CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

Comparing the Results of the Test of Gross Motor Development-2 (TGMD-2) to the Results of the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System (FG-COMPASS)

A thesis submitted in partial fulfillment of the requirements

For the degree of Master of Science in Kinesiology

By

Mackinsey Woolever

May 2016

The thesis of Mackinsey Woolever is approved:

Dr. Rosa Angulo-Barroso Date

Dr. Konstantinos Vrongistinos Date

Dr. Ovande Furtado, Jr., Chair Date

California State University, Northridge

Acknowledgements

I would like to extend the upmost appreciation to all those who supported, encouraged, and provided me with motivation during the completion of my master's degree at California State University Northridge.

I would first like to thank Dr. Ovande Furtado Jr., my thesis chair. Thank you for your endless support and feedback throughout this process. Your door was always open and without your guidance and support I would not have been able to accomplish what I have with my study.

To Dr. Jennifer L. Romack, thank you for providing me with the confidence to pursue and achieve my masters. You have instilled many great qualities in me and I will forever be grateful to have had the opportunity to have you as one of my committee advisors.

To Rosa Angulo-Barroso, thank you for your feedback and encouragement throughout my thesis project. Your knowledge and experience in this field is truly inspirational. I am so grateful to have you as one of my committee advisors.

Finally, I would like to thank my family. Mom, Mike, and Dad thank you for all the words of encouragement and showing me that I can overcome any challenges I face. I truly appreciate your never-ending support throughout my education. This accomplishment would not have been possible without you.

Dedication

My thesis is dedicated to Dr. Jennifer L. Romack. Her passion for the motor behavior program at California State University Northridge is what inspired me to pursue my Master's in Kinesiology. Thank you for giving me the strength to show me that I have the ability to make my passion a reality. You are my inspiration. Rest in Peace Dr. Romack.

"The best teachers are those who show you where to look, but don't tell you what to see"

- Alexandra K. Trenfor

Table of Contents

List of Tables

Table of Figures

Abstract

Comparing the Results of the Test of Gross Motor Development-2 (TGMD-2) to the Results of the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System (FG-COMPASS)

By

Mackinsey Woolever

Master of Science in Kinesiology

Fundamental movement skills (FMS) are considered the building blocks for the development of specialized skills. In addition, fundamental movement skill competency has been linked to decreased levels of obesity and increased levels of physical activity/sport participation. Thus, teachers and practitioners working with younger children must conduct regular assessments to gather evidence about the student's level of achievement in FMS development. This study aimed to collect evidence for concurrent validity, reliability of classification decisions, and inter/intra rater reliability for the FG-COMPASS. The FG-COMPASS is a process-oriented and criterion-referenced instrument that evaluates FMS development in children ages 5 to 10 years. The Test of Gross Motor Development 2 (TGMD-2) was used as "gold standard" for concurrent validity. Participants were 34 children between the ages of 5 and 10 years. Partial Pearson correlations (controlling for age) comparing the scores of both tests indicate a moderate to strong correlation for locomotion ($r_{xy.z}$ =.52, p < .01), object manipulation ($r_{xy.z}$ =.59, p < .001), and total scores ($r_{xy.z}$ =.63, p < .001). The reliability of classification decisions was assessed by comparing the live score ratings of five raters with the primary investigator's video scores on 10 randomly selected children per skill. The inter-rater reliability was assessed by comparing the scores across the 5 FG-COMPASS raters. The intra-rater reliability was assessed by comparing each FG-COMPASS rater's live scores to their video scores. The weighted kappa scores ranged from .51 to .83, .50 to .89, and .60 to .87 for reliability of classification decisions, inter and intra rater reliability, respectively. The results of this study provide further validity and reliability evidence for the FG-COMPASS. Further studies involving children with different ethnicity backgrounds and a larger sample size is recommended.

Chapter 1: Introduction

Fundamental gross motor skills (FMS) are thought of as the building blocks for future complex movements. Fundamental movement skills can be broken down into two subcategories; locomotor (i.e. running, jumping, hopping, etc.,) and object manipulation (i.e. throwing, catching, kicking, etc.,). There is a specific age range period where FMS's begin to emerge and is thought of as the fundamental motor pattern period (Clark, 2007). Gallahue and Ozmun (2006) state that there are developmental stages for each FMS that children progress with practice and experience.

Proficiency in gross motor skills is critical in young children and may present positive or negative effects on their development and lifestyle. Motor competence is a significant contributing factor in participation in physical activities (Castelli & Valley, 2007). Perceived motor competence has been shown to have a positive relationship with proficiency in FMS in children and adolescents (Biddle & Armstrong, 1992; Carroll & Loumidis, 2001; Woods, Bolton, Graber, & Crull, 2007). Physical activity participation is positively associated with proficiency of FMSs, especially if the activities are moderate to vigorous (Bellows, Davies, Anderson, & Kennedy, 2013; Carroll & Loumidis, 2001; Fisher et al., 2005; Lemos, Avigo, & Barela, 2012; McKenzie et al., 2002) and inversely associated with obesity (Bayer, Bolte, Morlock, Ruckinger, & von Kries, 2009; Graf et al., 2004; Lopes, Stodden, Bianchi, Maia, & Rodrigues, 2012). Many studies show that children who have poorer motor skill proficiency are less physically active (Fisher et al., 2005; McKenzie et al., 2002) and are more likely to be overweight or obese (Cliff et al., 2012).

Assessment tools are developed to assess development of gross motor skills and evaluate FMS proficiency. There are many assessment tools available to assess fundamental movement skill competency in children. In general, these instruments are not designed to be administered in a live setting by a single practitioner. A more practical assessment tool would allow physical education teachers to administer the test in a more time and cost efficient way to assess proficiency in FMS.

The Test of Gross Motor Development - Second Edition (TGMD-2) is considered to be the gold standard for assessing FMS competency in children between the ages of 3 and 10 years (Ulrich, 2000). The TGMD-2 has two gross motor subcategories (locomotor and object manipulation) and assesses a total of 12 skills. However, the TGMD-2 can be viewed as a time-consuming test that requires the use of video recording for accurate assessment of gross motor skills. The Furtado-Gallagher Computerized Observational Movement Pattern Assessment System (FG-COMPASS) was developed to assess gross motor development in real time in children between the ages of 5 and 10 years (Furtado, 2009). The FG-COMPASS is similar to the TGMD-2 in that it assesses both locomotor and object manipulation fundamental gross motor skills. The FG-COMPASS is designed to be administered in a live setting by a single practitioner without the need of video recordings.

Having a practical, valid, and reliable assessment tool to assess gross motor development is critical for identification of typical development and any potential motor delays in school-age children. There have been no attempts to compare the results of the FG-COMPASS to a valid assessment tool such as the TGMD-2. If the results of both instruments are found to be correlated, the FG-COMPASS may serve as a more sensible alternative to the TGMD-2 instrument. Thus, allowing physical education teachers to assess FMS development in a more time efficient manner. The purpose of this study was to collect criterion-related (concurrent) validity for the FG-COMPASS by comparing its results (total and both subtest scores) to the results of the TGMD-2. A second purpose of this study was to collect inter and intra-rater reliability for the FG-COMPASS. Three hypotheses were proposed for this study. First, there will be a positive "strong" (.50 or higher) correlation between the total scores of both assessment tests. Second, there will be a positive strong correlation between the locomotor and object manipulation scores between both assessment tests. Finally, there will be a "good" (weighted kappa =.61 or higher) inter-rater and intra-rater agreement for the total, locomotor, and object manipulation scores for the FG-COMPASS.

The following chapter will provide a background of the proposed research. The first three sections will provide a description of FMS, mastery of FMS, and the importance of assessing FMS. The following sections will discuss available FMS assessment tools, the two instruments being used in this study (TGMD-2 and FG-COMPASS), and the statistical concept of concurrent validity. The final section will review the need for a practical assessment tool, such as the FG-COMPASS in a school setting.

Chapter 2: Literature Review

What are fundamental motor skills?

Fundamental motor skills (e.g. run, jump, hop, throw, kick) have certain movement patterns (Gabbard, 2004) and believed to be the building blocks for future complex movements. FMS can be subdivided into two sub gross motor categories: locomotor and object manipulation. Locomotor skills involve the transportation of a child's body through space (e.g. walk, run, jump, hop, skip) which emerge between the ages of 1 and 7 years (Gabbard, 2004). Object manipulation skills involve applying a force to move an object (e.g. throw, kick, catch, dribble) (Haywood & Getchell, 2014). Gross motor skills are learned independently among locomotor and object manipulation categories'. However, once these skills are learned, they can be progressively combined to produce more complex movements (Gallahue & Ozmun, 2006).

Past research has defined specific inter-task sequences for the development of locomotor skills: beginning with walking, running, leaping, jumping, galloping, sliding, hopping, and ending with skipping (Haywood, Roberton, & Getchell, 2012). However, there is no definite inter-task sequence between leaping, jumping, galloping and sliding. Even though this is the proposed sequence of locomotor development, some children may develop faster or in a different sequence based on individual characteristics such as experiences and environments. No research has determined an inter-task sequence for object manipulation skills nonetheless; children can begin practicing and refining both locomotor and object manipulation skills in early childhood.

Mastery of FMS

Clark (2007) proposed a 'Mountain of Motor Development', which classifies stages of motor development from birth to death. The fundamental motor pattern period (1 to 7 years) is considered as the foundation period where a child learns the basic motor skills for future movements (Clark, 2007). During this period, movement behaviors begin to emerge, allowing young children to develop the basic motor skills. Children do not automatically acquire these skills, fundamental motor skills must be taught by an instructor and learned by the child (Clark, 2007). Mastering FMS plays a critical role in many factors of a child's active lifestyle.

Two terms are frequently used when describing gross motor development in young children: universality and variability. Universality is the concept that every child develops and learns the same FMS throughout their childhood. Variability refers to the different rates of FMS development each child experiences due to individual characteristics (Haywood & Getchell, 2014). Although every child will acquire and learn the same skills, previous studies have found age and sex differences across locomotor and object manipulation skills. As a child ages their movement patterns begin to change and become more advanced, and therefore age must be taken into consideration when looking at FMS in early childhood. To the best of our knowledge, sex differences have been found for object manipulation skills (Barnett, van Beurden, Morgan, Brooks, & Beard, 2009) but not for locomotor skills. Males tend to exhibit more proficient movement patterns in object manipulation skills than girls. This could be due to many factors including, participation in sport specific activities and participation in more vigorous physical activities (Barnett et al., 2009; Butterfield, Angell, & Mason, 2012). Knowing that females tend to be less proficient than males in object manipulation skills, activities should aim to develop these skills specifically in young females (Barnett et al., 2009).

Each FMS has a sequential process of development that children go through as they grow and practice skills. A child progresses to more advanced stages of development as they age and acquire experiences, but will not achieve an advanced level of performance without practice or experience (Gallahue & Ozmun, 2006). Developmental stages allow the observer to determine how well the individual performed the execution of a skill. The stages provide researchers with qualitative information regarding the progress of development of specific skills (Gallahue & Ozmun, 2006). The "Roberton method" uses a body component approach for developmental stages. This approach looks at body segments, such as the arms, legs, and sometimes the trunk independently, where each body segment will have a separate stage level (Roberton & Halverson, 1984). On the other hand, the "Seefeldt method" uses a total body component approach. This approach observes body components, such as the arms, legs, and trunk as a whole and classifies the performance under one stage of development (Seefeldt & Haubenstricker, 1982).

