

2017

Establishing the Reliability and Accuracy of Data Collection for the FIFA 11+ Pre-Participation Study

Tessa E. Kettelkamp

Follow this and additional works at: <https://scholarworks.uvm.edu/hcoltheses>

Recommended Citation

Kettelkamp, Tessa E., "Establishing the Reliability and Accuracy of Data Collection for the FIFA 11+ Pre-Participation Study" (2017). *UVM Honors College Senior Theses*. 196.
<https://scholarworks.uvm.edu/hcoltheses/196>

This Honors College Thesis is brought to you for free and open access by the Undergraduate Theses at ScholarWorks @ UVM. It has been accepted for inclusion in UVM Honors College Senior Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.

Establishing the Reliability and Accuracy of Data Collection for the

FIFA 11+ Pre-Participation Study

Tessa Kettelkamp

Faculty Advisor: Bruce Beynnon, PhD

University of Vermont

College of Nursing and Health Sciences

Table of Contents

<i>Key Terms</i>	3
<i>Abstract</i>	4
<i>Chapter 1-Introduction</i>	6
<i>Chapter 2-Literature Review</i>	10
<i>Chapter 3-Methods</i>	17
<i>Chapter 4-Results</i>	23
<i>Chapter 5-Discussion</i>	33
<i>Conclusion</i>	34
<i>Appendix</i>	37

Key Terms

Outcome Measure: A tool used to collect data, in this study the form developed to record observations of warm-ups.

Component: A warm-up exercise that appears on the form.

Intraclass Correlation Coefficient: The preferred index of correlation coefficients that reflects both correlation and agreement.

Cohen's Kappa Coefficient: A statistic that measures inter-rater agreement for qualitative (categorical) items.

Interrater Reliability: Variation between two or more raters who measure the same group of subjects.

Intrarater Reliability: The stability of data recorded by one individual across two or more trials.

Accuracy: The degree of closeness of measurements of a quantity to that quantity's true value. In this study it would be the degree of closeness of the examiners with the gold standard.

Gold Standard: The best tool for collecting the data, in this study the gold standard was the certified athletic trainer who created the data collection form and recorded the most observations for the FIFA 11+ study.

Sensitivity: Describes the proportion of positive results in relation to the actual positive results. In this study, the sensitivity would be the proportion of times the examiners recorded the component relative to the number of times the gold standard recorded it.

Specificity: Describes the proportion of negative results in relation to the actual negative results. In this study, the specificity would be the proportion of times the examiners did not record the component relative to the number of times the gold standard did not record it.

Abstract

The FIFA 11+ Pre-Participation study at the University of Vermont is investigating the impact that FIFA 11+ has on injuries to high school athletes. The FIFA 11+ Pre-Participation program has been shown to reduce the incidence of lower extremity injury in elite soccer athletes; however, it is unclear if the program has a similar effect on developing high school athletes or if it can reduce injury in other sports. Recognizing that an athletes' compliance with the FIFA 11+ program may be directly linked to the effectiveness of the program, an outcome measure that documented compliance was developed. The outcome measure that was developed was a form designed to record pre-participation components of a warm-up by observers for the FIFA 11+ study. The objective of this investigation was to establish the intrarater and interrater reliability and accuracy of this compliance outcome measure.

A repeated-measures study design was used to determine the reliability and accuracy of the outcome measure. The examiners who collected data for the FIFA 11+ study were asked to volunteer for this investigation, which involved attending two observation sessions that were two weeks apart. The observation sessions involved watching five warm-up videos, each one about ten minutes long, and then recording what occurred in the warm-up. They used the outcome measure to record their observations of the same five pre-participation warm-ups during each session. The outcome measure had 66 warm-up exercises, or components, that could be recorded. Intraclass Correlation Coefficients (ICCs) were used to determine the intrarater and interrater reliability for each component of the outcome measure. A component with an ICC above 0.60 was considered reliable for this study. The sensitivity and specificity of each component, as well as percent agreement of the examiners with the gold standard examiner for each component were used to determine the accuracy. A component was accurate if above 60% of the observations

were in agreement with the gold standard examiner that the component either was or was not present in the warm-up. If any components were proven unreliable or inaccurate the outcome measure was simplified by reducing the number of components. The new components, which were each a result of combining two unreliable components, had ICCs, sensitivity and specificity recalculated as if all observations of either of the original components counted toward the new component.

The outcome measure was established to be partially reliable and partially accurate. Out of the 34 components observed there were five components that were intrarater unreliable and 18 components that were interrater unreliable. All of the components that were intrarater unreliable were also interrater unreliable. Only one component was inaccurate with 58% of the observations of that component in agreement with the gold standard examiner's observations. Of the total 18 unreliable components, seven were combined with another component to simplify the outcome measure.

Chapter 1- Introduction

Background

As participation in sports increases, so do the injuries and the health care resources needed to provide medical treatment. The cost to treat these injuries is high, as is the cost of the associated morbidity and long-term sequela. The United States Consumer Product Safety Commission estimates that from 1997-2002 the United States spent \$69,504,000 on injuries associated with participation in sport (“Injury Statistics,” 2015). From 1997-2002, United States emergency departments treated 1,695,790 athletes for sport related injuries and 4,477,950 athletes sought medical treatment outside of the emergency department (“Injury Statistics”, 2015).

The need for injury prevention was glaring and, in response to that need, pre-participation injury prevention programs were developed. One of the more popular programs is FIFA 11+, a warm-up program that is focused on enhancing strength, flexibility, and neuromuscular control and has been proven effective in reducing the rate of injury in elite soccer teams (Bizzini, 2013; Silvers-Granelli, 2015; Soligard, 2008). The program was originally developed in 2003 as just the “11”, but updated in 2006 to the “11+” to be a more comprehensive warm-up and not just a strengthening program (FIFA 11+; a complete warm-up programme). FIFA 11+ was designed to target core muscles and gradually progress the athletes as they build strength and master each level of the program (“FIFA 11+; a complete warm-up programme”). The twenty-minute program integrates running, strength, agility, and balance training to prepare the athlete for the stress it experiences during participation in sport (Bizzini, 2013).

Injury prevention could benefit more than just the elite athletic population. A portion of those millions of athletes going to the emergency department, or seeking medical treatment outside the emergency department, are young teenagers who are competing at a more recreational level of sport (“Injury Statistics”, 2015). There was a need to prevent injuries for all athletes and although the eradication of athletic injuries is not possible, reducing the incidence of injuries is (Bizzini, 2013; Silvers-Granelli, 2015; Soligard, 2008).

The FIFA 11+ program has been studied with professional soccer athletes, but not with adolescent athletes participating in other sports (Bizzini, 2013; Silvers-Granelli, 2015; Soligard, 2008). The University of Vermont decided to conduct a study that was designed to determine the effectiveness of the FIFA 11+ program to reduce lower extremity injury in male and female high school athletes who participated in different sports (soccer, football, basketball, and lacrosse). The FIFA 11+ study was a three-year cluster-randomized control trial (with high school as the unit of randomization) that was designed to determine if the FIFA 11+ program could reduce the incidence of lower extremity injury in high school athletes.

