that could measure technical surgical skills in surgical residents using a similar format employed in OSCEs. The primary difference introduced by this group was they often compared checklists to global rating scales in the quest for high examination reliability (Martin et al., 1997; Reznick et al., 1998; Reznick et al., 1997; Ault et al., 2001). The concept of an objective structured assessment of technical skills (OSATS) evolved from the need to assess technical skill outside of the surgical suite with real patients to a more objective and reliable examination (Winckel et al., 1994). Winckel et al. (1994) studied inter-rater reliability and construct validity of structured technical skills assessment forms on three common general surgical procedures: cholecystectomy, inguinal hernia repair and bowel resection. The raters measured 26 residents' performances of the procedures using the 120 point checklist forms which used a three point scale: 0 = not performed; 1 = performed poorly; 2 = performed well. The rating form was separated into two components of the surgical procedure and the average score for those two parts was used for each examinee. The results showed a moderate level of internal consistency for their form: .78 and .73 for parts I and II respectively.

Regehr et al. (1998) studied the reliability of checklists as compared to global rating scale. Using established testing procedures outlined in previous studies (Martin et al., 1997; Reznick et al., 1997), two raters measured student performance on a wet bench surgical procedure, but examiners employed different assessment tools. One examiner/rater used a global rating form while another examiner used a global rating form but only after they had completed a detailed checklist. Finally, two blinded

surgeons rated the surgical result (from the bench exam) independently from the other examiners using a global score. The results showed that the most reliable method to measure the examinee performance was the combination of a global rating scale and a checklist (.89) followed by the global rating scale alone (.85) with the checklist alone (.79) producing the lowest reliability.

In subsequent iterations of the OSATS assessment tool, global rating scales and checklists were compared to determine the tool's construct validity (Ault et al., 2001). Researchers exported a previously established method (Martin et al., 1997; Reznick et al., 1997) to measure OSATS validity and reliability of surgical residents at Northwestern University and the University of Southern California. In order to measure construct validity, examiners used both checklists and global rating scales to determine residents' (or examinees) competence with surgical skills. Researchers discovered that checklists produced lower reliability than global rating scales: Northwestern - .68 and USC - .71 and .82 at both Northwestern and USC, respectively. The trend of higher reliability for global rating scales compared to checklists was supported in subsequent study (Goff et al., 2000).

Results from the University of Toronto surgical group support earlier claims (Van Luijk et al., 1990) that larger and more global scales can produce better reliability (Streiner et al., 2003). The level of examinee and examiner expertise began to emerge as an interesting trend in the data. In the Regehr et al. (1998) study, the researchers identified higher levels of reliability on global forms with advanced levels of examinee and examiner training. The combination of examiner and examinee experience should be a consideration in the choice of scales and scaling responses in future studies.

Practicality of the OSCE

In order for a measurement tool to be useful, it must be practically applied (Harden et al., 1979). The OSCE, if designed properly, takes a tremendous amount of work to create. There are number of studies which must test its validity and reliability and so the development process may, in and of itself, hinder its practicality. In addition, the resources to run an OSCE are substantial: people (physicians, experts, SPs, administrators); financial; time; space (Cusimano et al., 1994; Carpenter, 1995). However, some of these resources can be reduced if proper reliability testing is completed. For example, when most studies establish the reliability of an OSCE, they perform what Cronbach called a "G-study" (generalizability study) (Cronbach, Gleser, Harinder, & Rajaratnam, 1972). In this study, the reliability of the tool or inter-rater reliability is assessed. In a "d-study" (or decision study), the researcher can manipulate the number of raters or the number of stations to determine the optimal number while not reducing the measurements' overall reliability. However, even with an efficiently designed OSCE, a number of researchers have commented on the resource-intensive nature of this type of measurement (Cusimano et al., 1994; Carpenter, 1995). There have been some attempts to reduce the resource intensive nature using technology to tabulate and provide feedback to students (Schmidts, 2000; Treadwell, 2006). Those studies will be explored in greater detail below.

Phase III – Use of Valid Assessment Tools in Teaching, Learning and Assessment of Underlying Constructs

Assessment of learning has a number of purposes. Norton described five purposes for assessment that help frame issues relevant to this dissertation and phase of the research (Norton, 2004). Emphasis is placed on the description of formative assessment, summative assessment and assessment as a lever from a theoretical perspective, but also to provide practical examples within the OSCE literature.

Formative – to provide support for future learning

Summative – to provide information about performance at the end of a course

Certification – selecting by means of qualification

Evaluative – a means by which stakeholders can judge the effectiveness of the system as a whole

A lever - the effect or role assessment has on how students learn. This role of assessment acknowledges students' motivation to learn and their perspective that the curriculum is the assessment

It should be noted that Norton (2004) did not use the term lever, but rather, the term is introduced in this paper to present a succinct description of the concept. This concept of assessment is often forgotten and/or ignored. However, students live a world that rewards their learning by success in summative assessments through scholarships, entrance to further programs and specialties and ultimately continuance in a program (Cohen-Schotanus, 1999).

Newble and Jaeger (1983) studied the power of assessment and its influence on student learning. Instructors and curriculum leaders changed the curriculum so that

students would spend more time on wards learning clinical concepts and less time in classroom learning didactically. Rather than having year end clinical viva exams, students were graded generically on the wards by supervisors. No one failed during these assessments. The only remaining measure of students' clinical proficiency was a multiple choice examination (MCE) at the end of the year. Even though students had more time to spend in the wards for learning, they chose to spend their time studying for the MCE. In fact, students began to demand more didactic lectures. As a result, in the subsequent year researchers implemented the clinical viva exam once again and the ward presence subsequently increased. The study gathered preferences and satisfaction with the types of examination and study habits from the two groups. Generally, greater satisfaction was identified in groups with the addition of the clinical viva. This finding substantiated their anecdotal view that student presence in the wards increased when the clinical viva was part of the summative assessment process. Further, students stuck to studying with books for MCEs while ward work ranked most highly for those studying for clinical vivas (Newble & Jaeger, 1983). This study illustrates the dangers of ignoring the power of assessment as a lever, while also demonstrating what can happen if assessment as a lever is used. Constructivists might argue that assessment as a lever is inappropriate, so as a result, this argument will be explored from both an objectivist and constructivist perspective.

Earl's Model of Learning

The concept of using assessment to facilitate learning is not new, but Earl (2003) was the first to articulate the model differentiating between of, for and as learning. Assessment of learning is primarily summative in nature. This is the traditional type of assessment consisting of a final and/or mid-term examination. Assessment of learning is what those in post-secondary education are most familiar with, since this type of assessment is prevalent throughout this educational system. In professions such as medicine or athletic therapy, professional standards must be met (i.e., certification examinations). If standards are not met, future accreditation may be jeopardized. This requirement may make professional programs unique compared to other post-secondary education. Nonetheless, this type of assessment for professional programs will remain in place to uphold standards.

Formative assessment has two purposes according to Earl (2003): "for" learning and "as" learning. Assessment for learning is used by the teacher/professor/instructor (instructor herein) where information used is gathered from assessment for diagnostic purposes. The "diagnosis" provides insight to the instructor regarding the student's current understanding of a learning objective (s) preparing them to personalize teaching based on a gap between what the student knows and what they need to know (i.e., the learning objective). Black, Harrison, Lee, Marshall and Wiliam (2003) provided an apt metaphor that described this gap as a student going up a set of stairs. Some students may know how to climb the first three stairs, but they could become stuck on the fourth stair. A knowledge gap is created where the student needs to learn how to get past that fourth stair. The instructor's job would be to recognize that the student needs help with that

stair and subsequently devise a plan that best suits the student to move past that stair to the ultimate goal: "the top of the stairs." There is a constant interaction between instructor and student in the assessment for learning model (Black, Harrison, Lee, Marshall, & Wiliam, 2003).

Assessment as learning is also considered formative, but responsibility for learning shifts from the instructor to the student. The student takes responsibility and participates actively in the assessment process through critical thinking and meta-cognition. Continuing with the metaphor above, the student can see the top of the stairs, knows the all the steps to get there, recognizes that they are missing stairs one and five, (for example) and devises a plan to get past those stairs to achieve the goal. This perspective/paradigm is almost completely student-centred and does not involve the instructor unless planned by the student. Assessment as learning is the pinnacle of assessment according to Earl (2003), but she acknowledged that ignoring the other types of assessment is a mistake. Assessment as learning is clearly rooted in a constructivist learning theory, but the fact that we can not ignore the other types of assessment raises the question: *"Is Earl's (2003) model constructivism or objectivism?"*

Earl's Model: Constructivism or Objectivism?

A brief account of the two epistemologies is necessary to help answer this question. Roblyer and Schwier (2003) separate learning theory and epistemology as it relates to instruction into two main categories: constructivism and objectivism. Objectivism has its roots in radical behaviorism and as such was heavily influenced by early pioneers such as Thorndike and Skinner (Driscoll, 2005; Skinner, 1971). Objectivists believe that "knowledge has a separate, real existence of its own outside the human mind; learning

happens when this knowledge is transmitted to and acquired by learners" (Roblyer et al., 2003, p. 50). In other words, the locus of control exists outside the individual learner. Of late, this has certainly become a less popular theory, but there is no question that aspects of this epistemology continues to be used in instruction (Driscoll, 2005).

In contrast, constructivists believe that "humans construct all knowledge in their minds by participating in certain experiences; learning happens when one constructs both mechanisms for learning and his or her own unique version of the knowledge, colored by background, experiences and aptitudes" (Roblyer et al., 2003, p.50). In other words knowledge is not something that is espoused from someone, but rather something a student must take in and individually make sense of in order for it to actually become integrated as knowledge.

The origin of Earl's (2003) model began in a well rooted constructivist belief that learning was more than just acquisition of facts and figures. It would seem that assessment of learning has its roots in objectivism since summative assessments typically assess knowledge that an instructor has defined either explicitly or implicitly. Assessment as learning seems to clearly have its roots in constructivism since the locus of control for learning is directly related to the individual learner. Assessment for learning is less clearly pigeon-holed in one belief or the other. If the two epistemologies are on a continuum, then perhaps learning for assessment might lie somewhere in the middle? It may fit in the realm of "pragmatism" (Driscoll, 2005)? Driscoll (2005) defined pragmatism as the point where reality is interpreted through both internal and external signals (Driscoll, 2005). Rezaei & Katz (2002) introduced a model (the Inventive Model = IM) that supported assessment for learning and thus would fall somewhere in the

middle of this continuum. The IM calls for a diagnosis of students misconceptions so the instructor can subsequently, deliberately and directly address those in a lesson plan. In closing, assessment for learning may be best left as ill-defined rather than pigeon-hole it to either extreme of objectivism or constructivism (Rezaei & Katz, 2002).

OSCEs to Teach: Assessment OF, FOR and AS learning?

Traditionally, medical students have learned skills (history, physical exam, technical skills) through an apprenticeship method in their clerkship. However, there seems be little correlation between final summative OSCE examinations and the clinical experience (McManus, Richards, Winder, & Sproston, 1998; Martin, Stark, & Jolly, 2000; Chatenay et al., 1996). As a result, a more structured, planned and purposeful approach to using OSCEs should be employed in either real life or mock situations to be useful. A number of OSCE studies employed OSCEs to provide feedback of, as and for learning. Each will be explored further below.

Assessment OF Learning in OSCEs

Sloan et al. (1996) set out to measure the validity and reliability of a fifteen station, 210 minute, two part OSCE. Additionally, the two parts of the exam (part A and B) were separated by feedback station that could assist student learning. Part of their hypothesis and concern was that they would reduce a previously established reliability coefficient of .91 (Sloan, Donnelly, Schwartz, & Strodel, 1995). The reliability did not change (.91) and the feedback was positively received (Sloan et al., 1996). In this study, the impact of feedback on student learning was not measured nor was student opinion (in any great detail). Therefore, it is challenging to make further inferences other than that reliability

appeared unaffected by providing feedback in the middle of the OSCE. This study demonstrated that even assessment of learning has the potential to provide feedback to students with little psychometric impact to the final summative grade.

In a highly structured study, Black and Harden (1986) evaluated the type and timing of feedback on students' preferences. Three types of feedback were provided at various times throughout the OSCE and all students rotated through all OSCE stations with the various types of feedback. Unfortunately, there was no control group and thus OSCE results could not be compared to the various types of feedback (i.e., treatment groups). The results indicated that students preferred to receive feedback by checklist and expert video demonstration much more than checklist only or video only. With respect to the timing of feedback, students preferred to receive their feedback during an inter-station break rather than waiting until after completing all stations (Black & Harden, 1986). Again, this study would have been significantly strengthened had the students been split into quasi-experimental and control groups thus allowing their OSCE scores to be compared. Nonetheless, important information can be gleaned from this study regarding the timing and type of student feedback during an OSCE.

Assessment FOR Learning in OSCEs

No studies were identified which explicitly stated the research purpose was to investigate assessment for learning. However, one study set out to evaluate the impact of patient volume on learning as measured by an OSCE (Chatenay et al., 1996). The authors concluded that higher patient volume, as seen by undergraduate medical students, did not correlate to higher OSCE scores and was confirmed by two subsequent studies (McManus et al., 1998; Martin et al., 2000). In all three investigations, students' learning

style was found to be a good predictor of OSCE success (which supports assessment as learning theory). One of the variables measured in the Chatenay et al. (1996) study was supervisor feedback as measured by student log books. The researchers found that students who received the highest quality feedback generally performed better on the final OSCEs. Although this does not provide overwhelming support assessment for learning, it does demonstrate a relationship between the two factors.

Assessment AS Learning in OSCEs

Most of the OSCE studies that seem to focus on assessment as learning have used qualitative research gathering opinions and preferences of the students as well as student self assessment. Using a small group of students, Evans et al. (2005) studied the perceptions of students who self assessed and were evaluated by experts using objective structured assessment of technical skills (OSATS) checklists. In this qualitative study, the authors concluded that self assessment had a positive impact on learning surgical skills (Evans, McKenna, & Oliver, 2005). Unfortunately, the study stopped short of integrating the comments/feedback from the expert, but it was stated that students generally preferred feedback from experts. Correlation between some of the variables or covariates such as expert feedback, self assessment on practice OSATS exams and final OSATS exams would have strengthened this study tremendously.

Geddes and Crowe (1998) examined a unique concept of peer OSCEs in order to introduce students to the OSCE format and help teach content in a cardiorespiratory unit of a physiotherapy program. In this study session, three students rotated through positions of being a rater, a standardized patient (SP) and an examinee. Students were exposed to both the OSCE process and the OSCE marking scheme where peer feedback

was provided at the end of each station. Students were able to reflect on their performance and others' performances from multiple perspectives (Geddes & Crowe, 1998). It is interesting to note that all students in the three years of this study participated in this optional study session (1994 – 53 students; 1995 – 61 students; 1996 – 59 students). This result alone speaks volumes about the value students attached to this type of learning opportunity (i.e., assessment as a LEVER). Seventy four percent of the students felt the method was a great process to introduce them to the OSCE. Ninety four percent of the students felt the exercise was either fairly effective or that great learning took place during the exercise. Although qualitative and descriptive in nature, this research identifies how students felt about the OSCE and the impact it had on their learning.

Mavis, Turner, Lovell and Wagner (2006) studied innovative ways to introduce medical students to the clinical experience early (pre-clerkship) using OSCE-type checklists and global scales to assist in learning. Instructors acted as the SPs in the OSCE in order to gain insight to student progress. Students qualitative responses were quite positive as in the previous study (Geddes et al., 1998). Data gathered from the SPs lent support to the assessment FOR learning theory. For example, on a 5 point likert scale (5 = strongly agree), instructors felt quite strongly (4.29) that "by participating as a standardized patient [they] have gained insights about the abilities of the students" (Mavis et al., 2006, p.134). This study could have been strengthened had it employed a control group compared to the group who had their instructors as SPs (Mavis, Turner, Lovell, & Wagner, 2006). Then, a final OSCE could compare the two groups to determine the overall effect on learning the material.

Assessment AS, FOR and OF Learning in OSCEs

There is a paucity of material in the literature which attempts to capture all three aspects of Earl's (2003) model (i.e., as, for and of learning) except for a few (Rahman, 2001; Abraham, 1998). Rahman (2001) introduced the OSCE to Dhaka Medical College in Bangladesh to assist with teaching and assessing the pediatric curriculum. Content validity was loosely established by a group of senior pediatricians for use with multiple choice questions (MCQ) and OSCE stations. At the beginning of their pediatric rotation, students were introduced to the 11 core problems established by the expert group in a teaching and learning session (i.e., instructor or assessment for learning). Students were then exposed to a self assessment with key answers provided to them (i.e., assessment as learning). Next, students were sent to an inpatient department for four weeks to continue learning in this pediatric rotation. Finally, students were given an OSCE and MCQ exam based on the session taught and reviewed at the beginning of the rotation. Students were given immediate feedback after their OSCE and encouraged to discuss the results with their examiners. Students were then asked to complete a satisfaction survey to evaluate the self assessment questions (introduced at the beginning of the rotation) and the OSCE experience (performed at the end of the experience). OSCE results were provided in the paper, but researchers did not attempt to correlate these results with any of the student satisfaction surveys or their self assessment success. Student participant survey results indicated a "strongly positive" (4.4 on a 5-point Likert scale) response in favor of the self assessment tool introduced at the beginning of the rotation. Two other notable responses from Rahman's (2001) research included the students' beliefs that the self assessment tool assisted in learning the required skills (4.5 out of 5) and acted as a guide for the

exam (4.5 out of 5). The survey inquired as to the feedback provided after the OSCE and most responses were extremely positive. Students felt that learning could be improved substantially by receiving feedback (4.6 out of 5) and that this type of formative assessment should be adopted in undergraduate medical education for better learning (4.8 out of 5). Overall, this study illustrated how an assessment tool can function as, for and of learning. It also demonstrated how the tool could be used to assist students in learning how to learn through reflection (assessment as learning). The study helped instructors structure what and how they taught (assessment for learning). Finally, the study used the same tool employed to teach as a final summative assessment tool (assessment of learning). Although there was no empirical evidence to draw connections between the interventions (i.e., teaching session, self assessment session, feedback sessions) and the dependent variable (i.e., the final OSCE), qualitative results definitely provided some anecdotal support for this type of teaching and learning system.

Abraham (1998) introduced a model of assessment as, for and of learning to an obstetrics and gynecology nine week rotation. A week prior to the learning session students were provided with learning materials (teaching session information, aims of the teaching session and lessons, format of the sessions, objectives of each learning station, time taken to complete the learning sessions, group sizes, people involved as well as the formative and summative evaluation forms). Students were split into learning groups of three or four students with two gynecological assistants. Gynecological assistants were women who volunteered their time, bodies, and expertise for the teaching and assessment components of the final gynecological examination. Students were directed to work their way through five learning stations in two hours to achieve explicit objectives.

Essentially, each station included the knowledge, skills and abilities required to complete a full gynecological examination. In the fifth station, students were to role play a doctor and integrate all knowledge, skills and abilities from the previous stations and as such, actually perform a gynecological examination. When performing the examination, students were graded by the gynecological assistant, by an expert observer and by themselves using the same assessment form shared with them the week prior to the learning session. All parties debriefed after assessment forms had been completed. The very same form that was used in this learning session was also used in a year end OSCE. Assessment for learning took place across the learning stations (the gynecological assistant and observer provided feedback) while the final station also permitted the individual to reflect on their performance and use the tool as their guide (assessment as learning). Ultimately, the tool was used in the final summative examination (i.e., assessment of learning) which completed the cycle of all Earl's (2003) components of the assessment and learning (Abraham, 1998).

From a methodological perspective, this study evaluated students' opinions of the teaching method and the assessment tool (Abraham, 1998). However, to introduce some level of comparison, Abraham employed a control group (70 students) who were responsible for the same content, but relied on ward based teaching rather than this system (Abraham, 1998). Although the author did not mention the homogeneity of the groups, they did find that the group exposed to this teaching method and their practical experience was much more satisfied than the 'control group.' Once again, despite some methodological concerns, this study demonstrated the potential to practically apply Earl's (2003) model of assessment of, for and as learning.

Technology Adoption to Assist in Teaching, Learning and Assessment of Orthopedic Assessment Skills

Integrating technology into teaching, learning and assessment has been shown to be effective in both medicine and kinesiology (Liebermann et al., 2002; Carr, Reznick, & Brown, 1999; Vivekananda-Schmidt et al., 2005; Schmidts, 2000; Treadwell, 2006). The role that technology plays in an educational environment can be more central or peripheral to a research question. Few studies clearly define the role of technology in teaching and learning, thus forcing it to the periphery and making inferences about its use problematic.

Empirical evidence within this review is drawn from the literature in computer assisted learning (CAL), computer based instruction (CBI), computer assisted instruction (CAI), and computer based assessment (CAA) in the medical field. Theoretical underpinnings supporting the integration of CAL, CBI, CAI, CAA and assessment in general, are explored to better understand and appreciate why and how they impact learning and reciprocally, how research informs theory.

CAL, CBI, CAI and CAA Definitions and Associated Learning Theories

There does not seem to be a universally accepted definition in the literature that clearly outlines the various roles computers play (i.e., CAL, CBI, CAI, CAA) in the education. Greenhalgh (2001) defined CAL as computer applications that "generally require the student to follow the content without immediate or direct supervision of a tutor" (p. 40). Broudo et al (1997) defined CBI in the same realm where the student is independently learning material that is computer based and not supplemented by face to face instruction. Hayward (2004) defined CAI as instruction which is designed to augment face to face

instruction with the use of CD ROMS and software programs. It is obvious that based on the author, varying definitions for using technology in teaching and learning exist (Greenhalgh, 2001; Broudo et al., 1997; Hayward, 2004). Applications of these definitions are explored within the medical education environment below.

Review of the Literature for CAL, CBI and CAI in Medicine

Integration of technology into a medical education environment has become commonplace (Greenhalgh, 2001; Rootenberg, 1992), but the impetus for such a trend must be scrutinized. Bold statements for technology's potential to revolutionize learning have been proposed (Greenhalgh, 2001; Summers, Rinehart, Simpson, & Redlich, 1999; Keane, 1990). Caution must be taken to ensure enthusiasm of new technology does not cloud the empirical evidence to support its integration into education (Carr et al., 1999). For example, resource efficiencies have often been cited as the primary purpose to introduce technology rather than pedagogical rationale (Summers et al., 1999; Greenhalgh, 2001; Vivekananda-Schmidt et al., 2005; Broudo et al., 1997). Espousing cost savings and efficiencies is a common ploy to convince administrators to resource a project, but does little for the credibility of establishing the technology as a legitimate pedagogical method to aid in learning (Greenhalgh, 2001). Prospective studies with clearly articulated hypotheses and well established outcome measurements are critical. Greenhalgh (2001) searched in Medline and ERIC using the MESH term "medical education" and "computer based" and/or "computer assisted" and found twelve prospective studies (Greenhalgh, 2001). Three of the twelve studies employed performance based outcome measures related to psychomotor skills, technical skills or employed OSCE type measurements.

Rogers et al. (1998) designed a study to introduce a technical surgical skill (knot tying) to students in an attempt to make clinical and surgical time more efficient for supervisors. Students were randomly chosen for the experimental group which received instruction through a computer-based program created in Macromedia Director ©. Students could view the twelve step process watching slides and video. Students in the CAL group were provided the necessary equipment to practice independently, but they were not afforded any feedback from a surgical expert. The control group received face to face instruction and was permitted to view techniques via still images. It should be noted that the control group was given expert feedback when they practiced. All students were assessed using a valid and reliable tool (.79 Cronbach's alpha) (Rogers, Regehr, Yeh, & Howdieshell, 1998). Results indicated that the CAL group performed significantly worse than the control group leading the researchers to conclude that CAL was an inferior method to teach this skill.

In a subsequent study, researchers recognized some research design shortcomings in the Rogers et al. (1998) study and attempted to eliminate the flaws (Summers et al., 1999). Summers et al (1999) determined that the primary difference between the control and experimental group was the lack of feedback provided to the experimental group (CAL). They randomly assigned students into three groups:

Group 1) traditional didactic instruction;

Group 2) videotape instruction;

Group 3) multimedia computer based training (or CAL).

In all groups, content was planned and taught by the same person thus attempting to ensure educational continuity. Further, students were permitted to practice, but no

expert feedback was provided to any group. Outcome measures of success included both multiple choice questions (MCQ) and an objective structured assessment of technical skills (OSATS). Students were tested immediately following the instructional intervention and again one month later. Results indicated that group 1 (above) scored significantly better than groups two and three on the MCQ. No difference between any of the groups in the OSATS scores immediately following the instructional intervention. However, at the one month OSATS follow-up, group 3 scored significantly better than either of the other groups.

Based on the conflicting results between two similar studies (Summers et al., 1999; Rogers et al., 1998), it is challenging to make absolute conclusions about the impact of CAL/CAI on learning. It should be noted that neither study permitted students the opportunity to view the OSCE or OSATS marking sheets used by expert raters (prior to or after the exam). One of the major differences between the Roger et al (Rogers et al., 1998) and Summers et al (1999) study was that feedback was nullified by prohibiting feedback to all students. When designing research that involves technology, feedback may not be the primary variable one is studying. However, researchers should strive to standardize the volume and quality of feedback afforded to study participants across both experimental and control groups.

Carr et al. (1999) studied the impact of CAI on third year otolaryngology clerkship students' ability to diagnose and treat a common medical issue physicians are regularly confronted with in their practice: epistaxis. An identical instructional design for the CAI module developed in Toolbook II Instructor 5.01 and the small group seminar instructional group was employed (Carr et al., 1999). In the CAI group, students

experienced 60 content-specific pages of a self directed module that provided formative assessment through interactivities such as fill-in-the-blank questions, MCQ, true and false questions as well as animations. The control group experienced small class, seminar style teaching which involved small group discussion and didactically delivered material related to the same content. Measurement of knowledge was tested with a seventeen point short answer exam and a practical test done separately which assessed the student's management proficiency on a simulated epistaxis model. The practical exam was evaluated by an expert using a 16 point checklist. Results for the written and practical examinations revealed no difference between the two groups leading the researchers to conclude that CAI was equally effective compared to traditional methods of instruction. Contextually, authors saw this result as very positive since the self directed CAI saved resources compared to teaching the small group seminar. Authors did not expand on the level or type of feedback provided to the seminar group (face to face instruction), even though that seemed to be a major focus of the CAI group (Carr et al., 1999).

Feedback was a significant component of the CAL in a study that evaluated the impact of a CD ROM on learning musculoskeletal examination skills (Vivekananda-Schmidt et al., 2005). Two universities participated in the study whereby third year medical students were on a 12 or five week clinical skills module. Students were randomly assigned into a CD ROM group or a non CD ROM group. Feedback provided to the CAL experimental group was facilitated through typical (drill and practice) multiple choice questions which aligned with the content in the CD ROM. However, once again, researchers only reported the feedback and instructional delivery provided to the experimental group (CAL) and not the control group (Vivekananda-Schmidt et al.,

2005). As a result, with such limited information about the control group interventions, it is challenging to make any absolute conclusions about the effectiveness of the CAL. Nonetheless, researchers found the CAL groups performed better than a control group on previously validated OSCEs for the knee and shoulder assessment (Vivekananda-Schmidt et al., 2005).

There are a number of studies that have included CAL, CAI or CBI as a pedagogical intervention in medical education (Greenhalgh, 2001). Very few have met the stringent requirements of a prospective randomized research design while also employing a practical, OSCE type exam as their outcome measurement. Further, none used the combination of CAL, CBI, CAI and a computer assisted assessment by expert evaluators to aid in learning. It is difficult to make inferences about the impact of educational technology if there is little information about the control group interventions, including the level and type of feedback provided.

No one single, universally accepted definition of CAL, CBI or CAI exists in the literature, however it is possible to tease out the key principles. It is only necessary to define these terms so the research community explicitly views the role computers or technology may play in a research design. In other words, how generalizable are the results of a research study, and to what degree do these results contribute to the underlying theoretical constructs from which they are based? Three key issues have been identified to help frame the practical example that will be proposed later in this paper:

1) Whether the students are electronically learning the majority of material via distance (i.e., independently) or whether the electronic learning materials are considered supplementary to a traditional face to face class;

2) Is the technology central or peripheral to the research question (i.e., a causal factor versus co-variable?);

3) Is the technology intended to teach students, assess students or perform both?

Points two and three above will be explored in greater detail below. Point one is not relevant to this thesis since all teaching and assessment of practical skills must be done in face to face manner due to the nature of the activities and testing (i.e., practical, performance-based examinations that must be assessed in person by an expert rater or SP).

Technology's Role as Central or Peripheral?

In this question, one must ask whether the technology being employed in a study played a central or peripheral role in the research design. Educational research is complex enough with variables such as students, instructors and the learning environment while adding other variables makes the research even more complex. For example, if computers or technology are added, the complexity and interactions between all these variables becomes even greater. As a result, during design development, it is critical to determine if the technology is central to the research question so when the conclusions are drawn, one can confidently state that technology played a major role with student learning. For example, prospective, randomized designs with control groups that do not receive the technological intervention are critical to the overall research design (Greenhalgh, 2001).

Is the Technology Intended to Teach, Assess or Do Both?