Effects of motor proficiency

The development of FMS has many significant relationships that may positively or negatively affect a child's active lifestyle. Engagement in physical activity allows a child to develop their motor competence. Motor competence can be defined as the ability to enjoy and proficiently execute skills during participation of physical activity (Castelli & Valley, 2007). Castelli et al. (2007) state that building this motor competence at the basic level is critical for future participation in more complex skills. Encouraging participation in physical activities allows a child to acquire the basic skill movements. The child will then be able to combine multiple skills to participate in more complex and challenging activities in later childhood. Previous studies have shown that children who are more proficient movers are more physically active and participate in more physical activities compared to their less physically active counterparts (Bellows et al., 2013; Cliff et al., 2012; Lemos et al., 2012).

Differences between types of physical activity have also been looked at by previous research studies. Lemos et al., (2012) investigated the differences between recreational participation and structured physical education participation. They found that children who participate in structured physical activities developed more proficient movement patterns. Providing activities that specifically target gross motor skills at an early age allows a child to develop and refine those skills through practice and experience. Bellows et al. (2013) used a structured physical education program for preschoolers to assess proficiency in gross motor skills. They found that following the intervention with structured physical activity there was no immediate increase in the physical activity level, however, it may foreshadow an increase in participation at a later age (Bellows et al., 2013). Bellows et al. (2013) also found that providing structured physical activity for at least 60 minutes a week increased gross motor performances in preschool aged children. Structured physical activities can enhance the development of FMS in school-aged children. Clark (2007) emphasized the importance of physical education and the development of FMS. The central focus of physical education should be around motor skill development, and this idea is often overlooked (Clark, 2007). Physical educators play a key role in the development of gross motor skills; it is imperative that they have access to an assessment tool to assess gross motor development.

Perceived motor competence has a direct relationship with physical activity participation. Perceived motor competence can be thought of how the child views themselves as a mover. A child's inner motivation plays a key role in participation in physical activities. During physical activities children tend to feel more confident when they chose the activity for their personal significance (Woods et al., 2007). Inner motivation is increased when a child is able to perform a skill successfully and is lowered when a skill challenges the child's motor competence (Deci & Ryan 1985, cited in Woods et al., 2007). Challenging tasks may initially lower the child's perceived motor competence but over time with practice and experience the child will build on their perceived motor competence as the child becomes more proficient in the skill movement. Children and adolescents' perceived motor competence has been seen as a positive influence on participation in physical activities (Biddle & Armstrong, 1992; Sallis, Alcaraz, McKenzie, & Hovell, 1999).

Perceived motor competence can differ between sexes. Biddle & Armstrong, (1992) found that boys and girls, between the ages of 11 and 12 years, differ regarding their motivation to participate in physical activity. Boys' motives tend to be more intrinsically related, while girls' motives tend to be extrinsically related, and often seek feedback from a teacher (Biddle & Armstrong, 1992). Thus, one must consider the differences between boys and girls in perceived motor competence, when implementing physical activity programs. It is important to provide an environment that allows for practice, encourages participation, and provides instruction to increase the motivation and participation to strengthen proficient movement patterns in children (Gallahue & Ozmun, 2006). Participation in physical activities is also a critical component to enhance gross motor skill development.

Past studies have found that not only does a high level of motor competence and perceived motor competence increase physical activity participation, but these high levels of perceived motor and motor competence also have an effect on the amount of time spent in high intensity physical activities (Carroll & Loumidis, 2001; Fisher et al., 2005; McKenzie et al., 2002). Fisher et al. (2005) compared the relationship between physical activity intensities and gross motor performance. They found that children who participated in the least amount of moderate-vigorous physical activity (MVPA) had lower FMS scores, and that children who participated in the largest amounts of MVPA had higher FMS scores. Fisher et al. (2005) concluded that either the low participation in MVPA hinders motor skill development or that a delay in motor development may reduce the amount of time spent in MVPA. Children who perceived themselves as better movers did indeed participate in more physical activity as well as in higher intensity physical activities (Carroll & Loumidis, 2001). McKenzie et al. (2002) found that children who are skilled movers at an early age tend to participate in more physical activities in adolescence. In addition, Barnett et al. (2009) found that object manipulation skill proficiency in childhood may be an indicator of increased physical activity participation in adolescence. If a young child fails to master gross motor skills, it can have a negative effect on their motor development in later childhood (Lemos et al., 2012). Obesity is also an important aspect to consider.

Research has supported the idea that children who engage in more physical activities are less likely to be overweight or obese (Bayer et al., 2009; Graf et al., 2004). Bayer et al. (2009) looked at the relationships between physical activity levels and weight status completing the following motor task; how many times a child could jump over a bar in 15 seconds. They found that children who participated in more physical activities were not classified as overweight/obese and outperformed the children who were classified as overweight/obese in the simple motor task. Overweight/obesity can be an influential factor that delays gross motor development in young children (Graf et al., 2004). Being physically active is a critical aspect to a child's lifestyle. Due to the strong relationships between overweight/obesity with low participation and hindered development of FMS, if the child does not increase their participation in physical activities this relationship will most likely continue during adolescence and adulthood.

Assessment of fundamental motor skills

Due to the critical role mastering gross motor skills has on a child's lifespan, measuring FMS development is critical in preschool and elementary school aged children. According to

Ulrich (2000), gross motor skills are often overlooked in childhood education. Assessment tools are designed to detect motor delays that may range from minor to severe. If a child has a motor delay and it is not discovered early or treated, the child may struggle with future motor skills (Ulrich, 2000). According to Provost, Crowe, & McClain (2000), a developmental delay is defined as a 25% or greater difference between the child's actual age and their developmental age. If a motor delay is detected early, practitioners, parents, and educators can implement strategies to help the child.

Normative data ranging from geographic areas, genders, race, and ages are used to compare children's performances. Assessment tools that have updated normative data are important to finding trends and changes in different generations (Piek, Hands, & Licari, 2012). Administrators rely on valid and normative assessment tools to determine gross motor development in children (Provost et al., 2000). In addition to the normative data, assessment tools that use age equivalents can be helpful for determining a developmental delay (Provost et al., 2000). Gross motor assessment tools, such as the Movement Assessment Battery for Children Second Edition (Movement-ABC 2), the Peabody Developmental Motor Scales Second Edition (PDMS-2), and the Test of Gross Motor Development Second Edition (TGMD-2) have been developed to evaluate the quality of FMS performance.

Movement skill assessment tools

There are many components to the design and development of assessment tools. Assessment tools can either assess process or product scores. Process scores measures how well the child performed the specific skill versus product scores which measure the outcome of the skill (i.e., time, distance, or number of successful attempts) (Burton & Miller, 1998). Process scores provide more information on skill proficiency and should be considered when assessing gross motor development. An assessment tool that assesses specific performance criteria is known as criterion referenced. This type of assessment tool has pre-determined criterion that is used to compare a child's performance to the criteria Cools, De Martelaer, Samaey, and Andries (2009), which gives the administrator an idea of the child's skill proficiency. Assessment tools that use a criterion-referenced design can either use performance criteria as a component or composite approach (Burton & Miller, 1998). As stated before, the component approach looks at specific body parts during the performance of the skill execution, which can range from the arms, legs, and/or trunk. The component approach can result as the child being more proficient in one body component than the other due to the scoring being independent for each body component. A composite approach looks at the body components together as one. This approach results in an overall score of their skill performance, with the body components scores being combined into one final score.

In addition to criterion-referenced, an assessment tool can be normative referenced. Norm-referenced means that the child's performance is compared to a normative group (Cools et al., 2009). This provides information on where the child is at with development according to the normative data. Normative data can be used to provide administrators with age equivalents and percentiles.

Typically, gross motor assessments are broken up into sub categories, such as locomotor, object manipulation, and balance. Slater, Hillier, and Civetta (2010) found that the Movement-ABC 2 and the TGMD-2 were the two most widely used assessment tools to evaluate gross motor proficiency. The Movement-ABC 2 assesses eight skills in three categories; manual dexterity, balance, and object manipulation skills that fall into three age ranges, 3 to 6, 7 to 10, and 11 to 16 years (Brown & Lalor, 2009; Cools et al., 2009). This assessment tool is process oriented and criterion/norm-referenced and takes around 20 to 30 minutes to complete (Cools et al., 2009; Piek et al., 2012). Some limitations for the Movement-ABC 2 assessment are that there is no separate gender normative data for the younger children, and the different tasks for each of the age groups makes it difficult for researchers to compare a child's development across their age span (Piek et al., 2012).

An additional popular assessment tool is the PDMS-2 (Piek et al., 2012). The PDMS-2 assesses both gross and fine motor skills (e.g. reflexes, balance, locomotor, object manipulation, grasping, and visual-motor integration categories) in children from birth to 5 years and 11 months (Cools et al., 2009; Piek et al., 2012). According to Slater et al. (2010), the PDMS-2 is ranked the fifth assessment tool used out of the seven they reviewed. This assessment tool is process oriented and criterion norm-referenced. As discussed by Piek et al. (2012), there are two limitations to the PDMS-2. First, it only assesses development in children up to 5 years and 11 months. Therefore, you cannot retest the same child at a later age. Secondly, it takes a long period of time to complete the full assessment, typically lasting around 45 to 60 minutes, which is not practical for a school setting. Another widely used assessment tool is the TGMD-2.

TGMD-2

The TGMD-2 has been used in many studies and is considered to be the gold standard in assessing gross motor skills in children ages 3 to 10 years (Sun, Sun, Zhu, Huang, & Hsieh, 2011). The TGMD-2 was originally developed in 1985 by Dale Ulrich and revised in 2000 (Ulrich, 2000; Wong & Cheung, 2010). This assessment tool is process oriented and criterion/norm-referenced (Ulrich, 2000). In 2000 the normative data was updated when the assessment tool was revised (Sun et al., 2011; Ulrich, 2000; Wong & Cheung, 2010). The TGMD-2 uses a composite approach when scoring skill performance. Each performance criteria focuses on one body component, but when the actual skill is being scored, all performance criteria are included in the final score for the specific skill. The instrument is comprised of twelve gross motor skills divided into two subtests (locomotor and object manipulation). The locomotor skills are: running, galloping, jumping, leaping, hopping, and sliding. The object manipulation skills are: kicking, catching, overhand throwing, underhand rolling, dribbling, and striking. The TGMD-2 provides the administrators with age equivalents, percentiles, and a gross motor quotient value that Ulrich (2000) states is the most reliable value, because it is comprised of both subcategories.

The TGMD-2 has been found to be a reliable and valid assessment tool. Reliability refers to the consistency of scoring. The TGMD-2 was tested for reliability on three categories; test-retest (time sampling), content sampling, and inter-rater (Ulrich, 2000). Test-retest reliability was done to ensure that after assessing the child twice at different times the scores would be similar, the reliability value for time sampling was found to be .88 for locomotor, .93 for object manipulation, and .96 for gross motor quotient (Ulrich, 2000). Content sampling investigates if the category is measuring what it is supposed to. The reliability values were found to be .85 for locomotor, .88 for object manipulation, and .91 for gross motor quotient (Ulrich, 2000). According to Ulrich (2000), the three high scores across the subcategories and the gross motor category indicate that the TGMD-2 has little test error. Finally, inter-rater reliability refers to the consistency of scoring between multiple assessors (Vincent & Weir, 2012). The values for reliability for inter-rater scoring was .98 for locomotor, object manipulation, and gross motor quotient (Ulrich, 2000).