In the first year of the study, data were collected to determine what exercises high school teams used for their pre-participation warm-up. Eleven examiners including undergraduate students, medical students, a medical resident, and a certified athletic trainer traveled to the high schools that were a part of the study and observed the teams’ pre-practice warm-up to document what they did. Data were collected on an outcome measure developed by the research team and this included characterization of the type of activity, the duration of the activity, and how well the athletes performed the activity.

Research Question

The goal of this study was to answer the question: is the compliance outcome measure for the FIFA 11+ study reliable, between and within observers, and accurate when documenting high school athletic team warm-ups? If the outcome measure was proven to be unreliable and inaccurate, what changes were made to produce reliable and accurate data and how did those changes make a difference in the estimated results? The second part of this study was using the data to find the unreliable and inaccurate components and making them even simpler to obtain estimated reliable and accurate results. Establishing reliability and accuracy of the outcome measure prior to the FIFA 11+ study was essential.

Purpose

There is a potential for error with any data collection. The purpose of this study is to determine if the compliance outcome measure that the FIFA 11+ study used was reliable (both within and between observers) and accurate. A comprehensive understanding of how to document pre-participation warm-up with the FIFA 11+ program is essential to studies designed to determine the effectiveness of the program at reducing the incidence of sport related lower extremity injury. For this reliability and accuracy study the gold standard was the examiner who had created the outcome measure and obtained the majority of the data for the FIFA 11+ study.

Ascertaining the reliability and accuracy of an outcome that can be applied to document an athletic teams' compliance with the FIFA 11+ pre-participation program was an essential first step that needed to be accomplished prior to conducting a Randomized Control Trial (RCT) that evaluated the efficacy of the program to reduce the incidence of lower extremity injury. With an accurate and reliable outcome measure, future RCTs will benefit from the use of the outcome measure to quantify what proportion of the program the teams complete. These data can be used

to establish what the teams complete over the course of a season and help provide insight into what aspect of the program contributed to injury reduction.

Hypothesis

The FIFA 11+ research team at the University of Vermont created the outcome measure to record what the high school teams were doing as their pre-participation warm-up throughout the three years of the study. The same outcome measure was used to document FIFA 11+ warm-ups and control group warm-ups. An outcome measure has never been created for the purpose of observing high school warm-ups. We hypothesized that the outcome measure created to log compliance for the FIFA 11+ Pre-Participation study was reliable and accurate when used to document warm-ups of high school athletic teams. If the results proved the form to be unreliable and inaccurate, the simplifications the research team would make to the outcome measure would create an estimated reliable and accurate form.

Chapter 2- Literature Review

Search Strategy

To begin searching for research related to injury prevention, I went to the federal government's Consumer and Product Safety Commission website to view the injury statistics for 1997-2003. The information is located under the Research and Statistics tab and then under Sports Activities and Equipment (Excluding Major Team Sports). The chart reflects data collected in Emergency Rooms in the United States on injuries related to sports ("Injury Statistics", 2015).

Reviewing the literature came down to two major components, reviewing previous research done specifically on the efficacy of the FIFA 11+ warm up program at reducing the incidence of lower extremity injury in different populations and reviewing previous research done on reliability and accuracy of outcome measures unrelated to FIFA 11+ or injury prevention. For previous FIFA 11+ studies, going to the National Center for Biotechnology Information website, PubMed, and searching "FIFA 11+" resulted in 105 hits. Narrowing the search down by adding the term "injury prevention" resulted in 40 hits. When adding the word "compliance" to the search, there were only six hits.

One of the articles "Efficacy of the FIFA 11+ Injury Prevention Program in the Collegiate Male Soccer Player" referenced another article written on the implementation of FIFA 11+ titled, "Implementation of the FIFA 11+ football warm up program: How to approach and convince the Football associations to invest in prevention"(Bizzini, 2013; Silvers Granelli, 2015). References that were cited by articles found on PubMed was one of the methods used to discover more research on FIFA 11+ injury prevention.

The Journal of Athletic Training and the Athletic Training Education Journal published reliability and accuracy studies done on outcome measures unrelated to FIFA 11+ but which used similar research designs as this study. Issues released in the last ten years for the Journal of Athletic Training and the Athletic Training Education Journal both had studies pertaining to other data collection tools and were useful to prove that by using a specific research design, the tools can be proven reliable and accurate.

Previous Studies

The reliability or accuracy of the outcome measure that was developed to monitor compliance with the FIFA 11 program has not been established. Research has focused on establishing the reliability and accuracy of other outcome measures used to score athletes on functional movements. There have been other studies done on the FIFA 11+ program, but none were focused on the population that was used for this FIFA 11+ study and none discussed a compliance outcome measure being used.

There are three major studies published that had statistically significant evidence of FIFA 11+ reducing the incidence of lower extremity injuries. The only participants were already active athletes, none were previously sedentary. In 2008 a study was published from Norway that took an intervention group of over one thousand female soccer players and put them through a “comprehensive warm-up programme” that was really the new FIFA 11+ program (Soligard, 2008). Another study was done on NCAA Division I and Division II collegiate men’s soccer teams that used a randomized control trial study design and implemented the FIFA 11+ program weekly in 27 teams (Silvers-Granelli, 2015). In two different studies where the intervention groups participated in the FIFA 11+ program the same number of sessions during the two seasons, the females’ season was twice as long. This meant the females participated in the

program the same number of times as the males but over a greater period of time (Soligard, 2008; Silvers-Granelli, 2015). Since the FIFA 11+ study at the University of Vermont introduces a new population that has not been observed participating in the program, it is important to look at the populations that have already been observed in other studies. The study being done at the University of Vermont is looking at high school athletes of both genders participating in a variety of sports. The researchers have reason to believe FIFA 11+ will be successful in reducing injury rates because of the success the program has had with both genders in older age groups.

The results of the study done in Norway with only female athletes showed that the control group had more than twice as many injuries as the intervention group over the eight-month season (Soligard, 2008). The figure that shows the overall injury, severe injury, and overuse injury data concluded that the warm-up programme was effective in preventing injury among the 13-17 year old female soccer athletes (Soligard, 2008).

The study done with NCAA Division I and II men's soccer teams took place over the course of one season and analyzed injuries that were recorded by the team's certified athletic trainer. Within that study any complaints sustained during play were considered injuries. The data yielded a 46.1% decrease in injuries as well as reducing the amount of time lost from the sport due to injury by 28.6% (Silvers-Granelli, 2015). A drawback of the study was the way they defined injuries because it encompassed every type of injury and every level of severity. It can be good to encompass everything to be sure no injury was overlooked and the data was as accurate as possible, however it could also lead to information that was too broad.

The studies done with collegiate male athletes and teenage female athletes strongly supported the use of FIFA 11+ in athletic teams, however the intensity of the sport can change the results of the study. Collegiate athletes can be assumed to have more discipline than high

school athletes and perform at a much higher level of play. The collegiate teams were likely to be more invested in their well being and more dedicated to improving themselves physically. The time that FIFA 11+ requires and the necessary level of focus could be factors in why the program is effective in that setting. The University of Vermont FIFA 11+ study is conducted on high school athletes, a population that does not have that motivation and may go a few days without performing the warm-up correctly or at all.