The use of technology in teaching has been reviewed previously (Summers et al., 1999; Rogers et al., 1998; Carr et al., 1999; Vivekananda-Schmidt et al., 2005). However, little

has been published on assessment using technology with performance based exams. There appears to be a common and agreed upon definition for CAA in the literature. CAA can be defined as "a range of activities such as the collation, analysis and transmission of examination grades across networks and most desirably, the use of computer-based assessment where students complete assessment at workstations and their answers are automatically marked" (Bull, 1999, p. 123). Although this definition of assessment has been clearly defined, there are other forms of assessment using technology which do not meet this definition.

There seems to be two basic reasons for implementation of computer assisted assessment: formative and summative. There have been a number of studies that introduce performance-based assessments (i.e., OSCEs) as a means of learning or formative assessment where some studies have employed technology to assist them (Sloan et al., 1996; Black et al., 1986; Chatenay et al., 1996; Evans et al., 2005; Geddes et al., 1998; Mavis et al., 2006; Rahman, 2001; Abraham, 1998; Treadwell, 2006). Two of these studies attempted to evaluate student OSCE performance summatively using technology (Palarm et al., 2004; Treadwell, 2006).

The goal of the Palarm et al. (2004) study was to create an OSCE that could assess and capture student answers electronically in order to be graded at a later time by an expert. There were two drivers to create this e-OSCE. The first driver was lack of expert raters. Program administrators did not have enough experts to measure student performance at one sitting so as a result, students rotated from computer to computer rather than from station to station (physically) as happens with traditional OSCEs. The second driver was simplification of image reproduction. Still and moving images are

critical components within the radiology discipline (i.e;. x-rays, ultrasound, MRI, etc.) and so technology simplified and expedited this process. This study did not follow traditional quantitative procedures (outlined previously), but rather employed an action research method. Development of the tool took place over the course of six months and evolved based on feedback from multiple stakeholders such as students, instructors, clinical supervisors and investigators. Suggestions were made by all parties and the primary investigator tried to incorporate all this feedback into the final e-OSCE.

Unfortunately, no reliability statistics were completed, making it difficult to evaluate the OSCE's psychometric soundness. In attempt to gain clarification about the assessment process, the author was contacted to determine how the e-OSCE was graded and the author indicated that a marking "framework" was used (Griffith, 2006). Based on the interpretation of the email response, it seems this type of marking is similar to the concept of expert raters judging essays which can be commonly found in other disciplines such as English or Language Arts. Future research might evaluate the involvement of technology in learning assessment using traditional statistical measures such as an inter-rater reliability research study (Ebel et al., 1986).

Treadwell (2006) studied the impact of using personal digital assistants (PDAs) to assist in grading an OSCE. A primary goal adoption of PDAs was to automate the administration of the examination, including tabulation of examiners' objective and subjective evaluation. In addition, automation of the tabulation permits timely feedback to students which may have not been a priority in the past since most of the instructor's efforts went into the tabulation process. The study goals included studying the effectiveness, efficiency of the PDA and the user satisfaction as measured by subjective

feedback. Effectiveness was measured by student performance on the OSCE using traditional paper and pencil marking with one cohort followed by testing using the electronic version of the OSCE with another cohort and a final cohort that was tested with the electronic version of the OSCE, but this group were aware of the assessment criteria ahead of time. There was no difference between any of the experimental groups (three of them) when performing an independent t-test leading authors to conclude assessment using PDAs did not have an impact on the average score of students regardless of the assessment method employed. There was a trend whereby the students who were exposed to the tool had a higher average grade, but the statistical technique employed did not test differences between the lowest performing group and the highest. It is appropriate to employ an ANOVA in situations when there are three or more groups, something this group did not complete (Treadwell, 2006).

Efficiency was measured by the amount of time taken for the logistical operation of the OSCE (Treadwell, 2006). They found a 77% decrease in time the first year when the OSCE checklists were converted to digital format (from paper). They concluded this method of testing was superior since the original goal of the study was to reduce the workload and logistics associated with running the OSCE. The fact that the mean scores did not significantly change from experimental group to experimental group provided more credibility to the success of the efficiency outcome. Subjective comments about the adoption of the PDAs by examiners are difficult to capture, but generally they were quite supportive of the concept. The concept of an e-OSCEs is unique as is evidenced by the paucity of research on the topic. However, research related to OSCEs that employ

technology and can be used in both a formative and summative setting is even sparser making this area ripe for future research.

Review of Literature Summary

In summary, there have been a number of assessment tools that have been developed to measure underlying constructs in medicine to be used in OSCEs. There is less focus on the tool development for each individual station and more on the OSCE as a whole in measuring underlying constructs. Part of the reason Hodges (2003) questions the validity of the OSCEs is because hasty content validation may lead to poor overall validity. A focus on the individual station tool development such as the one proposed in this study may lead to greater validity and reliability. Further, once the tool has reasonable psychometric characteristics, its practicality for formative or summative assessment (i.e., assessment as, for and of learning) becomes possible. The current study will test this proposal.

Research Questions

1. Is it possible to create a psychometrically sound evaluation tool that can measure the orthopedic assessment clinical competency construct?
2. Does exposure to all educational stakeholders (students, instructors and clinical supervisors) with a psychometrically sound evaluation tool ensure greater success in learning the underlying constructs measured with the evaluation tool?

3. Does mobile technology (PDAs) have a significant impact on learning underlying construct measured by a psychometrically sound evaluation tool?
4. Does the e-SOAT have predictive validity to success in the underlying constructs as measured by a psychometrically sound evaluation tool?
5. Does supervisor and instructor feedback have an influence of student success with orthopedic assessment skills as measured by a psychometrically sound evaluation tool?

6. Do covariates such as age, sex, history of the number of post-secondary courses, or grade point average have an influence of student success with orthopedic assessment skills as measured by a psychometrically sound evaluation tool?

CHAPTER THREE: METHODS

Methods are outlined by the three separate phases of research. In the first phase, the primary goal was to undergo content validation of the Standardized Orthopedic Assessment Tool (SOAT). In the second phase, the primary goal was to perform reliability testing for the SOAT. In the final phase, validity and reliability measurements continued while also measuring the SOAT's impact on learning the underlying construct: orthopedic assessment clinical competence. The participants, adopted tools, procedures employed, statistical analyses performed and research questions are elaborated on for each phase independently. A schematic representation of the study design is outlined in Figure 3.2 and summarizes the third phase of research.

Phase I –SOAT Content Validation

Participants

The participants for the first phase consisted of local experts and a national representation of certified athletic therapists in Canada. In 2005, a twelve person expert panel of athletic therapists from across Canada was struck to participate in the content validation process. An email advertisement was sent to all Certified Athletic Therapists in Canada (approximately 1000). Selection of the twelve experts was based on the following basic criteria:

1. Must be a Certified Athletic Therapist for at least five years;
2. Expertise in teaching and/or examining orthopedic assessment skills.

Additionally, to ensure representation from all Canadian Athletic Therapists' Association (CATA) accredited programs, the primary investigator (PI) sorted applications

accordingly. There are six CATA accredited programs in Canada: Sheridan College, Concordia University (Montreal), York University, University of Winnipeg, University of Manitoba and Mount Royal College (with the University of Calgary). There was no representation from two schools (University of Manitoba and York) from the initial pool of applicants. Subsequently, the PI successfully sought experts from those institutions by contacting the program director or the CATA certification committee. Experts were identified from both those institutions, and then approached to determine if they were interested in participation in the study. Both experts agreed to participate in the study completing the twelve person panel from across Canada with representation from every accredited athletic therapy program in Canada. Other than the basic criteria to participate as an expert, no other demographic or psychographic information was collected for the group of experts.

Tools

The original SOAT forms were converted from a Microsoft Excel © spreadsheet into a web-based survey for expert analysis (Appendix A). Survey Pro 3.0 © was the web-based software used to collect expert opinions on the SOAT and tabulate the descriptive results.

Procedures

A modified Ebel procedure was used to create the SOAT. A detailed description of the methods employed in the current study is outlined in Butterwick et al. (2006) and in Violato et al. (2002). This method has been shown to be superior to other methods, particularly for performance based, practical examinations (Cantor, 1989).

Creation of an instrument to measure student performance in practical OSCE-type examinations at the Mount Royal College-University of Calgary athletic therapy program began in 1999. The examination consisted of seven stations, one of which was a 30 minute orthopedic assessment station. Practical internship clinical supervisors met once a month for five years and provided feedback on student performance in the clinic and feedback on the student assessment tools. This was an important step in the assessment tool's refinement and development: simple validation.

The tool consisted of 10 major categories commonly used in orthopedic assessments (history; observation; scanning examination; clearing joints above and below the lesion site; active range of motion; passive range of motion; isometric resisted testing; special testing; palpation; conclusion) with approximately 200-250 tasks (body region dependent) comprising the details under those major categories. Each major category (i.e., history, observation, etc.) measures a variety of skills (i.e., communication skills, clinical reasoning ability, observational skills, psychomotor & technical skills) that collectively contribute to the orthopedic assessment clinical competence. Students are provided 30 minutes to complete such an examination.

A unique SOAT representing the tasks necessary to complete a clinical orthopedic assessment for eight body regions (cervical spine, lumbar spine, shoulder, elbow, wrist/hand/thumb, hip, knee and lower leg/foot/ankle) were the focus of this study. Three diagnoses (scenarios) for each body region (8) were judged by the expert group (Table 3.1). Experts were asked to grade each task (listed in Appendix A) for each body part and scenario on importance (essential, important, not important, not applicable) and difficulty (hard, medium, easy, not applicable) (Figure 2.3). All grading was done using

an on-line survey tool housed on a Mount Royal College server. Experts were given six weeks to complete the on-line surveys.

Table 0.1 Body Regions and Diagnoses

Body Region	**Diagnosis**
Cervical Region	C 5-6 Facet Sprain
	Strained Sternocleidomastoid
	Degenerated Disc with Radicular Pain
Shoulder Region	Subacromial Bursitis
	Superior Labral Tear/Lesion
	Infraspinatus Tendinitis
Elbow Region	Olecranon Bursitis
	Tennis Elbow
	Ulnar Collateral Ligament Sprain
Wrist & Hand Region	Scaphoid Fracture
	Extensor Pollicis Tenosynovitis
	Ulnar Collateral Ligament Sprain
Lumbar Region	Lumbar 3-4 Facet Sprain
	Strained Abdominal Muscle
	Spondylolisthesis
Hip Region	Strained Hamstring
	Strained Iliopsoas
	Anterior Capsule Strain
Knee Region	Strained Rectus Femoris
	Anterior Cruciate Ligament and Medial Collateral Ligament Sprain
	Patellar Tendinitis
Ankle Region	Anterior Talofibular Ligament Sprain
	Achilles Tendinitis
	Deltoid Ligament Strain

Traditional use of the Ebel procedure has been to determine the minimum passing level in professional practical exams (Violato et al., 2003; Cantor, 1989). Normally, this process consists of two stages:

Stage 1 - Experts grade each SOAT based on the diagnosis provided using the difficulty (hard, medium, easy, not applicable) and importance (essential, important, not important, not applicable) scale.

Stage 2 - Face to face discussion attempted to build consensus on items that did not achieve 80 percent consensus for each task/item on the SOAT for each region on the Ebel grid (Figure 2.3).

SOAT, M. Lafave

Butterwick et al. (2006) established the benchmark of 80 percent consensus when attempting to identify which tasks or items should be included in the final instrument. Any task listed on the tool that did not achieve 80 percent consensus was considered for elimination from the tool. The current study used the same benchmark established by Butterwick et al (2006) for both the initial grading of importance and difficulty and the subsequent face to face consensus building process.

The second stage of the Ebel procedure required the expert validation committee to meet face to face to discuss items that did not achieve 80 percent consensus. Examiners were flown into Calgary, Alberta, Canada for a one day (10 hours) face to face meeting to discuss items that did not achieve 80 percent consensus. Discussion was facilitated by the PI while each major category (patient history, observation, etc.) and each task therein was debated for inclusion and assignment to one of the three importance scale categories (essential, important, not as important).

Statistical Analysis

Descriptive statistics were tabulated in the first stage of the Ebel procedure by the Survey Pro 3.0 © software. Descriptive statistics were collected in the second stage of the Ebel procedure by a show of hand (agree, disagree) and manually counted by the PI throughout the facilitated discussion. The final stage of the procedure required manual collection of expert opinions of agreement for the new model through email correspondence (i.e., descriptive statistics).

Research Question

Is it possible to create a content-valid assessment tool that can measure the orthopedic assessment clinical competency construct?

Phase II – Initial Reliability of the SOAT

Participants

Two types of participants were needed for the second phase of the study: students and examiners. There were two types of examiners: Raters and Standardized Patients (SPs). Selection rationale of these participants will be described separately.

Raters

Raters were solicited from the Alberta Athletic Therapists Association through an email distribution list to all certified members in Alberta (approximately 100 at the time of the study). The only criteria for raters to participate in this phase of the study were for them to be currently certified as athletic therapists in Canada. There was no requirement for raters to be certified as athletic therapists for a minimal number of years. Raters were chosen based on their availability over the course of a week (as advertised). A time block was set according to rater availability for examination of a minimum of five students at 45 minutes per student. Choosing raters and setting time blocks was done prior to student or scenario assignment. Two raters (and one SP) per student were needed to grade student performances.

Standardized Patients (SPs)

Clinical practicum course instructors at Mount Royal College (two instructors for three, three credit hour courses: PHED 3354, Clinical Practicum in Athletic Therapy) acted as the SPs for the examinations. The grades and results of the testing were being used as a summative grade for the students, so the instructors felt it was necessary to participate as SPs and be consistent for all students. One SP rated students on three time

blocks while the other rated students on two time blocks. The SPs were assigned to time blocks in such a way to limit fatigue (i.e., no back to back testing time blocks).

Students

Participants consisted of thirty-two students from the athletic therapy program at Mount Royal College (MRC) enrolled in their first clinical practicum class (PHED 3354). The MRC students were selected as participants for this study as they were familiar with the performance based testing procedures and the SOAT from their involvement with the clinical practicum class.

Tools

The original SOAT validation study focused on eight body regions (cervical spine region, shoulder, elbow, wrist and hand, lower back region, hip, knee, and the foot and ankle). In this phase of the study, three body regions became the foci; the ankle, knee, and shoulder. These regions are frequently injured and thus, assessed by athletic therapists/trainers, physiotherapists or sports medicine specialists (Caine, Caine, & Lindner, 1996). One scenario/diagnosis for each body region was employed in the current study (knee: pretibial bursitis; ankle: lateral ankle sprain; shoulder: subacromial bursitis). Appendix B illustrates the SP 'answer key' for each of the tasks outlined in the SOAT.

Procedures

A convenience block randomized sampling method was employed to ensure that at least 10 students were tested for each body region (i.e., at least 10 ankle, 10 knee and 10 shoulder scenarios). An objective structured clinical examination (OSCE) framework was established with five to seven students assigned to one of five designated time

blocks. Each student was assigned one of three body regions (i.e., ankle, knee or shoulder) and was assessed by two independent raters and the standardized patient (SP) using the corresponding SOAT.

All raters attended a three hour orientation and training session for the SOAT (i.e., the grading scheme and rating scales they would be using throughout the evaluation process, purpose of the assessment and the intended targeted audience). A standardized Microsoft Powerpoint © presentation was employed to ensure the training session could be consistent for future research (Appendix C). Moreover, great attention was paid to the rules for marking the iterative nature of the SOAT and more specifically, the special testing section and the palpation component of the orthopedic assessment (Table 3.2). Conceptually, flexibility in an examinee's decision making process must be carefully scrutinized by the raters thus making the SOAT more than just a dichotomous checklist of right and wrong, done or not done as was the original intent of OSCEs (Harden et al., 1979).

After the basic orientation was complete, each SOAT for the shoulder, knee, and ankle was reviewed with the specific 'answers' that the SP would provide when asked or tested (Appendix B). The training was quite interactive with many opportunities to stop the presenter and ask for clarification.

Table 0.2 Rules and/or Assumptions about the use of the Tool when Evaluating Student Performance

This checklist is intended to measure a basic level of competence in Athletic Therapy students and/or recent graduates. The examinee should be considered to be at a basic level of competence.
The grading rubric for the history and observation components of the orthopedic assessment are done with a 2 point scale:
0 = Not done or done incorrectly
1 = Done
The grading rubric is a 6 point scale with the following definitions for each
0 = **Not done.** The student failed to do the task when they should have done it. 1 = **Very poor performance**. Student performed the majority of the task incorrectly and/or inappropriately and/or inaccurately. 2 = **Poor performance**. Student performed some of the task incorrectly or inappropriately and/or inaccurately. 3 = **Basic performance**. Student performed the task at a basic level of competence appropriate for recent graduates of an athletic therapy curriculum. 4 = **Very good performance**. Student performed the task above the basic level of competence. 5 = **Outstanding performance**. Student performed the task well above the basic level of competence.
All tasks in all components of the tool will default to n/a if they are not graded by the examiner.
Students will lose grades on tasks that they missed (or completed incorrectly) which were thought to be relevant by the evaluator for a specific scenario or condition.
Students should be graded on tasks based on the index of suspicion they identified at the end of each section. If the student does not have the correct index of suspicion and they fail to complete tasks that would permit them to know more about the condition, the student will lose marks for those specific tasks. If they perform tasks that relate to their index of suspicion, they should receive grades for those tasks done correctly.
The model (standardized patient) will ask the student if the index of suspicion has changed at the end of each component of the assessment (i.e; history, observation, scanning exam, clearing joints above and below, examination, special tests, palpation).
All checklists have the same tasks for history, observation, scanning examination and examination (AROM, PROM, IR)
The checklist is used generically for all conditions. ALL motions will be listed for AROM, PROM and IR. The default will be n/a and should only be graded if the student did it or failed to do it when they should have.
Each checklist has unique tasks listed for the special tests and palpation sections

Students should not be required to complete a specific and discrete list of **special tests** for any given condition. Rather, the specific special tests that should be completed will be based on what the student/examinee completed throughout the preceded components of the orthopedic assessment (i.e., history, observation, scanning exam, clearing joints above and below and examination). All special tests will be defaulted as n/a. Special tests will only be marked if the student performed them (i.e., correctly or incorrectly) **OR** if they should have tested a structure, but failed to do so. In the case when a student/examine FAILS to complete a special test to which would identify or confirm the anatomical structure/lesion site, they will be given a grade of zero for all special tests which would have tested that structure.
If a student does not complete a special test at the basic level of competence (i.e., 3/5), then the remaining special tests listed on the form that are supposed to test that structure should be marked as 0. Said another way, if a student gets marked as a 1 or 2 with their performance on a special test, unless they go an attempt another special test that tests those structures, then all other tests that WOULD HAVE tested that anatomical structure should be marked as zero (0).
If there is a special test missing from the list, the examiner should mark add the test and then mark it based on its accuracy. If none of the examiners are familiar with the test, it should be highlighted and the lead examiner should research the test AND ask the examinee the name and purpose of the special test at the end of the entire test. If this is a valid and reliable test, it can be added to the database.
Special test expectations are that a student ONLY has to complete 1 special test on an anatomical structure in order to effectively clear that anatomical structure regardless of the test they choose.
Students should not be required to **palpate** a specific and discrete list of anatomical structures for any given condition. Rather, the specific anatomical structure that should be palpated will be based on what the student/examinee completed throughout the preceded components of the orthopedic assessment (i.e., history, observation, scanning exam, clearing joints above and below and examination). All anatomical structures in the palpation section will be defaulted as n/a. Palpation will only be marked if the student performed them (i.e., correctly or incorrectly) OR if they should have palpated a structure, but failed to do so. In the case when a student/examine FAILS to palpate an anatomical structure which would identify or confirm the lesion site, they will be given a grade of zero for all anatomical structures which would have tested that lesion site
Special tests are listed in alphabetical order
Anatomical structures for palpation are listed in alphabetical order
A final impression of the student's overall performance should be graded at the end of the form.

Statistical Analysis

Completing practical examinations in athletic therapy is extremely time consuming and thus presents a logistical challenge when designing a study and resulting statistical analysis that can be employed. It is acknowledged that the gold standard for testing reliability in performance based examinations is a fully crossed generalizability analysis (Brennan, 2000; Brennan, 1992). In a fully crossed design, each student should be tested by each rater and SP over the course of at least two scenarios/cases. In this case, generalizability coefficients can be calculated separating variance into rater, case and student. However, this study design was not possible in the current study based on the availability of examiners. As a result, Cronbach's alpha reliability coefficient was employed. Cronbach's alpha reliability coefficient has been shown to be effective and essentially the same (mathematically) as a special case, one-facet generalizability theory design (Shrout et al., 1979; Cronbach et al., 1972; Streiner, 1993). The facet being studied was rater reliability with the SOAT.

Research Question

Is it possible to establish the initial reliability of the SOAT, further building its overall validity? Strong reliability in this phase of the research question would permit further study of the SOAT for teaching, learning and assessment of orthopedic assessment clinical competency in the third phase of study in this thesis.

Phase III – Integration of the SOAT in Various Canadian Athletic Therapy Curricula

Participants

Three types of participants were needed for the third phase of the study: educational institutions, students, examiners. There were two types of examiners: Raters and SPs. Selection rationale of these participants will be described separately.

Educational Institutions

There are six accredited athletic therapy programs in Canada (Sheridan College, Concordia University [Montreal], York University, University of Winnipeg, University of Manitoba and Mount Royal College [with the University of Calgary]). Four participating institutions were targeted for this study due to the similarity in curricular design: Concordia University (Montreal), University of Winnipeg, University of Manitoba and Mount Royal College (with the University of Calgary). All students/subjects took an introductory orthopedic assessment class in the fall 2006 semester (with a structured lab component), followed by a clinical internship course in the winter 2007 semester. All institutions adopted the same textbook for their introductory course (Magee, 2002) and all programs followed the same basic competencies outlined by the professional governing body (the Canadian Athletic Therapists' Association – CATA). The competencies are used as part of the self study accreditation process (Canadian Athletic Therapists Association, 2005). All programs target the national certification examination as the gold standard for their students to achieve the orthopedic assessment construct.

Students

Third year students from four athletic therapy programs were targeted as subjects for this phase of the study:

Group 1 – University of Winnipeg: The comparison group.

Group 2 – Concordia University: Only the instructor received (i.e., no explicit exposure for students) exposure to the paper version of the SOAT.

Group 3 – Mount Royal College/University of Calgary: Student and instructor receive exposure to the e-SOAT via desktop computer.

Group 4 - University of Manitoba: Student and instructor receive exposure to e-SOAT via desktop computer or PDA.

Students were solicited for participation by the lead instructor for their orthopedic assessment classes at each institution. If students were enrolled in the third year (or greater) of their program at the commencement of the study (September, 2006) and they were taking the orthopedic assessment class in the fall, 2006 semester and the clinical internship class in the 2006-2007 academic year, they were considered as potential volunteers for the study.

Raters and Rater Training

The PI was the SP for all examinations (n = 58) while two local examiners were trained based on the procedure outlined below. Raters at each educational institution were chosen based on availability. Concordia University (Group 2) consisted of raters that were also the instructor for the primary orthopedic assessment class and lab. All students were graded by those two examiners across five potential, testing blocks. These

raters were trained using the standardized training Microsoft Powerpoint© presentation outlined in the methods from the second phase above (Appendix C).

The University of Winnipeg (Group 1) raters were solicited via an advertisement sent through an email distribution list from the regional chapter of professional athletic therapists: The Manitoba Athletic Therapists' Association (Appendix D). Six volunteers were trained, but only four raters were needed to test the students in Winnipeg. As a result, a convenience sample of raters was chosen based on availability with examining dates.

Raters from Calgary (Group 3) were solicited through a similar process as the University of Winnipeg students. An email advertisement (Appendix E) was sent through the regional chapter of professional athletic therapists via a membership distribution list: The Alberta Athletic Therapists' Association. Eight volunteers responded, but only five were needed based on the total number of students being tested in Calgary. As a result, a convenience sample was chosen based on rater availability, student examination schedule and PHED 3354 practicum instructor availability. The examination was being used as a final examination for the PHED 3354 course and thus, the two instructors wanted to be included as volunteers for the study.

Tools

A paper copy of the SOAT was sent via Canada Post to group 2 instructors. Groups three and four were provided access to the e-SOAT through a link in a Blackboard Learning Management System© course shell specifically designed for this study. The e-SOAT is an electronic version of the paper copy of the SOAT sent to group 2 (Appendix A). A sample of an e-SOAT page is displayed in Figure 3.1 below. Each

major category within the orthopedic assessment structure is provided as a link on the left hand side of the e-SOAT. When someone clicks on a major section, a detailed checklist pops up on the frame in the right hand side of the screen. The scaling responses employ a radio button design with six total options (from 0-5). The number of scaling responses was specifically chosen because seven was optimal (Streiner et al., 2003; McKelvie, 1978), but only six could fit within the personal digital assistant (PDA) screen size. Five Dell Axim X50 PDAs were deployed to group 4 for usage in their classroom, laboratory and/or clinical setting. The PDA's were wi-fi enabled permitting users to access and view the e-SOAT site in wi-fi hotspots (i.e., clinic, classroom and labs). The e-SOAT could also be accessed from any desktop computer for both groups three and four.

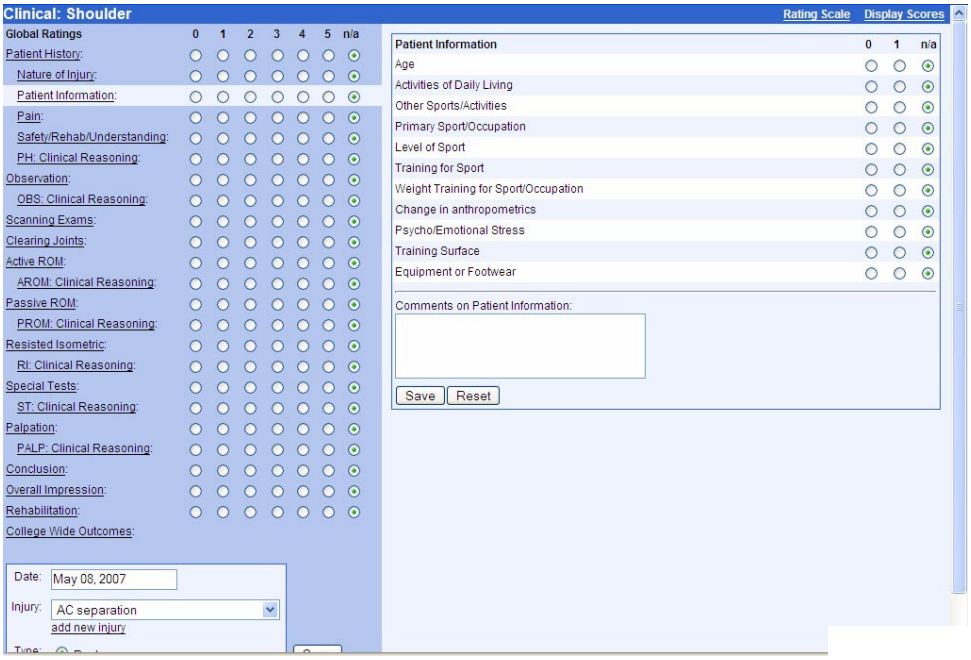


Figure 0.1. Screen capture of the e-SOAT.

Procedures

Each University was randomly assigned to one of four interventions one comparison group and three treatment groups (Figure 3.2):

Group 1 – University of Winnipeg: The comparison group.

Group 2 – Concordia University: Only the instructor received (i.e., no explicit exposure for students) exposure to the paper version of the SOAT.

Group 3 – Mount Royal College/University of Calgary: Student and instructor receive exposure to the e-SOAT via desktop computer.

Group 4 - University of Manitoba: Student and instructor receive exposure to e-SOAT via desktop computer or PDA.

Instructors, clinical supervisors and students (groups 3 and 4) from the treatment groups (above) were orientated to the e-SOAT by the PI prior to the commencement of classes in the fall, 2006 semester. Part of the orientation included instructions on how to access the e-SOAT. The e-SOAT was housed on an academic server at Mount Royal which had restricted access/traffic. As a result, links were created in a Blackboard Learning Management System© course shell for groups three and four above. Instructors, students and clinical supervisors were told to only access the link to the e-SOAT site through the Blackboard Learning Management System© course shell even though they could directly access the e-SOAT on the Mount Royal server. Each student and instructor in Groups three and four was provided with a unique username and login that would permit tracking of their website entries (i.e., hits to the website). There was no attempt to contact the instructors, clinical supervisors or students at any of the institutions by the PI throughout the year. However, if instructors or clinical supervisors

contacted the PI for clarification about the SOAT methodology, then the PI addressed those questions.

At the end of the winter 2007 semester, the PI traveled to each institution to test students on either knee and shoulder scenarios/cases. Student volunteers who signed up to participate were randomly assigned a knee or shoulder scenario (Appendix F) ensuring there was approximately equal number of cases tested at each institution (i.e., treatment group). Students were all asked to complete a questionnaire after they had been examined which included a number of psychographic and demographic items (Appendix G). All students were tested across a four week span between March and April, 2007.

At the completion of the third phase of research, more data was collected from the participating treatment group institutions (groups two, three and four above) regarding technology and SOAT adoption. A retrospective questionnaire was developed and distributed in May, 2007 (Appendix H).