Many studies have been conducted using the TGMD-2 to measure differences in typically developing children and children with special populations. A longitudinal study by Westendorp et al. (2014) used the TGMD-2 to assess the changes seen in locomotor and object manipulation skills in typically developing children and children with learning disabilities between the ages of 7 and 11 years. The TGMD-2 was found to be appropriate assessment tool for children with intellectual disabilities and visual impairments (Houwen, Hartman, Jonker, & Visscher, 2010; Simons, Daly, Theodorou, Caron, & Andoniadou, 2008). Finally, the TGMD-2 has been compared to newly developed assessment tools, testing concurrent validity (Sun et al., 2011).

Similar to the design of the proposed study, Sun et al. (2011) compared the validity of the TGMD-2 to the newly developed Preschool Gross Motor Quality scale (PGMQ). The PGMQ is similar to the TGMD-2 but assesses different skills. The PGMQ assesses 17 items in three subcategories; eight locomotor, five object manipulation, and four balancing tasks. This study compared the results of the total and subcategories scores of each assessment tool using Pearson moment-product correlations. They found that the total score (.86) and subtest scores (locomotor .82 and object manipulation .76) showed significant positive correlations between PGMQ and TGMD-2. These results indicate that the PGMQ can be considered as a valid assessment tool to assess gross motor development in preschool aged children.

Despite being popular among researchers, the TGMD-2 may be difficult to administer in a live setting without the use of video recordings to accurately assess a child's performance. Due to this it may take multiple days to complete a full assessment in a school setting for one child. There are alternatives to the TGMD-2, that can be used in school setting Gallahue & Cleland-Donnelly (2007), but they lack validity and reliability.

FG-COMPASS

Initially named Furtado-Gallagher Movement Skill Assessment Tool (Furtado, 2004), the FG-COMPASS was proposed in 2004, when content-related validity was collected. In 2009, the test was officially developed and reliability for classification decisions was tested (Furtado, 2009). The instrument assesses gross motor skills in children between the ages of 5 and 10 years. The FG-COMPASS was developed and designed to be administered without the use of video recordings for physical educators to assess gross motor development. This assessment tool uses a process-oriented and criterion-referenced design. The FG-COMPASS was developed through combining the ideas of the composite 3-stage approach with the observational plan approach (Furtado, 2009). The 3-stage approach results in the child only being able to fall into three stages; initial, elementary, and mature (McClenaghan, 1976 cited in Gabbard, 2004). There are many performance criteria in this approach, which makes it difficult for teachers to administer an assessment tool that uses a 3-stage approach. The developers of the FG-COMPASS believed if the number of the performance criteria in the 3-stage approach were lowered it would become a more practical assessment tool. The observational plan approach uses a decision tree design, with only a few performance criteria being included, and where the administrators scores the child's performance on either yes (they demonstrated the performance criteria) or no (they did not) (Haywood & Getchell, 2014). The developers of the FG-COMPASS defined their design as a composite decision-tree approach.

In 2012 the reliability of classification decisions of the FG-COMPASS was tested (Furtado & Gallagher, 2012). Participants (n=30) were trained on the composite decision-tree approach and assessed gross motor skills of previously video recorded children. The reliability was tested with weighted kappa statistics and found to range from .51 to .85, which is considered "good" (Altman, 1991). Eleven gross motor skills (five locomotor and six object manipulation skills) were tested in this study. The results indicated that six skills (side slide, horizontal jump, leap, kick, stationary dribble, and overhand throw) needed to be further investigated. Another reliability study was done in 2014 to further investigate the content validity of the six gross motor skills found to have discrepancies in the FG-COMPASS. The researchers found that side sliding was not able to distinguish between elementary and mature levels of development; in addition, leaping was not able to distinguish between initial and elementary levels of development. Therefore, both of these skills were removed from the FG-COMPASS (Furtado & Gallagher, 2016). The authors determined that the four skills left would remain on the FG-COMPASS. However, no studies have examined the validity of the FG-COMPASS in relationship to another assessment tool.

Concurrent Validity

The type of validity that should be used to compare and view the relationship of the results between two tests is concurrent validity (Campbell & Kolobe, 2000). There have been many studies done comparing two motor assessment tools to determine a relationship between both tests. It is important that when selecting an assessment tool to administer that the practitioner is aware of the strengths, restrictions, and limitations of the test (Provost et al., 2000). Concurrent validity studies allow assessment tools to be compared to each other and to determine if an assessment tool is valid at measuring specific characteristics. Testing concurrent validity on multiple assessment tools provides practitioners and examiners with the opportunity of a wide range of assessment tools to select from to make sure their chosen assessment tool is the best fit. The first step to determine validity of one assessment tool to another is to make sure that both assessment tools are similar in what they assess and the results produced from each test. The basic protocol from previous studies is that both assessment tests are administered to the participants and then running Pearson product-moment correlations to determine relationships between the scores from both tests (Campbell & Kolobe, 2000; Provost et al., 2000; Sun et al., 2011). The researchers will then determine the level of strength of the Pearson correlations. Typically a correlation of .85 to 1.0 determines a very high relationship between the two variables with 1.0 being an exact match (Provost et al., 2000; Sun et al., 2011).

Need for a practical assessment tool

Assessing gross motor skill development in young children is critical to encourage physical activity participation in later childhood and in adulthood. In school-aged children besides participating in sports, children are exposed to structured physical activity during the required physical education hours. Physical educators and activities have an influence on the development of skills and participation of movement activities. Physical education requirements vary for each state nationwide. Kim (2012) looked at the affects of state required policies on physical activity in children. The results showed that although states have required policies, some schools might not follow the requirements, which may result in no direct relationship between physical activity participation and state requirements. Another study compared the state requirements to the physical activity engagement of children, and found that although participating in physical education does not lower the risk of obesity, it increased the amount of time spent participating in physical activity (Cawley, Meyerhoefer, & Newhouse, 2007). However, school policies have a larger effect on physical activity in children than state policies.

Although school requirements have a larger impact on physical activity participation, if physical educators are not given extra time in addition to the implementation of their curricular plans, they will not be able to administer assessment tests. It is critical that physical educators have access to an assessment tool that can be administered in a reasonable amount of time and without the use of video equipment. An assessment tool, such as the FG-COMPASS provides just this. It can be administered and completed within ten to fifteen minutes, it does not require the use of a camera, and is cost efficient. The gap between time needed to administer a test and the need for assessment will be decreased, if the FG-COMPASS is found valid. It is important that physical educators have access to a test that provides them with information on where their students are at developmentally. This allows the physical educator to have a specific goal, so that they target FMS within their activities.

Chapter 3: Methods

Participants

A convenient sampling method was used to recruit participants for this study. Forty-one participants between the ages of 5 and 10 years were recruited. However, only thirty-four participants, 22 girls (M = 8.14, SD = 1.78) and 12 boys (M = 8.44, SD = 1.49) had full participation (see Table 1). Three participants dropped out on their own, two never attended the assessment sessions, one was injured outside of school hours, and one child moved out of the state during data collection. To control for even distribution of age ranges, one randomly selected classroom from kindergarten through fourth grade received a recruitment packet (see Appendix A). The Los Angeles Unified School District (LAUSD) Committee of External Research Review and the California State University of Northridge (CSUN) Committee for the Protection of Human Subjects approved the recruitment packet prior to distribution to the classrooms. Recruitment packets were given to additional classrooms if no sufficient consent forms were returned. A participant was excluded from the study if he/she: 1) was younger than 5 years or older than 10 years and 11 months, 2) had developmental delays or disabilities that may have affected their motor performance, 3) had no parental or verbal consent, or 4) had a “yes” response on any of the first five Physical Activity Readiness-Questionnaire (PAR-Q) questions.

Table 1. Participant Information

	Girls			Boys		
Grade	*N*	*M*	*SD*	*N*	*M*	*SD*
Kindergarten	7	5.94	0.16	3	6.30	0.12
1st Grade	1	7.32	-	-	-	-
2nd Grade	1	8.28	-	-	-	-
3rd Grade	6	9.09	0.30	5	9.17	0.34
4th Grade	7	10.08	0.19	4	9.84	0.34
Total	22	8.14	1.78	12	8.44	1.49

Notes. N = sample size. M = average age in years. SD = standard deviation in years.

Instrumentation

The Test of Gross Motor Development-Second Edition (TGMD-2) and the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System (FG-COMPASS) were used to collect data for this study.

TGMD-2. The TGMD-2 assesses twelve gross motor skills (six locomotion and six object manipulation) in children between the ages of 3 and 10 years. The TGMD-2 assesses a number of performance criteria that focuses on different body components, such as the arms, legs, and trunk (see Figure 1). For example, a performance criterion for the skill of hopping is: "Arms flexed and swing forward to produce force". The child is given two trials to complete each skill. For each trial and performance criterion, a score of 1 is given if the criterion is demonstrated; otherwise, the child is assigned a score of 0. Once the performance criterion has been scored for the two trials on each skill, the scores are then added up to provide an overall raw score for each subtest. The raw scores for each subtest are then converted to a standard score where the child's standard score can be compared to normative data. The standard scores from both subtests provide age equivalents and percentiles. When added together, the standard scores can be converted to a gross motor quotient value.

3. Hop	A minimum of 15 feet of clear space	Tell the child to hop three times on his or her preferred foot (established before testing) and then three times on the other foot. Repeat a second trial.	1. Non support leg swings forward in pendular fashion to produce force		
			2. Foot of nonsupport leg remains behind body		
			3. Arms flexed and swing forward to produce force		
			4. Takes off and lands three consecutive times on preferred foot		
			5. Takes off and lands three consecutive times on nonpreferred foot		
				Score	

Figure 1. Adapted score sheet for the TGMD-2

FG-COMPASS. The FG-COMPASS assesses eight gross motor skills (three locomotion and five object manipulation) in children between the ages of 5 and 10 years. The instrument is comprised of a locomotion and an object manipulation subtest. The FG-COMPASS scoring is based on a scale in the form of a decision tree (see Figure 2). Each decision tree has three performance criteria in the form of questions. However, only two are considered when scoring a given performance. The first criterion is intended to discriminate between levels 1 and 4, whereas the second criterion (either upper or lower) is intended to function as a confirmatory question. For instance, if the first criterion is present (i.e., scored as Yes), it is assumed that the child is at level 4. The second criterion is used to confirm this assumption. If the child fails to confirm it, then the child is scored at level 3.

Hopping	Is the thigh of the nonsupport leg in a vertical position with knee flexed at 90^0 or less?	**Y**	Does the foot of the nonsupport leg pass behind/in front of the support leg?	**Y**	4
				N	3
		N	Is the nonsupport leg held in front of the body?	**N**	2
				Y	1

Figure 2. Adapted decision tree for the FG-COMPASS

Research Assistants

Sixteen kinesiology undergraduate students were recruited as research assistants. Ten of the sixteen students were recruited as raters with each instrument having five raters to complete the assessment. Once selected, the raters were randomly assigned to either Group 1 (TGMD-2) or Group 2 (FG-COMPASS). At the conclusion of the study, the two raters from each group whose scores had the highest agreement with the expert's ratings received a $50 gift card. The remaining six recruited undergraduate students served as test administrators at the testing site. The test administrators were divided between the two groups. The TGMD-2 group had three test administrators and the FG-COMPASS group had two test administrators. One student never attended the site. This student and another student who was in charge of administering the FG-COMPASS were in charge of editing the videos for further analysis. Training was slightly different for each assessment tool. Each group of raters went through multiple training and review sessions prior to data collection. The test administrators went through multiple training and practice sessions on skill set-up, instructions, and demonstrations, in addition to camera placement and handling prior to data collection.