If the program yields similar results even though it is done half as often in one of the studies, it is reasonable to conclude that the implementation of the program is effective despite the frequency with which it is used (Soligard, 2008). The study in Norway showed that athletes complied with the program at a high rate and the more compliant they were with it the lower the rate of their injuries were (Soligard, 2008). There is an effort to make the program universally used by coaches in an effort to maintain the health of their athletes and in Switzerland and New Zealand these campaigns have been successfully implemented in smaller populations and have been assumed to lower the injury rate of the athletes (Bizzini, 2013).

One study was not able to show a decrease in injury rates but the researchers provided a theory why. FIFA 11+ is used to replace the warm-up that teams typically would otherwise do. It contains exercises and activities that focus on strength, plyometric, balance, and running which other warm-ups typically lack. In a study done in 2008 that researched the original “FIFA 11” program, the cluster-randomized control trial used female youth soccer players to test the new program over four months of their season (Steffen, 2008). Researchers believe the reason that the trial concluded with no reduction in injury rates was because a significant portion of the intervention group did not participate in the program with the frequency that the others did (Steffen, 2008). The researchers stated that fourteen out of the fifteen teams in the intervention

group participated in the program less than twenty times in the four months of season (Steffen, 2008).

Instead of specifically studying the incidence of lower extremity injuries, some studies have taken a physiological approach to prove that what happens at a cellular level can help an athlete protect themselves when abnormal stresses are placed on the body. In a previous study done by implementing FIFA 11+ for four weeks in eleven males, glucose uptake increased in relation to the muscle activity of five major muscles in the legs (Takata, 2016). Glucose uptake is the activity that occurs when a muscle is contracted and glucose is transported from storage to the plasma membrane of the muscle cells, powering the muscle to complete the contraction (Richter, 2012). The study has a small sample size and does not look at other factors in the men, such as activity outside of training, intensity of training, or level of performance. However, the study concludes that there is speculation that this glucose uptake could be a contribution to a decrease in injuries related to sports (Takata, 2016). This is not a direct correlation nor did the study acknowledge the number or type of injuries acquired before, during, or after the study took place but it does address the glucose uptake, a concept that could influence injuries.

It was clear that the FIFA 11+ study needed to be conducted, but no other studies done on FIFA 11+ used the outcome measure that was used by the researchers at the University of Vermont. Other data collection tools in the medical community have been researched to test their reliability and accuracy. Comparing the methods that the other studies used to the methods used to test the reliability and accuracy of the FIFA 11+ outcome measure can assist in interpreting the results. If the results of the FIFA 11+ study show that there was no injury reduction, the outcome measure can provide information on how compliant the teams were when they were part of the intervention group. If the outcome measure can reflect that the team did not complete

the warm-up in its entirety then there is a reason the incidence of lower extremity injury was not impacted.

Test-retest reliability was analyzed in two studies done to show results of medical screening, one involving the Functional Movements Screen and the other done on the SCAT3. These methods of recording the severity of an athlete's injuries are used clinically and taught at institutions to aspiring healthcare providers. Researchers at Stanford University conducted the study on the Functional Movements Screen (FMS) in 2013 and the test-retest and interrater reliability were analyzed by using 6 raters, 39 participants, and 2 trials (Shultz, Anderson, Matheson, Marcello, & Besier, 2013). The results were analyzed by using Intraclass Correlation Coefficients and setting standards for how high the correlations needed to be to be considered "poor", "fair", "good", or "excellent" (Shultz et al., 2013). The possible results in this study were scores ranging from 0 to 3 for each of the 7 tasks that comprise the FMS (Shultz et al., 2013). In a similar study done at the University of Toledo in 2013 the Star Excursion Balance Test (SEBT) was tested for interrater reliability by using 5 raters, 29 participants, and 2 trials (Gribble, Kelly, Refshauge, & Hiller, 2013).

The study conducted to look at the reliability of the SCAT3 was done over two hockey seasons using 179 professional ice hockey players (Hanninen, 2016). The study resulted in the SCAT3 showing slight inconsistencies but was still considered within the athletes' normal ranges and therefore determined that the testers were interchangeable and the SCAT3 would yield significant results if used correctly.

A study done on goniometry use among physical therapists, "Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example", uses intraclass correlation coefficients as the best statistic for

interrater reliability (Ellaszlw, 1994). There were raters who took more than one goniometric measurement at different times and the researchers wanted to look at all of the measurements in comparison to each other and not just one measurement from each rater. A minor hindrance they faced was that when using an ICC to determine the interrater reliability it is difficult, if not impossible, to compare that ICC to an ICC from other studies that researched the same measurements (Ellaszlw, 1994). The number of subjects in the study can influence the ICC significantly and that can make it difficult to compare across studies (Portney, 2000).

The studies that analyzed the reliability of a tool were pertinent to the reliability and validity study for FIFA 11+ because they successfully proved, using a test-retest reliability format, that their examiners were accurate.

Summary of what is known and not known about the research topic

Research on reliability and accuracy of data collection using the outcome measure for the FIFA 11+ study does not exist, which is why this study is necessary. Though the studies done on accuracy and reliability of other outcome measure in athletic testing show some reliability, they show how necessary the studies are to be sure the data recorded using those tools was as accurate as possible. Research does exist on implementing FIFA 11+, almost entirely done outside of the United States and almost all of which revolve around soccer programs but also show significant results in reduction of sports-related injuries. The studies that have been done provide a good starting point to show there was a correlation between a reduction in injuries and FIFA 11+ use and a need to delve more into how effective FIFA 11+ can be on different populations and sports.

Chapter 3- Methods

Research Design

This reliability and accuracy study included eleven examiners who were using this outcome measure to observe the high school teams for the FIFA 11+ Pre-Participation study. They were each asked to participate by coming to the University of Vermont and spending an hour and a half during each session to record their observations of a few warm-ups they watched.

The repeated-measures study design tested the intrarater and interrater reliability and accuracy of our examiners when using the outcome measure. The results would be collected and analyzed and if considered unreliable and/or inaccurate the form would be modified to be reliably and accurately used in the future. The modification meant combining the unreliable or inaccurate components and combining them with another component to increase the number of observations for that larger component, which would result in an estimated ICC that was higher than the original. The process meant that of the two components that were combined, if an examiner had recorded one of them then it would count as an observation for the new component, if the examiner had recorded both of the components then it would count as one observation for the new component, and if the examiner had not recorded either component that it would not count as an observation of the new component.

Study Samples

The sample used for this study were the eleven examiners that had volunteered from the University of Vermont community including five undergraduate students, two post-baccalaureate pre-medical students, one medical student, two medical residents, and one certified athletic trainer. The volunteers came to the research team for the FIFA 11+ study and the team determined if they were going to be dedicated throughout the study and if they were competent

individuals who had at least part of an undergraduate education. If they met those criteria then the research team accepted them as volunteers. The examiners used for this study were not given compensation for their time or transportation. They participated on a volunteer basis for the FIFA 11+ study and were asked again to volunteer to be a part of the reliability and accuracy study. They did not need to sign a consent form because nothing identifying them was going to be used for this study. There were no examiners that refused to be a part of this study, however, one examiner could only make it to the first observation and had to be dropped from the study. Therefore, even though the study began with twelve examiners, only the data collected from eleven examiners was used.