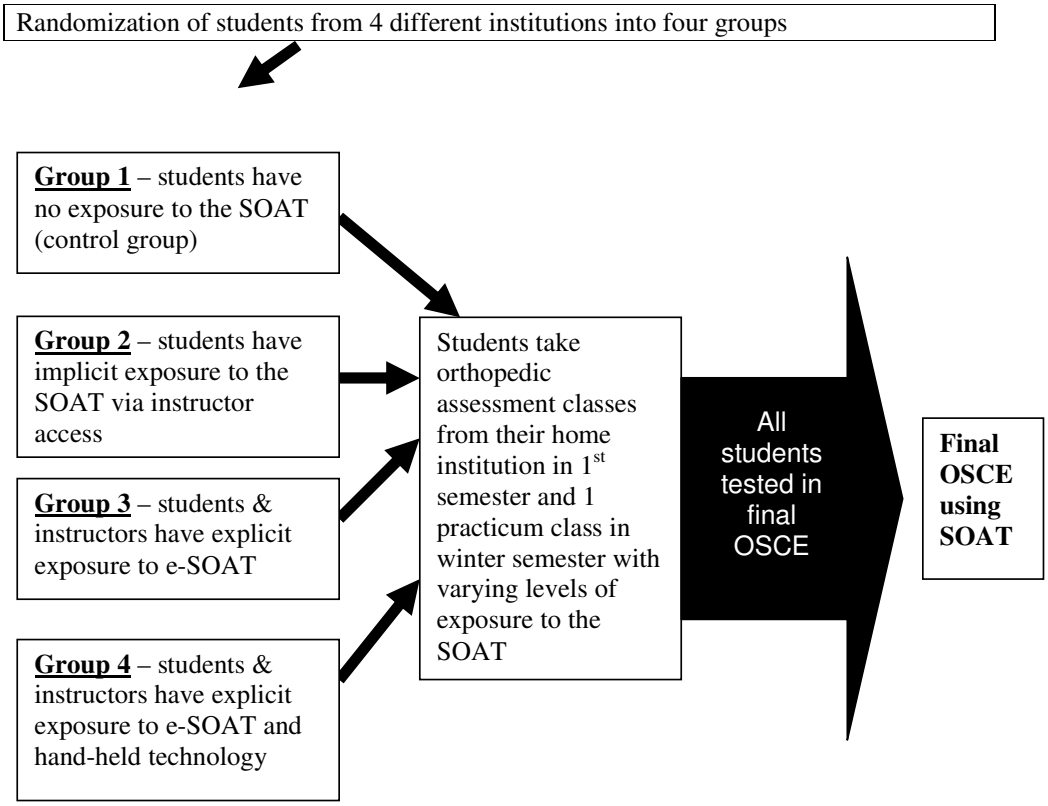


Figure 3.2. Schematic Representation of Study Design

Statistical Analyses

Exploratory simple correlation (Pearson r) was employed initially to determine the connection between covariates and each other, but more importantly between the covariates and the students' final exam scores. Once the simple correlation was completed, significant correlations warranted an analysis of covariance to be employed to measure the differences between quasi-experimental and comparison groups along with the influence of the "total number of courses" covariate. Post-hoc analysis was also completed to determine where significant differences between groups existed. A simple

correlation (Pearson r) was completed between the student scores in the examination using the SOAT and the number of internet hits measured in the Blackboard Learning Management System © course shell. Finally, descriptive statistics were calculated for student estimates of quality and quantity of instructor and clinical supervisor feedback.

Cronbach's alpha reliability coefficient was performed for all groups (comparison, group 2 and three above). This statistic was calculated for each of the two scenarios: knee and shoulder. In addition, student scores from all groups were summated and Cronbach's alpha reliability coefficient was computed. Finally, the group 2 examination site had the same examiners and SP for all students. Although students were not tested on more than one scenario as would be required for a fully crossed generalizability design, a nested design and subsequent analysis was completed. The scenario and raters were nested within the student thus permitting generalizability of the rater and scenario facets for this treatment group. All statistical procedures were calculated using SPSS 14.0 or manually, in the case of the generalizability coefficient.

Research Questions

The first research question in the third phase was to evaluate the impact that explicit exposure to the SOAT had on learning orthopedic assessment skills in athletic therapy students. The research design was such that students were not truly randomized into comparison and treatment groups and thus, covariates were collected in attempt to measure their correlation to students' scores and contribution to variance between groups. As a follow up to the first research question, it was further proposed that the greater the volume of exposure to the SOAT (as measured by total internet 'hits'), the better the score in the final. A final research goal was to continue to establish the validity and

reliability of the SOAT as a measurement instrument of the orthopedic assessment clinical construct. Building construct validity of a measure must take place over time and a number of studies (Violato et al., 1998). This phase of the research would be the third study contributing to the overall validity and reliability of the SOAT.

Ethics

Ethical approval for all participating institutions was obtained through their respective ethics approval bodies: the University of Calgary, Mount Royal College, the University of Manitoba, the University of Winnipeg and Concordia University. The instructor contacts at each institution were the primary investigators on each application. A copy of the University of Calgary Ethics approval letters is in Appendix I.

CHAPTER FOUR: RESULTS

The results for the three phases of research with the thesis are described separately. However, a brief outline of the results for each phase is described first. In the first phase, a modified Ebel procedure resulted in a content-valid SOAT. In the second phase, employing the SOAT for knee, shoulder and ankle scenarios across multiple examiners resulted in strong initial reliability for each. In the final phase, validity and reliability of the SOAT continued. In addition, treatment groups with varying levels of stakeholder exposure to the SOAT were compared using an analysis of covariance. Greater details of the results for all phases of research will be presented for each phase independently below.

Phase I –SOAT Content Validation

The Ebel procedure consisted of two stages: initial assessment of each item's importance and difficult followed by a second stage which included a face to face discussion to achieve consensus on those items that did not achieve 80% consensus in the initial stage (Butterwick et al., 2006). In the first stage of the content validation process, there was limited consensus for items and categories that achieved 80% or higher in grid square number one (easy and essential; Figure 2.3): active range of motion; passive range of motion; isometric resisted testing. Summarized task data is presented in Table 4.1 categorized on the Ebel grid square presented in Figure 2.3. The forearm, wrist and hand body region was 79% and considered close enough to the 80% consensus target to accept.

Table 0.1 Initial Consensus Results from Stage One of Ebel Procedure

Item	% Consensus		
	Essential (Grid #1)	**Important (Grid #2)**	**Less Important (Grid #3)**
Hip			
Active ROM	95	5	0
Passive ROM	81	19	0
Isometric Resisted	85	15	0
Knee			
Active ROM	82	18	0
Passive ROM	84	16	0
Isometric Resisted	82	18	0
Lower Leg, Foot & Ankle			
Active ROM	100	0	0
Passive ROM	100	0	0
Isometric Resisted	100	0	0
Shoulder			
Active ROM	80	18	2
Passive ROM	80	18	2
Isometric Resisted	80	18	2
Elbow			
Active ROM	83	17	0
Passive ROM	86	14	0
Isometric Resisted	90	10	0
Forearm, Wrist & Hand			
Active ROM	79	19	2
Passive ROM	79	19	2
Isometric Resisted	79	19	2

In the second stage of the validation process (the face to face component), the majority of the discussion focused on those items and categories that did not achieve initial consensus: history; observation; scanning exam; clearing joints above and below; special testing; palpation. There was little consensus on any history items/tasks before or during the face to face meeting (i.e., stage 1 or 2). As a result, thematic clusters for the

history evolved from the discussion, which were intended to permit greater consensus on individual items so experts felt their preferences for each task were addressed. As a result of the thematic clusters, there was 100 percent consensus on all thematic clusters and subsequent consensus on the vast majority of tasks under each thematic cluster. Three exceptions did not meet the 80 percent target of consensus (pain scale, sport and sleeping position), but the PI decided to keep these tasks in the final SOAT nonetheless.

The observation component achieved 80 percent consensus for all tasks during the face to face discussion. Experts felt that if an upper extremity lesion was being evaluated, then only the upper extremity had to be observed (and vice versa with the lower extremity). Additionally, clearing joints above and below and the scanning exam components achieved consensus with relatively little discussion.

The special testing and palpation components were extremely controversial areas and it was obvious after hours of discussion that achieving 80 percent consensus was not attainable. As a result, a third stage to the process was added due to time constraints of the expert group (they had to fly out of town). The group discussed the merits and potentially negative impact of a new model that linked the special testing and palpation sections throughout the entire assessment procedure which the PI later captured, articulated and captured as “clinical reasoning.” Discussion focused on the concept of deductive and inductive approaches to orthopedic assessment and which of these should be expected for athletic therapy graduates. In the end, a new model of the SOAT included a clinical reasoning component that tied each of the 10 components together (i.e., history linked to observation, and then linked to the scanning exam, and then linked to the clearing joints, and then linked to the active range of motion, and then linked to the

passive range of motion, and then linked to the isometric resisted testing, and then linked to the special testing, and then linked to the palpation and then linked to the conclusion). This new model was sent (via email) to the expert group two weeks after the face to face evaluation. The new model achieved the 80% consensus target for all clinical reasoning components, and the originally controversial special testing and palpation components. Final versions of the SOAT can be found in Appendix A.

Phase II – Initial Reliability of the SOAT

Scale reliability analysis (Cronbach's alpha) was completed on the SOAT for each of the three body regions (knee, shoulder and ankle). The SOAT's sum total for individual body region for the two raters and one SP were used to calculate the students' performance scores. The SP's completed a global rating scale for each major section while the raters completed the detailed checklist in the SOAT followed by completion of a global rating scale (Appendix J and A, respectively). Raters' scores are actually a sum total of the detailed tasks combined with the global ratings for each category. In contrast, the SP data consisted of only the global ratings for each major section and sub-section within the SOAT. Major section headings include history, observation, scanning examination, clearing joints above and below the lesion site, examination (including active range of motion, passive range of motion, isometric resisted testing), special testing, palpation and conclusion. Two separate reliability analyses were completed: one with the SP included in the analysis and the other with only the scores of the two raters. As shown in Table 4.2, the mean overall reliability of all three SOAT body regions (i.e., knee, ankle and shoulder) was similar whether SP's scores were included in the analysis

($\alpha = 0.85$) or not ($\alpha = 0.86$). An inter-rater analysis between the two examiners and SP raters show an overall agreement rate of 99 percent.

Table 0.2 Reliability Statistics by Body Region

Body Region & Diagnosis	**Reliability with SP Data Included (Cronbach's Alpha)**	**Reliability without SP Data Included (Cronbach's Alpha)**
Ankle	.91	.91
Knee	.83	.91
Shoulder	.82	.76

Reliability data were also organized and analysed by the five student-randomized blocks according to the three separate body regions. As shown in Table 4.3, the average number of years of raters' experience influenced the reliability of the SOAT between blocks. In particular, the greater the number of years of experience (e.g., those with 10 or more years of experience) the less likely the raters were found to be consistent in their evaluation of student performance on the scenarios.

Table 0.3 Reliability Statistic by Randomized Block

Rater # (1 - 10)	**Average Rater Years of Experience**	**SP (1 or 2)**	**# Ankle Scenarios**	**# Knee Scenarios**	**# Shoulder Scenarios**	**Block Reliability (Cronbach's Alpha)**
1, 2	21	1	1	3	2	.45
3, 4	2	1	2	1	3	.93
5, 6	4	2	3	3	1	.84
7, 8	6.5	2	3	2	1	.84
9, 10	15.5	1	1	2	4	.55

Phase III – Integration of the SOAT into Various Canadian Athletic Therapy Curricula

The results of this phase of the research are separated into the various research questions. However, Table 4.4 outlines psychographic and demographic information on the student-volunteers who participated in this study. There were a total of 58 subjects tested in this phase of the research project. There was one SP who tested all subjects, but there were varying numbers of raters with various groups that were tested: group 1 = four raters; group 2 = two raters; group 3 = five raters.

Table 0.4 Psychographic and Demographic Student Information

Institution	Total n Male	Total n Female	Mean Age	Mean GPA on 4.0 Scale	Mean # courses completed
Group 1	4	5	26.22	3.44	48.22
Group 2	10	14	23.13	3.17	33.21
Group 3	5	20	22.12	3.09	41
Mean or total across all groups	**19**	**39**	**23.82**	**3.23**	**40.81**

The psychometric research question was to confirm the reliability of the SOAT. Cronbach's alpha reliability coefficient was calculated using SPSS 14.0 ©. The results for both knee and shoulder scenarios broken down by institution are listed in Table 4.5. When all institution scores were combined into one analysis and Cronbach's alpha was calculated, the reliability coefficients were .90 for the shoulder and .93 for the knee. The

fourth treatment group (PDA assigned) does not have any data because no student volunteers signed up to be tested.

Table 0.5 SOAT Reliability of Knee and Shoulder Cases

Institution	Cronbach Alpha Reliability Coefficient **SHOULDER (n)**	Cronbach Alpha Reliability Coefficient **KNEE (n)**
1. Comparison Group	.71 (4)	.87 (5)
2. Instructor only (i.e., no explicit exposure for students) exposure to the paper version of the SOAT	.85 (12)	.85 (12)
3. Student and instructor exposure to the e-SOAT via desktop computer	.86 (13)	.89 (11)
4. Student and instructor exposure to e-SOAT via desktop computer or PDA	n/a (0)	n/a (0)
Mean reliability by case (across 3 groups)	**.81**	**.87**

The research design was such that it was not feasible to employ a fully crossed design and perform generalizability analysis. However, at one institution (group 2 above), all students were evaluated by the same raters and the same SP. This was a nested design (in students) to permit generalizability analysis. Variance estimates were calculated using SPSS 14.0 for the raters and scenarios with the students as the dependent variable. The results are listed in Table 4.6 below. Generalizability coefficients were manually calculated using the following formula and resulted in .73 for the rater facet and .89 for the scenario facet:

$$Ep^2_{(r)} = \frac{538.1}{538.1+198.5}$$

$$Ep^2_{(s)} = \frac{1581.6}{1581.6+198.5}$$

Table 0.6 Factor Level Information and Variance Estimates for the Rater and Scenario Facets

Factor Level Information

		N
Rater#	1.00	24
	2.00	24
	3.00	24
Scenario	1.00	36
	2.00	36

Dependent Variable: Stud.Scores

Variance Estimates

Component	Estimate
Var(Rater#)	538.137
Var(Scenario)	1581.548
Var(Error)	198.544

Dependent Variable: Stud.Scores
Method: Minimum Norm Quadratic Unbiased Estimation
(Weight = 1 for Random Effects and Residual)

In addition, a number of covariates were collected with a survey (Appendix G) when the students were tested on their orthopedic assessment skills. Relevant descriptive statistics are listed in Table 4.4 above. A Pearson r, simple correlation was computed using SPSS 14.0 © for all covariates collected: age, sex, grade point average (GPA), total number of post-secondary courses completed at the time of testing and, estimates of the quality and quantity of instructor and clinical supervisor contact and feedback related to orthopedic assessment skills. There was only one significant correlation between the covariates and the dependent variable (i.e., student scores using the SOAT): Total number of courses taken to date. The resultant significant correlation is listed in Table 4.7.

Table 0.7 Pearson r Correlation Coefficient for Student Scores and the Total Number of Courses Completed at Time of Testing

		Stud.Scores	Total.Courses
Stud.Scores	Pearson Correlation	1	.322(*)
	Sig. (2-tailed)		.018
	N	58	54
Total.Courses	Pearson Correlation	.322(*)	1
	Sig. (2-tailed)	.018	
	N	54	54

* Correlation is significant at the 0.05 level (2-tailed).

The significant correlation between the student scores and the total number of courses warranted the use of a one-way analysis of covariance to compare the influence of explicit exposure to the SOAT on learning. The independent variable was the institution (Mount Royal/U of C, Concordia University and University of Winnipeg) and the dependent variable consisted of the scores of the practical exams using the SOAT for both knee and shoulder regions. The total numbers of courses students have taken in post-secondary were used as the covariate in this analysis. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes and reliability of the covariate. However, tests for linearity and homogeneity did show assumptions for employing an ANCOVA were violated (Appendix K). Although there was a significant difference between groups (post hoc analysis below), the influence of the covariate did not demonstrate a significant impact on the difference between groups.

The research question that compared instructional methods for learning orthopedic assessment skills was evaluated by comparison of groups one, two and three. Descriptive statistics of student scores using the SOAT in a performance-based, practical

examination for the comparison group and two treatment groups are listed in Table 4.8. A one-way analysis of covariance was conducted to explore the impact of exposure to the SOAT (i.e., explicitly or implicitly) on learning orthopedic assessment skills. There was a statistically significant difference in SOAT scores for the three groups [F (2, 110.4) = 28.6, p = .01]. The effect size, calculated using eta squared, was .51. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for quasi-experimental group 1 (M = 77.36, SD = 7.9) was significantly different from the treatment group 2 (M = 55.85, SD = 12.37) and the comparison group (M = 57.61, SD = 11.44). There was no difference between the treatment group 2 and group 1 (comparison group).

Table 0.8 Descriptive Statistics for Quasi-experimental and Comparison Groups

Quasi-experimental and Comparison Groups	**N**	**Mean**	**Std. Deviation**	**Std. Error**
Group 1	9	57.6	11.4	3.8
Group 2	24	55.8	12.4	2.5
Group 3	25	77.4	7.9	1.6
Total	58	65.4		

One insignificant correlation should be noted since it was central to the research question that was measuring the impact of technology on students' scores. The correlation between the total number of hits (i.e., e-SOAT website entries) and the final orthopedic examination grade is outlined in Table 4.9. This correlation was only run with group 3. Groups one and two did not have access to the e-SOAT site and group 4 had access, but no students volunteered to take the final examination. There was no significant correlation between the SOAT score and the number of hits on the e-SOAT site.

Table 0.9 Correlation Between SOAT Score and e-SOAT Website Hits

		SOAT Grade	Actual Hits
Student	Pearson Correlation	-.068	.363
	Sig. (2-tailed)	.752	.082
	N	24	24
SOAT Grade	Pearson Correlation	1	.330
	Sig. (2-tailed)		.116
	N	24	24
Actual Hits	Pearson Correlation	.330	1
	Sig. (2-tailed)	.116	
	N	24	24

Descriptive statistics related to instructor and clinical supervisor quality and quantity of feedback as perceived by the students is summarized in Table 4.10 and 4.11. Questions for these qualitative questions were posed in the psychographic and demographic information collected at the time of SOAT testing (Appendix G, questions 12, 13, 14 and 15 inclusively).

Table 4.10 Descriptive statistics related to instructor and clinical supervisor quantity of feedback as perceived by students

Survey Questions	**Estimated Number of Hours of Feedback**		
	0-5 hours	6-20 hours	21-60 hours
Survey Question 12: In your opinion, estimate the total contact with the instructor related to feedback on your performance of orthopedic assessment skills in class or out of class time (over the course of a semester/term):	16	29	13
Survey Question 13: In your opinion, estimate the total contact with the practicum supervisor/preceptor related to feedback on your performance of orthopedic assessment skills out of class time (over the course of a semester/term):	20	25	10
Total	36	54	23

Table 4.11 Descriptive statistics related to instructor and clinical supervisor quality of feedback as perceived by students

Survey Questions	**Qualitative Ranking**				
	Poor	Below Average	Average	Above Average	Out-standing
Survey Question 14: I would rank the feedback provided to me by my instructor as:	1	2	14	30	10
Survey Question 15: I would rank the feedback provided to me by my practicum supervisor/preceptor as:	4	4	16	24	5
Total	5	6	30	54	15

Retrospective Analysis with Technology Adoption

A retrospective analysis for technology adoption was completed because the treatment group that was given the PDA's (group 4) did not participate in the final testing. Initially, upon commencement of the study in September, 2006, there was complete support with 100 percent of the students, faculty and supervisors stating they were interested in participating in the study. Student, instructors and clinical supervisors were orientated on the use of the tool in September, 2006. In April, 2007 during the final testing period, no students agreed to participate in the study.

The total number of hits from the PDA treatment group was eight for 19 students (i.e., who stated their interest in participation). In contrast, the total number of hits for the other group who had access to the e-SOAT (i.e., group 3: Mount Royal-University of Calgary) was 497 for 24 students who participated in the study.

The retrospective analysis consisted of a post-study questionnaire (Appendix H) that underwent an additional ethics committee review and approval. Qualitatively, results indicated the SOAT, as a tool was well accepted as a measure of orthopedic assessment clinical competence (validity). One hundred percent of the five instructors who responded to the post-study survey made positive comments about the SOAT to measure this construct. In contrast, of the four respondents who had access the e-SOAT, 100 percent stated at least one negative comment about the technology.

It was interesting to note that the group that received the PDA's (group 4) did not use them very often due to logistical problems associated with the PDA's themselves (i.e., battery life, wi-fi network problems, screen size) rather than the e-SOAT site. There

were two primary people at this institution who agreed to participate as instructors or clinical supervisors. The instructor had zero hits on the e-SOAT site while the clinical supervisor has 19 hits over an eight month period. In contrast, the instructor from group 3 had 17 hits while the supervisor data was missing. The instructor and supervisor from group 4 complained that the battery life was too short to perform an entire assessment and the screen size was too small to view the assessment tool. Moreover, the groups that had access to the e-SOAT (groups three and four) site commented that the technology was a burden. In contrast, the instructor from group 2 (Concordia University) projected that they would have preferred a lap top to grade the students when testing them in the final examination.

CHAPTER FIVE: DISCUSSION

Discussion pertaining to the results is separated into three phases of research completed. In addition, limitations in study design and analysis have also been separated by each phase of research. A discussion for all three phases of research in their totality is summarized at the end of this chapter.

Phase I –SOAT Content Validation

Content validation is a crucial step in overall creation of a valid and reliable tool to measure orthopedic assessment clinical skills (Hopkins, 1998). Content validation of the SOAT followed a modified Ebel procedure (Ebel et al., 1986; Violato et al., 2002; Butterwick et al., 2006) but also added a third stage of consensus building in order to create a more valid tool that could measure clinical orthopedic skills at multiple levels of expertise. It is quite common for assessment instruments to go through many iterations to improve validity, particularly after more expert feedback or reliability analysis has been completed (Austin et al., 2003; Palarm et al., 2004). Even though the beta versions of the SOAT were tested over a long period of time locally (in Calgary, Alberta, Canada), consensus was not strong enough at a national level and hence, further development had to take place. More stages in the content validation process help build stronger evidence for measuring the underlying construct even though traditional descriptions of the Ebel procedure did not describe these additional stages (Butterwick et al., 2006; Cantor, 1989). The additional steps also support the notion that classroom assessment typically has great face validity, but their validity across a wider audience is much poorer unless it has

undergone further scientific study (Violato et al., 1998; Hopkins, 1998; Krathwohl, 1998).

There are many who believe that attempting to measure clinical competence through performance based examinations trivializes the complexities of clinical competence (Joorabchi & Devries, 1996; Cox, 2000; Norman, Tugwell, Feightner, Muzzin, & Jacoby, 1985; Hodges, 2003; Hodges et al., 1999). Contributing to the debate is the appropriateness of using detailed checklists as opposed to global rating scales to measure the clinical competence construct (Van der Vleuten et al., 1991; Norman et al., 1991; Reznick et al., 1998; Van Luijk et al., 1990; Regehr et al., 1998). This debate was considered throughout the creation of a content-valid SOAT.

The SOAT evolved from an earlier conception with checklists that only measured a deductive reasoning diagnostic approach to a tool that permitted greater flexibility in the choice and order of the assessment protocol. The result of this evolution was the creation of a tool that is robust enough to permit either a deductive reasoning or an inductive reasoning approach from the student (or practitioner) while still measuring the same construct of clinical competence along a novice-expertise continuum (Boshuizen, 2004). The tool accomplishes this by adding a "clinical reasoning" metric that attempts to link each of the major components of an orthopedic assessment method together. By going through the rigorous content validation process (including adding a third stage to the Ebel procedure), a tool was created and should permit flexibility in student's decisions throughout their orthopedic evaluation thus leading to the possibility of higher fidelity at multiple levels of expertise. Testing varying levels of expertise was not the focus throughout the next two phases of research since the cohorts that were tested were

both third year athletic therapy students. Nonetheless, the design of the SOAT is such that it may address some of the issues and concerns that OSCE checklists have a tendency to trivialize the construct being measured (Van der Vleuten et al., 1991; Norman et al., 1991).

Limitations in Phase I

The expert panel struggled with the lack of definition and clear boundaries between the terms essential, important and less important. This lack of clarity and definition seemed to prevent 100% consensus with all tasks. In fact, one person questioned their own reliability from one moment to the next. As a result, the final SOAT task list used a binary scale for inclusion or exclusion rather than a scale that split the vote between essential, important and less important. This variation of the Ebel procedure was justified by the researchers because the original intent of the Ebel procedure was to establish a minimal passing score in performance based examinations and not to establish the inclusion or exclusion of tasks which was the original goal of this study. In the future, if the SOAT was to be used as a tool to measure a minimal passing score for orthopedic assessment, then an expert panel would need to measure each task for the level of importance and difficulty and a mathematical equation which weighted each task would be assigned to provide a passing test score.

A third phase of content validation was added which breaks from the traditional method described by Violato et al (2002) and Butterwick et al (2006). However, this seemed to be necessary to make the tool truly content-valid. During this process, two of the experts on the panel dropped out of the study due to time commitment. It may not be

evident from the consensus statistics, but in the third stage of the validation process, the consensus figures were reported based on 10 rather than 12 experts.

The intended use of the content-validated SOAT is quite different than the traditional OSCE originally described by Harden et al. (1979). Traditional OSCE's have a number of stations that, combined, are supposed to represent an underlying construct. The intention for the SOAT is to combine stations (history, observation, physical examination, etc.) into one station. Instead of three 10-min stations, the students would be permitted 30 min in total to go through all the steps necessary to make a final diagnosis/conclusion. This could be considered a significant paradigm shift in performance-based, practical assessments as it relates to measuring clinical competence, particularly as compared to traditional OSCE's.

Phase II – Initial Reliability of the SOAT

OSCE's were originally created in order to objectively measure students in performance-based, practical examinations and thus, created a more reliable measure (Harden et al., 1979; Harden et al., 1975). As was shown by a series of studies (Van Luijk et al., 1990; Martin et al., 1997; Martin et al., 2000; Ault et al., 2001; Reznick et al., 1997), the most reliable model to measure an underlying construct is a combination of both binary/dichotomous and continuous scales. The SOAT scales were added in this phase of the research project by identifying the characteristics of those tasks and labelling them as either continuous or dichotomous. The final product was a tool that was a hybrid of both dichotomous and continuous based on the task characteristic. This may be one of the explanations of the good reliability results in the second phase of the research.

The initial results indicate good inter-rater reliability for the SOAT for the specific scenarios that were used. Most psychometricians agree that in order to use a measurement instrument for high stakes examinations, a reliability coefficient of .70 or greater should be targeted (Streiner et al., 2003). The ankle assessments had the highest reliability coefficients ($\alpha = 0.91$) followed by the knee ($\alpha = 0.83$) and shoulder ($\alpha = 0.82$) when SPs are included in the grading process. The reliability of the ankle assessment may have been highest because it has the least number of special tests and likely the least complex scenario compared to the shoulder and knee joints. However, overall the reliability for each site specific SOAT was good enough to continue to study the tool and its impact on learning. The final phase of this research project focused on the shoulder

and knee regions with different scenarios and thus, provides further evidence for the generalizability of the SOAT to new cases.

Many have criticized traditional OSCEs stating they have a tendency to trivialize the underlying construct they attempt to measure thus, calling into question the overall validity (Norman et al., 1991; Van der Vleuten et al., 1991; Van Luijk et al., 1990). In contrast, a rater using the SOAT has the ability to mark the examinee differently on various tasks within the SOAT yet grade them globally on general skills. In other words, the SOAT requires expert judgement when grading examinees and changes how raters evaluate each individual examinee relative to another. For example, if one examinee decided to use a Lachman special test for anterior cruciate ligament stability, but another student chose to use an anterior drawer special test, both could be considered correct if applied appropriately. This level of discretion is often required in clinical environment for clinicians to accurately diagnose pathology or a lesion. As a result, this type of flexibility in measurement may also get closer to measuring the pinnacle of clinical competence or as Miller (1990) stated, “does” (Figure 2.1).

Conceptually, the SOAT has a slightly different approach than was proposed by Denegar and Fraser (2006). They recommended evidence-based decisions for special tests to be employed in a physical examination procedure (Denegar & Fraser, 2006). In fact, some meta-analyses have proposed superiority of sensitivity and specificity for some special tests over others in the knee, for example (Solomon, Simuel, Bates, Katz, & Schaffer, 2001; Benjaminse, Gokeler, & van der Schans, 2006). However, even those meta-analyses are problematic since they do not account for multiple ligament or tissue damage which is often the case in real life. However, even if one special test

demonstrated superior diagnostic power over another, determining how each individual special test fits into an overall orthopedic assessment procedure remains unexplored in the literature until the current study. Ideally, a recognized standardized orthopedic assessment protocol should be developed for the evaluation of all patients with knee injuries. However, those protocols will likely be limited by the multitude of factors associated with each individual case (types of tissue damage, degree of tissue damage, individual response to injury, etc.). In summary, the SOAT was designed to address some of the concerns that OSCEs tend to trivialize the content and brings into question the measurements' overall reliability and validity (Norman et al., 1991; Van der Vleuten et al., 1991; Van Luijk et al., 1990).