Training for the TGMD-2. The raters completed six training sessions prior the start of data collection. Each training session lasted approximately one hour and focused on two skills. Due to scheduling issues, the primary investigator trained the raters in two groups on separate days. The raters met with the primary investigator where they reviewed the performance criteria for the two skills, and then completed assessments from previously recorded children. Each rater was given their own TGMD-2 score sheet where they recorded their scores and wrote any comments or questions they had during the training session. The score sheets were allowed to be taken home with them so they could review the criteria at any given time. While watching the first few videos, the primary investigator discussed with the raters as a group whether or not they felt the child met each performance criteria for the two trials. As the raters became comfortable, they watched and scored the child's performance independently without any group discussion. After each rater scored the two trials, the primary investigator and the group went over the scores to ensure a clear understanding of the performance criteria and accurate scoring.

After all training sessions were completed, each rater completed one review session, which was done individually and lasted approximately 90 minutes. The review session was comprised of new videos that the raters had not seen before. Each rater had their own randomized list that they were to follow during their assessments. The raters assessed a total of 60 skills during their review session (five children for each skill). The primary investigator selected and scored all videos used for the training and review sessions to ensure that the raters observed skill performances that ranged between novice and proficient patterns. The training and review sessions were done to confirm a clear understanding of the performance criteria and scoring of the TGMD-2.

The test administrators completed two training sessions that lasted between 45 and 60 minutes. The TGMD-2 assesses a total of 12 skills, which required three test administrators to ensure effective administration. The first training session focused on the six locomotion skills and the second session focused on the six object manipulation skills. During each training session, the test administrators were taught each of the six skills set-up, instructions and demonstrations along with the placement and handling of the camera. Each test administrator was given a packet with pictures of the set up along with a copy of the TGMD-2 score sheet which provided written set up directions and verbal instructions. In addition to the training sessions, the test administrators completed two live practice sessions. During the practice sessions the test administrators provided instructions and demonstrations for all 12 sills to live participants, which were the raters for the TGMD-2 group.

Training for the FG-COMPASS. The raters completed four training sessions. Each training session lasted approximately one hour and focused on two skills. The training and review sessions for the FG-COMPASS group followed the same procedures as the TGMD-2 group, with the exception of the videos used for assessments. The primary investigator ensured that all raters understood the decision trees for each skill and asked questions prior to assessing the videos. The first few assessments were done as a group to check for understanding and then done independently, followed by a group discussion. The raters completed one review session in the lab separately from one another. The review session used new videos the raters had not seen. The raters assessed a total of 70 videos during the review session. The primary investigator selected and scored all videos used for the training and review sessions to ensure that the raters observed skill performances that ranged between levels one through four on the decision tree. Similar to the TGMD-2 group this was done to confirm a clear understanding of the performance criteria and scoring of the FG-COMPASS.

The test administrators completed two training sessions that lasted approximately 45 to 60 minutes. The training was slightly different from the TGMD-2, as the FG-COMPASS has different skill set ups, instructions, and camera placements for each skill. The raters and test administrators had two additional live practice sessions prior to data collection. This allowed the raters to experience scoring in a live setting.

Procedures

Data collection. Although the assessment tools evaluate the same construct, the TGMD-2 and the FG-COMPASS differ in a number of ways. However, some common practices were used when administrating both tests. Participants were assessed during either their morning recess or lunch recess. Prior to the first day of data collection, the primary investigator received verbal assent from each participant. Each participant who provided verbal assent to participate in the study was assigned an identification number that ranged from 01 to 41 for confidentially reasons. At the testing site, the primary investigator handed the identification number cards upon the participants' arrival and sent them to either the TGMD-2 or FG-COMPASS based on a randomized list. Participants were randomly selected to begin at either the TGMD-2 or FG-COMPASS each week to guarantee there was no learning effect in the event the instruments were measuring the same skill. Once participants arrived at a station, they followed a three-step process. First, the test administrators had each participant hold up their identification number card and state "I am (ID number)" to the camera before any instructions, demonstrations, or any skill performances began. Next, the test administrators provided each participant with instructions and demonstrations for the skill and asked for any questions. Finally, the participant completed their performance trials, which were used for scoring. The test administrators for both the TGMD-2 and FG-COMPASS were given checklists, which were used as reference when administering the assessment tools. Upon being assessed on the first station (either TGMD-2 or FG-COMPASS), participants moved to the second station to complete the assessment. In some cases, participants completed multiple skills for each instrument; therefore, they rotated back and forth between the two stations. Due to absences or forgetting to meet with the primary investigator, not all children completed the skill(s) each week; therefore, some had to attend make up sessions to test remaining skills in upcoming weeks. The schedule for data collection is listed in Table 2 and was expanded over an eight-week period. Each week was designed to match skills as close as possible for both instruments.

Table 2. Schedule of Data Collection

Week	FG-COMPASS	TGMD-2
1	Skip	Run/Horizontal Jump
2	Hop/Horizontal Jump	Hop/Gallop
3	No Data Collection	Slide/Leap/Make Up
4	Throw/Dribble	Throw/Dribble
5	Kick/Make Up	Kick/Make Up
6	Catch/ Make Up	Underhand Roll/Catch/Strike
7	Strike/Make Up	Make Up
8	Make Up	Make Up

TGMD-2. The test administrators attended the testing site each week to administer the TGMD-2 test. Test administration procedures, which are outlined in the test's user manual (Ulrich, 2000), were followed closely including suggested materials and directions. Each child completed three trials. The first trial was a practice trial and the second and third trials were used as testing trials. All performances were videotaped.

Data coding. The raters for the TGMD-2 watched the videotapes and completed their skill assessments at their scheduled time with the primary investigator present in the lab. The skill assessment sessions occurred within one week of the initial performances. At each assessment session, the raters referred to their folder that contained score sheets and a flash drive to complete the assessment for the skill(s). Each participant was assigned their own score sheet where the raters would record the scores for both trials on the 12 skills. All edited videos were transferred to a flash drive that the raters referred to during their video assessments. The folders were kept in a drawer in the primary investigator's desk. The skill assessment sessions lasted approximately one to two hours each week depending on how many skills were completed.

FG-COMPASS. Unlike the TGMD-2, the performances for the FG-COMPASS were assessed live at the site. The raters and test administrators attended the testing site each week to assess performances and administer the FG-COMPASS. The raters and test administrators attended the testing site for approximately four hours each week. In addition, the raters attended two reassessment sessions, which took approximately 30 minutes to one hour. Each rater was given the scoring sheet (decisions trees) for the skill(s) being tested. Each scoring sheet was comprised of the performance criteria (questions) displayed as a decision-tree and the four levels for the raters to circle (see Appendix B). Each participant completed one practice and three testing trials per skill. If raters were uncertain between two scores, they could request the participant to complete two additional trials for a maximum of five testing trials.

Live and Video Data coding. All live skill performances for the FG-COMPASS were videotaped. This was done to check for raters vs. expert consensus and intra-rater reliability. The raters reassessed twenty videos per skill. This was done twice to verify for potential discrepancies between the live and video assessments. The videos were randomly selected from the existing pool of videos. Therefore, each rater had three scores per skill for 20 randomly selected participants. The video reassessment sessions were done independently among the raters with the primary investigator present at each video assessment. The raters were given a new score sheet for each reassessment session so they could not see their previous recorded scores. When assessing fundamental movement skill performance, it is common practice (Ulrich, 2000) to test for intra-rater reliability by comparing the ratings of two assessments one or two week(s) apart. To check for consistency, we required the raters to re-assess the same videos a second time, one week following the first video assessment (Video 2). The protocol for the reassessments was determined based on the time constraint of the study. Chien, Scanlon, Rodger, & Copley (2014), used similar procedures, with having the two retests done within a two week period. Other studies measured intra-rater reliability using a 2-week and 1-month retest procedure (Cano-Cappellacci et al., 2015; Houwen et al., 2010; Simons et al., 2008). Therefore, all the reassessment sessions occurred between 8 and 9 days and 15 and 16 days after the initial live performance was scored. The schedule of the raters live and reassessment sessions can be found in Table 3.

Table 3. Schedule of Live and Video Reassessments for FG-COMPASS Raters

Week	*Live*	*Video 1*	*Video 2*
1	Skip	None	None
2	Hop/Horizontal Jump	Skip	None
3	No Data Collection	Hop/Horizontal Jump	Skip/Hop/Horizontal Jump
4	Throw/Dribble	None	None
5	Kick	Throw/Dribble	None
6	Catch	Kick	Throw/Dribble
7	Strike	Catch	Kick
8	No Data Collection	Strike	Catch/Strike

Data entering. The primary investigator selected two undergrad students to enter in the scores from both instruments. One student entered in all recorded scores for the FG-COMPASS (live and two video scores) and the other student entered in all recorded scores for both trials for the TGMD-2. These students used the computer in the lab to enter in the recorded scores. At the end of the study, the primary investigator checked all data entry and corrected any errors made before running the analysis.

Dealing with missing data. Before comparing the two instruments and investigating concurrent validity for the FG-COMPASS, the data were reduced because of missing values. A total sample of 28 subjects was used. Six subjects were eliminated from the concurrent validity analysis because they failed to complete all skills for both tests. Note that these subjects were included when investigating the raters vs. expert consensus, inter, and intra-rater reliability. Following the removal of six subjects, the Little's Missing Completely at Random test Little (1988) was conducted since there were some missing scores for some of the remaining subjects. This was done in the case a given participant had a missing score, for example, running from rater 2 (TGMD-2), or throwing from rater 4 (FG-COMPASS).

After running the analysis, we found that there was 1 missing value for the FG-COMPASS (dribbling for rater 4) and two missing values for the TGMD-2 (running for rater 3, dribbling for rater 3). The test indicated that the data were missing at random (chi-square=.000, DF=357, $p > .05$). This allowed us to replace missing values with predicted values using the Expectation-Maximization technique Dempster, Laird, & Rubin (1977) that is part of the IBM SPSS computer statistical package.

Data Analysis

This study employed a cross-sectional design. Statistical analyses were performed using the IBM SPSS for Windows (*IBM SPSS Statistics for Windows*, 2013).

Inter/intra rater analysis for the FG-COMPASS. The inter/intra rater reliability was assessed with weighted kappa statistics. Kappa is an index of reliability that takes into account the proportion of agreement with correction for chance classification (Fleiss & Cohen, 1973).

Kappa ranges from -1 to +1. A negative score indicates poorer than chance agreement. A score of zero indicates exactly chance agreement and a positive value indicates better than chance agreement. The weighted kappa index is recommended for cases in which more than 2 coders independently classify entities into 3 or more discrete (ordinal level) categories. Unlike kappa, "weighted" kappa assigns different weights depending on the level of the disagreement. Consider the scales (decision trees) for the FG-COMPASS, for instance. The scales range from 1 through 4, with 1 being least advanced and 4 being most advanced with regards to fundamental movement skill development. The seriousness of the disagreement is related to the level of classification. If coder 1 assigned a score of 1 to participant "Y" and coder 2 assigned a score of 2 to the same participant, then a disagreement exists between coders. Such a disagreement is considered less serious compared to a situation in which coders would assign values of 1 and 3, or 1 and 4. The weighted kappa values will be interpreted as follows: >0.80 very good, 0.61-0.80 good, 0.41-0.60 moderate, 0.21-0.40 fair, and <0.21 poor (Altman, 1991).