The only personal information we needed from the examiners was their names and the level of education they had. When we recorded their names from the form onto the spreadsheet of data all of the data had to be coded into numbers. Each examiner was assigned a number and coded accordingly. The statistician who analyzed the data was the only other person who knew the numbers assigned to the examiners.

Instruments

There were five videos made by recording high school athletes participating in a warm-up that they did on a regular basis. The videos were between seven and ten minutes long. The Institutional Review Board approved recording the videos of the athletes as long as their faces were blurred to the point of not being able to be identified because most of the athletes were minors. The athletes that were in the videos were all a part of an organized school team from Milton High School in Vermont. Notices were sent out to the guardians of the athletes informing them that this study was going to be happening and if they did not want to participate they could opt out of the videotaping. The videos were projected to the examiners via laptop. The only

person who knew the number of the video relating to the team that was being watched was the author of this thesis.

The outcome measure used for the FIFA 11+, located in Appendix A, was created by the research team, referred to as the pre-participation warm-up form, and was used by every examiner at every warm-up they observed. The outcome measure was used even for warm-ups that did not occur, as there was a part of it to indicate that there was no warm-up. The outcome measure was modified many times before this study to reflect what the research team felt could capture the most accurate and detailed information. Information on the outcome measure included the name of the data collector and the date the data was taken to prevent any mix up of data. The outcome measure had mostly yes or no questions and numbers for recording the order that the activities were done in, but also gave the opportunity for write-in components. On the second page of the outcome measure there is component labeled “high knees”. As an example, if that component was performed first during a warm-up the examiner would write a “1” to the left of the component and then check off if it was done at a walking pace or a jogging pace. The write-in components allowed for a lot of freedom which is up to the examiner’s discretion to determine if the movement could fall into another category or needs a whole new category.

The front page of outcome measure consisted of entirely FIFA 11+ movements and was formatted that way to give the examiner an opportunity to collect information for schools or teams who claimed to conduct FIFA 11+ warm-ups. Most of the teams that claimed to run through FIFA 11+ every day were actually only going through the first section of the program and doing it incorrectly. However, the outcome measure had a section to record how well or complete they were doing it.

Procedures

The examiners watched five recordings of anonymous high school students participating in a pre-participation warm up and recorded their observations on the outcome measure that was used during the study. They recorded what the majority of the students were doing, in case a few decided to stray from the activity. Two weeks later they watched the same five recordings and recorded that their observations on the outcome measure. The examiners were given instructions to record the components just as though they were watching the warm up at the high school. They were not given the information for which high school the videos were taken at, to protect the identity of the team and the athletes.

After the data had been collected from the two sessions of video watching it was entered from the outcome measures into an excel spreadsheet. There were instructions on how to code each part of the outcome measure so it could be entered the same way data throughout the FIFA 11+ study were entered. Each yes recorded was a one and each no was a two and the only section of the outcome measure that was recorded without a number was the open ended section allowing examiners to record any components that may not have already appeared on the outcome measure.

Data Analysis

The intrarater and interrater reliability were analyzed by computing the intraclass correlation coefficient (ICC) between observers and within observers for each component that was recorded during all of the five observations. ICCs were chosen for analysis over kappa coefficients because kappa coefficients only apply to nominal data, such as yes/no or true/false results. Since the interrater and intrarater reliability of this study also include the time that each

component was performed, as well as other interval measurements, ICCs were considered the best measurement tool. The closer to 1.0 the ICC was, the more reliable it was considered.

Percent agreement with the gold standard, sensitivity, and specificity were used to measure the accuracy of each component that was observed. There were a total of 66 components on the outcome measure, 34 of those components were observed and used to determine reliability. In table 1, the components observed have an interrater and intrarater reliability calculated in the column on the right. The percent agreement took into consideration every time the gold standard recorded the component and during which videos, and every time the examiner recorded the component and during which videos. This was shown as a percent out of 100. The four components that were not a part of the percent agreement analysis were the duration components, because they would have shown an agreement of 0% if they did not have the exact time that the gold standard examiner recorded, which would have been skewed. A total percent agreement was calculated based on how many components the examiners agreed with the gold standard on.

The sensitivity and specificity, which measure how often the technique identifies the true positive and true negative results when compared to the gold standard, were measured for 26 components that were also used for the percent agreement analysis, for each examiner. Only 26 of the components were measured because they were nominal components and not interval measurements. Sensitivity is a number that describes the proportion of positive results that a technique finds compared to the actual positive results. In this study the technique would be the different examiners and the sensitivity for each component reflects how often the examiner identified that component relative to how often the gold standard identified it. Specificity, on the other hand, is descriptive of the proportion of negative results that a technique finds compared to

the actual negative results. Therefore, specificity in this study reflected how often the examiner did not identify the component relative to how often the gold standard did not identify the component.

After performing the analysis, any components that produced ICCs below what was considered reliable by the research team, 0.60, were extracted from the results. The research team decided 0.60 would be the lowest ICC they would want to consider reliable in this study. Through re-watching the videos and checking the components with the low ICCs and when they were recorded, it was determined that if two components were recorded by a majority of the examiners at the same point in the videos then they were likely similar in their appearance. Components with insufficient ICCs could be simplified if they had a counter-component that was observed at the same time and was considered very similar in appearance by the research team. Those similar components were collapsed to perform a new analysis of the data. This helped the researchers modify the data collection form and make it simpler than the original to ensure more accurate data collection for the remainder of the FIFA 11+ study.

Chapter 4- Results

Reliability

The results of the reliability portion were analyzed through intrarater ICCs and interrater ICCs as shown on Table 1. Table 1 contains all of the components on the form, the proportion of observations each component was recorded in, and the intrarater ICCs and interrater ICCs for each component. The components that show a proportion of observations above zero, meaning they were observed during at least one video, but do not show any ICC had an ICC so close to zero that it was negligible. The interrater ICCs ranged from 0.15 to 1.00 but only 12 components had an interrater ICC above 0.60 and were considered reliable. The intrarater ICCs ranged from 0.25 to 1.00 and 25 components had an intrarater ICCs above 0.60 and were considered reliable. Components 6, 7, 27, 28, and 46 were completely reliable with ICCs at 1.00. Every examiner recorded observations for every video, which meant there were 110 total warm-up observations being analyzed.

The components on the form that fell below the minimum reliable ICC were traced back and determined to have a counterpart component that was similar in appearance and also fell below the minimum reliable ICC. Not every component had another component that was similar to it. When two components were matched up this way, it was decided to combine them into one component to produce an estimated reliability, based on recalculations, would have been the ICC if the components were the same from the beginning. When the data was collapsed this way it was analyzed again and produced results that gave much higher intrarater and interrater ICCs as shown in Table 4. For example, components 53 and 54 were both variations on a lunge and when they were standing alone 53 had an interrater ICC of 0.77 and an intrarater ICC of 0.80 and 54

had an interrater ICC of 0.57 and an intrarater ICC of 0.57. However, after the components were combined into “Front Lunge” they had an interrater ICC of 0.77 and an intrarater ICC of 0.81. In the data that was collapsed the lowest interrater ICC was 0.69 and the highest was 0.94. The lowest intrarater ICC in the collapsed data was then 0.81 and the highest was 1.00. The collapsed data all fell above the threshold for the ICC that is considered reliable.