The strong SOAT reliability may be attributable to one or more of the following: the extensive content validation steps taken initially in the rating scale development process employed; and/or the establishment of the grading scheme for the SOAT; and/or the thorough training sessions provided to the raters. Performance based examinations do not always report the validation process employed which leads readers to infer this step in the overall validation may not have been established prior to measuring the reliability of the tools used in the assessment of examinees' abilities. Ignoring the initial content validation phase in the development of competency measures may result in lower internal consistency and hence, overall scale reliability (Hopkins, 1998). In contrast, good reliability does not ensure good validity. A balance and constant interaction between validity and reliability is critical to an instrument's construct validity evolution (Violato et al., 1998).

A final explanation of the tool's reliability could be the extensive training required for each rater and SP prior to the examination. Part of the training process included a review of the rules associated with the SOAT (Table 3.2). The examiner/rater training session was standardized through the use of a common Microsoft PowerPoint © presentation and a set of detailed instructions on how the tool should be used during the examination (Appendix C). The training session was interactive, permitting rater trainees to ask questions and gain clarification based on the specific scenarios. Although the training session took approximately three hours to complete, the explicit nature of the training on the procedures for using the SOAT may have contributed to the strong reliability coefficients obtained.

There seems to be room for improvement and perhaps reliability of each randomized block scheduled and the types of raters identified to participate in the examination process should be a consideration. For example, the two blocks with relatively low overall reliability included raters with the highest mean number of years of practical experience as certified athletic therapists in Canada. One explanation of this trend may include a specific interaction with the scenario/case examined. In the case of shoulder and knee, both injuries were bursitis diagnoses. Typically, special tests are inconclusive and thus, irrelevant within the physical examination for bursitis diagnoses. Although not conclusive, further information about the effects of raters' experiences, demographic and psychographic information, the interaction between this data and the scenario as well as the use of rating/global scales for assessment should be collected and analyzed in future studies so as to effectively determine the impact on the reliability.

Limitations in Phase II

The total number of participants in this study was limited to students enrolled in the clinical practicum class at Mount Royal College in the winter semester of 2006. This convenience sample was too small (n = 32) and the research design was not fully crossed. Therefore, it was inappropriate to perform a generalizability theory statistical analysis. As a result, Cronbach's alpha was used to obtain a baseline of the reliability of the tool. It is acknowledged that Cronbach's alpha alone may be insufficient as a measure of reliability for performance-based examinations (Brennan, 2000).

Phase III – Integration of the SOAT in Various Canadian Athletic Therapy Curricula

SOAT Psychometric Properties

Performance-based, practical, OSCE-type examinations require valid and reliable instruments to measure student performance (Streiner et al., 2003). In the two preceding phases of research, the SOAT underwent content validation and initial reliability testing. Analysis demonstrated good reliability for the SOAT across examiners, students and cases. The third phase of research was the second study where the reliability of the SOAT for testing the shoulder and knee regions remained stable and strong. Data from all institutions in the final phase of research were combined and the reliability coefficient for the knee and shoulder regions were .93 and .90, respectively. Stable reliability over time with multiple examiners/raters, examination sites, standardized patients, students and cases builds a stronger argument for construct validity of the SOAT to measure its underlying construct: clinical competence with the orthopedic assessment skills.

Further, generalizability analysis of treatment group 2 (Concordia University) demonstrated strong reliability across cases/scenarios and raters. The strong generalizability coefficients ($Ep^2_{(rater)} = .73$; $Ep^2_{(scenario)} = .89$) indicate the SOAT may be generalized across a wider sample of raters and scenarios. Moreover, future research with the SOAT should expect relatively stable reliability for new raters and scenarios other than the knee and shoulder diagnoses that were part of the third phase of this study. Generalizability beyond the scenarios tested with the Concordia is further evidence since the first reliability study (phase II) also consisted of different scenarios with good reliability. Further, results indicate reliability may also be generalized across other body regions such as the ankle, elbow or wrist regions. This is possible since initial content validation study (phase I) designed the SOAT for all body regions. Future research with a fully crossed study would best test this theory.

Scale development takes time, resources and a well structured research plan (Violato et al., 1998; DeVellis, 2003). The third phase of this research project was the third step in the research plan designed to create a valid and reliable assessment tool: the SOAT. Strong, stable reliability over time can be attributable to following a structured plan consisting of content validation followed by reliability analysis (Hopkins, 1998).

SOAT's Impact on Learning

Comparison of quasi-experimental and comparison groups in the third phase of this thesis demonstrated a significant difference for the group of students who were explicitly exposed to the SOAT. Students who were implicitly exposed to the SOAT (group 2) by the instructors performed equally as well (or poorly) as the comparison group (group 1). Potential implications for instructional methods support the need to expose students to the

assessment methods and procedures as part of the learning process if specific learning objectives and competencies are required by accrediting bodies. The "number of courses taken in their post-secondary career" covariate was not a major factor in explaining the variance between groups. However, this analysis may not be completely accurate based on the violations of assumption adoption of the ANCOVA (Appendix K). Further complicating matters, there was a significant correlation between this covariate and the student scores (dependent variable) which originally lead to the ANCOVA being employed. This variable should be considered in future studies considering the conflicting evidence in this study.

Teaching, learning and assessment are difficult concepts to separate and prioritize when considering instructional strategies (Bowen et al., 2002). This challenge is particularly true with teaching clinical skills in human service professions such as medicine or athletic therapy (Spencer, 2003). Earl (2003) proposed a model whereby assessment is employed as a modality for teaching and learning. Earl (2003) reluctantly recognized that all types of assessment have their place in learning and that understanding how and when to apply them is the critical factor. In her words, instructors must "recognize the inevitable contradictions among them" (Earl, 2003, p.22). With this contradiction in mind, Earl (2003) suggested a re-emphasis to assessment AS learning and thus moved it from the peak to the base of the assessment pyramid (Figure 5.1). To challenge this position, a reframed paradigm is proposed in Figure 5.2: *recognize the inevitable connections among them and use those* ***connections*** *to facilitate learning*.

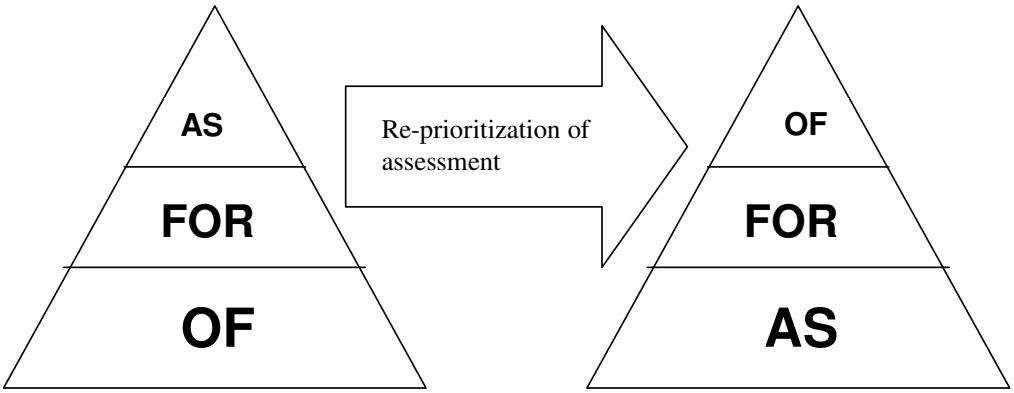


Figure 0.1. Earl's Model of Assessment AS, FOR and OF Learning. The re-prioritization.

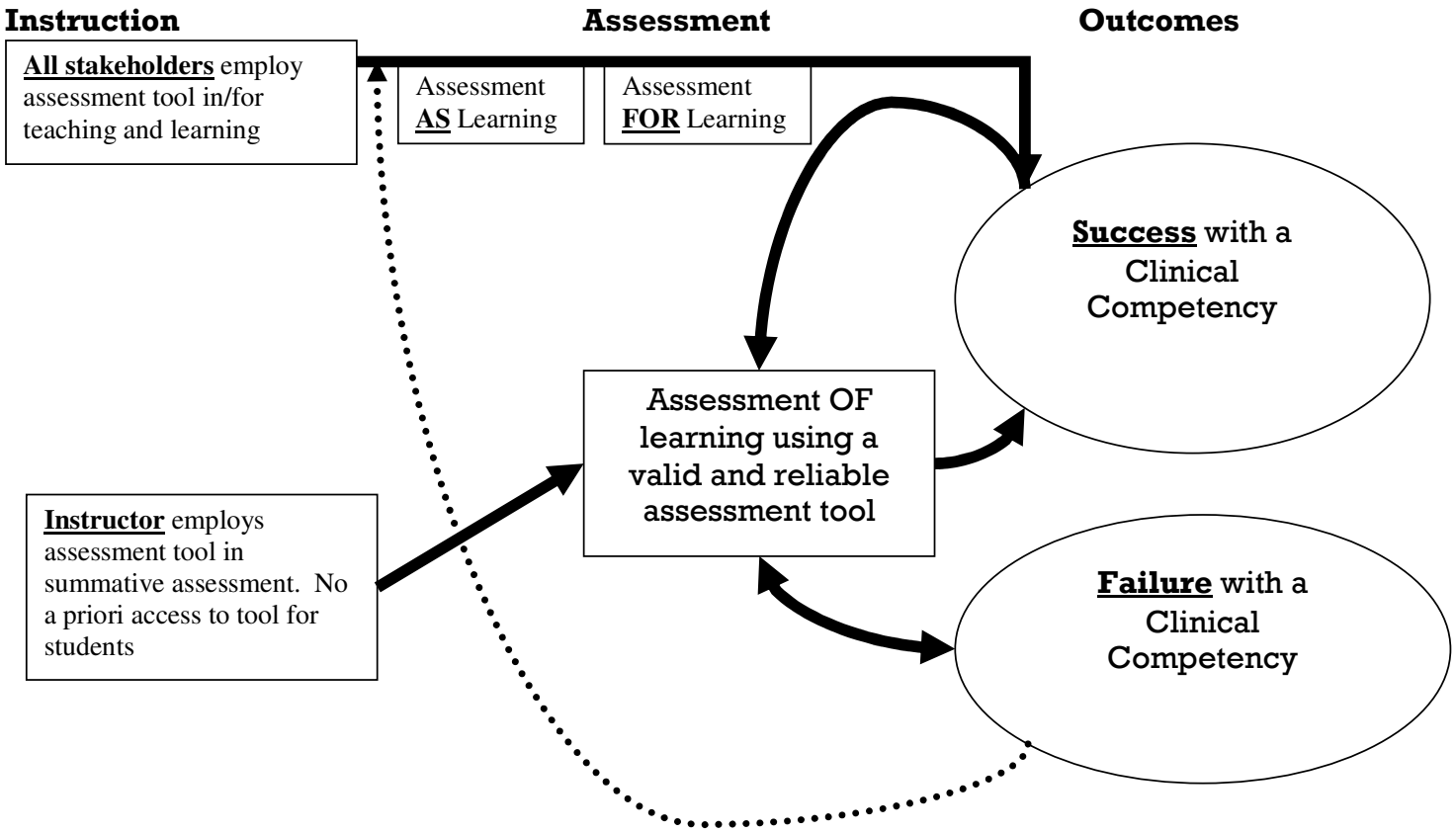


Figure 0.2. Predictive Learning Assessment Model (PLAM)

The PLAM illustrates that students who use an assessment tool that has undergone rigorous psychometric testing can use the tool to learn the underlying construct (i.e., assessment as learning). An important factor in the model is that instructors must refer to the assessment tool in their teaching methods (i.e., assessment for learning) and they must employ those same assessment tools in final, summative examinations (i.e., assessment of learning). This model also helps support the notion of assessment as a lever to assist learning (Norton, 2004). Study results from the SOAT integration (Phase III) study help support the creation and introduction of this model.

The SOAT was designed in such a way that instructors or clinical supervisors/preceptors could use it formatively to evaluate students' progress. The instructor could use the SOAT as a mechanism to diagnose the student's ability and provide feedback for future learning (i.e., assessment for learning) (Rezaei et al., 2002). The SOAT was designed in such a way that a student could refer to the evaluation of their performance from a clinical setting or mock scenario, reflect on it, and potentially, make change (i.e., assessment as learning). Finally, the SOAT was a tool that was used to assess students in a summative examination counting towards their final grade in a course (i.e., assessment of learning). It is possible that this is how the SOAT was used by group 3 (i.e., the treatment group who had access to it) which may account for the difference in scores when compared with group 1 (i.e., comparison group) or group 2 (i.e., the other treatment group). Ultimately, the pathway of learning outlined by a straight line to "success" illustrates greater success with the orthopedic assessment clinical competence prior to being formally tested. In comparison, those students who did not have access to the SOAT prior to being tested (i.e., groups one and two) with it

are more likely to fail the orthopedic assessment clinical competency. This theory was supported by the mean scores for treatment groups one and two (57.6% and 55.9% , respectively).

Once students have experienced the process of being tested by the SOAT, the PLAM illustrates a supplemental pathway of success or failure of the clinical competence. If a student is merely re-tested without exposure and feedback, then it may be possible they could fail once again (as indicated by the solid two-way arrow). However, if students are offered feedback with their performance through exposure to the assessment tool (the dotted line to assessment as and for learning), then they may have a greater chance of success in learning the underlying construct.

Educators have questioned the concept of "teaching to the exam" (Cohen-Schotanus, 1999). However, students live in a world that values and rewards grades which could explain why students focus on grades rather than learning itself (Thomson & Falchikov, 1998; Skinner, 1971; Cohen-Schotanus, 1999). Critics would say that teaching to the exam (and thus catering to student focus on grades) merely de-values the education and learning while it encourages student passivity (Knowles, 1978; Sambell & McDowell, 1998). This epistemology is clearly based in a constructivist perspective of learning. However, to make learning explicit and avoid issues around "the hidden curriculum," it may be important to make objectives and expectations of student performance very clear, transparent and accessible (Cohen-Schotanus, 1999; Sambell et al., 1998). Sambell and McDowell (1998) defined the hidden curriculum as "the shadowy, ill-defined and amorphous nature of that which is implicit and embedded in educational experiences in contrast with the formal statements about curricula and the

surface features of education interaction" (p. 391). Technology and the e-SOAT enabled orthopedic assessment competencies and learning objectives to be explicit due to the accessibility of the tool while concomitantly employing the SOAT as lever to assist learning. In contrast, the students who were exposed to the SOAT implicitly (i.e., group 2 – Concordia University) performed at the same level on the summative performance-based exam as the group who did not have any exposure at all (group 1 – University of Winnipeg). In other words, even if the instructor knows how the students will be assessed, but does not provide explicit mechanisms for them to learn, the curriculum will remain hidden. Thus, the value of making learning objectives explicit can not be over-stated.

The Role of Feedback on Learning

Interestingly, even though the e-SOAT provided a mechanism for instructors or clinical supervisors to give feedback to students (in the case of group 3), feedback did not seem to be a covariate that could explain differences between group scores in the SOAT. Students from all groups seemingly had the perception of feedback (quality and quantity) that was quite positive (Figure 4.1 to 4.4 inclusively). Generally, the majority of students from all groups were quite satisfied with instructor and supervisor feedback as it related to the orthopedic assessment clinical competency. The majority of students felt they had an above average level of supervision or higher with 40 out of 57 for instructors (Question 14) and 29 out of 53 for clinical supervisors/preceptors. It is difficult to extrapolate any conclusions from these qualitative data, particularly since only one group (group 3) actually received feedback through the SOAT. Nonetheless, this is an

interesting area for future research, thus providing more support for the PLAM (Figure 5.2).

The Role of Technology with Learning Orthopedic Assessment Skills

Results indicated that the technology (e-SOAT) did not have a significant, direct impact on learning orthopedic assessment skills. This was supported due to the lack of correlation between the total number of e-SOAT website hits relative to the grade produced in the final performance based exam. Ideally, comparison of another treatment group who only had access to a paper version of the SOAT would help separate the learning that took place with the e-SOAT compared to a paper version of the SOAT.

The results of the retrospective survey circulated to instructors and supervisors (Appendix H) revealed limitations of the technology, but not necessarily the SOAT itself. All participants (five respondents) stated they liked the SOAT, but those participants who had access to the PDA's or the e-SOAT through a desktop computer were not happy with the technology and/or design. Group 4 instructors and clinical supervisors (i.e., University of Manitoba) liked the desktop version of the e-SOAT and preferred using it to assess students compared to using the PDA's. When applying the results to the "technology acceptance model", the perceived "ease of use" was a key component associated with PDA's adoption in this study (Davis, 1989). The technology acceptance model is quite prescriptive requiring ease of use for technology prior to perceived usefulness. Again, applying this model to the current study, it stands that even if the SOAT was perceived as useful, it would be a moot point if there was no ease-of-use (Davis, 1989). In fact, this was actually the case whereby instructors found the SOAT useful, but there was no ease of use, thus rendering it useless for the PDA group (i.e.,

group 4)! The lack of total website hits from group 4's instructor and clinical supervisors support this notion. Future research that employs the e-SOAT must address the technological and design issues in order to make it easier to use and increase the likelihood of adoption.

It was puzzling that the treatment group that received the Dell Axim X50 did not follow through with the testing. Participation was completely voluntary, so perhaps that is one explanation for lack of participation. Another potential explanation is that neither the instructor from this institution nor the students valued the SOAT, the e-SOAT and/or the PDAs. This explanation is supported by the paucity of internet hits on the e-SOAT website from this group, particularly with comparison to the other group that had access to the e-SOAT through their desktop computers only: 13 hits with a total n of 19 students compared to the treatment group number one who had 497 hits with 24 students. There was limited personal contact between the PI and the instructor from group 4 throughout the eight months when teaching the orthopedic assessment skills were taking place. In contrast, the instructor from group 3 had approximately ten personal conversations with the PI to gain clarification of the purpose and use of the SOAT.

Further contrast of the value placed on understanding and adoption of the SOAT by groups three and four can be found when comparing the total number of hits on the e-SOAT website. The instructor from group 4 (i.e., University of Manitoba) had zero hits on the e-SOAT website over an eight month period! The instructor from group 3 (Mount Royal/University of Calgary) had 23 hits on the e-SOAT website over an eight- month period. Comments written in the post-study survey from the group 4 instructor (University of Manitoba) lend further credence to the argument that they did not value the

SOAT or e-SOAT. The instructor could only reference a conversation that the PI had with him in the summative examination orientation meetings which was eight months after originally being introduced to the tool. The instructor espoused positive value of the SOAT as a teaching tool and went on to explain how he would integrate those concepts into his teaching in the future. This discovery was too late to have an impact on the results of the current study. Applying the technology adoption model, both perceived ease of use (the complexity of the technology or the SOAT itself) and the perceived usefulness (of the SOAT conceptually) were both lacking for this individual (Davis, 1989). Future studies that employ this technology should include a pre-screening that could evaluate the technology adoption or readiness prior to formal study.

Limitations in Phase III

The quasi-experimental was such that four athletic therapy programs were chosen based on similarity in their curricular design. In fact, there are only six athletic therapy programs in Canada. There was limited contact with all instructors who had access to the SOAT or e-SOAT. No attempts were made to seek those instructors out throughout the year. This lack of contact may have lead to a lower number of participants across all institutions, but was particularly true at the institution where no students volunteered.

A fully crossed generalizability design whereby all students were tested twice (i.e., two cases) by all raters and an SP would have been optimal to evaluate the SOAT's reliability. However, it was challenging to obtain volunteers for the amount of hours that it took to examine the students who did participate let alone doubling that number. In total, it took approximately 58 man hours to evaluate the students in this study. As a result, generalizability theory was completed, where possible, using a nested design.

When measuring the total number of hits on the e-SOAT website, a portal of entry system was employed. The Blackboard Learning Management System © at Mount Royal College in Calgary was used to track the total number of hits from a hyperlink created within this shell. However, the e-SOAT was actually housed on a different server with a completely separate URL. Any students with technical savvy would discover they could get to the e-SOAT site directly and skip the step of going through the Blackboard Learning Management System © shell. The PI requested that students avoid taking this shortcut for the study purposes, but it is possible that they entered the site more often than the statistics demonstrated. In addition, students may have printed the sheets from the e-SOAT, thus, skewing the statistics collected in the Blackboard Learning Management System © shell. Therefore, the true impact of the e-SOAT may not be accurate.

The positive results when comparing each group in the final phase of research may have alternative explanations vis a vis those hypothesized in this thesis. It is possible that since the groups were not randomly assigned to each of the comparison or treatment group that they may not be a representative sample in the population of athletic therapy students. However, a number of covariates were collected through the questionnaire and correlation analyses were computed. There was only one positive correlation with the dependent variable (SOAT exam score): total number of post-secondary courses completed. Other covariates such as grade point average, for example, did not have a significant impact on the final exam results. There may be other covariates that were not collected in this study that may have had an impact on the results as well.

SUMMARY AND CONCLUSIONS

Educational assessment is challenging, particularly when one is attempting to measure practical skills in performance based examinations (Rethans et al., 2002; Daelmans et al., 2004). The industry standards in medical education for performance based examinations has become the objective structured clinical examinations (OSCEs) and they have undergone considerable evolution since it was first reported over thirty years ago (Harden et al., 1975; Harden, 1990; Harden et al., 1979). The objectives of the current study were to create a valid and reliable, or in other words, a psychometrically sound instrument that could measure an orthopedic assessment clinical competency construct of athletic therapy students. Further, the study was designed to measure the impact of this tool on learning the same underlying clinical competency.

The first phase of research employed a modified Ebel procedure to create a content-valid Standardized Orthopedic Assessment Tool (SOAT). Experts from across Canada, representing a sample of graduate athletic therapists from each of the six nationally accredited programs in Canada, provided feedback on the importance of each task that should be included in an orthopedic assessment of a clinically competent individual. Traditionally, students are taught to employ a deductive approach in their history and physical examination. Usually, a prescriptive number of standard questions are asked, followed by observation and a physical exam. However, experts typically do not continue to employ this approach over time as they gain experience. Rather, experts typically use an inductive reasoning approach whereby their experience permits them to ask a few general questions which lead to other more relevant questions based on their "hunch" or index of suspicion (Boshuizen, 2004). The rest of the history and physical

examination proceed to confirm or deny those indices systematically. The SOAT was designed in such a way as to allow flexibility in the approach the student decides to take: a deductive or inductive approach. Deductive approaches would dictate an extremely prescribed list of tasks that must be completed, likely in a linear fashion. However, an inductive approach would permit the student to follow their own pathway of discovery to complete a diagnosis. Employment of an inductive or deductive approach may address concerns with validity and the OSCE (Hodges, 2003; Van der Vleuten et al., 1991; Norman et al., 1991). The SOAT may address validity with a unique combination of dichotomous tasks, global rating scales and a series of "clinical reasoning" tasks that link the various components of orthopedic clinical competence together (Hodges & McIlroy, 2003).

The SOAT was designed to increase the validity that has traditionally been missing in some OSCE formats and measurement instruments (Hodges, 2003; Van der Vleuten et al., 1991; Norman et al., 1991). The challenge is the apparent paradoxical relationship between validity and reliability with performance-based assessments (Brennan, 2001). In other words, if validity increases, does reliability have to decrease? Based on the results of reliability assessment across multiple sites, examiners/raters, SPs and cases in this study, it seems that it is possible to strike a balance between validity and reliability. This study demonstrated that the SOAT, by the nature of tool itself (described previously) and the process employed to develop it, has found a balance of validity and reliability.

Typically, evaluation instruments rarely go beyond face validation (Violato et al., 1998) which would, in turn, prevent them from being used universally across various

curricula. However, the SOAT did undergo more rigorous validation which permitted integration into various curricula in Canada. Validity and reliability testing were important initial steps and justification to integrate the as a teaching and learning tool into a variety of athletic therapy curricula. A significant difference between the treatment group (i.e., where the instructors and students had free access to the SOAT) and the remaining two groups (i.e., no exposure or comparison group and implicit exposure where only instructors had access) lends support to the proposed Predictive Learning Assessment Model (PLAM) introduced in Chapter Five. It was proposed that students who have an assessment tool that was used summatively to measure a construct, but also have access to the tool formatively (assessment as and for learning) may learn faster and more efficiently than students who did not have access to the tool. If the instructor was aware of the tool that will be used to assess student performance, but did not explicitly share it with students, learning was proposed to take much longer (i.e., lower efficiency). Moreover, if the tool has not undergone the psychometric development outlined in phases I and II of this study, the predictive relationship may be nullified. The mere nature of the validation process forces educators to be clear about the intended outcomes, thus, making the learning more clear and explicit.

There were a number of limitations that may prevent complete validation of this predictive model, but future research can improve on these limitations through continuous testing of the model. In addition, future research should attempt to statistically measure the predictive nature of the PLAM. Research designs that include specific correlations between causal factors in the successful performance must be measured separately so as to account for their overall contribution. Regardless of the

limitations within the current study design, the current study does provide a good foundation for future evolution and development of the model.

Reference List

Abraham, S. (1998). Gynaecological examination: a teaching package integrating assessment with learning. *Medical Education, 32,* 76-81.

Ault, G., Reznick, R., Macrae, H., Leadbetter, W., DaRosa, D., Joehl, R. et al. (2001). Exporting of technical skills evaluation technology to other sites. *The American Journal of Surgery, 182,* 254-256.

Austin, Z., O'Byrne, C., Pugsley, J., & Munoz, L. (2003). Development and validation processes for an objective structured clinical exam (OSCE) for entry to practice certification in pharmacy: the Canadian experience. *American Journal of Pharmaceutical Education, 67,* 1-8.

Barrows, H. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine, 68,* 443-453.

Benjaminse, A., Gokeler, A., & van der Schans, C. (2006). Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. *Journal of Orthopaedic Sports Physical Therapy, 36,* 267-288.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32,* 364.

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education and Research, 18,* 57.

Black, N. & Harden, R. (1986). Providing feedback to students on clinical skills by using the objective structured clinical examination. *Medical Education, 20,* 48-52.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Berkshire, England: Open University Press, McGraw Hill Education.

Boshuizen, H. (2004). Does practice make perfect? A slow and discontinuous process. In *Professional Learning: Gaps and transitions on the way from novice to expert.* (pp. 73-95). Springer Netherlands.

Bowen, J. & Irby, D. (2002). Assessing quality and costs of education in ambulatory settings: a review of literature. *Academic Medicine, 77,* 621-680.

Brennan, R. (1992). Generalizability Theory. *Educational Measurement: Issues and Practice, 11,* 27-34.

Brennan, R. (2000). Performance Assessment from the Perspective of Generalizability Theory. *Applied Psychological Measurement, 24,* 339-353.

Brennan, R. (2001). Some problems, pitfalls and paradoxes in educational measurement. *Educational Measurement: Issues and Practice, Winter,* 6-18.

Bridge, P., Musial, J., & Frank, R. (2003). Measurement practices: methods for developing content-valid student examinations. *Medical Teacher, 25,* 414-421.

Broudo, M., White, M., Rodenburg, D., Arseneau, R., Chalmers, A., Wright, J. et al. (1997). The effectiveness of interactive media as an instructional aid for learning basic clinical skills and knowledge. In A.Scherpbier, C. van der Vleuten, J. Rethans, & A. van der Steeg (Eds.), *Advances in Medical Education* (pp. 321-326). Dordrecht: Kluwer Academic Publishers.

Brown, G., Manogue, M., & Martin, M. (1999). The validity and reliability of an OSCE in dentistry. *European Journal of Dental Education, 3,* 117-125.

Butterwick, D., Paskevich, D., Vallevand, A., & Lafave, M. (2006). The Development of Content Valid Technical Skill Assessment Instruments for Athletic Taping Skills. *Journal of Allied Health, 35,* 149-157.

Caine, D., Caine, C., & Lindner, K. (1996). *Epidemiology of Sports Injuries.* Champaigne: Human Kinetics.

Canadian Athletic Therapists Association (2005). CATA Competencies. http://www.athletictherapy.org [On-line]. Available: http://www.athletictherapy.org

Cantor, JA. (1989). A validation of Ebel's method for performance standard setting through its application with comparison approaches to a selected criterion-referenced test. *Educational and Psychological Measurement, 49,* 709-721.

Carpenter, J. (1995). Cost analysis of OSCEs. *Academic Medicine, 70,* 828-833.

Carr, M., Reznick, R., & Brown, D. (1999). Comparison of computer assisted instruction and seminar instruction to acquire psychomotor and cognitive knowledge of epitaxis management. *Otolaryngology Head and Neck Surgery, 121,* 430-434.

Cavanaugh, S. (1991). Response to a legal challenge: five steps to defensible credentialing examinations. *Evaluation and the Health Professions, 14,* 13-40.

Chatenay, M., Maguire, T., Skakun, E., Chang, G., Cook, D., & Warnock, G. (1996). Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? *American Journal of Surgery, 172,* 366-372.

Coady, D., Kay, L., & Walker, D. (2003). Regional Musculoskeletal Examination. *Clinical Rheumatology, 9,* 67-71.