Inter-rater analysis for the TGMD-2. For the TGMD-2, reliability was assessed using the two-way random, average measures intraclass correlation coefficient ICC(2,5) (McGraw & Wong, 1996). The ICC is intended to assess rating reliability by comparing the variability of different ratings of the same participant to the total variation across all ratings and all participants. The ICC values were interpreted as follows: >0.75 excellent, 0.40-0.75 fair to good, and <0.40 poor (Fleiss, 2011).

Raters vs. expert consensus for the FG-COMPASS and TGMD-2. Consensus estimates are based on the premise that two raters should share a common interpretation of the construct being assessed (Stemler, 2004). In the present study, the ratings made by the five raters were compared with those of an expert (gold standard). For the FG-COMPASS, agreement was estimated using weighted kappa. The two-way random, single measure (absolute agreement) intraclass correlation coefficient ICC(2,1) was used to estimate consensus between the raters and the expert for the TGMD-2.

Concurrent validity. Both the partial (controlling for age) and zero-order Pearson product-moment correlation coefficients were calculated to evaluate the concurrent validity of the FG-COMPASS with the TGMD-2 test operating as the criterion. The correlations were interpreted as follows: $0.10 < |r| < 0.30$ small correlation; $0.30 < |r| < 0.50$ medium/moderate correlation; $|r| > 0.50$ large/strong correlation (Cohen, 1988).

Chapter 4: Results

The results are presented next. First, we present the results of inter-rater reliability for both tests, followed by the results of intra-rater reliability for the FG-COMPASS. Then, the degree of consensus between combined ratings of raters and the expert's ratings are presented for both tests, which is followed by the results of concurrent validity for the locomotion, object manipulation subtests as well as for the total test. Finally, correlations between similar skills of both tests are presented.

Inter-rater reliability for the FG-COMPASS

The inter-rater reliability of the FG-COMPASS was assessed by comparing the ratings of all five raters involved in the study. The results can be found in Table 4. Weighted kappa was computed for each rater pair then averaged to provide a single index of reliability (Light, 1971). Weighted kappa values ranged from .50 to .74 (M = .66) for the locomotion subtest and from .62 to .89 (M = .77) for the object manipulation subtest. The averaged weighted kappa for all skills (total test) was .73. The resulting weighted kappa indicated a "good" agreement for the locomotion, object manipulation subtests as well as for the total test. Note that rater #5 had major disagreements with all other raters for overhand throw, which is indicated by the asterisks in Table 4. If rater #5 is removed from the analysis, the M_{kw} for overhand throw changes from .62 to .79. If the weighted kappa of .79 is used for overhand throw, the total test (all skills combined) inter-rater reliability for the FG-COMPASS for the current study changes from .73 to .75.

Inter-rater reliability for the TGMD-2

The inter-rater reliability of the TGMD-2 was assessed with the two-way random, average measures intraclass correlation coefficient ICC(2,5). The results are presented in Table 5. The ICC values ranged from .74 to .95 (*M* = .89) for the total test, .74 to .95 (*M* = .87) for the locomotion and from .84 to .95 (*M* = .91) for the object manipulation subtests. When compared to the adopted criteria, these results indicate an "excellent" agreement (Fleiss, 2011).

Table 4. Weighted Kappa Statistics for the Inter-Rater Analysis – FG-COMPASS

	Raters										
Skill	*1x2*	*1x3*	*1x4*	*1x5*	*2x3*	*2x4*	*2x5*	*3x4*	*3x5*	*4x5*	M_{kw} (M_{SE})
Hopping	.56	.43	.43	.50	.42	.49	.46	.53	.64	.52	.50(.13)
Horizontal Jump	.93	.91	.7	.65	.84	.73	.62	.6	.6	.69	.73(.09)
Skipping	.78	.74	.60	.63	.87	.77	.76	.69	.71	.83	.74(.10)
Overhand Throw	.85	.79	.85	.45*	.80	.73	.35*	.71	.31*	.40*	.62(.11)*
Kicking	.83	.68	.64	.82	.63	.59	.73	.50	.72	.64	.68(.10)
Stationary Dribbling	.86	.87	.83	.92	.88	.84	.91	.88	.93	.93	.89(.05)
Catching	.87	.91	.87	.90	.83	.84	.90	.80	.80	.89	.86(.05)
Batting	.81	.80	.72	.72	.87	.80	.80	.69	.78	.79	.78(.09)

Notes. K_w= weighted kappa; SE= asymptotic standard error. 1x2, 1x3, …= rater pairs agreement; M_k= arithmetic mean of rater pairs agreement; M_{SE}= arithmetic mean of the raters' standard error values. Asterisks indicate major disagreement between rater 5 and all other raters.

Table 5. Intraclass Correlation Statistics for the Inter-Rater Analysis – TGMD-2

Skill	*ICC*	*95% CI*
Run	.74	.58-.86
Gallop	.84	.74-.91
Hop	.89	.82-.94
Leap	.87	.78-.93
Horizontal Jump	.95	.91-.97
Slide	.90	.83-.94
Striking a Stationary Ball	.91	.86-.95
Catch	.92	.87-.96
Kick	.84	.74-.91
Overhand Throw	.93	.88-.96
Underhand Roll	.92	.86-.95
Dribble	.95	.92-.97

Notes. N = 34. Values computed using two-way random, average measures ICC(2,5).

Intra-rater reliability for the FG-COMPASS

As mentioned earlier, all five raters independently assessed all performances in a live setting. Filming the live performances allowed us to test for consistency of ratings over time (intra-rater reliability). One week following the live assessment, each rater re-assessed the performances (a batch of 20 randomly selected participants) by watching and rating the recorded videos (Live vs. Video 1). The intra-rater reliability was assessed by comparing the live assessment ratings with those obtained from the video assessment.

The weighted kappa was computed for each rater (and each skill) then, similarly to the inter-rater analysis, the values were averaged to provide a single index of reliability (see values under M_{kw} (M_{SE}) in Table 6). The averaged weighted kappa values (Live vs. Video 1) ranged from .60 to .80 (*M* = .69) for the locomotion subtest, and .64 to .87 (*M* = .79) for the object manipulation subtest. The averaged weighted kappa values for the total test (all skills combined) was .76. Similar to the inter-rater reliability analysis, the resulting weighted kappa values for the intra-rater reliability indicated a “good” agreement for the locomotion, object manipulation subtests as well as for the total test (Altman, 1991).

When comparing “Live vs. Video 1” (see Table 6) and “Live vs. Video 2” (see Table 7), one can see that the averaged weighted kappa values are very similar, though slightly lower for Live vs. Video 2. Overall, the consistency between Video 1 vs. Video 2 is relatively high with averaged scores ranging from .65 to .89 (*M* = .78) for the locomotion subtest and from .73 to .96 (*M* = .84) for the object manipulation subtest. The total test average score was .81.

Table 6. Weighted Kappa Statistics for the Intra-Rater Analysis (Live vs. Video 1) – FG-COMPASS

	Raters					
Skill	*1*	*2*	*3*	*4*	*5*	M_{kw} (M_{SE})
Hopping	.72(.15)	.44(.15)	.64(.16)	.69(.12)	.84(.11)	.67(.14)
Horizontal Jump	.56(.21)	.46(.20)	.74(.18)	.56(.21)	.67(.18)	.60(.20)
Skipping	.96(05)	.95(.04)	.92(.04)	.51(.25)	.66(.15)	.80(.11)
Overhand Throw	.81(.11)	.63(.18)	.52(.16)	.69(.12)	.54(.17)	.64(.15)
Kicking	.80(.09)	.87(.09)	.73(.10)	.65(.13)	.82(.07)	.77(.10)
Stationary Dribbling	.94(.04)	.82(.11)	.88(.06)	.85(.11)	.86(.07)	.87(.08)
Catching	.97(.03)	.81(.11)	.83(.06)	.82(.06)	.84(.05)	.85(.06)
Batting	.85(.07)	.94(.05)	.83(.08)	.71(.15)	.85(.08)	.84(.09)

Note. Numbers outside the parentheses indicate weighted kappa values. Numbers inside the parentheses indicate asymptotic standard error values. M_k= arithmetic mean agreement of 5 raters. M_{SE}= arithmetic mean of the raters' standard error values.

Table 7. Weighted Kappa Statistics for the Intra-Rater Analysis (Live vs. Video 2) – FG-COMPASS

	Raters					
Skill	*1*	*2*	*3*	*4*	*5*	M_{kw} (M_{SE})
Hopping	.83(.12)	.55(.12)	.62(.19)	.40(.26)	.63(.13)	.61(.16)
Horizontal Jump	.49(.21)	.68(.22)	.38(.21)	.28(.18)	.67(.18)	.50(.20)
Skipping	.96(.05)	.79(.12)	.87(.06)	.68(.20)	.52(.19)	.76(.12)
Overhand Throw	.52(.15)	.62(.17)	.54(.19)	.76(.11)	.66(.15)	.62(.15)
Kicking	.85(.08)	.70(.16)	.48(.27)	.65(.12)	.85(.08)	.71(.14)
Stationary Dribbling	.91(.06)	.82(.11)	.92(.05)	.88(.10)	.88(.06)	.88(.08)
Catching	.97(.03)	.83(.06)	.83(.07)	.86(.05)	.65(.22)	.83(.09)
Batting	.81(.07)	.84(.08)	.79(.09)	.74(.15)	.64(.19)	.76(.12)

Notes. Numbers outside the parentheses indicate weighted kappa values. Numbers inside the parentheses indicate asymptotic standard error values. M_k= arithmetic mean agreement of 5 raters. M_{SE}= arithmetic mean of the raters' standard error values.

Table 8. Weighted Kappa Statistics for the Intra-Rater Analysis (Video 1 vs. Video 2) – FG-COMPASS

	Raters					
Skill	*1*	*2*	*3*	*4*	*5*	M_{kw} (M_{SE})
Hopping	.89(.11)	.75(.12)	.88(.08)	.62(.25)	.80(.09)	.79(.13)
Horizontal Jump	.83(.13)	.39(.18)	.49(.21)	.56(.19)	1.00(.00)	.65(.14)
Skipping	1.00(.00)	.86(.11)	.89(.05)	.86(.12)	.83(.11)	.89(.08)
Overhand Throw	.53(.19)	.97(.03)	.60(.13)	.70(.10)	.86(.09)	.73(.11)
Kicking	.95(.05)	.74(.16)	.56(.19)	.91(.06)	.86(.06)	.80(.10)
Stationary Dribbling	.97(.03)	1.00(.00)	.97(.03)	.97(.03)	.90(.06)	.96(.03)
Catching	1.00(.00)	.82(.06)	.76(.10)	.86(.06)	.71(.21)	.83(.09)
Batting	.97(.03)	.83(.08)	.83(.07)	.89(.07)	.77(.15)	.86(.08)

Notes. Numbers outside the parentheses indicate weighted kappa values. Numbers inside the parentheses indicate asymptotic standard error values. M_k= arithmetic mean agreement of 5 raters. M_{SE}= arithmetic mean of the raters' standard error values.

Raters vs. expert consensus for the FG-COMPASS

The degree to which raters agreed with expert ratings for the FG-COMPASS was assessed by comparing the live ratings of the five raters with the expert's video scores on 20 randomly selected participants per skill. This was done to improve the internal validity of the study. The results are presented in Table 9. Weighted kappa scores ranged from .51 to .73 (*M* = .63) for the locomotion subtest and from .52 to .83 (*M* = .72) for the object manipulation subtest. The mean weighted kappa value for the total test was .69. The resulting weighted kappa indicated a "good" agreement Altman (1991) for the locomotion, object manipulation subtests as well as for the total test since the values were within .61 and .80.