Accuracy

The samples in this study consisted of five undergraduate students, two post-baccalaureate pre-medical students, two medical students, one medical resident and one certified athletic trainer. The certified athletic trainer was the gold standard for this study. For the accuracy component of the study, all the examiners fell between 80 total percent agreements and 90 total percent agreements with the gold standard except for one examiner, an undergraduate student, who had 69.6 total percent agreements with the gold standard. The results for the total accuracy of the examiners were very close in range except for the one outlier.

Table 3, below, represents the percent agreement portion of the accuracy results of the study. For each component that was recorded, the percent of examiners who agreed with the gold standard that it was or was not performed is recorded. The total percent agreement between the examiner and the gold standard appear at the bottom of the table. The lowest percent agreement for a component was 58% for the lean hold exercise and the highest percent agreement for a component was 100% for the knee to chest exercise. For each of the examiners their percent agreement with the gold standard overall was calculated and appears at the bottom of table 3.

Table 1-Interrater and Intrarater reliability for each component expressed in the form of an ICC; Proportion of observations that the component appeared in.

COMPONENT	PROPORTION OF OBSERVATIONS	Interrater ICC	Intrarater ICC
1.FIFA COMPONENTS USED	0.46	0.56*	0.97
2.FIFA PART 1 RUNNING-ANY COMPONENTS	0.42	0.69	0.87
3.FIFA PART 2 STRENGTH-ANY COMPONENTS	0.00		
4.FIFA PART 3 RUNNING- ANY COMPONENTS	0.00		
5.NON-FIFA RUNNING COMPONENTS-ANY	0.82	0.26*	0.73
6.NON-FIFA DYNAMIC MOBILITY-ANY	1.00	1.00	1.00
7.NON-FIFA DYNAMIC STRETCH-ANY	1.00	1.00	1.00
8.NON-FIFA STATIC STRETCH-ANY	0.00		
9.NON-FIFA STRETCH ON OWN	0.00		
10.NON-FIFA STRENGTH COMPONENT-ANY	0.04		
11.NON-FIFA PLYOMETRIC COMPONENT-ANY	0.00		
12.NON-FIFA AGILITY/BALANCE COMPONENT-ANY	0.00		
13.NON-FIFA SPORT SPECIFIC-ANY	0.05		
14.NON-FIFA SPORT SKILLS	0.02		
15.NON-FIFA JOGGING	0.64	0.66	0.73
16.NON-FIFA RUN-STRAIGHT AHEAD	0.21	0.22*	0.64
17.NON-FIFA RUN-BACKWARDS	0.18	0.81	0.90
18.NON-FIFA RUN-SIDE SHUFFLE	0.19	0.77	0.77
19.NON-FIFA RUN-KARAOKE	0.15	0.75	0.94
20.NON-FIFA RUN-SKIPPING	0.00		
21.NON-FIFA RUN-VERTICAL JUMP	0.00		
22.NON-FIFA RUN-INCR PACE	0.53	0.33*	0.67
23.NON-FIFA RUN-FRONT/BACK	0.05		
24.NON-FIFA RUN- SIDE TO SIDE	0.03		
25.NON-FIFA RUN- DIAGONALS	0.00		
26.NON-FIFA SS RUN/SPRINT	0.05		
27.NON-FIFA HIGH KNEE	0.99	1.00	1.00
28.NON-FIFA BUTT KICK	0.98	1.00	1.00
29.NON-FIFA LEG FRNT BK	0.29	0.30*	0.68
30.NON-FIFA LEG FRNT DIAG	0.68	0.26*	0.65
31.NON-FIFA LEG BK DIAG	0.08		
32.NON-FIFA LEG FORW TOUCH	0.30	0.72	0.74
33.NON-FIFA LEG ADD ABD	0.00		
34.NON-FIFA LEG ADD	0.00		
35.NON-FIFA LEG ABD	0.00		
36.NON-FIFA HIP IN	0.55	0.44*	0.67
37.NON-FIFA HIP OUT	0.55	0.48*	0.74
38.NON-FIFA HIP INT ROT	0.05		
39.NON-FIFA HIP EXT ROT	0.08		
40.NON-FIFA POWER KARAOKE	0.04		
41.NON-FIFA LUNGE	0.20	0.34*	0.58*
42.NON-FIFA SQUAT	0.00		

43.NON-FIFA ANKLES	0.00		
44.NON-FIFA OTHER	0.12	0.15*	0.25*
45.NON-FIFA OTHER2	0.01		
46.NON-FIFA KNEE CHEST	0.40	1.00	1.00
47.NON-FIFA HEEL BUTT	0.22	0.92	0.92
48.NON-FIFA HL BTT BEND	0.02		
49.NON-FIFA LEAN HOLD	0.23	0.25*	0.47*
50.NON-FIFA LEAN SCOOP	0.37	0.44*	0.69
51.NON-FIFA LEAN SWEEP	0.00		
52.NON-FIFA GLUT HOLD	0.18	0.90	0.90
53.NON-FIFA FRNT LUNGE HOLD	0.62	0.77	0.80
54.NON-FIFA FRNT LUNGE UB ROT	0.14	0.57*	0.57*
55.NON-FIFA BK LUNGE HOLD	0.01		
56.NON-FIFA BK LUNGE UB ROT	0.00		
57.NON-FIFA SIDE LUNGE	0.45	0.50*	0.71
58.NON-FIFA QRT LUNGE	0.00		
59.NON-FIFA COMB MOVE	0.00		
60.NON-FIFA OTHER	0.18	0.81	0.90
61.NON-FIFA OTHER	0.09		
62.TOTAL WARM UP	7.66	0.96	0.96
63.FIFA PART 1 RUNNING	0.55	0.30*	0.77
64.NON-FIFA RUNNING	1.05	0.36*	0.41*
65.NON-FIFA JOGGING	0.63	0.52	0.78
66.NON-FIFA DYNAMIC MOBILITY AND STRETCH	3.99	0.08*	0.81

*ICCs that fell below the threshold of reliability, 0.60

Table 2- Sensitivity and Specificity for each nominal component observed reflecting the accuracy for each- only 10 examiners because the 11th is the gold standard that these examiners are being compared to.