Cohen, R., Reznick, R., Taylor, B., Provan, J., & Rothman, A. (1990). Reliability and validity of the objective structured clinical examination in assessing surgical residents. *The American Journal of Surgery, 160,* 302-305.

Cohen-Schotanus, J. (1999). Student assessment and examination rules. *Medical Teacher, 21,* 318-321.

Cook, C., Heath, F., Thompson, R., & Thompson, B. (2001). Score reliability in web - or internet based surveys: unnumbered graphical rating scales versus Likert-type scales. *Educational and Psychological Measurement, 61,* 697-706.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104.

Cox, K. (2000). Examining and recording clinical performance; a critique and some recommendations. *Education for Health, 13,* 45-52.

Cronbach, L. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6,* 475-494.

Cronbach, L., Gleser, G., Harinder, N., & Rajaratnam, N. (1972). The multifacet concept of observational procedures. In *The dependability of behavioral measurements: thoery of generalizability for scores and profiles* (pp. 1-32). New York: John Wiley & Sons Inc.

Cusimano, M., Cohen, R., Rucker, W., Murnaghan, J., Kodama, R., & Reznick, R. (1994). A comparative analysis of the costs of administration of an OSCE. *Academic Medicine, 69,* 571-576.

Daelmans, H., Hoogenboom, R., Donker, A., Scherpbier, A., Stehaouwer, C., & Van der Vleuten, C. (2004). Effectiveness of clinical rotations as a learning environment for achieving competences. *Medical Teacher, 26*.

Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13,* 319.

Denegar, C. & Fraser, M. (2006). How useful are physical examination procedures? Understanding and applying likelihood ratios. *Journal of Athletic Training, 41,* 201-206.

DeVellis, R. (2003). *Scale Development: Theory and Applications*. (Second ed.) Thousand Oaks: Sage Publications.

Driscoll, M. (2005). *Psychology of Learning for Instruction*. (3 ed.) Toronto: Pearson.

Dupras, D. & Li, J. (1995). Use of an objective structured clinical examination to determine clinical competence. *Academic Medicine, 70,* 1029-1034.

Ebel, R. & Frisbie, D. (1986). *Essentials of Educational Measurement*. (4 ed.) Englewood Cliffs, NJ: Prentice Hall.

Evans, A., McKenna, C., & Oliver, M. (2005). Trainees' perspectives on the assessment and self assessment of surgical skills. *Assessment and Evaluation in Higher Education, 30,* 163-174.

Fox, R., Dacre, J., & Mclure, C. (2001). The impact of formal instruction in clinical examination skills on medical student performance - the example of peripheral nervous system examination. *Medical Education, 35,* 371-373.

Gagne, R. (1985). *The conditions of learning and theory of instruction*. (Fourth ed.) New York: Hold, Rinehart and Winston Inc.

Geddes, E. & Crowe, J. (1998). Peer rated objective structured clinical exam. *Physiotherapy Canada, Fall,* 268-274.

Goff, B., Lentz, G., Lee, D., Fenner, D., Morris, J., & Mandel, L. (2001). Development of a bench station objective structured assessment of technical skills. *Obstetrics and Gynecology, 98,* 412-416.

Goff, B., Lentz, G., Lee, D., Houmard, B., & Mandel, L. (2000). Development of an objective structured assessment of technical skills for obsetric and gynecology residents. *Obstetrics and Gynecology, 96,* 146-150.

Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education. *British Medical Journal, 322,* 40-44.

Griffith, M. (2006). Questions about the e-OSCE. Mark Lafave. 7-17-2006. Ref Type: Personal Communication

Gronlund, N. (1981). *Measurement and Evaluation in Teaching*. (Fourth ed.) New York: Macmillan Publishing.

Haladyna, T. (1994). A research agenda for licensing and certification testing validation studies. *Evaluation and the Health Professions, 17,* 242-256.

Hall, W., Violato, C., Lewkonia, R., Lockyer, J., Fidler, H., Toews, J. et al. (1999). Assessment of physician performance in Alberta: the Physician Achievement Review. *Canadian Medical Association Journal, 161,* 52-57.

Harden, R. (1990). The OSCE - A 15 year Retrospective. In I. Hart, R. Harden, & Des Marchais (Eds.), *Ottawa Conference* Montreal: Can Heal Publications.

Harden, R. & Gleeson, F. (1979). Assessment of clinical competence using an objective structured clinical exam (OSCE). *Medical Education, 13,* 41-54.

Harden, R., Stevenson, M., Downie, W., & Wilson, G. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal, 1,* 447-451.

Hayward, L. (2004). Integrating web-enhanced instruction inot a research methods course: examination of student experiences and perceived learning. *Journal of Physical Therapy Education, 18,* 54-63.

Hodges, B. (2003). Validity and the OSCE. *Medical Teacher, 25,* 250-254.

Hodges, B. & McIlroy, J. (2003). Analytic global OSCE ratings are sensitive to level of training. *Medical Education, 37,* 1012-1016.

Hodges, B., Regehr, G., Hanson, M., & McNaughton, N. (1998). Validation of an objective structured clinical examination in psychiatry. *Academic Medicine, 73,* 910-912.

Hodges, B., Regehr, G., McNaughton, N., Tiberius, R., & Hanson, M. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine, 74,* 1129-1134.

Hopkins, K. (1998). *Educational and Psychological Measurement and Evaluation*. (8 ed.) Toronto: Allyn & Bacon.

Jaques, D. (2003). ABC of teaching and learning in medicine: Teaching small groups. *British Medical Journal, 326,* 492-494.

Joorabchi, B. & Devries, J. (1996). Evaluation of clinical competence; the gap between expectation and performance. *Pediatrics, 97,* 179-186.

Kaufman, D., Mann, K., Muijtjens, A., & van der Vleuten, C. (2000). A comparison of standard setting procedures for an OSCE in undergraduate medical education. *Academic Medicine, 75,* 267-271.

Kay, L. & Walker, D. (1998). Improving musculoskeletal clinical skills teaching. A regionwide audit and intervention study. *British Medical Journal, 57,* 656-659.

Keane, D. (1990). The inadequacy of recent research on computer assisted instruction. *Pedagogue Perspectives on Health Sciences Education, 2,* 1-7.

Kerlinger, F. & Lee, H. (1986). *Foundations of Behavioral Research*. (3 ed.).

Knowles, MS. (1978). *The Adult Learner: A Neglected Species*. (2nd Edition ed.) Houston: Krieger Pub. Co.

Kramer, A., Jansen, J., Zuithoff, P., Dusman, H., Tan, L., Grol, R. et al. (2002). Predictive validity of written knowledge test of skills for an OSCE in postgraduate training for general practice. *Medical Education, 36,* 812-819.

Krathwohl, D. (1998). *Methods of Educational and Social Science Research*. (2 ed.) Long Grove: Waveland Press Inc.

LaDuca, A. (1994). Validation of professional licensure examinations. *Evaluation and the Health Professions, 17,* 178-197.

Lawson, D. (2002). Report to the profession. *Journal of the Canadian Chiropractic Association, 46,* 201-205.

Liebermann, D., Katz, L., Hughes, M., Bartlett, R., McClements, J., & Franks, I. (2002). Advances in the application of information technology to sports performance. *Journal of Sports Sciences, 20,* 755-769.

Likert, R. (1952). A technique for the development of attitude scales. *Educational and Psychological Measurement, 12,* 313-315.

Magee, D. (2002). *Orthopedic Physical Assessment*. (4th ed.) Toronto: Saunders.

Mager, R. (1997). *Measuring Instructional Results. How to find out if your instructional objectives have been achieved*. (Third Edition ed.) Centre for Effective Instruction Inc.

Martin, I., Stark, P., & Jolly, B. (2000). Benefiting from clinical experience: the influence of learning style and clinical experience on performance of an undergraduate objective structured clinical exam. *Medical Education, 34,* 530-534.

Martin, J., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchinson, C. et al. (1997). Objective structured assessment of technical skills (OSATS) for surgical residents. *British Journal of Surgery, 84,* 273-278.

Mavis, B., Turner, J., Lovell, K., & Wagner, D. (2006). Faculty, students and actors as standardized patients: expanding opportunities for performance assessment. *Teaching and Learning in Medicine, 18,* 130-136.

McKelvie, S. (1978). Graphical rating scales - how many categories? *British Journal of Psychology, 69,* 185-202.

McManus, I., Richards, P., Winder, B., & Sproston, K. (1998). Clinical experience, performance in final examinations nad learning style in medical students: prospective study. *British Medical Journal, 316,* 345-350.

Messick, S. (1998). Test Validity: A Matter of Consequence. *Social Indicators Research, 45,* 35-44.

Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65 Supplemental,* S63-S67.

Nayer, M. (1993). An overview of the objective structured clinical exam. *Physiotherapy Canada, 45,* 171-178.

Neufeld, V. & Norman, G. (1985). *Assessing Clinical Competence*. (vols. Volume 7) New York: Springer.

Newble, D., Hoare, J., & Elmslie, R. (1981). The validity and reliability of a new examination of the clinical competence of medical students. *Medical Education, 15,* 46-52.

Newble, D. & Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Medical Education, 17,* 165-171.

Norman, G., Tugwell, P., Feightner, W., Muzzin, L., & Jacoby, L. (1985). Knowledge and clinical problem solving. *Medical Teacher, 19,* 344-356.

Norman, G., van der Vleuten, C., & De Graaff, E. (1991). Pitfalls in the pursuit of objectivity; issues of validity, efficiency and acceptability. *Medical Education, 25,* 119-126.

Norton, L. (2004). Using assessment criteria as learning criteria: a case study in psychology. *Assessment and Evaluation in Higher Education, 29,* 687-702.

Palarm, T., Griffiths, M., & Phillips, R. (2004). The design, implementation and evaluation of electronic objective structured clinical examinations in diagnotic imaging: an "action" research strategy. *Journal of Diagnostic Radiography and Imaging, 5,* 79-87.

Pratt, D., Arseneau, R., & Collins, J. (2001). Theoretical foundations: reconsidering good teaching across the continuum of medical education. *Journal of Continuing Education in the Health Professions, 21,* 70-81.

Probert, C., Cahill, D., McCann, G., & Ben-Shlomo, Y. (2003). Traditional finals and OSCEs in predicting consultant and self reported clinical skills of PRHOs': a pilot study. *Medical Education, 37,* 597-602.

Rahman, S. (2001). Promoting learning outcomes in paediatrics through formative assessment. *Medical Teacher, 23,* 467-470.

Regehr, G., MacRae, H., Reznick, R., & Szalay, D. (1998). Comparing the Psychometric Properties of Checklist and Global Rating Scales for Assessing Performance on an OSCE-format Examination. *Academic Medicine, 73,* 993-997.

Rethans, J. J., Norcini, J. J., Bar+¦n-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T. et al. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical Education, 36,* 901-909.

Rezaei, A. & Katz, L. (2002). Using computer assisted instruction to compare the inventive model and the radical constructivist approach to teaching physics. *Journal of Science Education and Technology, 11,* 367-380.

Reznick, R., Regehr, G., Macrae, H., Martin, J., & McCulloch, W. (1997). Testing technical skill via an innovative "bench station" examination. *American Journal of Surgery, 172,* 226-230.

Reznick, R., Regehr, G., Yee, G., Rothman, A., Blackmore, D., & Dauphinee, D. (1998). High Stakes Examinations; What do we know about measurement? *Academic Medicine, 73,* S 97-S 99.

Rogers, D., Regehr, G., Yeh, K., & Howdieshell, T. (1998). Computer assisted learning versus a lecture and feedback seminar for teaching a basic surgical technical skill. *American Journal of Surgery, 175,* 508-510.

Rootenberg, J. (1992). Information technologies in US medical schools; clinical practices outpace academic applications. *Journal of the American Medical Association, 268,* 3106-3107.

Sambell, K. & McDowell, L. (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education, 23,* 391-402.

Schmidts, M. (2000). OSCE logistics - handheld computers replace checklists and provide automated feedback. *Medical Education, 34,* 947-958.

Shrout, P. & Fleiss, J. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Simon, S., Volkan, K., Hamann, C., Duffey, C., & Fletcher, S. (2002). The relationship between second year medical students' OSCE scores and USMLE Step 1 scores. *Medical Teacher, 24,* 535-539.

Skinner, B. (1971). *Beyond Freedom and Dignity*. New York: Alfred Knopf.

Sloan, D., Donnelly, M., Schwartz, R., Felts, J., Blue, A., & Strodel, W. (1996). The use of the objective structured clinical examination (OSCE) for evaluation and instruction in graduate medical education. *Journal of Surgical Research, 63,* 225-230.

Sloan, D., Donnelly, M., Schwartz, R., & Strodel, W. (1995). The objective structured clinical exam (OSCE): the new gold standard for evaluating resident performance. *Annals of Surgery, 222,* 735.

Smee, S. (2003). Skill based assessment. *British Medical Journal, 326,* 703-706.

Solomon, D., Simuel, D., Bates, D., Katz, J., & Schaffer, J. (2001). Does this patient have a torn meniscus or ligament of the knee? *JAMA, 286,* 1610-1621.

Spencer, J. (2003). Learning and teaching in clinical environment. *British Medical Journal, 326,* 591-594.

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103,* 667-680.

Streiner, D. (1993). A checklist for evaluating the usefulness of rating scales. *Canadian Journal of Psychiatry, 38,* 140-148.

Streiner, D. & Norman, G. (2003). *Health Measurement Scales*. (Third ed.) New York: Oxford University Press.

Summers, A., Rinehart, G., Simpson, D., & Redlich, P. (1999). Acquisition of surgical skills: a randomized trila of didactic, videotape and computerized based training. *Surgery, 126,* 330-336.

Svensson, E. (2000). Comparison of the quality of assessments using continuous and descrete ordinal rating scales. *Biometrical Journal, 42,* 417-434.

Swanson & Stillman, P. (1990). Use of standardized patients for teaching and assessing clinical skills. *Evaluation in Health Professions, 13,* 79-103.

Swanson, D., Clauser, B., & Case, S. (1999). Clinical skills assessment with standardized patients in high stakes tests: a framework for thinking about score precision, equating and security. *Advances in Health Science Education, 4,* 67-106.

Symonds, P. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology, 7,* 456-461.

Thomson, K. & Falchikov, N. (1998). "Full on until the sun comes out:" the effects of assessment on student approaches to studying. *Assessment and Evaluation in Higher Education, 23,* 379-390.

Thorndike, R. (1971). *Educational Measurement*. (Second ed.) Washington, DC: American Council on Education.

Treadwell, I. (2006). The usability of personal digital assistants (PDA's) for assessment of practical performance. *Medical Education, 40,* 855-861.

Van der Vleuten, C., Norman, G., & De Graaff, E. (1991). Pitfalls in the pursuit of objectivity: issues of reliability. Medical Education 25, 110-118.
Ref Type: Journal (Full)

Van Luijk, S. & Van der Vleuten, C. (1990). A comparison of checklists and rating scales in performance based testing. In I. Hart, R. Harden, & J. Des Marchais (Eds.), *Current Developments in Assessing Clinical Competence* Montreal: Can-Heal Publications.

Velleman, P. & Wilkinson, L. (1993). Nominal, Ordinal, Interval and Ratio Typologies Are Misleading. *The American Statistician, 47,* 65-72.

Violato, C., Marini, A., & McDougall, D. (1998). *Assessment of Classroom Learning*. Calgary, AB, Canada: Detselig Enterprises LTD.

Violato, C., Salami, L., & Muiznieks, S. (2002). Certification Examinations for Massage Therapists: A Psychometric Analysis. *Journal of Manipulative Physiological Therapeutics, 25,* 111-115.

Violato, C., Marini, A., & Lee, C. (2003). A validity study of expert judgement procedures for setting cutoff scores on high stakes credentialing examinations using cluster analysis. *Evaluation and the Health Professions, 26,* 59-72.

Vivekananda-Schmidt, P., Lewis, M., & Hassell, A. (2005). Cluster radomized controlled trial of the impact of a computer assisted learning package on the learning of musculoskeletal examination skills by undergraduate medical students. *Arthritis and Rheumatism, 53,* 764-771.

Waltz, C., Strickland, O., & Lenz, E. (2005). *Measurement in Nursing and Health Research*. (3 ed.) Springer Publishing Company.

Wass, V., van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet, 357,* 945-949.

Winckel, C., Reznick, R., Cohen, R., & Taylor, B. (1994). Reliability and construct validity of a structured technical skills assessment form. *The American Journal of Surgery., 167,* 423-427.

Deleted: ¶
¶

Appendix A The SOAT for 8 body regions

Knee SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	**1**					
Heat		0	**1**					
Altered Function		0	**1**					
Redness		0	**1**					
Discoloration		0	**1**					
Lower Extremity:								
Pelvic Height		0	**1**					
ASIS		0	**1**					
PSIS		0	**1**					
Medial Malleoli		0	**1**					

Patellae	0	**1**					
Genu Valgum/Varum	0	**1**					
Tibial Valgum/Varum	0	**1**					
Foot - Static	0	**1**					
Foot - Dynamic (Gait)	0	**1**					
Symmetry							
Foot rotation/position	0	**1**					
Posture							
Observes from side/side	0	**1**					
Observes from rear/front	0	**1**					
Observes Pelvic Position/Tilt	0	**1**					
Scars, cuts, abrasions	0	**1**					
Other	0	**1**					
Other	0	**1**					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?	0	**1**					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?	0	**1**					
Overall Observation Global Rating Scale	0	1	2	3	4	5	
Clearing Joints Above and Below							
Did the student clear the joint **below** the lesion site?	0	**1**					
Did the student clear the joint **above** the lesion site?	0	**1**					
Did the student **change their index of suspicion** appropriately after clearing the joints above and below the lesion site?	0	**1**					
Scanning Exams							
Did the student ask inquire about **radicular pain** in the history?	0	**1**					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?	0	**1**					
Did the student perform **active ROM** in the **lower back**?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?	0	**1**					
Active ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**

External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	**1**					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	**1**					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plantarflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	**1**					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Anterior drawer	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Apley's distraction / compression	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Brush/sweep test	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Clarke's sign	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Dermatomes	0	**1**	**2**	**3**	**4**	**5**	**n/a**

Ely's	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Jerk test of hughston	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Kendall	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Lachman's	0	**1**	**2**	**3**	**4**	**5**	**n/a**
McMurray's	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Myotomes	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Obers	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Patellar Apprehension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Pivot Shift	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Posterior drawer	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Posterior sag	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Reflexes	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Slocum	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Straight leg raise	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Superior Tib-Fib Jt. Play	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Thomas	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Valgus stress	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Varus stress	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing special tests?	0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Overall Special Testing Global Rating Scale	0	1	2	3	4	5	
Palpation							
Adductor tubercle	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Anterior Superior Iliac Spine	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Biceps femoris and lateral gastrocnemius tendons	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Fat pad	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Fibular head	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Gastrocnemius muscle bellies	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Gerdy's tubercle/IT Band Insertion	0	**1**	**2**	**3**	**4**	**5**	**n/a**

Hamstring muscle bellies	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Iliotibial band	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Infrapatellar tendon	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Joint Play	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Lateral femoral condyle and epicondyle	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Lateral joint line	0	**1**	**2**	**3**	**4**	**5**	**n/a**
LCL	0	**1**	**2**	**3**	**4**	**5**	**n/a**
MCL	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Medial femoral condyle and epicondyle	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Medial hamstring and gastrocnemius	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Medial joint line	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Patella	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Patellar retinaculum	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Pes anserine insertion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plica	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Popliteal fossa	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Quadriceps (whole)	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Quadricepts (each individually)	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Sartorius	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Superficial bursa	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Suprapatellar pouch	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Suprapatellar tendon	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Tibial plateau	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Tibial tuberosity	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
other	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing palpations?	0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Overall Palpation Global Rating Scale	0	1	2	3	4	5	

Conclusion							
Is the conclusion/diagnosis correct?	0	**1**	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?	0	1					
Did the student refer to a physician for a complete diagnosis?	0	**1**					
Overall Impression of the Student's/Examinee's Assessment Skills							
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?	0	**1**	**2**	**3**	**4**	**5**	

Ankle SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	**1**					
Heat		0	**1**					
Altered Function		0	**1**					
Redness		0	**1**					
Discoloration		0	**1**					
Lower Extremity:								
Pelvic Height		0	**1**					
ASIS		0	**1**					
PSIS		0	**1**					
Medial Malleoli		0	**1**					

Patellae	0	**1**					
Genu Valgum/Varum	0	**1**					
Tibial Valgum/Varum	0	**1**					
Foot - Static	0	**1**					
Foot - Dynamic (Gait)	0	**1**					
Symmetry							
Foot rotation/position	0	**1**					
Posture							
Observes from side/side	0	**1**					
Observes from rear/front	0	**1**					
Observes Pelvic Position/Tilt	0	**1**					
Scars, cuts, abrasions	0	**1**					
Other	0	**1**					
Other	0	**1**					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?	0	**1**					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?	0	**1**					
Overall Observation Global Rating Scale	0	1	2	3	4	5	
Clearing Joints Above and Below							
Did the student clear the joint **below** the lesion site?	0	**1**					
Did the student clear the joint **above** the lesion site?	0	**1**					
Did the student **change their index of suspicion** appropriately after clearing the joints above and below the lesion site?	0	**1**					
Scanning Exams							
Did the student ask inquire about **radicular pain** in the history?	0	**1**					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?	0	**1**					
Did the student perform **active ROM** in the **lower back**?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?	0	**1**					
Active ROM							
Plantarflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Dorsiflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**

Inversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Eversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion appropriately** after completing **AROM**?	0	**1**					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Plantarflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Dorsiflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Inversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Eversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	**1**					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Plantarflexion - knee extended	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Dorsiflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plantarflexion - knee flexed	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Inversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Eversion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion appropriately** after completing **IR**?	0	**1**					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Anterior drawer	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Eversion stress test	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Hoffa's	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Homan's	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Inversion stress test (neutral)	0	**1**	**2**	**3**	**4**	**5**	**n/a**

Inversion stress with plantarflexion		0	1	2	3	4	5	n/a
Joint Play		0	1	2	3	4	5	n/a
Kleiger		0	1	2	3	4	5	n/a
Morton's		0	1	2	3	4	5	n/a
Percussion		0	1	2	3	4	5	n/a
Posterior Drawer		0	1	2	3	4	5	n/a
Potts compression		0	1	2	3	4	5	n/a
Syndesmosis		0	1	2	3	4	5	n/a
Talar tilt		0	1	2	3	4	5	n/a
Thompson test		0	1	2	3	4	5	n/a
Tinel's		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	1					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Achilles tendon		0	1	2	3	4	5	n/a
Anterior & posterior talofibular ligaments		0	1	2	3	4	5	n/a
Anterior dome of talus		0	1	2	3	4	5	n/a
Anterior tibialis muscle & tendon		0	1	2	3	4	5	n/a
Anterior tibiofibular ligament		0	1	2	3	4	5	n/a
Base of 5th metatarsal		0	1	2	3	4	5	n/a
Calcaneal tuberosity		0	1	2	3	4	5	n/a
Calcaneofibular ligament		0	1	2	3	4	5	n/a
Calcaneus		0	1	2	3	4	5	n/a
Cuboid		0	1	2	3	4	5	n/a
Cuneiforms		0	1	2	3	4	5	n/a
Deltoid ligament		0	1	2	3	4	5	n/a
Dorsalis pedis pulse		0	1	2	3	4	5	n/a
Extensor digitorum brevis		0	1	2	3	4	5	n/a

Extensor digitorum tendons		0	1	2	3	4	5	n/a
Extensor hallucis tendon		0	1	2	3	4	5	n/a
Fibula		0	1	2	3	4	5	n/a
First metatarsal		0	1	2	3	4	5	n/a
Flexor digitorum tendon		0	1	2	3	4	5	n/a
Flexor hallucis longus tendon		0	1	2	3	4	5	n/a
Gastrocnemius		0	1	2	3	4	5	n/a
Lateral malleolus		0	1	2	3	4	5	n/a
Medial border of tibia		0	1	2	3	4	5	n/a
Medial cuneiform		0	1	2	3	4	5	n/a
Medial malleolus		0	1	2	3	4	5	n/a
Metatarsal heads		0	1	2	3	4	5	n/a
Metatarsals & phalanges		0	1	2	3	4	5	n/a
Navicular tubercle		0	1	2	3	4	5	n/a
Peroneal muscle & tendons		0	1	2	3	4	5	n/a
Plantar fascia		0	1	2	3	4	5	n/a
Posterior tibial artery		0	1	2	3	4	5	n/a
Sesamoid bones		0	1	2	3	4	5	n/a
Sinus tarsus		0	1	2	3	4	5	n/a
Soleus		0	1	2	3	4	5	n/a
Tibial crest		0	1	2	3	4	5	n/a
Tibial shaft		0	1	2	3	4	5	n/a
Tibialis posterior tendon		0	1	2	3	4	5	n/a
Toes		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	1					
Did the student **change their index of suspicion appropriately** after completing palpations?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	1	2	3	4	5	

Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?		0	1					
Did the student refer to a physician for a complete diagnosis?		0	**1**					
Overall Impression of the Student's/Examinee's Assessment Skills								
If you had to provide an overall grade out of 10, what would you give to the student/ examinee?		0	**1**	**2**	**3**	**4**	**5**	

Hip SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	**1**					
Heat		0	**1**					
Altered Function		0	**1**					
Redness		0	**1**					
Discoloration		0	**1**					
Lower Extremity:								
Pelvic Height		0	**1**					
ASIS		0	**1**					
PSIS		0	**1**					
Medial Malleoli		0	**1**					

Patellae	0	**1**					
Genu Valgum/Varum	0	**1**					
Tibial Valgum/Varum	0	**1**					
Foot - Static	0	**1**					
Foot - Dynamic (Gait)	0	**1**					
Symmetry							
Foot rotation/position	0	**1**					
Posture							
Observes from side/side	0	**1**					
Observes from rear/front	0	**1**					
Observes Pelvic Position/Tilt	0	**1**					
Scars, cuts, abrasions	0	**1**					
Other	0	**1**					
Other	0	**1**					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?	0	**1**					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?	0	**1**					
Overall Observation Global Rating Scale	0	1	2	3	4	5	
Clearing Joints Above and Below							
Did the student clear the joint **below** the lesion site?	0	**1**					
Did the student clear the joint **above** the lesion site?	0	**1**					
Did the student **change their index of suspicion** appropriately after clearing the joints above and below the lesion site?	0	**1**					
Scanning Exams							
Did the student ask inquire about **radicular pain** in the history?	0	**1**					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?	0	**1**					
Did the student perform **active ROM** in the **lower back**?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?	0	**1**					
Active ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**

External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	**1**					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	**1**					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plantarflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	**1**					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Abdominal strength test	0	1	2	3	4	5	n/a
Dermatomes	0	1	2	3	4	5	n/a
Ely's	0	1	2	3	4	5	n/a
Fulcrum test	0	1	2	3	4	5	n/a
Gillet's Test	0	1	2	3	4	5	n/a

Gluteus medius Strength Test		0	1	2	3	4	5	n/a
Gluteus minimus Strength Test		0	1	2	3	4	5	n/a
Gracilis Strength Test		0	1	2	3	4	5	n/a
Hamstring - strength lateral (biceps femoris)		0	1	2	3	4	5	n/a
Hamstring - strength medial (semin tend.memb)		0	1	2	3	4	5	n/a
Hamstring tighness (SLR)		0	1	2	3	4	5	n/a
Hip joint play		0	1	2	3	4	5	n/a
Iliopsoas Strength Test		0	1	2	3	4	5	n/a
Kendall's		0	1	2	3	4	5	n/a
Long axis compression & distraction		0	1	2	3	4	5	n/a
Myotomes		0	1	2	3	4	5	n/a
Noble compression test		0	1	2	3	4	5	n/a
Ober's		0	1	2	3	4	5	n/a
Patrick's test		0	1	2	3	4	5	n/a
Piriformis muscle testing		0	1	2	3	4	5	n/a
Quadrant test		0	1	2	3	4	5	n/a
Reflexes L1-S2		0	1	2	3	4	5	n/a
Tensor Fascia latae		0	1	2	3	4	5	n/a
Thomas test		0	1	2	3	4	5	n/a
Tripod sign		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Abdominal muscles		0	1	2	3	4	5	n/a
Adductor longus		0	1	2	3	4	5	n/a
Adductor magnus		0	1	2	3	4	5	n/a
Anterior superior iliac spine		0	1	2	3	4	5	n/a

Femoral artery		0	1	2	3	4	5	n/a
Gluteus maximus		0	1	2	3	4	5	n/a
Gluteus medius		0	1	2	3	4	5	n/a
Gracilis		0	1	2	3	4	5	n/a
Greater trochanter		0	1	2	3	4	5	n/a
Hamstrings		0	1	2	3	4	5	n/a
Iliac crest		0	1	2	3	4	5	n/a
Iliopsoas		0	1	2	3	4	5	n/a
Inguinal ligament		0	1	2	3	4	5	n/a
Interior lateral angle of sacrum		0	1	2	3	4	5	n/a
Ischial Tuberosity		0	1	2	3	4	5	n/a
Joint Play		0	1	2	3	4	5	n/a
Pectineus		0	1	2	3	4	5	n/a
Piriformis		0	1	2	3	4	5	n/a
Posterior superior iliac spine		0	1	2	3	4	5	n/a
Rectus femoris		0	1	2	3	4	5	n/a
Sacral sulcus		0	1	2	3	4	5	n/a
Symphysis pubis		0	1	2	3	4	5	n/a
Vastus lateralis		0	1	2	3	4	5	n/a
Vastus medialis		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing palpations?		0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	**1**	**2**	**3**	**4**	**5**	

Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?	0	1					
Did the student refer to a physician for a complete diagnosis?	0	**1**					
Overall Impression of the Student's/Examinee's Assessment Skills							
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?	0	**1**	**2**	**3**	**4**	**5**	

Low Back SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	**1**					
Heat		0	**1**					
Altered Function		0	**1**					
Redness		0	**1**					
Discoloration		0	**1**					
Lower Extremity:								
Pelvic Height		0	**1**					
ASIS		0	**1**					
PSIS		0	**1**					
Medial Malleoli		0	**1**					

Patellae	0	**1**					
Genu Valgum/Varum	0	**1**					
Tibial Valgum/Varum	0	**1**					
Foot - Static	0	**1**					
Foot - Dynamic (Gait)	0	**1**					
Symmetry							
Foot rotation/position	0	**1**					
Posture							
Observes from side/side	0	**1**					
Observes from rear/front	0	**1**					
Observes Pelvic Position/Tilt	0	**1**					
Scars, cuts, abrasions	0	**1**					
Other	0	**1**					
Other	0	**1**					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?	0	**1**					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?	0	**1**					
Overall Observation Global Rating Scale	0	1	2	3	4	5	
Active ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	**1**					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**

Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	**1**					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left	0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plantarflexion	0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	**1**					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Approximation test	0	1	2	3	4	5	n/a
Bilateral straight leg raise	0	1	2	3	4	5	n/a
Bowstring	0	1	2	3	4	5	n/a
Braggard's	0	1	2	3	4	5	n/a
Brudzinski-kernig	0	1	2	3	4	5	n/a
Cephalad movement of ilium	0	1	2	3	4	5	n/a
Cephalad movement of sacrum	0	1	2	3	4	5	n/a
Dermatomes	0	1	2	3	4	5	n/a
Gluteus medius	0	1	2	3	4	5	n/a
Gluteus minimus	0	1	2	3	4	5	n/a
Goldthwait's test	0	1	2	3	4	5	n/a
Gracilis	0	1	2	3	4	5	n/a
Hibb's gapping test	0	1	2	3	4	5	n/a
Iliopsoas	0	1	2	3	4	5	n/a
Joint play - TVP	0	1	2	3	4	5	n/a
Joint play PA –CVP	0	1	2	3	4	5	n/a
Joint play PA-UVP	0	1	2	3	4	5	n/a
Laguere's	0	1	2	3	4	5	n/a
Lower abdominal test	0	1	2	3	4	5	n/a
McKenzie slide glide	0	1	2	3	4	5	n/a
Myotomes / Reflexes L1-S2	0	1	2	3	4	5	n/a
Naffziger's	0	1	2	3	4	5	n/a
One legged standing lumbar extension test	0	1	2	3	4	5	n/a

Pheasant		0	1	2	3	4	5	n/a
Piriformis muscle testing		0	1	2	3	4	5	n/a
Prone knee bending / Ely's		0	1	2	3	4	5	n/a
Quadrant test		0	1	2	3	4	5	n/a
Sacral apex pressure		0	1	2	3	4	5	n/a
Segmental instability		0	1	2	3	4	5	n/a
SI joint play		0	1	2	3	4	5	n/a
SI rocking		0	1	2	3	4	5	n/a
Squish tests		0	1	2	3	4	5	n/a
Straight leg raise		0	1	2	3	4	5	n/a
Supine gapping		0	1	2	3	4	5	n/a
Tensor Fascia latae		0	1	2	3	4	5	n/a
Trendelenberg		0	1	2	3	4	5	n/a
Valsalva		0	1	2	3	4	5	n/a
Yeoman's test		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Erector Spinae Muscle Group		0	1	2	3	4	5	n/a
External Oblique Muscle		0	1	2	3	4	5	n/a
Internal Oblique Muscle		0	1	2	3	4	5	n/a
Lumbosacral Junction		0	1	2	3	4	5	n/a
Quadratus Lumborum		0	1	2	3	4	5	n/a
Rectus Abdominis		0	1	2	3	4	5	n/a
Sacroiliac Joints		0	1	2	3	4	5	n/a
Spinous Processes of T12 to L5		0	1	2	3	4	5	n/a
Thoracolumbar Fascia		0	1	2	3	4	5	n/a
Transverse Processes of T12 to L5		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a

other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing palpations?		0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	**1**	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?		0	1					
Did the student refer to a physician for a complete diagnosis?		0	**1**					
Overall Impression of the Student's/Examinee's Assessment Skills								
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		0	**1**	**2**	**3**	**4**	**5**	

Shoulder SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	1					
Heat		0	1					
Altered Function		0	1					
Redness		0	1					
Discoloration		0	1					
Upper Extremity:								
Shoulder Height		0	1					
Clavicles		0	1					
AC (Step Deformity)		0	1					
Sulcus		0	1					

Scapulae		0	1					
Shoulders Rounded		0	1					
Symmetry								
Nose in line Sternum		0	1					
Head Tilt		0	1					
Deformity		0	1					
Scars, cuts, abrasions		0	1					
Posture								
Observes from side/side		0	1					
Observes from rear/front		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		0	1					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		0	1					
Overall Observation Global Rating Scale		0	1	2	3	4	5	
Clearing Joints Above and Below								
Did the student clear the joint **below** the lesion site?		0	1					
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		0	1					
Scanning Exams								
Did the student ask inquire about **radicular pain** in the history?		0	1					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		0	1					
Did the student perform **active ROM** in the **cervical spine**?		0	1					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?		0	1					
Active ROM								
Flexion		0	1	**2**	**3**	**4**	**5**	**n/a**
Extension		0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction		0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction		0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation		0	1	**2**	**3**	**4**	**5**	**n/a**

Internal Rotation	0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Scapular Protraction/Retraction	0	1	**2**	**3**	**4**	**5**	**n/a**
Scapular Elevation/Depression	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	1					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	1	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Scapular Protraction/Retraction	0	1	**2**	**3**	**4**	**5**	**n/a**
Scapular Elevation/Depression	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	1					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation	0	1	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation	0	1	**2**	**3**	**4**	**5**	**n/a**
External Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation (at 90 degrees)	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Horizontal Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**

Scapular Protraction/Retraction	0	1	**2**	**3**	**4**	**5**	**n/a**
Scapular Elevation/Depression	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1	**2**	**3**	**4**	**5**	
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	1					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
AC compression	0	1	**2**	**3**	**4**	**5**	**n/a**
AC shear	0	1	**2**	**3**	**4**	**5**	**n/a**
Active impingement	0	1	**2**	**3**	**4**	**5**	**n/a**
Anterior drawer	0	1	**2**	**3**	**4**	**5**	**n/a**
Apprehension & relocation	0	1	**2**	**3**	**4**	**5**	**n/a**
Clunk test	0	1	**2**	**3**	**4**	**5**	**n/a**
Drop arm	0	1	**2**	**3**	**4**	**5**	**n/a**
Empty can	0	1	**2**	**3**	**4**	**5**	**n/a**
Hawkin's Kennedy	0	1	**2**	**3**	**4**	**5**	**n/a**
Joint Play	0	1	**2**	**3**	**4**	**5**	**n/a**
Load & shift	0	1	**2**	**3**	**4**	**5**	**n/a**
Ludington's test	0	1	**2**	**3**	**4**	**5**	**n/a**
Neer impingement	0	1	**2**	**3**	**4**	**5**	**n/a**
Pec major contracture test	0	1	**2**	**3**	**4**	**5**	**n/a**
Posterior drawer	0	1	**2**	**3**	**4**	**5**	**n/a**
Speed's	0	1	**2**	**3**	**4**	**5**	**n/a**
Sulcus sign	0	1	**2**	**3**	**4**	**5**	**n/a**
Thoracic outlet syndrome tests	0	1	**2**	**3**	**4**	**5**	**n/a**
Winging Scapula	0	1	**2**	**3**	**4**	**5**	**n/a**
Yergason's	0	1	**2**	**3**	**4**	**5**	**n/a**
other	0	1	**2**	**3**	**4**	**5**	**n/a**
other	0	1	**2**	**3**	**4**	**5**	**n/a**
other	0	1	**2**	**3**	**4**	**5**	**n/a**
other	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicion?	0	1					
Did the student **change their index of suspicion** appropriately after completing special tests?	0	1					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	0	1					

Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
AC joint		0	1	**2**	**3**	**4**	**5**	**n/a**
Acromion process		0	1	**2**	**3**	**4**	**5**	**n/a**
Anterior serratus		0	1	**2**	**3**	**4**	**5**	**n/a**
Bicipital groove		0	1	**2**	**3**	**4**	**5**	**n/a**
Brachial artery		0	1	**2**	**3**	**4**	**5**	**n/a**
Clavicle		0	1	**2**	**3**	**4**	**5**	**n/a**
Coracoid process		0	1	**2**	**3**	**4**	**5**	**n/a**
Costicartilage		0	1	**2**	**3**	**4**	**5**	**n/a**
Greater tubercle		0	1	**2**	**3**	**4**	**5**	**n/a**
Inferior and superior angles of the scapula		0	1	**2**	**3**	**4**	**5**	**n/a**
Infraspinatus tendon		0	1	**2**	**3**	**4**	**5**	**n/a**
Latissimus dorsi		0	1	**2**	**3**	**4**	**5**	**n/a**
Lesser tubercle		0	1	**2**	**3**	**4**	**5**	**n/a**
Lower cervical spinous processes		0	1	**2**	**3**	**4**	**5**	**n/a**
Pectoralis major		0	1	**2**	**3**	**4**	**5**	**n/a**
Ribs		0	1	**2**	**3**	**4**	**5**	**n/a**
SC joint		0	1	**2**	**3**	**4**	**5**	**n/a**
Spine of the scapula		0	1	**2**	**3**	**4**	**5**	**n/a**
Sternum		0	1	**2**	**3**	**4**	**5**	**n/a**
Subacromial bursa		0	1	**2**	**3**	**4**	**5**	**n/a**
Subscapularis tendon		0	1	**2**	**3**	**4**	**5**	**n/a**
Supraspinatus tendon		0	1	**2**	**3**	**4**	**5**	**n/a**
Triceps tendon		0	1	**2**	**3**	**4**	**5**	**n/a**
Upper and middle thoracic spinous processes		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicion?		0	1					
Did the student **change their index of suspicion** appropriately after completing palpations?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					

Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	0	1					
Overall Palpation Global Rating Scale	0	1	2	3	4	5	
Conclusion							
Is the conclusion/diagnosis correct?	0	1	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?	0	1					
Did the student refer to a physician for a complete diagnosis?	0	1					
Overall Impression of the Student's/Examinee's Assessment Skills							
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?	0	1	**2**	**3**	**4**	**5**	

Elbow SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	1					
Heat		0	1					
Altered Function		0	1					
Redness		0	1					
Discoloration		0	1					
Upper Extremity:								
Shoulder Height		0	1					
Clavicles		0	1					
AC (Step Deformity)		0	1					
Sulcus		0	1					

Scapulae		0	1					
Shoulders Rounded		0	1					
Symmetry								
Nose in line Sternum		0	1					
Head Tilt		0	1					
Deformity		0	1					
Scars, cuts, abrasions		0	1					
Posture								
Observes from side/side		0	1					
Observes from rear/front		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		0	1					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		0	1					
Overall Observation Global Rating Scale		0	1	2	3	4	5	
Clearing Joints Above and Below								
Did the student clear the joint **below** the lesion site?		0	1					
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		0	1					
Scanning Exams								
Did the student ask inquire about **radicular pain** in the history?		0	1					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		0	1					
Did the student perform **active ROM** in the **cervical spine**?		0	1					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?		0	1					
Active ROM								
Flexion		0	1	**2**	**3**	**4**	**5**	**n/a**
Extension		0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation		0	1	**2**	**3**	**4**	**5**	**n/a**
Supination		0	1	**2**	**3**	**4**	**5**	**n/a**

Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	1					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation	0	1	**2**	**3**	**4**	**5**	**n/a**
Supination	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	1					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation	0	1	**2**	**3**	**4**	**5**	**n/a**
Supination	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	1					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Dermatome	0	1	**2**	**3**	**4**	**5**	**n/a**
Elbow flexion test	0	1	**2**	**3**	**4**	**5**	**n/a**
Elbow flexion test	0	1	**2**	**3**	**4**	**5**	**n/a**
Joint play	0	1	**2**	**3**	**4**	**5**	**n/a**
Lateral epicondylitis active	0	1	**2**	**3**	**4**	**5**	**n/a**
Lateral epicondylitis passive	0	1	**2**	**3**	**4**	**5**	**n/a**
Lateral epicondylitis test	0	1	**2**	**3**	**4**	**5**	**n/a**
Medial epicondylitis active	0	1	**2**	**3**	**4**	**5**	**n/a**
Medial epicondylitis passive	0	1	**2**	**3**	**4**	**5**	**n/a**
Myotome	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronator teres test	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronator teres test	0	1	**2**	**3**	**4**	**5**	**n/a**

Radioulnar joint stress test		0	1	**2**	**3**	**4**	**5**	**n/a**
Reflexes		0	1	**2**	**3**	**4**	**5**	**n/a**
Tinel's sign		0	1	**2**	**3**	**4**	**5**	**n/a**
Ulnar nerve entrapment		0	1	**2**	**3**	**4**	**5**	**n/a**
Valgus stress		0	1	**2**	**3**	**4**	**5**	**n/a**
Varus stress		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicion?		0	1					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Annular ligament		0	1	**2**	**3**	**4**	**5**	**n/a**
Biceps tendon		0	1	**2**	**3**	**4**	**5**	**n/a**
Brachial artery		0	1	**2**	**3**	**4**	**5**	**n/a**
Brachioradialis		0	1	**2**	**3**	**4**	**5**	**n/a**
Cubital Fossa		0	1	**2**	**3**	**4**	**5**	**n/a**
Extensor / supinator muscle mass & origin		0	1	**2**	**3**	**4**	**5**	**n/a**
Flexor / pronator muscle mass & origin		0	1	**2**	**3**	**4**	**5**	**n/a**
Lateral epicondyle		0	1	**2**	**3**	**4**	**5**	**n/a**
LCL		0	1	**2**	**3**	**4**	**5**	**n/a**
Medial epicondyle		0	1	**2**	**3**	**4**	**5**	**n/a**
Medial supracondylar ridge		0	1	**2**	**3**	**4**	**5**	**n/a**
Median nerve		0	1	**2**	**3**	**4**	**5**	**n/a**
Olecranon bursa		0	1	**2**	**3**	**4**	**5**	**n/a**
Olecranon process		0	1	**2**	**3**	**4**	**5**	**n/a**
Radial collateral ligament		0	1	**2**	**3**	**4**	**5**	**n/a**
Radial head		0	1	**2**	**3**	**4**	**5**	**n/a**
Triceps insertion		0	1	**2**	**3**	**4**	**5**	**n/a**
Ulnar collateral ligament		0	1	**2**	**3**	**4**	**5**	**n/a**
Ulnar nerve		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**

other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
other		0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicion?		0	1					
Did the student **change their index of suspicion** appropriately after completing palpations?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	1	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?		0	1					
Did the student refer to a physician for a complete diagnosis?		0	1					
Overall Impression of the Student's/Examinee's Assessment Skills								
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		0	1	**2**	**3**	**4**	**5**	

Forearm, Wrist and Hand SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	1					
Heat		0	1					
Altered Function		0	1					
Redness		0	1					
Discoloration		0	1					
Upper Extremity:								
Shoulder Height		0	1					
Clavicles		0	1					
AC (Step Deformity)		0	1					
Sulcus		0	1					

Scapulae		0	1					
Shoulders Rounded		0	1					
Symmetry								
Nose in line Sternum		0	1					
Head Tilt		0	1					
Deformity		0	1					
Scars, cuts, abrasions		0	1					
Posture								
Observes from side/side		0	1					
Observes from rear/front		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		0	1					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		0	1					
Overall Observation Global Rating Scale		0	1	2	3	4	5	
Clearing Joints Above and Below								
Did the student clear the joint **below** the lesion site?		0	1					
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		0	1					
Scanning Exams								
Did the student ask inquire about **radicular pain** in the history?		0	1					
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		0	1					
Did the student perform **active ROM** in the **cervical spine**?		0	1					
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?		0	1					
Active ROM								
Flexion		0	1	**2**	**3**	**4**	**5**	**n/a**
Extension		0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction		0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction		0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation		0	1	**2**	**3**	**4**	**5**	**n/a**

Supination	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **AROM**?	0	1					
Overall Active Range of Motion Global Rating Scale	0	1	2	3	4	5	
Passive ROM							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation	0	1	**2**	**3**	**4**	**5**	**n/a**
Supination	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1					
Did the student **change their index of suspicion** appropriately after completing **PROM**?	0	1					
Overall Passive Range of Motion Global Rating Scale	0	1	2	3	4	5	
Isometric Resisted							
Flexion	0	1	**2**	**3**	**4**	**5**	**n/a**
Extension	0	1	**2**	**3**	**4**	**5**	**n/a**
Adduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Abduction	0	1	**2**	**3**	**4**	**5**	**n/a**
Pronation	0	1	**2**	**3**	**4**	**5**	**n/a**
Supination	0	1	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?	0	1	**2**	**3**	**4**	**5**	
Did the student **change their index of suspicion** appropriately after completing **IR**?	0	1					
Overall Isometric Resisted Testing Global Rating Scale	0	1	2	3	4	5	
Special Tests							
Allen test (hand)	0	1	**2**	**3**	**4**	**5**	n/a
Axial compression of the 1st metacarpal	0	1	**2**	**3**	**4**	**5**	n/a
Bunnel-Littler test	0	1	**2**	**3**	**4**	**5**	n/a
Carpal & Metacarpal fracture test	0	1	**2**	**3**	**4**	**5**	n/a
Extensor hood rupture	0	1	**2**	**3**	**4**	**5**	n/a
Extensor tendon avulsion test	0	1	**2**	**3**	**4**	**5**	n/a
Finkelstein/DeQuervains	0	1	**2**	**3**	**4**	**5**	n/a

Flexor tendon avulsion test		0	1	**2**	**3**	**4**	**5**	n/a
Glide tests		0	1	**2**	**3**	**4**	**5**	n/a
Joint Play		0	1	**2**	**3**	**4**	**5**	n/a
Lunatotriquetral ballottement test		0	1	**2**	**3**	**4**	**5**	n/a
Murphy's Sign		0	1	**2**	**3**	**4**	**5**	n/a
Phalangeal fracture test		0	1	**2**	**3**	**4**	**5**	n/a
Phalen's		0	1	**2**	**3**	**4**	**5**	n/a
Thoracic Outlet Syndrome		0	1	**2**	**3**	**4**	**5**	n/a
Tinel's		0	1	**2**	**3**	**4**	**5**	n/a
Upper Limb Tension Test		0	1	**2**	**3**	**4**	**5**	n/a
Valgus stress test		0	1	**2**	**3**	**4**	**5**	n/a
Varus stress test		0	1	**2**	**3**	**4**	**5**	n/a
Watson's Test (Scaphoid instability)		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
Did the student specifically test structures that were identified in their previous index of suspicion?		0	1					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Abductor Pollicis		0	1	**2**	**3**	**4**	**5**	n/a
Anatomical snuff box		0	1	**2**	**3**	**4**	**5**	n/a
Carpi radialis longus and brevis		0	1	**2**	**3**	**4**	**5**	n/a
Dorsal hand		0	1	**2**	**3**	**4**	**5**	n/a
Extensor digiti minimi		0	1	**2**	**3**	**4**	**5**	n/a
Extensor indicis		0	1	**2**	**3**	**4**	**5**	n/a
Extensor Pollicis Brevis		0	1	**2**	**3**	**4**	**5**	n/a
Extensor Pollicis Longus		0	1	**2**	**3**	**4**	**5**	n/a
Flexor carpi radialis		0	1	**2**	**3**	**4**	**5**	n/a
Flexor carpi ulnaris		0	1	**2**	**3**	**4**	**5**	n/a
Flexor Pollicis Longus		0	1	**2**	**3**	**4**	**5**	n/a
Guyon's tunnel		0	1	**2**	**3**	**4**	**5**	n/a
Hook of the hamate		0	1	**2**	**3**	**4**	**5**	n/a

Hypothenar eminence		0	1	**2**	**3**	**4**	**5**	n/a
Lister's tubercle		0	1	**2**	**3**	**4**	**5**	n/a
Lunate		0	1	**2**	**3**	**4**	**5**	n/a
Metacarpals and metacarpal heads		0	1	**2**	**3**	**4**	**5**	n/a
Palmar hand		0	1	**2**	**3**	**4**	**5**	n/a
Palmaris longus		0	1	**2**	**3**	**4**	**5**	n/a
Pisiform		0	1	**2**	**3**	**4**	**5**	n/a
Radial styloid process		0	1	**2**	**3**	**4**	**5**	n/a
Scaphoid		0	1	**2**	**3**	**4**	**5**	n/a
Thenar eminence		0	1	**2**	**3**	**4**	**5**	n/a
Triquetrum		0	1	**2**	**3**	**4**	**5**	n/a
Ulnar styloid process		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
other		0	1	**2**	**3**	**4**	**5**	n/a
Did the student specifically test structures that were identified in their previous index of suspicion?		0	1					
Did the student **change their index of suspicion** appropriately after completing palpations?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	1					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	1					
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	1	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?		0	1					
Did the student refer to a physician for a complete diagnosis?		0	1					
Overall Impression of the Student's/Examinee's Assessment Skills								
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		0	1	**2**	**3**	**4**	**5**	

Neck SOAT

History Themes	Tasks							
Nature of Injury	Chief Complaint	0	**1**					
	Date of Injury/When Did Injury Occur	0	**1**					
	MOI/Recreation of MOI	0	**1**					
	Previous Injury to this Site	0	**1**					
	Previous Injury to joint above/below	0	**1**					
	Functional Limitations	0	**1**					
	History Back or Neck Injury/Pain	0	**1**					
	Subjective/patient SHARP	0	**1**					
	Unusual Sounds/Sensations	0	**1**					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information	Age	0	**1**					
	Primary Sport/occupation	0	**1**					
	Level of Sport	0	**1**					
	Training for Sport	0	**1**					
	Weight Training for Sport/occupation	0	**1**					
	Change in anthropometrics	0	**1**					
	Psycho/Emotional Stress	0	**1**					
	Training Surface	0	**1**					
	Equipment or Footwear	0	**1**					
	Activities of Daily Living	0	**1**					
	Other Sports/Activities	0	**1**					
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions	Pain: Pain Location	0	**1**					
	Pain: Describe Pain	0	**1**					
	Pain: Pain Scale	0	**1**					
	Pain: Relieves Pain	0	**1**					
	Pain: Exacerbates Pain	0	**1**					

SOAT, M. Lafave

	Pain: Morning or night Pain	0	**1**					
	Pain: Pain before, during or after activity	0	**1**					
	Pain: Radiating Pain Proximal or Distal	0	**1**					
	Pain: Increase with coughing, sneezing or bearing-down	0	**1**					
	Pain: Sleeping and Position	0	**1**					
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding	Treatment	0	**1**					
	Reviewed by Physician	0	**1**					
	X-rays/diagnostic imaging	0	**1**					
	Medications	0	**1**					
	Allergies	0	**1**					
	Goals for RETURN	0	**1**					
	Overall Health	0	**1**					
	Medical Conditions	0	**1**					
	Family History of Medical Conditions	0	**1**					
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		0	**1**					
Did the student **hone** into their **index of suspicion** in their history?		0	**1**					
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Swelling		0	1					
Heat		0	1					
Altered Function		0	1					
Redness		0	1					
Discoloration		0	1					
Upper Extremity:								
Shoulder Height		0	1					
Clavicles		0	1					
AC (Step Deformity)		0	1					
Sulcus		0	1					

Scapulae		0	1					
Shoulders Rounded		0	1					
Symmetry								
Nose in line Sternum		0	1					
Head Tilt		0	1					
Deformity		0	1					
Scars, cuts, abrasions		0	1					
Posture								
Observes from side/side		0	1					
Observes from rear/front		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Other		0	1					
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		0	**1**					
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		0	**1**					
Overall Observation Global Rating Scale		0	1	2	3	4	5	
Active ROM								
Flexion		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left		0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing **AROM**?		0	**1**					
Overall Active Range of Motion Global Rating Scale		0	1	2	3	4	5	
Passive ROM								
Flexion		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left		0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**

Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing **PROM**?		0	**1**					
Overall Passive Range of Motion Global Rating Scale		0	1	2	3	4	5	
Isometric Resisted								
Flexion		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Extension		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Right		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Side Bending Left		0	**1**	**2**	**3**	**4**	**5**	**n/a**
External Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Internal Rotation		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Plantarflexion		0	**1**	**2**	**3**	**4**	**5**	**n/a**
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing **IR**?		0	**1**					
Overall Isometric Resisted Testing Global Rating Scale		0	1	2	3	4	5	
Special Tests								
1st thoracic nerve root test		0	1	2	3	4	5	n/a
Bakody		0	1	2	3	4	5	n/a
Brachial plexus tension test		0	1	2	3	4	5	n/a
Cranial nerve function		0	1	2	3	4	5	n/a
Dermatome		0	1	2	3	4	5	n/a
Distraction		0	1	2	3	4	5	n/a
Foraminal compression		0	1	2	3	4	5	n/a
Hautants		0	1	2	3	4	5	n/a
Jackson		0	1	2	3	4	5	n/a
Joint play PA –CVP		0	1	2	3	4	5	n/a
Joint play PA –CVP Joint play - TVP		0	1	2	3	4	5	n/a
Joint play PA-UVP		0	1	2	3	4	5	n/a
Myotome		0	1	2	3	4	5	n/a
Passive scapular approximation		0	1	2	3	4	5	n/a
Reflexes C3-T1		0	1	2	3	4	5	n/a
Rib springing (anterior/posterior)		0	1	2	3	4	5	n/a
Shoulder depression		0	1	2	3	4	5	n/a
Slump test		0	1	2	3	4	5	n/a
Soto-Hall		0	1	2	3	4	5	n/a
Swallowing		0	1	2	3	4	5	n/a
ULTT		0	1	2	3	4	5	n/a
Vertebral artery test		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a

other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
other		0	1	2	3	4	5	n/a
Did the student specifically test structures that were identified in their previous index of suspicions?		0	**1**					
Did the student **change their index of suspicion** appropriately after completing special tests?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		0	**1**					
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Base of occiput								
Costovertebral junction								
Erector Spinae								
Facet joints								
Hyoid muscles								
Latissimus Dorsi								
Ribs								
Scalene muscles								
Scapula								
Semispinalis capitis								
Spinous processes								
Splenius capitis								
Sternocleidomastoid								
Sternocostal articulations								
Sternum								
Thoracolumbar fascia								
Transverse processes								
Trapezius muscle group								
other								
other								
other								
other								
other								
other								

Did the student specifically test structures that were identified in their previous index of suspicions?	0	**1**					
Did the student **change their index of suspicion** appropriately after completing palpations?	0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	0	**1**					
Overall Palpation Global Rating Scale	0	1	2	3	4	5	
Conclusion							
Is the conclusion/diagnosis correct?	0	**1**	**2**	**3**	**4**	**5**	
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?	0	1					
Did the student refer to a physician for a complete diagnosis?	0	**1**					
Overall Impression of the Student's/Examinee's Assessment Skills							
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?	0	**1**	**2**	**3**	**4**	**5**	

Appendix B SP answer key from phase II

Ankle: Inversion sprain

Examiner Name:

Student Name:

History		
Themes	**Tasks**	**Patient Information/Answer Key**
Nature of Injury	Chief Complaint	Sore ankle
	Date of Injury/When Did Injury Occur	Twisted yesterday when running on the grass
	MOI/Recreation of MOI	Stepped on uneven ground and twisted ankle.
	Previous Injury to this Site	Yes, I have sprained both ankles previously.
	Previous Injury to joint above/below	No other injuries other than sore knees every now and again…nothing I needed to see anyone about though.
	Functional Limitations	Can barely walk….significant limp. They should recognize you need to be on crutches.
	History Back or Neck Injury/Pain	None
	Subjective/patient SHARP	Swelled immediately with a big goose egg around the ankle bone. Warm and very tender to touch.
	Unusual Sounds/Sensations	Felt a crunching sensation
Patient Information	Age	21
	Primary Sport/occupation	Cross country running & running
	Level of Sport	Recreational
	Training for Sport	Run 5-6 days per week with varied intensities and times depending on how I feel.
	Weight Training for Sport/occupation	No weight training
	Change in anthropometrics	None
	Psycho/Emotional Stress	Nothing major
	Training Surface	Indoor track, indoor treadmill, outside on grass, side walks and into the back woods depending on the weather.
	Equipment or Footwear	Work with Gord's running room to get the right shoes every six months