Table 9. Weighted Kappa Statistics for the Raters vs. Expert Agreement – FG-COMPASS

Skill	*K_w(N)*	*SE*	*95% CI*
Hopping	.66*(17)*	.06	[.553, .775]
Horizontal Jump	.73*(18)*	.08	[.578, .879]
Skipping	.51*(18)*	.09	[.346, .680]
Overhand Throw	.70*(18)*	.06	[.574, .817]
Kicking	.73*(19)*	.06	[.613, .847]
Stationary Dribbling	.83*(17)*	.04	[.745, .914]
Catching	.83*(20)*	.03	[.768, .890]
Batting	.52*(19)*	.09	[.338, .692]

Notes: K_w= weighted kappa. N = sample size. SE= asymptotic standard error. CI = confidence interval.

Raters vs. expert consensus for the TGMD-2

Consensus estimates for the TGMD-2 was assessed by comparing the ratings of the five raters with the expert's video scores on 10 randomly selected participants per skill. The ICC values ranged from .50 to .95 (*M* = .68) for the locomotion subtest and from .67 to .92 (*M* = .83) for the object manipulation subtest. The total test ICC value was .75 (see Table 10). The resulting ICC averaged value for the locomotion subtest indicates "fair to good" agreement while the value for the object manipulation subtest indicates an "excellent" agreement. The averaged ICC value for all skills (total test) indicates a "fair to good" agreement (Fleiss, 2011).

Table 10. Intraclass Correlation Statistics for Raters vs. Expert Agreement – TGMD-2

Skills	*ICC*	*95% CI*
Run	.51(.72)	-.04-.85
Gallop	.59(.73)	-.04-.88
Hop	.50(.75)	-.06-.84
Leap	.73(.90)	.11-.93
Horizontal Jump	.95(.97)	.82-.99
Slide	.79(.87)	.33-.94
Striking a Stationary Ball	.92(.97)	.71-.98
Catch	.67(.86)	.14-.91
Kick	.77(.90)	.27-.94
Overhand Throw	.90(.94)	.65-.98
Underhand Roll	.85(.92)	.53-.96
Dribble	.87(.92)	.56-.97

Notes. N = 34. Values computed using two-way random, average measures ICC(2,5).

Correlations between the FG-COMPASS and the TGMD-2

Concurrent validity was investigated by comparing the locomotion, object manipulation, and total raw scores of the FG-COMPASS with those of the TGMD-2. Pearson partial product-moment correlations (controlling for age) comparing the scores of both tests indicate a moderate to strong correlation (see Table 11) for locomotion ($r_{xy.z}$ (31) =.52, $p < .01$), object manipulation ($r_{xy.z}$ (31) =.59, $p < .001$), and total scores ($r_{xy.z}$ (31) =.63, $p < .001$).

Table 11. Correlations Between FG-COMPASS and TGMD-2

FG-COMPASS	TGMD-2 (criterion)		
	Locomotion	Object manipulation	Total score
Locomotion	.52* (.60**)		
Object manipulation		.59** (.81**)	
Total score			.63** (.82**)

Notes: numbers inside of parentheses denote Pearson Zero-Order correlations, whereas numbers outside parentheses denote Pearson Partial correlations with age (in years) as the covariate.
* $p < .01$; ** $p < .001$

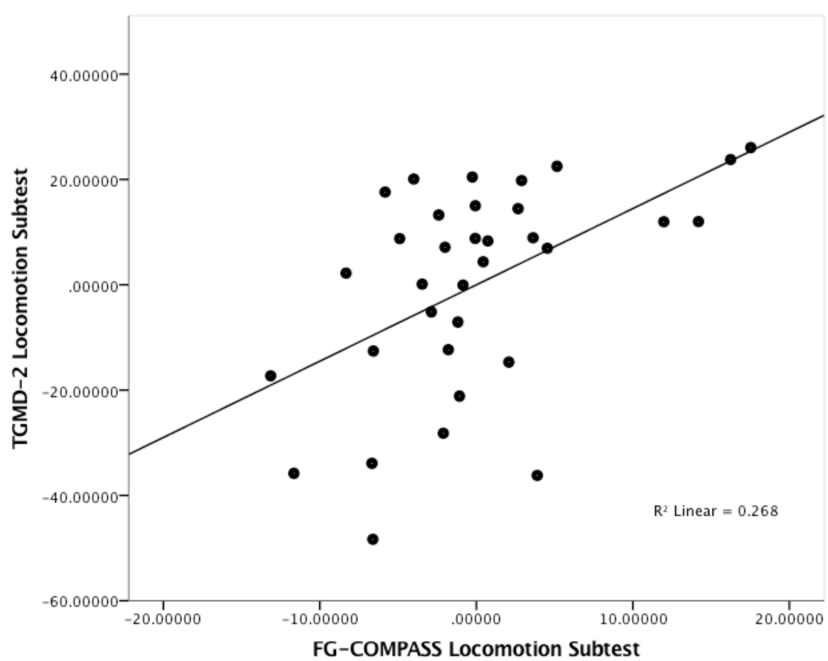


Figure 3. Scatter plot of FG-COMPASS locomotion subtest residual scores vs. TGMD-2 locomotion subtest residual scores (n = 34, $r_{xy.z}$ = .52).

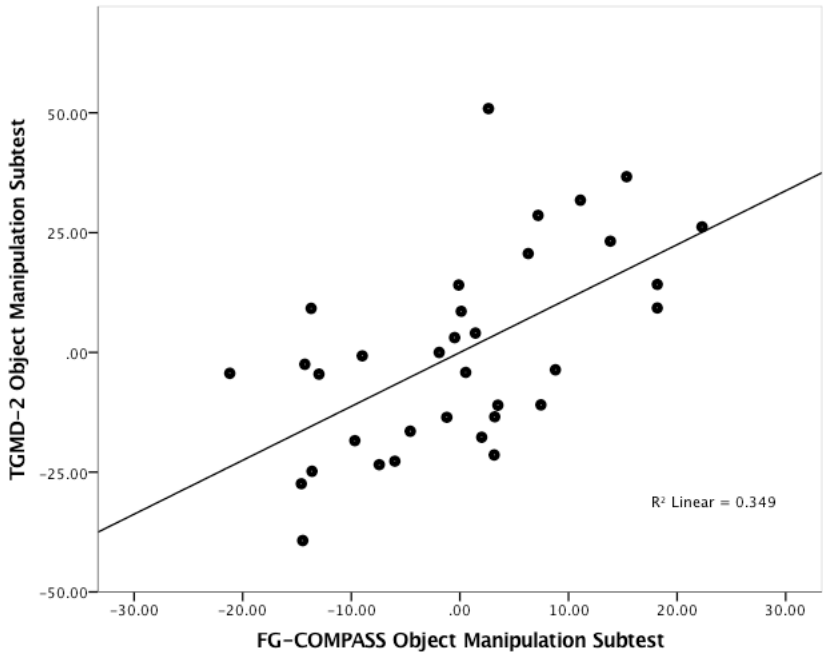


Figure 4. Scatter plot of FG-COMPASS object manipulation subtest residual scores vs. TGMD-2 object manipulation subtest residual scores (n = 34, $r_{xy.z}$ = .59).

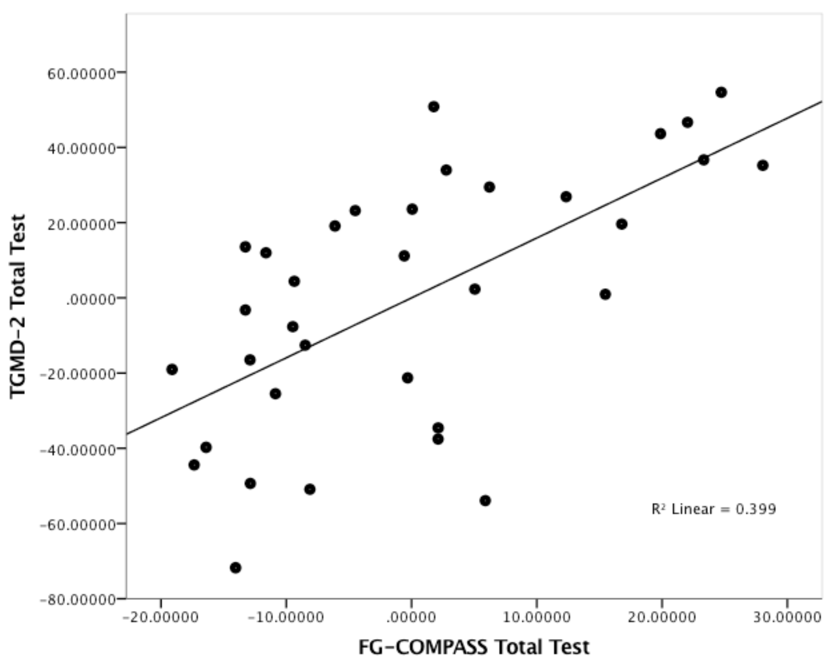


Figure 5. Scatter plot of FG-COMPASS total test residual scores vs. TGMD-2 total test residual scores (n = 34, $r_{xy.z}$ = .63).

Correlations between similar skills for the FG-COMPASS and TGMD-2

There are seven common skills between the FG-COMPASS and the TGMD-2. Two of the skills belong to the locomotion subtest (hop and jump) and five belong to the object manipulation subtest (overhand throw, kick, hand dribble, catch, and batting). Pearson partial product-moment correlations (controlling for age) comparing the total raw scores (from all 5 raters) indicate a "small" correlation for jump ($r_{xy.z}$ (31) =.12, $p > .05$) and kick ($r_{xy.z}$ (31) =.24, $p > .05$). "Moderate" correlations were found for overhand throw ($r_{xy.z}$ (31) =.43, $p < .05$), hand dribble ($r_{xy.z}$ (31) =.37, $p < .05$), catch ($r_{xy.z}$ (31) =.46, $p < .05$), and batting ($r_{xy.z}$ (31) =.37, $p < .05$). A "strong" correlation was found for the skill of hop ($r_{xy.z}$ (31) =.60, $p < .001$).

Table 12. Correlations Between Similar Skills for the FG-COMPASS and TGMD-2

	Partial Correlation		*Zero-order Correlation*	
Skill	$r_{xy.z}$	*df*	*r*	*df*
Hop	$.60^{**}$	31	$.60^{**}$	32
Jump	.12	31	.26	32
Overhand Throw	$.43^{*}$	31	$.66^{**}$	32
Kick	.24	31	.13	32
Hand Dribble	$.37^{*}$	31	$.69^{**}$	32
Catch	$.46^{*}$	31	$.50^{*}$	32
Batting	$.37^{*}$	31	$.40^{*}$	32

Notes: numbers inside of parentheses denote Pearson Zero-Order correlations, whereas numbers outside parentheses denote Pearson Partial correlations with age (in years) as the covariate.
$^{*}p < .01$; $^{**}p < .001$

Chapter 5: Discussion

The main purpose of this study was to collect concurrent validity evidence for the FG-COMPASS by comparing its results to the results of the TGMD-2. In addition, this study aimed to collect inter and intra-rater reliability for the FG-COMPASS. The results supported the first hypothesis that there would be a strong correlation between the subtests (i.e., locomotion and object manipulation) as well as between the total scores (i.e., both scales combined) of the two instruments. The results also supported the second and third hypotheses that there would be a "good" agreement for both inter and intra-rater reliability for the FG-COMPASS on the scores of both subtests and total scores.