COMPONENT	PERFORMED		NOT PERFORMED		TOTAL (N=100) % CORRECT
	N	SENSITIVITY	N	SPECIFICITY	
1. FIFA COMPONENTS USED	40	90.0	60	81.7	85.0
2.FIFA PART 1 RUNNING	40	90.0	60	90.0	90.0
5.NON-FIFA RUNNING COMPONENTS-ANY	80	90.0	20	50.0	82.0
15.NON-FIFA JOGGING	60	93.3	40	80.0	88.0
16.NON-FIFA RUN-STRAIGHT AHEAD	40	32.5	60	90.0	67.0
17.NON-FIFA RUN-BACKWARDS	20	85.0	80	98.8	96.0
18.NON-FIFA RUN-SIDE SHUFFLE	20	85.0	80	97.5	95.0
19.NON-FIFA RUN-KARAOKE	20	75.0	80	100.0	95.0
22.NON-FIFA RUN-INCR PACE	80	62.5	20	100.0	70.0
29.NON-FIFA LEG FRNT BK	0		100	68.0	68.0
30.NON-FIFA LEG FRNT DIAG	100	65.0	0		65.0
32.NON-FIFA LEG FORW TOUCH	20	90.0	80	83.8	85.0
36.NON-FIFA HIP IN	60	78.3	40	80.0	79.0
37.NON-FIFA HIP OUT	60	80.0	40	82.5	81.0
41.NON-FIFA LUNGE	20	60.0	80	90.0	84.0
44.NON-FIFA OTHER	0		100	88.0	88.0
46.NON-FIFA KNEE CHEST	40	100.0	60	100.0	100.0
47.NON-FIFA HEEL BUTT	20	100.0	80	97.5	98.0
49.NON-FIFA LEAN HOLD	40	22.5	60	81.7	58.0
50.NON-FIFA LEAN SCOOP	20	50.0	80	62.5	60.0
52.NON-FIFA GLUT HOLD	20	90.0	80	100.0	98.0
53.NON-FIFA FRNT LUNGE HOLD	60	95.0	40	90.0	93.0
54.NON-FIFA FRNT LUNGE UB ROT	20	65.0	80	98.8	92.0
57.NON-FIFA SIDE LUNGE	40	82.5	60	80.0	81.0
60.NON-FIFA OTHER	20	85.0	80	98.8	96.0
61.NON-FIFA OTHER	0		100	91.0	91.0

Table 3- Percent agreement by each examiner with the gold standard for each nominal component observed- only 10 examiners because the 11th is the gold standard that these examiners are being compared to.

COMPONENTS	Examiner									
	2	3	4	5	6	7	8	9	10	12
FIFA: MAJOR COMPONENTS										
1.FIFA COMPONENTS USED (CHECKED ON FORM)	100	100	100	40	60	100	100	60	90	100
2.FIFA PART 1 RUNNING – ANY COMPONENTS	100	100	100	60	90	100	100	60	90	100
NON-FIFA: MAJOR COMPONENTS										
5.NON-FIFA RUNNING COMPONENTS-ANY	70	90	80	60	80	100	70	80	100	90
NON-FIFA: INDIVIDUAL RUNNING COMPONENTS										
15.NON-FIFA JOGGING	100	90	90	70	80	100	70	80	100	100
16.NON-FIFA RUN-STRAIGHT AHEAD	60	60	70	60	60	60	80	80	70	70
17.NON-FIFA RUN-BACKWARDS	100	100	100	70	100	100	100	100	100	90
18.NON-FIFA RUN-SIDE SHUFFLE	100	100	100	90	100	100	100	70	100	90
19.NON-FIFA RUN-KARAOKE	100	100	100	80	100	100	80	100	100	90
22.NON-FIFA RUN-INCR PACE	90	100	90	20	50	90	50	100	80	30
NON-FIFA: INDIVIDUAL DYNAMIC MOBILITY COMPONENTS										
29.NON-FIFA LEG FRNT BK	60	70	70	80	30	70	100	50	60	90
30.NON-FIFA LEG FRNT DIAG	60	60	70	70	30	70	90	50	60	90
32.NON-FIFA LEG FORW TOUCH	100	90	80	80	90	80	80	80	80	90
36.NON-FIFA HIP IN	80	90	80	70	90	100	30	70	80	100
37.NON-FIFA HIP OUT	80	90	80	70	90	100	40	80	80	100
41.NON-FIFA LUNGE	70	80	100	50	80	90	80	90	100	100
44.NON-FIFA OTHER	100	100	100	60	100	80	80	80	100	80
NON-FIFA: INDIVIDUAL DYNAMIC STRETCH COMPONENTS										
46.NON-FIFA KNEE CHEST	100	100	100	100	100	100	100	100	100	100
47.NON-FIFA HEEL BUTT	90	100	100	90	100	100	100	100	100	100
49.NON-FIFA LEAN HOLD	60	60	50	50	80	60	40	40	70	70
50.NON-FIFA LEAN SCOOP	60	60	50	60	80	60	40	50	70	70
52.NON-FIFA GLUT HOLD	100	100	100	100	90	100	100	90	100	100
53.NON-FIFA FRNT LUNGE HOLD	100	90	100	70	90	100	100	100	90	90
54.NON-FIFA FRNT LUNGE UB ROT	90	90	100	90	90	100	100	90	90	80
57.NON-FIFA SIDE LUNGE	70	60	100	50	70	90	80	90	100	100
60.NON-FIFA OTHER	100	100	100	80	100	100	80	100	100	100
61.NON-FIFA OTHER	80	100	80	90	90	100	100	90	100	80
All of the above components	85.4	87.7	88.1	69.6	81.5	90.4	80.4	80.0	88.8	88.5

Table 4- Interrater and intrarater reliability after the components were collapsed

COMPONENTS COMBINED	COMPONENT	PROPORTION OF OBSERVATIONS	ICC	
			INTERRATER	INTRARATER
53/54	FRONT LUNGE	0.75	0.77	0.81
31/32	LEG SWINGS	0.38	0.94	0.94
41/57	SIDE LUNGE	0.65	0.86	0.97
2/37	HIP OUT	0.73	0.69	1.00
2/36	HIP IN	0.73	0.70	0.92

Table 5- Percent agreement with the gold standard after the components were collapsed

COMPONENT	EXAMINER										
	2	3	4	5	6	7	8	9	10	12	
FRONT LUNGE	90	100	100	60	100	100	100	90	100	100	
LEG SWINGS	100	100	100	80	100	100	100	100	100	100	
SIDE LUNGE	100	80	100	80	90	100	100	100	100	100	
HIP OUT	100	100	100	100	100	100	20	100	100	100	
HIP IN	100	100	100	100	100	100	30	90	100	100	

Table 6- Sensitivity and specificity after the components were collapsed

COMPONENT	PERFORMED		NOT PERFORMED		TOTAL (N = 100)
	N	SENSITIVITY	N	SPECIFICITY	% CORRECT
FRONT LUNGE	80	92.5	20	100.0	94.0
LEG SWINGS	40	95.0	60	100.0	98.0
SIDE LUNGE	60	100.0	40	87.5	95.0
HIP OUT	80	90.0	20	100.0	92.0
HIP IN	80	90.0	10	100.0	92.0

Chapter 5- Discussion

The data analysis determined how reliable and accurate the examiners were when using the outcome measure to observe five different warm-ups at two different times. The goal of this study was to determine the interrater and intrarater reliability of the outcome measure and the accuracy of the examiners when compared to the gold standard. The analysis of data that was extracted from this study reflects the reliability and accuracy of the data collected for the FIFA 11+ Pre-Participation Study. From the information that this study collects, the FIFA 11+ Pre-Participation Study will modify the outcome measure used to collect data to be as accurate as possible. The outcome measure was hypothesized to be reliable and accurate. The results found, however, proved that the hypothesis was partially correct. About half of the components recorded from the form were found to be unreliable, but all but one of the examiners were accurate when compared to the gold standard.