	Activities of Daily Living	Can not do anything without a major limp since this happened.
	Other Sports/Activities	Hiking, biking periodically
Global Pain Questions	Pain: Pain Location	Lateral malleolus are (ATF, CF)
	Pain: Describe Pain	Felt sharp when it first happen, now it is just a dull ache with a really tight joint.
	Pain: Pain Scale	7/10 pain scale
	Pain: Relieves Pain	Not moving it..iced it once since yesterday
	Pain: Exacerbates Pain	Walking on it.
	Pain: Morning or night Pain	Really stiff when I woke up this AM.
	Pain: Pain before, during or after activity	n/a
	Pain: Radiating Pain Proximal or Distal	Mostly just around the ankle.
	Pain: Increase with coughing, sneezing or bearing-down	No
	Pain: Sleeping and Position	n/a
Safety/Rehab/Understanding	Treatment	Iced once
	Reviewed by Physician	No
	X-rays/diagnostic imaging	No
	Medications	No
	Allergies	Red Wine and cheese
	Goals for RETURN	ASAP
	Overall Health	Great
	Medical Conditions	None
	Family History of Medical Conditions	Father has high blood pressure and cholersterol.
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		Differential diagnosis may be: lateral ankle sprain, fracture, strained peroneals, talar fracture, tib-fib sprain, etc..
Did the student **hone** into their **index of suspicion** in their history?		
Observation		

Swelling		Should say that you see lots of swelling around the lateral malleolus and show the student.
Heat		Feels warm campared to other side.
Altered Function		Can barely walk
Redness		Seems red
Discoloration		None
Lower Extremity:		
Pelvic Height		WSWG
ASIS		WSWG
PSIS		WSWG
Patellae		WSWG
Genu Valgum/Varum		WSWG
Tibial Valgum/Varum		WSWG
Foot - Static		WSWG
Foot - Dynamic (Gait)		WSWG
Symmetry		
Foot rotation/position		WSWG
Posture		
Observes from side/side		WSWG
Observes from rear/front		WSWG
Observes Pelvic Position/Tilt		WSWG
Scars, cuts, abrasions		WSWG
Other		
Other		
Other		
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		
Did the student **add** any more anatomical structures to their **index of suspicion**?		
Clearing Joints Above and Below		
Did the student clear the joint **below** the lesion site?		Toes motion is normal…slight pain
Did the student clear the joint **above** the lesion site?		No pain with knee movement
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		
Scanning Exams		
Did the student ask inquire about **radicular pain** in the history?		No radicular pain

Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		Traumatic
Did the student perform **active ROM** in the **lower back**?		n/a because of traumatic MOI
Did the student **change their index of suspicion** after completing the **scanning exam**?		
Active ROM		
Plantarflexion		Painful and only approximately 50% of the good side
Dorsiflexion		Painful and only approximately 50% of the good side
External Rotation		Slight pain
Internal Rotation		Lots of lateral pain
Inversion		Extremely limited ROM (50%) and painful
Eversion		Slight pain and limited ROM
Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing **AROM**?		
Passive ROM		
Plantarflexion		Painful and only approximately 50% of the good side
Dorsiflexion		Painful and only approximately 50% of the good side
External Rotation		Slight pain
Internal Rotation		Lots of lateral pain
Inversion		Extremely limited ROM (50%) and painful
Eversion		Slight pain and limited ROM
Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing **PROM**?		
Isometric Resisted		
Plantarflexion - knee extended		No weakness only slight pain
Dorsiflexion		No weakness only slight pain
Plantarflexion - knee flexed		No weakness only slight pain
External Rotation		Slight weakness (4/5), slight pain
Internal Rotation		No weakness only slight pain
Inversion		Slight weakness (4/5), slight pain
Eversion		Weakness (3/5), moderate pain (5/10)

Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing **IR**?		
Special Tests		
Anterior drawer		Grade 2 laxity with extreme pain (8/10)
Eversion stress test		Slight pain, no laxity
Hoffa's		Negative
Homan's		Negative
Inversion stress with plantarflexion		Positive for pain (8/10) and laxity (empty end feel if they could get past the pain…but they can't
Kleiger/Syndesmosis		Pain (2/10), but no laxity
Morton's		Negative
Percussion		Negative
Posterior Drawer		Negative
Potts compression		Slightly painful (2/10) but negative
Talar tilt		Laxity and painful (8/10)
Thompson test		Negative
Tinel's		Negative
other		
other		
other		
other		
Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing special tests?		
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		
Palpation		
Achilles tendon		Feels like there is swelling back there
Anterior & posterior talofibular ligaments		ATF very sore (9/10)

Anterior dome of talus		Tender, but not bad (2/10)
Anterior tibialis muscle & tendon		
Anterior tibiofibular ligament		Normal
Base of 5th metatarsal		Normal
Calcaneal tuberosity		Normal
Calcaneofibular ligament		Very sore (9/10)
Calcaneus		Feel relatively normal outside ligamentous attachments on the lateral side
Cuboid		Normal
Cuneiforms		Normal
Deltoid ligament		Normal
Dorsalis pedis pulse		Normal
Extensor digitorum brevis		Normal
Extensor digitorum tendons		Normal, but slight swelling around them
Extensor hallucis tendon		Normal, but slight swelling around them
Fibula		Pain around lateral malleolus, but rest is normal
First metatarsal		Normal
Flexor digitorum tendon		Normal
Flexor hallucis longus tendon		Normal
Gastrocnemius		Tight and spasmed from limping
Lateral malleolus		Sore, but no deformity
Medial border of tibia		Normal
Medial cuneiform		Normal
Medial malleolus		Normal
Metatarsal heads		Normal
Metatarsals & phalanges		Normal
Navicular tubercle		Normal
Peroneal muscle & tendons		Tight and spasmed from limping
Plantar fascia		Normal
Posterior tibial artery		Normal
Sesamoid bones		Normal
Sinus tarsus		Sore and point tender (6/10)
Soleus		Tight and spasmed from limping
Tibial crest		Normal
Tibial shaft		Normal
Tibialis posterior tendon		Normal
Toes		Normal
other		
other		
other		
Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing palpations?		

Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		
Conclusion		
Is the conclusion/diagnosis correct?		2 degree ATF and CF sprain, potential for avulsion fraction, but need to send to physician and x-rays to discover it is not broken
Did the student follow the correct (ie. thorough and accurate) process and still not get the correct conclusion/diagnosis?		
Did the student refer to a physician for a complete diagnosis?		
Overall Impression of the Student's/Examinee's Assessment Skills		
If you had to provide an overall grade out of 10, what would you give to the student/ examinee?		

Diagnosis: Subacromial Bursitis

History		
Themes	**Tasks**	**Patient Information/Answer Key**
Nature of Injury	Chief Complaint	shoulder pain
	Date of Injury/When Did Injury Occur	Has been bothering me for 3 weeks
	MOI/Recreation of MOI	Hurts with overhead motion, not just sport, but anything in life. Did not have one incident, just starting hurting over time. Insidious onset, perhaps the end of a long season?
	Previous Injury to this Site	Never had problems before
	Previous Injury to joint above/below	bruised elbow from diving in vball
	Functional Limitations	can not lift arm above head, can not play v ball without pain
	History Back or Neck Injury/Pain	No history of back or neck pain
	Subjective/patient SHARP	patient did not notice any swelling, heat, redness, but had pain in shoulder
	Unusual Sounds/Sensations	felt some mild clicking, but nothing else
Patient Information	Age	21
	Primary Sport/occupation	v ball
	Level of Sport	varsity sport at college, jr. national team hopeful, middle hitter, they are at the end of the season
	Training for Sport	practice 5 days a week
	Weight Training for Sport/occupation	weights once a week during season and 3 times per week in off season
	Change in anthropometrics	gain a little weight over the last year: about 10 pounds…mostly muscle
	Psycho/Emotional Stress	no stress other than examinations
	Training Surface	volleyball court
	Equipment or Footwear	court shoes

	Activities of Daily Living	hurts to reach above head, getting dishes out of cupboard, reaching on top of fridge, brushing hair, etc..
	Other Sports/Activities	no other sports
Global Pain Questions	Pain: Pain Location	tip of shoulder, deep and around deltoid muscle area
	Pain: Describe Pain	thobbing in the AM, sharp pain when lift arm above head
	Pain: Pain Scale	when lifting arm, it is an 8/10, when trying to play v ball, it is an 8/10, otherwise it is a 5-6/10
	Pain: Relieves Pain	not lifting my arm, ice, advil
	Pain: Exacerbates Pain	motions with arm above head and sleeping on it
	Pain: Morning or night Pain	throbbing in AM and sometimes wakes me at night
	Pain: Pain before, during or after activity	pain during activity, seems to subside a little, but then really hurts after practice and game
	Pain: Radiating Pain Proximal or Distal	mostly localized except goes into deltoid muscle area
	Pain: Increase with coughing, sneezing or bearing-down	NO
	Pain: Sleeping and Position	hurts to sleep on that shoulder
Safety/Rehab/Understanding	Treatment	ice, advil
	Reviewed by Physician	no
	X-rays/diagnostic imaging	no
	Medications	advil
	Allergies	dust
	Goals for RETURN	have playoffs in 3 weeks
	Overall Health	great health
	Medical Conditions	no conditions
	Family History of Medical Conditions	mother has diabetes, grandmother has breast cancer

Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		Possible differential diagnoses are: rotator cuff tear, SLAP lesion, long head bicep tendenitis, tenosynovitis, cervical disc herniation, brachial plexus injury, thoracic outlet syndrome, etc…
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		The model (SP) should draw this information out of the student through conversation to make it seem less intrusive and "examination like"
Did the student **hone** into their **index of suspicion** in their history?		
Observation		
Swelling		no obvious swelling
Heat		feels slightly warmer than the other side
Altered Function		trouble getting shirt off
Redness		no obvious
Discoloration		none present
Upper Extremity:		
Shoulder Height		right handed and right side lower than left
Clavicles		WSWG
AC (Step Deformity)		WSWG
Sulcus		WSWG
Scapulae		WSWG
Shoulders Rounded		yes, rounded
Symmetry		
Nose in line Sternum		WSWG
Head Tilt		WSWG
Deformity		WSWG
Scars, cuts, abrasions		WSWG
Posture		
Observes from side/side		WSWG plus rounded shoulders
Observes from rear/front		WSWG
Other		
Other		
Other		
Other		
Other		

Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		Student should be observing contributing factors that may lead to this injury. Tight pec major evidenced by rounded shoulders and poor shoulder mechanics as result. Tight internal rotators as evidenced by arm internally rotated.
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		
Clearing Joints Above and Below		
Did the student clear the joint **below** the lesion site?		No problems with active elbow motion or overpressure
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.
Scanning Exams		
Did the student ask inquire about **radicular pain** in the history?		No radicular pain
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		Did not give the MOI
Did the student perform **active ROM** in the **cervical spine**?		No pain with cervical motion, students can be permitted to perform a quadrant test to clear the neck. This should be marked as a 3 whereas full, complete AROM is greater than a 3.
Did the student **change their index of suspicion** after completing the **scanning exam**?		The model should pull this information out of the student by asking if they think what they orginally thought it might be is still

	the same.
Active ROM	
Flexion	Pain above 75 degrees
Extension	No pain
Adduction	Slight pain when lowering arm to side, particularly around 110-75 degrees
Abduction	Lots of pain, particularly between 75 and 110 degrees
External Rotation	If arm adducted, not very much pain.
Internal Rotation	Painful either with arm adducted or abducted.
External Rotation (at 90 degrees)	Painful
Internal Rotation (at 90 degrees)	Very painful
Horizontal Abduction	No pain
Horizontal Adduction	Painful
Scapular Protraction/Retraction	No pain
Scapular Elevation/Depression	No pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **AROM**?	The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.
Passive ROM	
Flexion	Pain above 75 degrees
Extension	No pain
Adduction	Slight pain when lowering arm to side, particularly around 110-75 degrees
Abduction	Lots of pain, particularly between 75 and 110 degrees
External Rotation	If arm adducted, not very much pain.
Internal Rotation	Painful either with arm

	adducted or abducted.
External Rotation (at 90 degrees)	Painful
Internal Rotation (at 90 degrees)	Very painful
Horizontal Abduction	No pain
Horizontal Adduction	Painful
Scapular Protraction/Retraction	No pain
Scapular Elevation/Depression	No pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **PROM**?	The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.
Isometric Resisted	
Flexion	No real pain 1/10
Extension	No pain
Adduction	No real pain 1/10
Abduction	Pain and weakness at both 1-15 degrees and 90 degrees. 6/10
External Rotation	Pain
Internal Rotation	No pain
External Rotation (at 90 degrees)	Pain 6/10
Internal Rotation (at 90 degrees)	Pain 4/10
Horizontal Abduction	No pain
Horizontal Adduction	Pain
Scapular Protraction/Retraction	No pain
Scapular Elevation/Depression	No pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **IR**?	The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.

Special Tests	
AC compression	No pain
AC shear	No pain
Active impingement	Very painful
Anterior drawer	No pain, but pay attention to hand position
Apprehension & relocation	Painful, not apprehensive
Clunk test	Painful, not apprehensive and no odd sensations
Drop arm	Painful, but no sign of torn rotator cuff
Empty can	Very painful
Hawkin's Kennedy	Very painful
Joint Play	No pain
Load & shift	No pain
Ludington's test	No pain
Neer impingement	Very painful
Pec major contracture test	tight and sore to get into position
Posterior drawer	No pain
Speed's	Painful but not weak
Sulcus sign	None
Thoracic outlet syndrome tests	Negative
Winging Scapula	No
Yergason's	No
other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing special tests?	The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.

Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		
Palpation		
AC joint		
Acromion process		
Anterior serratus		
Bicipital groove		Pain, particularly up into the joint and near the subacromial bursa
Brachial artery		
Clavicle		
Coracoid process		
Costicartilage		
Greater tubercle		
Inferior and superior angles of the scapula		
Infraspinatus tendon		Painful around insertion and along tendon
Latissimus dorsi		
Lesser tubercle		
Lower cervical spinous processes		
Pectoralis major		
Ribs		
SC joint		
Spine of the scapula		
Sternum		
Subacromial bursa		very painful…particularly if they expose the bursa by rotating the humerus, then internally rotating arm and exposing the bursa
Subscapularis tendon		

Supraspinatus tendon		Very painful along tendon and muscle is tight and sore
Triceps tendon		
Upper and middle thoracic spinous processes		
other		
other		
other		
other		
other		
Did the student specifically test structures that were identified in their previous index of suspicions?		
Did the student **change their index of suspicion** after completing palpations?		The model should pull this information out of the student by asking if they think what they orginally thought it might be is still the same.
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		
Conclusion		
Is the conclusion/diagnosis correct?		

Did the student follow the correct (ie. thorough and accurate) process and still not get the correct conclusion/diagnosis?		
Did the student refer to a physician for a complete diagnosis?		If they do not refer to a physician, they should get 1
Overall Impression of the Student's/Examinee's Assessment Skills		
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		

Knee: Pre-Tibial Bursitis		
Examiner Name:		
Student Name:		
History Themes	**Tasks**	**Patient Information/Answer Key**
Nature of Injury	Chief Complaint	Knee pain
	Date of Injury/When Did Injury Occur	Has been bothering them for a couple of weeks, really bad the last week
	MOI/Recreation of MOI	No idea? It just started hurting after practice more and more. I thought it was a bruise, but it hasn't gone away.
	Previous Injury to this Site	Little aches and pains from time to time, but nothing serious. I feel like I have bruised it before from either falling on it or banging it with someone else.
	Previous Injury to joint above/below	Sprained ankle, no hip pain
	Functional Limitations	Unable to jump, land, expload when running, etc..
	History Back or Neck Injury/Pain	None
	Subjective/patient SHARP	I did not see any bruising, but it feels like a bruise. I hurts going up stairs and I limp a fair bit.
	Unusual Sounds/Sensations	None
Patient Information	Age	18
	Primary Sport/occupation	Basketball
	Level of Sport	Varsity HS and hopefully College and Jr. Provincial team member
	Training for Sport	Practice 5 days a week and play 1-2 times per week.
	Weight Training for Sport/occupation	Lift weights 1-2 times per week. Squats, leg extensions, leg curls, calf raises for lower body. Abs.
	Change in anthropometrics	Has grown 5 inches in the last 6 months and put on 10 pounds
	Psycho/Emotional Stress	None
	Training Surface	Basketball court

	Equipment or Footwear	Having to change shoes every 3-4 months because feet are growing. Uses basketball shoes and whatever the sports store recommends..if they look good.
	Activities of Daily Living	Up and down stairs hurt. Sitting for long periods of time throughout the day.
	Other Sports/Activities	Gym class, skate boarding
Global Pain Questions	Pain: Pain Location	Anterior knee
	Pain: Describe Pain	Throbbing after activity or sitting for long time. Sharp pain when jumping or pushing off.
	Pain: Pain Scale	5/10 for throbbing pain and 8/10 for sharp pain
	Pain: Relieves Pain	Rest.
	Pain: Exacerbates Pain	Jumping and running.
	Pain: Morning or night Pain	Throbs at night after long day of practicing or skate boarding.
	Pain: Pain before, during or after activity	Pain before is throbbing, during activity is sharp pain…tends to get better as the activity goes on, but still notice it and then after the activity (within 2 hours), it really starts to hurt.
	Pain: Radiating Pain Proximal or Distal	Localized to front of knee
	Pain: Increase with coughing, sneezing or bearing-down	no
	Pain: Sleeping and Position	Does not notice with sleep.
Safety/Rehab/Understanding	Treatment	None to date
	Reviewed by Physician	No
	X-rays/diagnostic imaging	No
	Medications	No
	Allergies	None
	Goals for RETURN	Want to continue to play, I just want to the pain to go away.
	Overall Health	great
	Medical Conditions	None
	Family History of Medical Conditions	Don't know?

Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		Potential diffirential diagnoses: suprapatellar bursa, PFS, Meniscal, Patellar Tendinitis, Osgood Schlatters Disease, fat pat bruise, Pes Anserinus Bursitis, ITB friction syndrome, MCL sprain
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		Examiner should evaluate on BOTH quality and quantity of options provided by the candidate
Did the student **hone** into their **index of suspicion** in their history?		
Observation		
Swelling		Slightly bigger compared to the contralateral side (only soft tissue, not bone)
Heat		Slightly warmer than the contralateral side
Altered Function		Slight limp
Redness		None
Discoloration		None
Lower Extremity:		
Pelvic Height		Level
ASIS		Level
PSIS		Level
Patellae		Normal
Genu Valgum/Varum		WSWG
Tibial Valgum/Varum		WSWG
Foot - Static		WSWG
Foot - Dynamic (Gait)		WSWG
Symmetry		
Foot rotation/position		WSWG
Posture		
Observes from side/side		WSWG
Observes from rear/front		WSWG
Observes Pelvic Position/Tilt		WSWG
Scars, cuts, abrasions		Old scrapes around knee from skate board wipe out
Osgood-Schlatter's Bump		No bump or extended tibial tub. present
Other		
Other		
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		

Clearing Joints Above and Below	
Did the student clear the joint **below** the lesion site?	No pain with active motion of the ankle
Did the student clear the joint **above** the lesion site?	No pain with active motion of the hip
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?	
Scanning Exams	
Did the student ask inquire about **radicular pain** in the history?	No radicular pain
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?	It is insidious, but can be narrowed to falling on knees when skate boarding and irriating the pre-tibial bursa. In addition, the constant jumping and running does not allow it to recover so it is slowly getting worse.
Did the student perform **active ROM** in the **lower back**?	They should be asking the athlete to complete back motion. There would be no pain if they did check AROM in the back.
Did the student **change their index of suspicion** after completing the **scanning exam**?	
Active ROM	
Flexion	pain at extreme flexion
Extension	slight pain
Adduction	No pain
Abduction	No pain
External Rotation	Slight pain
Internal Rotation	Slight pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **AROM**?	
Passive ROM	
Flexion	No pain
Extension	No pain
Adduction	No pain
Abduction	No pain
External Rotation	No pain
Internal Rotation	No pain

Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **PROM**?	
Isometric Resisted	
Flexion	5/5 strength, Slight pain (1/10)
Extension	5/5 strength, 7-8/10 pain during contraction
Adduction	No pain, full strength
Abduction	No pain, full strength
External Rotation	No pain, full strength
Internal Rotation	No pain, full strength
Plantarflexion	No pain, full strength
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing **IR**?	
Special Tests	
Anterior drawer	No pain
Apley's distraction / compression	No pain
Brush/sweep test	No Swelling/effusion
Clarke's sign	No pain
Dermatomes	No deficits
Ely's	Negative
Jerk test of hughston	Negative
Kendall	Tight quadriceps
Lachman's	Negative
McMurray's	Negative
Myotomes	No weakness
Obers	Negative
Patellar Apprehension	Negative
Pivot Shift	Negative
Posterior drawer	Negative
Posterior sag	Negative
Reflexes	Normal reflexes
Slocum	Negative
Straight leg raise	Tight Hamstrings (65 degrees)
Superior Tib-Fib Jt. Play	Negative
Thomas	Tight quadriceps
Valgus stress	Negative
Varus stress	Negative

other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing special tests?	
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	
Palpation	
Adductor tubercle	
Anterior Superior Iliac Spine	
Biceps femoris and lateral gastrocnemius tendons	
Fat pad	Close to where you have pain
Fibular head	
Gastrocnemius muscle bellies	
Gerdy's tubercle/IT Band Insertion	
Hamstring muscle bellies	
Iliotibial band	
Infrapatellar tendon	Close to where you have pain
Joint Play	
Lateral femoral condyle and epicondyle	
Lateral joint line	
LCL	
MCL	
Medial femoral condyle and epicondyle	
Medial hamstring and gastrocnemius	

Medial joint line	
Patella	
Patellar retinaculum	
Pes anserine insertion	
Plica	
Popliteal fossa	
Quadriceps (whole)	
Quadricepts (each individually)	
Sartorius	
Superficial bursa	Pre-tibial bursa is VERY point tender
Suprapatellar pouch	Normal
Suprapatellar tendon	Normal
Tibial plateau	
Tibial tuberosity	VERY sore because it's proximity to the bursa
other	
other	
other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** after completing palpations?	
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	
Conclusion	
Is the conclusion/diagnosis correct?	

Did the student follow the correct (ie. thorough and accurate) process and still not get the correct conclusion/diagnosis?		
Did the student refer to a physician for a complete diagnosis?		
Overall Impression of the Student's/Examinee's Assessment Skills		
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		

Appendix C Standardized rater training Microsoft Powerpoint © presentation

(outline view)

- **Standardized Orthopedic Assessment Tool (SOAT)**
- **Examiner Preparation Seminar**
- Introduction and Purpose
- **Thank you** to all volunteers for helping
- Purpose of this study is to determine the validity and reliability of two SOAT forms
- Previously, this tool was content validated by a group of experts from across Canada
- The tool has undergone initial reliability testing
- Results indicate good reliability (.82 & .83)
- **New twist on the forms**
- SOAT marking sheets have a list of items which is what people have likely been exposed to in the past
- **Novices** typically go through everything on a list because they don't have the experience eliminate items (i.e., cut corners)
- This is frustrating to **EXPERTS** because of the student's inefficiency
- However, it is important to allow students the flexibility to complete their assessments using all items on the form OR…..
- **New twist on the forms**
- The tool will also permit those students who want to eliminate items from the form by asking them to **verbalize their thought process** (i.e., we can understand their intentions)
- At the end of each section of the orthopedic assessment, students will be cued to give the patient their **index of suspicion**
- More later….
- How many scenarios are there and which examiners have each of them?
- There are only two scenarios/cases:
- 2nd degree MCL sprain (isolated)
- Supraspinatus tear (complete)
- Students will be randomly assigned a case to a maximum of half the n for each institution
- i.e., MRC 30 students = 15 knee and 15 shoulder
- **Who is this form intended to test?**
- This form is meant to test a **basic level of competence**
- **What does that mean?**
- Students who have taken a theoretical course in orthopedic assessment
- Students who have SOME experience with applying the theory in a practical setting
- **Basic Level of Competence?**
- **Why is this important?**
- Context, context, context!!!!

- You need to think back to when you were a student and not expect the students to be EXPERTS
- Putting yourself into this mindset will help you understand the grading scale we are asking you to employ
-
- Grading Scheme
- **History & Observation**
- Dichotomous data
- Did the student ask the history question or make an observation?
- **0 = Not Done**
- **1 = Done**
- **Observation**
- Model needs to cue the student to verbalize if they do not do it spontaneously
- Clearing Joints, Scanning Examination, Examination, Special Testing and Palpation
- More grading assumptions
- All tasks in all components of the tool will default to n/a if they are not graded by the examiner
- Students will lose grades on tasks that they missed (or completed incorrectly) which were thought to be relevant by the evaluator for a specific scenario or condition
- More grading assumptions
- Students should be graded on tasks based on the index of suspicion they identified at the end of each section. If the student does not have the correct index of suspicion and they fail to complete tasks that would permit them to know more about the condition, the student should lose marks for those specific tasks in the subsequent section. If they perform tasks that relate to their index of suspicion, they should receive grades for those tasks done correctly
- More grading assumptions
- The model (standardized patient) will ask the student if the index of suspicion has changed at the end of each component of the assessment (i.e; history, observation, scanning exam, clearing joints above and below, examination, special tests, palpation)
- More grading assumptions
- All checklists have the same tasks for history, observation, scanning examination and examination (AROM, PROM, IR)
- More grading assumptions

Special Testing

- Students should not be required to complete a specific and discrete list of **special tests** for any given condition. Rather, the specific special tests that should be completed should be based on what the student/examinee completed throughout the preceded components of the orthopedic assessment (i.e., history, observation, scanning exam, clearing joints above and below and examination). All special tests will be defaulted as n/a. Special

tests will only be marked if the student performed them (i.e., correctly or incorrectly) OR if they should have tested a structure, but failed to do so. In the case when a student/examinee FAILS to complete a special test which would identify or confirm the anatomical structure/lesion site, they will be given a grade of zero for all special tests which would have tested that structure

•

•HUH????

•You will need to provide examples when we go through the forms

•More grading assumptions

•If there is a **special test** missing from the list, the examiner should add the test and then mark it based on its accuracy. If none of the examiners are familiar with the test, it should be highlighted and the lead examiner should research the test AND ask the examinee the name and purpose of the special test at the end of the entire test. If this is a valid and reliable test, it can be added to the database

•More grading assumptions: **Palpation**

•Students should not be required to **palpate** a specific and discrete list of anatomical structures for any given condition. Rather, the specific anatomical structure that should be palpated will be based on what the student/examinee completed throughout the preceded components of the orthopedic assessment (i.e., history, observation, scanning exam, clearing joints above and below and examination). All anatomical structures in the palpation section will be defaulted as n/a. Palpation will only be marked if the student performed them (i.e., correctly or incorrectly) OR if they should have palpated a structure, but failed to do so. In the case when a student/examinee FAILS to palpate an anatomical structure which would identify or confirm the lesion site, they will be given a grade of zero for all anatomical structures which would have tested that lesion site

•

•HUH????

•You will need to provide examples when we go through the forms

•More grading assumptions

•Special tests are listed in alphabetical order

•Anatomical structures for palpation are listed in alphabetical order

•At the bottom of each major section, there is a global rating scale

•The global rating scale should be completed ONLY after you have finished marking each of the tasks in that section

•Can we talk after the student has left the room?

•Yes, you can talk the other examiner or the model, however….

•You can only ask for clarification about whether they completed a test, not as to it's value

•You can get feedback from the SP to assist you in determining the value of any testing or palpation

•**Let's get to the Forms and Scenarios**

Appendix D Advertisement for examiners/raters in Manitoba

UNIVERSITY OF CALGARY

Calling all Certified Athletic Therapists

Are you interested in helping develop a scientifically sound (valid and reliable) orthopedic assessment instrument? This is your chance with the Standardized Orthopedic Assessment Tool (**SOAT**).

Your participation in this study will require the following:

Training Session: a 3 hour training session will help orientate you to the SOAT. This interactive session will review the tool and the scenario. The training session will be held

Friday, April 13 from 6:30pm to 9:30 pm at the University of Winnipeg, Duckworth Centre.
Dinner will be provided and please keep your parking receipt for reimbursement.

Examination Blocks: in order to participate in this study, you must commit to at least TWO, 4 hour time blocks. The schedule of potential time slots are attached in the excel spreadsheet (**April 14-20**).

What is in it for you?
This is a great professional development experience. You will be granted at 1 CEU for your minimum requirement of two time blocks and the training. Development of tools that measure student performance are critical to the continuing evolution of the AT profession and education.

Interested?

****Please respond by March 23, 2007 if you are interested.****

Please submit your interest (dates and training availability) directly to Mark Lafave at mrlafave@ucalgary.ca or call 403-481-9059.

SOAT, M. Lafave

Appendix E Advertisement for examiners/raters in Calgary

UNIVERSITY OF CALGARY

Calling all Certified Athletic Therapists

Are you interested in helping develop a scientifically sound (valid and reliable) orthopedic assessment instrument? This is your chance with the Standardized Orthopedic Assessment Tool (**SOAT**).