Inter-rater reliability for the FG-COMPASS

The results for inter-rater reliability indicate a "good" agreement for the subtests' scores and the total score for the FG-COMPASS. Agreement for the object manipulation skills was slightly higher when compared to the locomotor skills, this trend has been seen in previous research when measuring inter-rater reliability (Houwen et al., 2010; Valentini, 2012). Dribbling ($K_w = .89$) and skipping ($K_w = .74$) resulting with the highest weighted kappa values. Hopping ($K_w = .50$) and throwing ($K_w = .62$) were the two skills with the lowest agreement among the five raters.

The performance criteria for hopping are solely based on the angle and movement of the non-support leg (also called the swing leg). There are a few possible explanations for why the agreement was moderate for hopping. First, there may have been insufficient training regarding the performance criteria, which could have led to guessing when classifying the performers. The raters' poor understanding of the criteria may also have led to more guessing, and in turn, poor agreement among the raters. Thirdly, borderline performances may have led to increased disagreement among raters. Performances from individuals who are transitioning between levels are difficult to classify. After a close inspection of the recorded performances, it was noted that two thirds of the performers could be considered "borderline".

Similarly, the weighted kappa value for overhand throwing ($K_w = .62$) was lower than expected. Recently, the inter-rater reliability for overhand throwing was investigated in a study Furtado & Gallagher (2015) that compared the ratings of three raters. The estimated weighted kappa was .93, which according to the adopted criteria is considered "very good". The obtained weighted kappa of .62 would increase to .79 if rater #5 was removed from the analysis. Rater #5 showed the largest disagreement when compared to the other four raters (see Table 4). The estimated agreement values were .45, .35, .31, and .40 between rater #5 and raters 1, 2, 3, and 4 respectively. These values are significantly lower when compared to the pair agreement values for the other raters, which ranged from .71 to .85. Such a discrepancy suggests that rater #5 possibly misunderstood the criteria associated with overhand throwing.

Intra-rater reliability for the FG-COMPASS

The current study assessed intra-rater reliability by having the raters reassess recorded performances of randomly selected children one and two weeks following the initial assessment. The results showed consistency between the scores of the raters across the live score and two reassessment scores. The averaged kappa values for both subtests were above .61 when comparing "Live vs. Video 1", "Live vs. Video 2" and "Video 1 vs. Video 2". This suggests a "good" agreement Altman (1991) for intra-rater reliability for the FG-COMPASS. These results are in accordance with the results of a previous study Furtado & Gallagher (2015) that investigated the intra-rater reliability of the FG-COMPASS involving three trained raters. The averaged weighted kappa scores were .91 and .92 for the locomotion and object manipulation subtests, respectively. In the current study, the weighted kappa scores ranged from .60 to .87 when the re-assessment was done within one week (Video 1) from the live assessment, and ranged from .50 to .88 for the re-assessment done within two weeks (Video 2) following the live assessment. The "moderate" agreement (K_w = .50) was observed for the skill of horizontal jumping. Individual kappa values were .68 and .67 for raters #2 and #5, respectively; but .49, .38, and .28 for raters #1, #3, #4, respectively. It is unclear at this point why the test-retest agreement for these three rates were so low when comparing Live vs. Video 2. It should be noted that the second re-assessment (Video 2) served as a controller to show that there were no significant discrepancies between scoring a child in a live performance and in a recorded performance.

Raters vs. expert consensus for the FG-COMPASS

The results showed a "good" agreement for the locomotion, object manipulation scores and total scores between the five raters and the primary investigator for the FG-COMPASS. Skipping (K_w = .51) and batting (K_w = .52) were the two skills with the lowest weighted kappa agreement. Agreement estimates for the remaining six skills were above .65.

After a careful analysis, it was noted that the disagreement was higher at the discriminatory performance criterion (i.e., first question of the decision tree) for skipping when comparing the scores from the five raters with the expert's scores. This performance criterion refers to arms moving in opposition to the legs during the performance. After reviewing the participants' recorded performances, it can be determined that there may have been some confusion when assessing this performance criterion. This could be due to either insufficient training or understanding by the raters. Some of the participants performed the skill with their arms extended and were not showing complete bi-lateral opposition. It appears that the raters inflated their scores by concluding the behavior was present (i.e., "yes" rating) as opposed to saying “no”. We suggest that further clarification should be provided to the raters who undergo training in future studies.

Batting was another skill that resulted in “moderate” agreement when comparing the scores of the five raters with the expert’s scores. Again, limited training and/or poor understanding of the criteria may have been the cause for the poor agreement observed with batting. During training, the raters were told to rate the attempts that resulted from a “good” toss, regardless whether the participant made contact with the ball or not. Some of the participants who did not made contact with the ball received a low score, despite the fact that the behavior was present. In addition, while reviewing the videos, it was noted that some of the participant’s performances were affected by receiving poor tosses. The participants should have received additional trials for every toss that was made poorly, which was not the case in many instances during the test administration. This should also be noted for future studies using the FG-COMPASS.

Correlations between FG-COMPASS and TGMD-2.

The results support the hypothesis of concurrent validity between the FG-COMPASS and TGMD-2. A moderate to strong and significant correlation was found between the total and both subtests (locomotion and object manipulation) scores for the two tests. Some studies Sun et al. (2011) have used the Pearson Zero-Order Product-Movement correlation index when comparing the results of their tests with the TGMD-2. In comparing the scores of the FG-COMPASS with those of the TGMD-2, we took a more conservative approach by relying on the Pearson Partial Product-Moment correlation index and controlling for age in years. We believe this a more accurate estimation of the correlation between the two tests as age may influence the association.

Correlations between similar skills from FG-COMPASS and TGMD-2.

There are a total of seven skills that are similar between both assessment tools. Hopping had the highest correlation ($r_{xy.z}$.60) when comparing the two instruments, which was strong and significant. Throwing, dribbling, catching, and batting all showed a significant moderate correlation with indexes ranging from .37 to .46. Kicking ($r_{xy.z}$ .24) and jumping ($r_{xy.z}$ .12) showed a small correlation between the two tests. It should be noted that the two instruments have many differences when comes to testing instructions and set up. Possible reasons for the low correlation of some of the skills are provided next.

Dribbling and batting for the FG-COMPASS mimic more 'realistic' situations than the TGMD-2. For dribbling, the FG-COMPASS has the child dribble a ball continuously for 20 seconds. Compare this with the instructions for the TGMD-2 where the child is asked to dribble the ball four times, catch it, and then repeat for two more trials. Arguably, dribbling the ball continuously for 20 seconds is considered a more difficult task. This could be used as an explanation for the moderate correlation of dribbling between the two tests.

The set up for batting (striking with a bat) is also very different between the two instruments. For the FG-COMPASS, the child is asked to strike a tossed ball, whereas the instructions for the TGMD-2 limit the child to strike a stationary ball off a tee. Similar to dribbling, striking a moving ball may more accurately capture the "true" performance level of a striker, making this a more difficult task when comparing to the TGMD-2's. This difference may explain the moderate correlation observed for the skill of striking with a bat when comparing the two tests.

Kicking had the second lowest correlation between the tests. The instructions provided by the test administrator for the FG-COMPASS are: "I want to see you kick the ball against the wall." Note that the instructor does not tell the child to kick the ball by taking multiple steps prior to making contact with the ball. To be considered a level 4 on the FG-COMPASS, we would expect the performer to naturally back away from the ball, and then approach the ball while running before ball contact. To elicit this behavior, the performer is asked to pick a ball from inside a basket that is 10 feet behind the spot where the ball is to be placed and then kick it against the wall. The performer must go back to the basket after each trial. With the TGMD-2, the performer starts on a line 10 feet from the ball, and is asked to run to the ball and kick it towards the wall. Therefore, for children who would not naturally perform the kicking with a running approach would score on the lower bracket of the FG-COMPASS decision tree. This may explain the weak, not significant, correlation for kicking between the two tests.

Jumping showed the lowest correlation between the FG-COMPASS and the TGMD-2. After reviewing the videos, it was concluded that the test administrators may have failed to follow the instructions from the test manual when testing horizontal jumping for the FG-COMPASS. The test set up is to have two lines two feet away from each other. As per the instructions, the performer is asked to "Jump over the second line as far as possible." There may have been insufficient accurate instructions because the majority of the participants were jumping on the second line, not over the line, therefore inhibiting their true performance. The second line could have been viewed as a target for the participant where some of them just simply stepped to the other line and did not even perform an actual jump. Jumping for the TGMD-2 has the child begin by standing at one line and the test administrator asks the child to jump forward as far as they can. When comparing the participants on both assessments, it is clear that most participants are capable of performing an advanced jump, but due to inaccurate instructions, the participants scored extremely low on the FG-COMPASS.

Agreement Estimates for the TGMD-2

Inter-rater reliability for the TGMD-2. The inter-rater reliability results for TGMD-2 were found to all be “excellent” >.75 with the exception of running “fair to good” .74 (Fleiss, 2011). This evidence demonstrates that the five raters were in agreement with the performance criteria following training with the primary investigator. Although inter-rater reliability for the TGMD-2 was not a main purpose of this study, it was done to measure agreement between the five raters to support accurate training and scoring when determining validity between the FG-COMPASS and TGMD-2.

Raters vs. expert consensus for the TGMD-2. Overall, the consensus between the raters and the expert was high for the TGMD-2. For seven out of the twelve skills, the agreement was considered “excellent” when compared to the adopted criteria (Fleiss, 2011). The agreement was considered “fair to good” for the remaining five skills with hopping (ICC = .50) and running (ICC = .51) being the lowest scores. There may have been a limited understanding and/or training on the performance criteria for hopping. When assessing the skill of hopping with the TGMD-2, one needs to pay close attention leg being assessed. One complete trial for hopping includes the child hopping on their right and left leg. Therefore, if a performance criterion was present for the right leg but absent for the left leg, the participant would receive a “0” for that performance criteria. This may have caused some confusion for the raters as they could have possibly scored the child incorrectly if the performance criterion was only seen on either the right or left leg.

Conclusion

The results of this study provide further estimates of validity and reliability for the FG-COMPASS. When combined with the results of previous studies (Furtado, 2004; Furtado &

Gallagher, 2012, 2016), these estimates support the premise that the FG-COMPASS can be used in a live setting to estimate levels of fundamental movement skill development in children ages 5 through 10. In addition, having a more practical valid assessment tool could potentially enhance physical education programs, which would increase skill development in children at an early age. Children with higher motor competence will be more likely to participate in more structured physical activities, which could increase the participation in physical activities at a later age (Bellows et al., 2013; Cliff et al., 2012; Lemos et al., 2012). Therefore, an assessment tool like the FG-COMPASS could be a critical component for physical education teachers.

This was the first study to test the concurrent validity of the FG-COMPASS; therefore, more studies are needed to support the results. Future studies should consider conducting the pilot study and training sessions, assessing the FG-COMPASS in a live setting versus training on recorded performances. They should also emphasize the instructions and demonstrations given to the participants. Each participant should have a clear understanding of what the skill is and should be shown level 4 performances by a test administrator prior to completing any practice or testing trials.