Within the components that were unreliably recorded, some were very similar in movement or were exactly the same in movement but categorized differently which resulted in lower interrater and intrarater ICCs. After an analysis of those particular components, the data was re-configured to show what the results would have looked like for those components if they had been the same from the beginning. The components with low ICCs were combined with other components if there was another similar component on the outcome measure. They were combined by assuming that the observations for either component were counted as an observation for the new combined component. The ICCs and accuracy were then determined based on the new observations.

The components that were combined were done so because they had unreliable ICCs and were similar enough in action to be considered the same component. During the warm-ups that were watched for this study, and during warm-ups in general, high school athletes had trouble all completing the same exact component as opposed to two different components that appear to be the same to them. It would frequently occur that a portion of the athletes completed one component at the same time as the other portion completed another component. In this study it led to significantly lower ICCs for two components that were very similar. For future use of the form, it would be modified to have the components shown only once on the form.

On the form that was used for this study, the first page contained a FIFA warm-up form, to be filled out when the team was completing the FIFA 11+ warm-up. There were two components that were located on both the FIFA section of the outcome measure and the non-FIFA section of the outcome measure. The components, hip out and hip in, appear the same regardless of the type of warm-up the team completed. The components varied in distance and technique but were often confused by the examiners and caused lower ICCs for those components.

There were eleven examiners that participated in this study and five warm-up videos that were observed. Each examiner watched the videos twice, with at least ten days between the first and second observations. The total number of observations completed was 110. The sample size was determined by the volunteers already participating in the FIFA 11+ study and since there were only eleven, that became the sample size of this study. The study could have been stronger if there had been more examiners from even more backgrounds than we had. For example, having people in the general community use the outcome measure instead of people in the University of Vermont community could provide a larger spectrum of reliability and accuracy

but that was not the case in this study. There was one examiner in our study that had a very low percent agreement with the gold standard and, when there are only eleven examiners in total, having one outlier can have a larger impact on the results.

In the study mentioned in the literature review of the Functional Movement Screening, only 6 examiners were used but they were able to prove reliability because they each observed the participants twice and there were 39 participants in the study (Shultz, Anderson, Matheson, Marcello, & Besier, 2013). Despite the smaller number of examiners, they had a larger number of observations because they each observed more participants. In our study we had more examiners but less observations because the examiners watched fewer warm-ups.

Conclusion

The outcome measure used for the FIFA 11+ pre-participation study was proven to be only partially reliable and partially accurate. There were modifications made to the outcome measure that resulted in improved measures of reliability and accuracy. Although the outcome measure was not established in its reliability and accuracy, there is now research done on a compliance outcome measure of FIFA 11+, which had not existed before. This study was a necessary step to aid in injury prevention studies. If the outcome measure is modified in the future and established to be reliable and accurate, then this study acted as a catalyst for the injury prevention compliance necessities. With more information on injury prevention, there is the potential to make strides in reducing the incidence of injuries related to sports participation.

References

- Bizzini, Mario, Astrid Junge, and Jiri Dvorak. "Implementation of the FIFA 11+ Football Warm up Program: How to Approach and Convince the Football Associations to Invest in Prevention." *British Journal of Sports Medicine* 47.12 (2013): 803–806. *PMC*. Web. 4 May 2016.
- Ellaszlw, M., Young, S. L., Woodbury, M. G., & Fryday-Field, K. (1994). Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example. *Physical Therapy*, 74(8), 777-788.
- FIFA 11+; a complete warm-up programme. (n.d.). Retrieved March 31 2016, from <http://f-marc.com/11plus/home/>.
- Fredrik R. Johansson, Eva Skillgate, Mattis L. Lapauw, Dorien Clijmans, Valentijn P. Deneulin, Tanneke Palmans, Human Kinetic Engineer, and Ann M. Cools (2015) Measuring Eccentric Strength of the Shoulder External Rotators Using a Handheld Dynamometer: Reliability and Validity. *Journal of Athletic Training*: July 2015, Vol. 50, No. 7, pp. 719-725.
- Hänninen, Timo, Jari Parkkari, Markku Tuominen, Grant L. Iverson, Juha Öhman, Matti Vartiainen, and Teemu M. Luoto. "Interpreting Change on the SCAT3 in Professional Ice Hockey Players." *Journal of Science and Medicine in Sport* (2016): n. pag. Web.
- Injury Statistics. (2015, March 09). Retrieved March 31, 2016, from <https://www.cpsc.gov/Research--Statistics/Injury-Statistics>
- Kujala, U. M. et al. "Acute Injuries in Soccer, Ice Hockey, Volleyball, Basketball, Judo, and Karate: Analysis of National Registry Data." *BMJ : British Medical Journal* 311.7018 (1995): 1465–1468. Print.

- Menzel, Paul T. "Economic Concepts in Healthcare." *Encyclopedia of Bioethics*. Ed. Stephen G. Post. 3rd ed. Vol. 2. New York: Macmillan Reference USA, 2004. 697-704. *Gale Virtual Reference Library*. Web. 10 Apr. 2016.
- Owoeye, Oluwatoyosi B. A. et al. "Efficacy of the FIFA 11+ Warm-Up Programme in Male Youth Football: A Cluster Randomised Controlled Trial." *Journal of Sports Science & Medicine* 13.2 (2014): 321–328. Print.
- Phillip A. Gribble, Sarah E. Kelly, Kathryn M. Refshauge, and Claire E. Hiller (2013) Interrater Reliability of the Star Excursion Balance Test. *Journal of Athletic Training*: Sep/Oct 2013, Vol. 48, No. 5, pp. 621-626.
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: applications to practice* (2nd ed.). Upper Saddle River: Prentice Hall Health.
- Rebecca Shultz, Scott C. Anderson, Gordon O. Matheson, Brandon Marcello, and Thor Besier (2013) Test-Retest and Interrater Reliability of the Functional Movement Screen. *Journal of Athletic Training*: May/June 2013, Vol. 48, No. 3, pp. 331-336.
- Richter, Erik A., and Mark Hargreaves. "Exercise, GLUT4, and Skeletal Muscle Glucose Uptake." *American Physiological Society* 93.3 (2012): 993-1017. *APS Journals*. Web. 3 May 2016.
- Silvers-Granelli, H., B. Mandelbaum, O. Adeniji, S. Insler, M. Bizzini, R. Pohlig, A. Junge, L. Snyder-Mackler, and J. Dvorak. "Efficacy of the FIFA 11 Injury Prevention Program in the Collegiate Male Soccer Player." *American Journal of Sports Medicine* 43.11 (2015): 2628-637. *PubMed*. Web. 2 May 2016.