Your participation in this study will require the following:

Training Session: a 3 hour training session will help orientate you to the SOAT. This interactive session will review the tool and the scenario. The training session will be held

Thursday, April 5 from 6:30pm to 9:30 pm at Mount Royal College, AT Lab. Dinner will be provided and you can arrange for a parking pass by contacting Cheryl Barker in Phys. Ed. at cbarker@mtroyal.ca

Examination Blocks: in order to participate in this study, you must commit to at least TWO, 4 hour time blocks. The schedule of potential time slots are attached in the excel spreadsheet (**April 10, 11 & 12**).

What is in it for you?

This is a great professional development experience. You will be granted at least 0.5 for your minimum requirement of two time blocks and the training. A request for more CEUs has been made to the Certification Committee. Development of tools that measure student performance are critical to the continuing evolution of the AT profession and education.

Interested?

Please submit your interest (dates and training availability) directly to Mark Lafave at mrlafave@ucalgary.ca or call 403-481-9059.

SOAT, M. Lafave

Appendix F Knee and shoulder answer key from phase III research

Knee SOAT Answer Key

History		
Themes	**Tasks**	
Nature of Injury	Chief Complaint	Knee sore
	Date of Injury/When Did Injury Occur	3 days ago
	MOI/Recreation of MOI	Playing soccer, went to kick ball and the ball was also contacted by another player at the same time. Ball was on inside of foot.
	Previous Injury to this Site	None
	Previous Injury to joint above/below	Yes, sprained ankle once before
	Functional Limitations	Can not walk up or down stairs without pain, walking with limp
	History Back or Neck Injury/Pain	None
	Subjective/patient SHARP	+2 swelling/knee effusion, '+2 redness and warmth
	Unusual Sounds/Sensations	felt a crunch when both players contacted the ball
Patient Information	Age	18
	Primary Sport/occupation	soccer
	Level of Sport	Premier
	Training for Sport	Practice 2x/week and play 2x/week
	Weight Training for Sport/occupation	no
	Change in anthropometrics	yes, he has grown 4 inches in the last 3 months, gained 10 pounds
	Psycho/Emotional Stress	none
	Training Surface	played on grass
	Equipment or Footwear	use cleats
	Activities of Daily Living	can not walk normally right now since injury, trouble sitting on toilet or getting into car
	Other Sports/Activities	Yes, high school sports
Global Pain Questions	Pain: Pain Location	whole knee, particularly inside
	Pain: Describe Pain	sharp at first, now just throbbing
	Pain: Pain Scale	9/10 when first happened, now 6/10
	Pain: Relieves Pain	less walking around, putting it up, ice
	Pain: Exacerbates Pain	walking, sitting, bumping it
	Pain: Morning or night Pain	wakes me up at night when I try to move around

	Pain: Pain before, during or after activity	have not done any activity since injury
	Pain: Radiating Pain Proximal or Distal	no
	Pain: Increase with coughing, sneezing or bearing-down	no
	Pain: Sleeping and Position	yes, pain with sleeping…tough finding a comfortable position
Safety/Rehab/Understanding	Treatment	ice, rest, elevation
	Reviewed by Physician	no
	X-rays/diagnostic imaging	no
	Medications	Tylenol 2 x
	Allergies	no
	Goals for RETURN	ASAP..middle of season
	Overall Health	good
	Medical Conditions	none
	Family History of Medical Conditions	father has type II diabetes, grandmother had colon cancer
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		Student should have systematically eliminated items from the list of history questions
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		Potential injuries may include: MCL, ACL, PCL, Med. Meniscus, Subluxed patella
Did the student **hone** into their **index of suspicion** in their history?		
Observation		
Swelling		++ effusion
Heat		++yes
Altered Function		yes
Redness		yes
Discoloration		no
Lower Extremity:		WSWG
Pelvic Height		WSWG
ASIS		WSWG
PSIS		WSWG
Medial Malleoli		WSWG
Patellae		WSWG
Genu Valgum/Varum		WSWG
Tibial Valgum/Varum		WSWG
Foot - Static		WSWG

Foot - Dynamic (Gait)	significant limp
Symmetry	
Foot rotation/position	WSWG
Posture	
Observes from side/side	WSWG
Observes from rear/front	WSWG
Observes Pelvic Position/Tilt	WSWG
Scars, cuts, abrasions	WSWG
Other	
Other	
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?	
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?	
Clearing Joints Above and Below	
Did the student clear the joint **below** the lesion site?	should clear ankle joint
Did the student clear the joint **above** the lesion site?	should clear hip joint
Did the student **change their index of suspicion** appropriately after clearing the joints above and below the lesion site?	
Scanning Exams	
Did the student ask inquire about **radicular pain** in the history?	
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?	traumatic injury
Did the student perform **active ROM** in the **lower back**?	should perform AROM in back to clear it
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?	

Active ROM	
Flexion	signficantly reduced (25 degrees of flexion from neutral)
Extension	painful, but can get into full extension
Adduction	hip motion is okay
Abduction	hip abduction makes the patient feel uncomfortable because someone unstable
External Rotation	++ pain and limited ROM (feels unstable)
Internal Rotation	no real pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing **AROM**?	
Passive ROM	
Flexion	signficantly reduced (25-30 degrees of flexion from neutral)
Extension	painful, but can get into full extension
Adduction	hip motion is okay
Abduction	hip abduction makes the patient feel uncomfortable because someone unstable
External Rotation	++ pain and limited ROM (feels unstable)
Internal Rotation	no real pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing **PROM**?	
Isometric Resisted	
Flexion	strong, but slight pain (5/5)
Extension	somewhat strong, but painful (4/5)
Adduction	painful*** particularly based on how they test (below knee or above knee)
Abduction	not painful
External Rotation	not painful
Internal Rotation	somewhat strong, but painful and feels unstable (3.5/5)
Plantarflexion	no pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion**	

appropriately after completing **IR**?	
Special Tests	
Anterior drawer	negative, but can be painful based on hand position
Apley's distraction / compression	slight pain because of position of knee for test…can not get into 90 degree s of flexion
Brush/sweep test	positive..thick effusion
Clarke's sign	slight pain
Dermatomes	negative
Ely's	can not get into much knee flexion to test properly
Jerk test of hughston	negative
Kendall	uncomfortable when leg hanging off table..unstable
Lachman's	negative, but can be painful based on hand position
McMurray's	pain with external rotation, not positive and no click
Myotomes	negative
Obers	normal
Patellar Apprehension	negative
Pivot Shift	pain with valgus motion, but no positive shift
Posterior drawer	negative
Posterior sag	negative
Reflexes	fine
Slocum	pain with external rotation, but not positive
Straight leg raise	negative
Superior Tib-Fib Jt. Play	negative
Thomas	negative
Valgus stress	++ pain, plus gapping/movement
Varus stress	negative
other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing special tests?	
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential	It would be important to eliminate the following structures from being injured: ACL, medial meniscus, patellar apprehension/subluxation

diagnosis strategy(ies))?	
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	Should have done a valgus stress test
Palpation	
Adductor tubercle	pain just around this spot and more inferior to it
Anterior Superior Iliac Spine	no pain
Biceps femoris and lateral gastrocnemius tendons	no pain
Fat pad	no pain
Fibular head	no pain
Gastrocnemius muscle bellies	no pain
Gerdy's tubercle/IT Band Insertion	no pain
Hamstring muscle bellies	no pain
Iliotibial band	no pain
Infrapatellar tendon	no pain
Joint Play	n/a
Lateral femoral condyle and epicondyle	no pain
Lateral joint line	no pain
LCL	no pain
MCL	++ pain throughout its substance
Medial femoral condyle and epicondyle	+ pain
Medial hamstring and gastrocnemius	no pain
Medial joint line	pain on the MCL only
Patella	no pain
Patellar retinaculum	no pain
Pes anserine insertion	no pain
Plica	no pain
Popliteal fossa	no pain
Quadriceps (whole)	no pain or spasm
Quadriceps (each individually)	no pain or spasm
Sartorius	no pain
Superficial bursa	no pain
Suprapatellar pouch	painful from effusion
Suprapatellar tendon	painful from effusion

Tibial plateau	no pain
Tibial tuberosity	no pain
other	
other	
other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing palpations?	
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	Should have palpated meniscus, pes anserine bursa, medial retinaculum
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	Should have identified the entire length of the MCL to get full marks
Conclusion	
Is the conclusion/diagnosis correct?	2nd degree MCL
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?	
Did the student refer to a physician for a complete diagnosis?	
Overall Impression of the Student's/Examinee's Assessment Skills	
If you had to provide an overall grade out of 5, what would you give to the	

student/ examinee?		

Shoulder SOAT Answer Key

History		
Themes	**Tasks**	
Nature of Injury	Chief Complaint	Pain in the shoulder
	Date of Injury/When Did Injury Occur	No particular time, it has been hurting more so for the last 4 months, but there has been pain on and off over the years
	MOI/Recreation of MOI	Pitcher in baseball
	Previous Injury to this Site	Pain over the years since he was 18
	Previous Injury to joint above/below	Some elbow pain over the years
	Functional Limitations	Difficulty raising arm above head
	History Back or Neck Injury/Pain	No
	Subjective/patient SHARP	No swelling, redness or heat, but pain on an off
	Unusual Sounds/Sensations	feel crunching noise periodically..more so lately
Patient Information	Age	22
	Primary Sport/occupation	Baseball
	Level of Sport	Minor League Pitcher
	Training for Sport	Pitch once a week at most, practice daily,
	Weight Training for Sport/occupation	weights for shoulders, lats, chest, abs, back, legs
	Change in anthropometrics	Increase in size of large muscle groups
	Psycho/Emotional Stress	none
	Training Surface	baseball field
	Equipment or Footwear	n/a
	Activities of Daily Living	trouble reaching above head
	Other Sports/Activities	roller blading
Global Pain Questions	Pain: Pain Location	tip of shoulder
	Pain: Describe Pain	throbbing
	Pain: Pain Scale	4/10 except just after pitching, then 8/10
	Pain: Relieves Pain	no pitching
	Pain: Exacerbates Pain	Pitching

SOAT, M. Lafave

	Pain: Morning or night Pain	yes, throbs in morning and keeps up at night
	Pain: Pain before, during or after activity	pain before, during and now, really bad after activity
	Pain: Radiating Pain Proximal or Distal	yes, seems to be all over shoulder
	Pain: Increase with coughing, sneezing or bearing-down	no
	Pain: Sleeping and Position	Yes, hard to sleep on it and hard to get comfortable
Safety/Rehab/Understanding	Treatment	Ice, stretch
	Reviewed by Physician	no
	X-rays/diagnostic imaging	no
	Medications	yes, advil
	Allergies	no
	Goals for RETURN	ASAP
	Overall Health	good
	Medical Conditions	heart mumur
	Family History of Medical Conditions	not that you are aware of
Did the student employ differential diagnosis strategy(ies) in the history to identify/eliminate other possible injuries/conditions?		
Did the student list at least **three** potential anatomical structures/injuries as an **index of suspicion**?		Some potential injuries could be bursitis, SLAP lesion, supraspinatus tendinitis, infraspinatus tendinitis, subacromial bursitis
Did the student **hone** into their **index of suspicion** in their history?		
Observation		
Swelling		nothing that you can see
Heat		none
Altered Function		yes, difficulty taking shirt off
Redness		none
Discoloration		none
Upper Extremity:		
Shoulder Height		right side higher
Clavicles		WSWG
AC (Step Deformity)		WSWG
Sulcus		WSWG
Scapulae		WSWG
Shoulders Rounded		Yes, particularly right side

Symmetry		
Nose in line Sternum		WSWG
Head Tilt		WSWG
Deformity		WSWG
Scars, cuts, abrasions		WSWG
Posture		
Observes from side/side		WSWG
Observes from rear/front		WSWG
Other		
Other		
Other		
Other		
Other		
Did the student employ differential diagnosis strategy(ies) in the observation to identify/eliminate other possible injuries/conditions?		
Did the student specifically **OBSERVE** for their **index of suspicion** identified after the history?		
Clearing Joints Above and Below		
Did the student clear the joint **below** the lesion site?		No pain in the elbow
Did the student **change their index of suspicion** after clearing the joints above and below the lesion site?		
Scanning Exams		
Did the student ask inquire about **radicular pain** in the history?		no radicular pain
Did the student clearly determine a mechanism of injury in the history? Did they establish the MOI was **traumatic and not insidious**?		insidious onset
Did the student perform **active ROM** in the **cervical spine**?		No pain with AROM in the C spine
Did the student **change their index of suspicion** appropriately after completing the **scanning exam**?		
Active ROM		
Flexion		pain throughout movement, and awkward looking
Extension		no pain
Adduction		slight pain
Abduction		definite pain and abnormal movement, lead by scapula past 15 degrees
External Rotation		slight pain

Internal Rotation	slight pain
External Rotation (at 90 degrees)	slight pain
Internal Rotation (at 90 degrees)	slight pain
Horizontal Abduction	slight pain
Horizontal Adduction	slight pain
Scapular Protraction/Retraction	no pain
Scapular Elevation/Depression	no pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing **AROM**?	
Passive ROM	
Flexion	slight pain
Extension	slight pain
Adduction	no pain
Abduction	slight pain
External Rotation	slight pain
Internal Rotation	slight pain
External Rotation (at 90 degrees)	slight pain
Internal Rotation (at 90 degrees)	slight pain
Horizontal Abduction	no pain
Horizontal Adduction	slight pain
Scapular Protraction/Retraction	no pain
Scapular Elevation/Depression	no pain
Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing **PROM**?	
Isometric Resisted	
Flexion	slight pain, but good strength (5/5)
Extension	slight pain, but good strength (5/5)
Adduction	slight pain, but good strength (5/5)
Abduction	complete weakness at 0 degrees (1/5), but 4/5 at 90 degrees with pain
External Rotation	3.5/5 strength and painful
Internal Rotation	slight pain, but good strength (5/5)
External Rotation (at 90 degrees)	slight pain, but good strength (5/5)
Internal Rotation (at 90 degrees)	no pain
Horizontal Abduction	slight pain, but good strength (5/5)
Horizontal Adduction	no pain
Scapular Protraction/Retraction	no pain
Scapular Elevation/Depression	no pain

Did the student specifically test structures that were identified in their previous index of suspicions?	
Did the student **change their index of suspicion** appropriately after completing **IR**?	
Special Tests	
AC compression	negative
AC shear	negative
Active impingement	slight pain
Anterior drawer	negative
Apprehension & relocation	negative
Clunk test	slight pain, but negative
Drop arm	positive
Empty can	positive
Hawkin's Kennedy	slight pain, but negative
Joint Play	negative
Load & shift	negative
Ludington's test	negative
Neer impingement	negative, but slight pain
Pec major contracture test	negative
Posterior drawer	negative
Speed's	slight pain, but negative
Sulcus sign	negative
Thoracic outlet syndrome tests	negative
Winging Scapula	negative
Yergason's	negative
other	
other	
other	
other	
Did the student specifically test structures that were identified in their previous index of suspicion?	
Did the student **change their index of suspicion** appropriately after completing special tests?	
Did the student identify and complete **special tests** that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?	Should have eliminated painful arc (subacromial bursitis), SLAP lesion, subluxating GH joint, long head of bicep tendinitis
Did the student identify and complete **special tests** that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?	Should have completed EITHER drop arm OR empty can

Palpation		
AC joint		no pain
Acromion process		no pain
Anterior serratus		no pain
Bicipital groove		no pain
Brachial artery		no pain
Clavicle		no pain
Coracoid process		no pain
Costicartilage		no pain
Greater tubercle		painful along supraspinatus attachment
Inferior and superior angles of the scapula		no pain
Infraspinatus tendon		no pain
Latissimus dorsi		no pain
Lesser tubercle		no pain
Lower cervical spinous processes		no pain
Pectoralis major		no pain
Ribs		no pain
SC joint		no pain
Spine of the scapula		slight pain just superior to it
Sternum		no pain
Subacromial bursa		slight pain, but only because of the supraspinatus tendon
Subscapularis tendon		no pain
Supraspinatus tendon		++ pain
Triceps tendon		no pain
Upper and middle thoracic spinous processes		no pain
other		
other		
other		
other		
other		
Did the student specifically test structures that were identified in their previous index of suspicion?		
Did the student **change their index of suspicion** appropriately after completing palpations?		
Did the student identify and **palpate** anatomical structures that permitted them to clearly **eliminate** structures that MAY be involved (i.e., employ a differential diagnosis strategy(ies))?		Should palpate all rotator cuff muscles and attachments, subacromial bursa, long head of biceps
Did the student identify and **palpate** anatomical structures that permitted them to clearly **identify** structures that MAY be involved/lesion site(i.e., employ a differential diagnosis strategy(ies))?		Should feel the supraspinatus tendon and muscle throughout its length

Conclusion		
Is the conclusion/diagnosis correct?		Complete supraspinatus rupture/tear
Did the student follow the correct (ie. thorough and accurate) **PROCESS** even though they may or may not have come up with the correct conclusion/diagnosis)?		
Did the student refer to a physician for a complete diagnosis?		
Overall Impression of the Student's/Examinee's Assessment Skills		
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		

Appendix G Student questionnaire with psychographic and demographic information

Student Orthopedic Assessment Tool Study Questionnaire

1. Name:
2. Date:
3. Student ID:
4. University or College:
5. Sex: M F
6. AGE:
7. Undergraduate GPA (on 4 point scale):
8. Total Number of Courses Taken To Date:
9. Total number of estimated entries into practicum websites:
    a. 0 - 5 times between September 2006 and April 2007
    b. 6 - 10 times between September 2006 and April 2007
    c. 11 - 15 times between September 2006 and April 2007
    d. 16 - 20 times between September 2006 and April 2007
    e. 21 - 25 times between September 2006 and April 2007
    f. 26 - 30 times between September 2006 and April 2007
    g. 31 - 35 times between September 2006 and April 2007
    h. 36 - 40 times between September 2006 and April 2007
    i. greater than 40 times between September 2006 and April 2007
10. Mid-Term exam for orthopedic assessment class(es) (report as a percent-grade):
11. Final Exam for orthopedic assessment class(es) (report as a percent grade):

12. In your opinion, estimate the total contact with the **instructor** related to feedback on your performance of orthopedic assessment skills in class or out of class time (over the course of a semester/term):

    1. low 0-5 hours
    2. medium 6-20 hours
    3. high 21-60 hours (or greater)

13. In your opinion, estimate the total contact with the practicum **supervisor/preceptor** related to feedback on your performance of orthopedic assessment skills out of class time (over the course of a semester/term):

    1. low 0-5 hours
    2. medium 6-20 hours
    3. high 21-60 hours (or greater)

14. I would rank the feedback provided to me by my **instructor** as:

    1. poor
    2. below average
    3. average
    4. above average
    5. outstanding

15. I would rank the feedback provided to me by my practicum **supervisor/preceptor** as:

    1. poor
    2. below average
    3. average
    4. above average
    5. outstanding

By signing this form, I agree to obtain mid-term and final grades from the MRC instructors for PHED 3303, 3301 (please initial beside this)_____
By signing this form, I agree that the primary investigator can obtain website tracking statistics for purposes of this study._____

Appendix H Retrospective questionnaire for instructors and clinical supervisors

Post-SOAT Study Survey
Rationale for Technology Adoption

The purpose of this survey is to follow up on the original SOAT study that integrated the SOAT into various curricula in Canada. The follow-up survey will evaluate the rationale for adoption of the SOAT technology including the website usage and PDA adoption. There are no correct answers, so please try to answer as honestly as you can. *Your participation is completely voluntary. If you choose to return the survey via email, we will consider it your informed consent to participate in this aspect of the study.*

1. In your estimation, how many times did you access the SOAT website (e-SOAT)? Please circle the response that most closely represents the number of times between September 2006 and April 2007:

0 - 5	6 – 10	11 – 15	16 – 20	21 – 25
26 – 30	31 – 35	36 – 40	more than 40	

2. What did you like about the e-SOAT site? (or the hard copy SOAT for Concordia University)

3. What changes would you recommend?

4. What did you like about using the PDA's to assess student performance on the e-SOAT site? (n/a for MRC and Concordia participants)

5. What changes would you recommend?

6. Overall, did you feel the technology was a barrier to using the SOAT more often? Please circle YES or NO.

YES NO

If yes, would you have used a paper version of it more often?

If no, would you have used the paper version just as often as the e-SOAT site?

Appendix I – Ethics Approval Letters

FACULTY OF MEDICINE | UNIVERSITY OF CALGARY

2005-12-15

OFFICE OF MEDICAL BIOETHICS
Room 93, Heritage Medical Research Bldg
3330 Hospital Drive NW
Calgary, AB, Canada T2N 4N1
Telephone: (403) 220-7990
Fax: (403) 283-8524
Email: omb@ucalgary.ca

Dr. L. Katz
Faculty of Kinesiology
University of Calgary
KNB 245
Calgary, Alberta

Dear Dr. Katz:

RE: Research Question: " Does explicit exposure to the Student Orthopedic Assessment Tool (SOAT) compared to non-exposure have an affect on students' performance in performance-based examination?"

Grant ID: 18898

Student: Mark Lafave

The above-noted proposal, including the Research Proposal (Revised Version September 2005), the RN Consent Form (Version 1.0, dated October 11, 2005), the B.S.N Student Consent Form (Version 1.0, dated October 11, 2005), the BSN Student Demographics, the Patient Consent Form (Version 1.0, dated October 11, 2005), the Patient Demographics, the Invitation to Participate in Nursing Research and the RN Key Information Demographics has been submitted for Committee review and found to be ethically acceptable.

Please note that this approval is subject to the following conditions:
(1) appropriate procedures for consent for access to identified health information has been approved;
(2) a copy of the informed consent form must have been given to each research subject, if required for this study;
(3) a Progress Report must be submitted by 2007-12-15, containing the following information:
- i) the number of subjects recruited;
- ii) a description of any protocol modification;
- iii) any unusual and/or severe complications, adverse events or unanticipated problems involving risks to subjects or others, withdrawal of subjects from the research, or complaints about the research;
- iv) a summary of any recent literature, finding, or other relevant information, especially information about risks associated with the research;
- v) a copy of the current informed consent form;
- vi) the expected date of termination of this project.

(4) a Final Report must be submitted at the termination of the project.

Please note that you have been named as a principal collaborator on this study because students are not permitted to serve as principal investigators. Please accept the Board's best wishes for success in your research.

Yours sincerely,

Ian Mitchell, MA, MB, FRCPC

Acting Chair, Conjoint Health Research Ethics Board

GG/mh

c.c. Dr. B. MacIntosh (information) Research Services Mr. M. Lafave (Student)
Office of Information & Privacy Commissioner Ms. Gail Corbett (Communications & Fund Development)

FACULTY OF MEDICINE | UNIVERSITY OF CALGARY

April 26, 2007

OFFICE OF MEDICAL BIOETHICS
Room 93, Heritage Medical Research Bldg
3330 Hospital Drive NW
Calgary, AB, Canada T2N 4N1
Telephone: (403) 220-7990
Fax: (403) 283-8524
Email: omb@ucalgary.ca

Dr. Larry Katz
Faculty of Kinesiology
University of Calgary, KNB 245
Calgary, Alberta

Dear Dr. Katz:

RE: Research Question: " Does explicit exposure to the Student Orthopedic Assessment Tool (SOAT) compared to non-exposure have an affect on students' performance in performance-based examination?"

Ethics ID: 18898

Your request to modify the above-named protocol with the addition of a short questionnaire has been reviewed and approved.

I am pleased to advise you that it is permissible for you to use the revised protocol, based on the information contained in your correspondence of April 24, 2007.

A progress report concerning this study is required annually, from the date of the original approval (2005-11-08). The report should contain information concerning:

(i) the number of subjects recruited;
(ii) a description of any protocol modification;
(iii) any unusual and/or severe complications, adverse events or unanticipated problems involving risks to subjects or others, withdrawal of subjects from the research, or complaints about the research;
(iv) a summary of any recent literature, finding, or other relevant information, especially information about risks associated with the research;
(v) a copy of the current informed consent form;
(vi) the expected date of termination of this project.

Thank you for the attention which I know you will bring to these matters.

Yours sincerely,

Glenys Godlovitch, BA(Hons), LLB, PhD
Chair, Conjoint Health Research Ethics Board

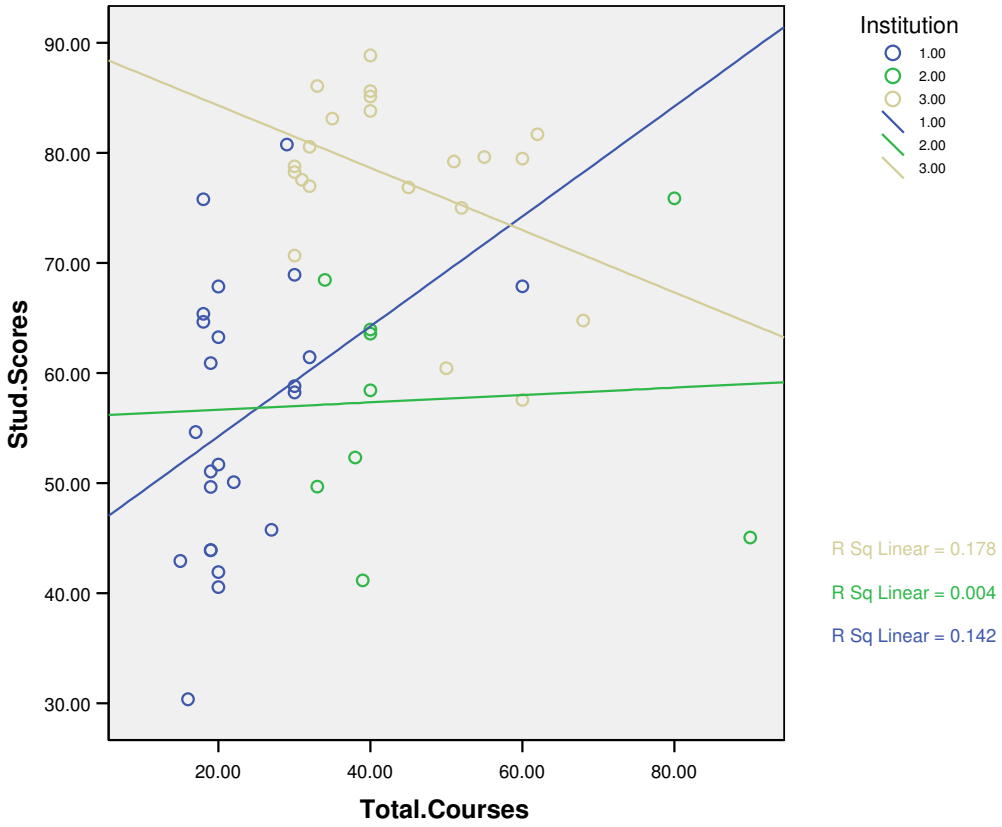
GG/jh

c.c. Adult Research Committee
Mr. Mark Lafave (student)

Appendix J SP global rating scale

Body Region & Condition								
Examiner Name:								
Student Name:								
History								
Nature of Injury Global Rating Scale		0	1	2	3	4	5	
Patient Information Global Rating Scale		0	1	2	3	4	5	
Global Pain Questions Global Rating Scale		0	1	2	3	4	5	
Safety/Rehab/Understanding Global Rating Scale		0	1	2	3	4	5	
Overall History Global Rating Scale		0	1	2	3	4	5	
Observation								
Overall Observation Global Rating Scale		0	1	2	3	4	5	
Active ROM								
Overall Active Range of Motion Global Rating Scale		0	1	2	3	4	5	
Passive ROM								
Overall Passive Range of Motion Global Rating Scale		0	1	2	3	4	5	
Isometric Resisted								
Overall Isometric Resisted Testing Global Rating Scale		0	1	2	3	4	5	
Special Tests								
Overall Special Testing Global Rating Scale		0	1	2	3	4	5	
Palpation								
Overall Palpation Global Rating Scale		0	1	2	3	4	5	
Conclusion								
Is the conclusion/diagnosis correct?		0	**1**	**2**	**3**	**4**	**5**	
If you had to provide an overall grade out of 5, what would you give to the student/ examinee?		0	**1**	**2**	**3**	**4**	**5**	

Appendix K ANCOVA assumption results


Institution
1.00
2.00
3.00
1.00
2.00
3.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
Stud.Scores
20.00
40.00
60.00
80.00
Total.Courses
R Sq Linear = 0.178
R Sq Linear = 0.004
R Sq Linear = 0.142

Preliminary check of homogeneity of variances is listed below with the interaction between institution and total courses. The interaction between the two breaches the assumption to run ANCOVA (.041).

Dependent Variable: Stud.Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power(a)
Corrected Model	6578.973(b)	5	1315.795	12.232	.000	.560	61.158	1.000
Intercept	20613.032	1	20613.032	191.617	.000	.800	191.617	1.000
Institution	2106.173	2	1053.086	9.789	.000	.290	19.579	.977
Total.Courses	56.108	1	56.108	.522	.474	.011	.522	.109
Institution * Total.Courses	733.420	2	366.710	3.409	.041	.124	6.818	.614
Error	5163.562	48	107.574					
Total	237165.415	54						
Corrected Total	11742.535	53						

a Computed using alpha = .05
b R Squared = .560 (Adjusted R Squared = .514)