- Soligard, Torbjørn et al. "Comprehensive Warm-up Programme to Prevent Injuries in Young Female Footballers: Cluster Randomised Controlled Trial." *The BMJ* 337 (2008): a2469. *PMC*. Web. 4 May 2016.
- Steffen, K., Myklebust, G., Olsen, O. E., Holme, I. and Bahr, R. (2008), Preventing injuries in female youth football – a cluster-randomized controlled trial. *Scandinavian Journal of Medicine & Science in Sports*, 18: 605–614. doi: 10.1111/j.1600-0838.2007.00703.x
- Takata, Y., J. Nakase, A. Inaki, T. Mochizuki, H. Numata, T. Oshima, S. Kinuya, and H. Tsuchiya. "Changes in Muscle Activity after Performing the FIFA 11 Programme Part 2 for 4 Weeks." *Journal of Sports Sciences* 7.1 (2016). *National Center for Biotechnology*. PubMed, 25 Feb. 2016. Web.
- Taz Kicklighter, Mary Barnum, Paul R. Geisler, and Malissa Martin (2016) Validation of the Quantitative Diagnostic Thinking Inventory for Athletic Training: A Pilot Study. *Athletic Training Education Journal*: January-March 2016, Vol. 11, No. 1, pp. 58-67.

Appendix A

Observation of Team Warm Up Form
University of Vermont FIFA 11+ Injury Prevention Study

Date of Observation: _____ **School:** _____

Sport: Soccer / Football **Team:** Freshman / JV / Varsity **Sex:** _____ Boys _____ Girls

Data Collector: _____ **Total Duration of Warm Up:** _____

Were components of the FIFA 11+ program used as a warm up? _____ Yes _____ No

Was the FIFA 11+ program followed in order from start to finish? _____ Yes _____ No

Was the field set up for FIFA 11+? _____ Yes _____ No

Were there modifications in distance, cones, repetitions, exercises etc? _____ Yes _____ No

Components of FIFA 11+ Exercises Observed:

Part 1 Running: (8mins) Duration of time spent on Part 1 running exercises: _____

	Focus	Cueing	Technique	% Completed
_____ Straight ahead	_____	_____	_____	yes /no /partial
_____ Hip Out	_____	_____	_____	yes /no /partial
_____ Hip In	_____	_____	_____	yes /no /partial
_____ Circling Partner	_____	_____	_____	yes /no /partial
_____ Shoulder Contact	_____	_____	_____	yes /no /partial
_____ Quick Forwards and Backwards	_____	_____	_____	yes /no /partial

Part 2 Strength / Plyometrics / Balance: (10 mins) Time Spent on Part 2 _____

	Focus	Cueing	Technique	% Completed
The Bench				
_____ Static	_____	_____	_____	yes /no /partial
_____ Alternate Legs	_____	_____	_____	yes /no /partial
_____ One Leg Lift and Hold	_____	_____	_____	yes /no /partial
Sideways Bench				
_____ Static	_____	_____	_____	yes /no /partial
_____ Raise & Lower Hips	_____	_____	_____	yes /no /partial
_____ With Leg Lift	_____	_____	_____	yes /no /partial
Hamstrings				
_____ Beginner	_____	_____	_____	yes /no /partial
_____ Intermediate	_____	_____	_____	yes /no /partial
_____ Advanced	_____	_____	_____	yes /no /partial
Single Leg Stance				
_____ Hold the Ball	_____	_____	_____	yes /no /partial
_____ Throwing Ball Partner	_____	_____	_____	yes /no /partial
_____ Test your Partner	_____	_____	_____	yes /no /partial
Squats				
_____ With Toe Raise	_____	_____	_____	yes /no /partial
_____ Walking Lunges	_____	_____	_____	yes /no /partial
_____ One-Leg Squats	_____	_____	_____	yes /no /partial
Jumping				
_____ Vertical Jumps	_____	_____	_____	yes /no /partial
_____ Lateral Jumps	_____	_____	_____	yes /no /partial
_____ Box Jumps	_____	_____	_____	yes /no /partial

Part 3 Running: (2 mins) Time spent on Part 3 Running Exercises: _____

	Focus	Cueing	Technique	% Completed
_____ Across the field/court	_____	_____	_____	yes /no /partial
_____ Bounding	_____	_____	_____	yes /no /partial
_____ Plant & Cut	_____	_____	_____	yes /no /partial

Components of NON-FIFA 11+ Warm up Observed

Part 1 Running Components: Time Spent on Running Components: _____

____ Jogging: _____ Time _____ Straight ahead _____ Backwards
 ____ Side Shuffles ____ Karaoke ____ Vertical jump with jogging ____ Increase Pace ____
 Change of Direction: ____ Front/Back ____ Side/Side ____ Diagonals

Dynamic Mobility: Time Spent on Dynamic Mobility & Stretch: _____

	Walking	Jogging
____ High Knees	_____	_____
____ Butt Kicks	_____	_____
____ Leg Swings: front/back	_____	_____
____ Leg Swings: front or front diagonal	_____	_____
____ Leg Swings: back or back diagonal	_____	_____
____ Leg Swings: back with forward touch	_____	_____
____ Leg Swings: add/abd	_____	_____
____ Leg Swings: add	_____	_____
____ Leg Swings: abd	_____	_____
____ Leg Swings: Diagonals	_____	_____
____ Hip In _____ Hip Out	_____	_____
____ Hip Int Rot _____ Hip Ext Rot	_____	_____
____ Power Karaoke	_____	_____
____ Side lunge- side to side	_____	_____
____ Ankles	_____	_____
____ _____	_____	_____
____ _____	_____	_____

Dynamic Stretch:

____ Knee to chest	_____	_____
____ Heel to butt	_____	_____
____ Heel to butt, bend to touch toes	_____	_____
____ Heel on ground forward lean- hold	_____	_____
____ Heel on ground forward lean- scoop ground	_____	_____
____ Ext Rot- glut- hold	_____	_____
____ Front Lunge-hold _____ Front Lunge with UB Rot	_____	_____
____ Side Lunge- hold	_____	_____
____ _____	_____	_____

____ **Static Stretch:** Time Spent on Static Stretch: _____ **Seated** **Standing**

____ "Stretch on your own"

Part 2 Strength/Plyometrics/Balance Components:

Strength: **Time:** _____

- ____ Sit ups
- ____ Push ups
- ____ Front Plank
- ____ Side Planks

Bridging
 Jumping Jacks
 Squats- Double Leg
 Squats- Single Leg
 Squats- with toe raise
 Lunges
 Hamstrings- (partner holds)

Plyometrics: **Time:** _____
 Single Leg Double Leg Combined (SL-DL or DL-SL)

Agility/Balance: **Time:** _____
 Agility/Balance Static: Single Leg Double Leg

Agility/Balance Dynamic: Single Leg Double Leg Combined (SL-DL or DL-SL)

Part 3 Sports Specific and Progression of Running Components: Time: _____

Running/sprinting across the field/court

Bounding

Plant & Cut

Other Sports Specific Warm Up _____ **Time:** _____

No Team Warm Up Done

General Questions/Observations:

Who was leading the warm up? coach captain no-one other

Was the warm up done as: a team individuals (on own) by position

Was the team generally focused through out the warm up? Yes / No / Partial

Was there Cueing of form/technique done throughout warm up? Yes / No / Partial

Was there correct form/technique done throughout warm up? Yes / No / Partial

Did the warm up run continuously? Yes / No / Partial

Did the warm up have significant stop/stand time? Yes / No / Partial

Time of Total Warm Up: _____