Limitations

This study is not without limitations. First, this study encompassed a small sample size compared to other studies investigating validity and reliability (Cano-Cappellacci et al., 2015; Houwen et al., 2010; S. Kim, Kim, Valentini, & Clark, 2014; Simons et al., 2008; Sun et al., 2011; Valentini, 2012). The primary investigator projected to use a sample size of 40 participants; however, only 34 completed the study with full participation. Within the small sample size there were almost twice as many girls than boys that completed the study. Gender differences with regards to FMS proficiency have been found in previous studies (Barnett et al., 2009). However, we do not feel that the larger female group compromised the interpretation of the results. Secondly, in many instances, the PI was not present when the tests were being administrated, since data were collected on two different locations inside the school and at about the same time. The insufficient standardization when collecting the data across all participants, especially in the case of the FG-COMPASS, made comparisons between the tests more difficult.

References

Altman, D. (1991). *Practical statistics for medical research* (1st ed.). New York: Chapman and Hall.

Barnett, L. M., van Beurden, E., Morgan, P. J., Brooks, L. O., & Beard, J. R. (2009). Childhood motor skill proficiency as a predictor of adolescent physical activity. *Journal of Adolescent Health*, *44*(3), 252–259.

Bayer, O., Bolte, G., Morlock, G., Ruckinger, S., & von Kries, R. (2009). A simple assessment of physical activity is associated with obesity and motor fitness in pre-school children. *Public Health Nutrition*, *12*(8), 1242–1247.

Bellows, L. L., Davies, P. L., Anderson, J., & Kennedy, C. (2013). Effectiveness of a physical activity intervention for head start preschoolers: A randomized intervention study. *American Journal of Occupational Therapy*, *67*(1), 28+.

Biddle, S., & Armstrong, N. (1992). Children's physical activity: An exploratory study of psychological correlates. *Social Science & Medicine*, *34*(3), 325–331. http://doi.org/10.1016/0277-9536(92)90274-T

Brown, T., & Lalor, A. (2009). The Movement Assessment Battery for Children—Second Edition (MABC-2): A review and critique. *Physical & Occupational Therapy in Pediatrics*, *29*(1), 86–103. http://doi.org/10.1080/01942630802574908

Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Champaign, IL: Human Kinetics.

Butterfield, S. A., Angell, R. M., & Mason, C. A. (2012). Age and sex differences in object control skills by children ages 5 to 14. *Perceptual and Motor Skills*, *114*(1), 261–274. http://doi.org/10.2466/10.11.25.PMS.114.1.261-274

Campbell, S. K., & Kolobe, T. H. (2000). Concurrent validity of the Test of Infant Motor Performance with the Alberta Infant Motor Scale. *Pediatric Physical Therapy*, *12*, 2–9.

Cano-Cappellacci, M., Leyton, F. A., Carreño, J. D., Cano-Cappellacci, M., Leyton, F. A., & Carreño, J. D. (2015). Content validity and reliability of test of gross motor development in Chilean children. *Revista de Saúde Pública*, *49*. http://doi.org/10.1590/S0034-8910.2015049005724

Carroll, B., & Loumidis, J. (2001). Children's perceived competence and enjoyment in physical education and physical activity outside school. *European Physical Education Review*, *7*(1), 24–43. http://doi.org/10.1177/1356336X010071005

Castelli, D. M., & Valley, J. A. (2007). Chapter 3: The relationship of physical fitness and motor competence to physical activity. *Journal of Teaching in Physical Education*, *26*(4), 358–374.

Cawley, J., Meyerhoefer, C., & Newhouse, D. (2007). The impact of state physical education requirements on youth physical activity and overweight. *Health Economics*, *16*(12), 1287–1301. http://doi.org/10.1002/hec.1218

Chien, C.-W., Scanlon, C., Rodger, S., & Copley, J. (2014). Intra-and inter-rater reliability of the assessment of children's hand skills based on video recordings. *British Journal of Occupational Therapy*, *77*(2), 82+.

Clark, J. E. (2007). On the problem of motor skill development. *Journal of Physical Education, Recreation & Dance (JOPERD)*, *78*(5), 39–44.

Cliff, D. P., Okely, A. D., Morgan, P. J., Jones, R. A., Steele, J. R., & Baur, L. A. (2012). Proficiency deficiency: Mastery of fundamental movement skills and skill components in overweight and obese children. *Obesity*, *20*(5), 1024–1033. http://doi.org/10.1038/oby.2011.241

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press.

Cools, W., De Martelaer, K., Samaey, C., & Andries, C. (2009). Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. *Journal of Sports Science & Medicine*, *8*(2), 154–168.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Fisher, A., Reilly, J. J., Kelly, L. A., Montegomery, C., Williamson, A., Paton, J. Y., & Grant, S. (2005). Fundamental movement skills and habitual physical activity in young children. *Medicine & Science in Sports & Exercise April 2005*, *37*(4), 684–688.

Fleiss, J. L. (2011). *Design and analysis of clinical experiments*. John Wiley & Sons.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619. http://doi.org/10.1177/001316447303300309

Furtado, O., Jr. (2004). *Furtado-Gallagher Movement Skill Assessment: Collecting evidence for content-related validity* (Master's Thesis). University of Pittsburgh, Pittsburgh, PA. Retrieved from http://etd.library.pitt.edu/ETD/available/etd-12042004-093510/

Furtado, O., Jr. (2009). *Development and initial validation of the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System - FG-COMPASS*

(Doctoral dissertation). Retrieved from http://etd.library.pitt.edu/ETD/available/etd-08052009-134650/

Furtado, O., Jr, & Gallagher, J. D. (2012). The reliability of classification decisions for the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System - FG-COMPASS. *Research Quarterly for Exercise and Sport*, *83*(3), 383–390. http://doi.org/10.5641/027013612802573021

Furtado, O., Jr, & Gallagher, J. D. (2016). *Collecting further validity and reliability evidence for the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System*. Manuscript in preparation.

Gabbard, C. P. (2004). *Lifelong motor development* (4th Ed.). San Francisco, CA: Pearson Education Inc.

Gallahue, D. L., & Cleland-Donnelly, F. (2007). *Developmental physical education for all children*. Champaign, IL: Human Kinetics.

Gallahue, D. L., & Ozmun, J. C. (2006). *Understanding motor development infants, children, adolescents, adults* (6th Ed.). New York, NY: McGraw-Hill.

Graf, C., Koch, B., Kretschmann-Kandel, E., Falkowski, G., Christ, H., Coburger, S., … Dordel, S. (2004). Correlation between BMI, leisure habits and motor abilities in childhood (CHILT-Project). *International Journal of Obesity*, *28*(1), 22–26. http://doi.org/10.1038/sj.ijo.0802428

Haywood, K. M., & Getchell, N. (2014). *Lifespan motor development* (6th Ed.). Champaign, IL: Human Kinetics.

Haywood, K. M., Roberton, M. A., & Getchell, N. (2012). *Advanced analysis of motor development*. Champaign, IL: Human Kinetics.

Houwen, S., Hartman, E., Jonker, L., & Visscher, C. (2010). Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. *Adapted Physical Activity Quarterly*, *27*, 143–159.

IBM SPSS Statistics for Windows. (2013). (Version 22.0) [Windows]. Armonk, NY: IBM Corp.

Kim, J. (2012). Are physical education-related state policies and schools' physical education requirement related to children's physical activity and obesity? *Journal of School Health*, *82*(6), 268–276. http://doi.org/10.1111/j.1746-1561.2012.00697.x

Kim, S., Kim, M. J., Valentini, N. C., & Clark, J. E. (2014). Validity and reliability of the TGMD-2 for South Korean children. *Journal of Motor Behavior*, *46*(5), 351–356. http://doi.org/10.1080/00222895.2014.914886

Lemos, A. G., Avigo, E. L., & Barela, J. A. (2012). Physical education in kindergarten promotes fundamental motor skill development. *Advances in Physical Education*, *02*(01), 17–21. http://doi.org/10.4236/ape.2012.21003

Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, *76*(5), 365–377. http://doi.org/10.1037/h0031643

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. http://doi.org/10.1080/01621459.1988.10478722

Lopes, V. P., Stodden, D. F., Bianchi, M. M., Maia, J. A. R., & Rodrigues, L. P. (2012). Correlation between BMI and motor coordination in children. *Journal of Science and Medicine in Sport*, *15*(1), 38–43. http://doi.org/10.1016/j.jsams.2011.07.005

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46. http://doi.org/10.1037/1082-989X.1.1.30

McKenzie, T. L., Sallis, J. F., Broyles, S. L., Zive, M. M., Nader, P. R., Berry, C. C., & Brennan, J. J. (2002). Childhood movement skills: Predictors of physical activity in Anglo American and Mexican American adolescents? *Research Quarterly for Exercise and Sport*, *73*(3), 238–244. http://doi.org/10.1080/02701367.2002.10609017

Piek, J. P., Hands, B., & Licari, M. K. (2012). Assessment of motor functioning in the preschool period. *Neuropsychology Review*, *22*(4), 402–413. http://doi.org/10.1007/s11065-012-9211-4

Provost, B., Crowe, T. K., & McClain, C. (2000). Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales in two-year-old children. *Physical & Occupational Therapy in Pediatrics*, *20*(1), 5–18.

Roberton, M. A., & Halverson, L. (1984). *Developing children - their changing movement: A guide for teachers*. Philadelphia: Lea & Febiger.

Sallis, J. F., Alcaraz, J. E., McKenzie, T. L., & Hovell, M. F. (1999). Predictors of change in children's physical activity over 20 months: Variations by gender and level of adiposity. *American Journal of Preventive Medicine*, *16*(3), 222–229. http://doi.org/10.1016/S0749-3797(98)00154-8

Seefeldt, V., & Haubenstricker, J. (1982). Patterns, phases, or stages: An analytical model for the study of developmental movement. In J.A.S. Kelso & J. Clark (Eds.), *The development of movement control and coordination* (pp. 309-318). New York: Wiley.

Simons, J., Daly, D., Theodorou, F., Caron, C., & Andoniadou, E. (2008). Validity and reliability of the TGMD-2 in 7-10-year-old Flemish children with intellectual disability. *Adapted Physical Activity Quarterly*, *25*(1).

Slater, L. M. Bp., Hillier, S. L., & Civetta, L. R. Bp. (Hons). (2010). The clinimetric properties of performance-based gross motor tests used for children with developmental coordination disorder: A systematic review. *Pediatric Physical Therapy Summer 2010*, *22*(2), 170–179. http://doi.org/10.1097/PEP.0b013e3181dbeff0

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, *9*(4).

Sun, S.-H., Sun, H.-L., Zhu, Y.-C., Huang, L., & Hsieh, Y.-L. (2011). Concurrent validity of preschooler gross motor quality scale with Test of Gross Motor Development-2. *Research in Developmental Disabilities*, *32*(3), 1163–1168. http://doi.org/10.1016/j.ridd.2011.01.007

Ulrich, D. A. (2000). *Test of gross motor development examiner's manual* (2nd Ed.). Austin, TX: Pro-ed.

Valentini, N. C. (2012). Validity and reliability of the TGMD-2 for Brazilian children. *Journal of Motor Behavior*, *44*(4), 275–280. http://doi.org/10.1080/00222895.2012.700967

Vincent, W. J., & Weir, J. P. (2012). *Statistics in kinesiology* (4th Ed.). Champaign, IL: Human Kinetics.

Westendorp, M., Hartman, E., Houwen, S., Huijgen, B. C. H., Smith, J., & Visscher, C. (2014). A longitudinal study on gross motor development in children with learning disorders. *Research in Developmental Disabilities*, (35), 357–363.

Wong, K. Y. A., & Cheung, S. Y. (2010). Confirmatory factor analysis of the Test of Gross Motor Development-2. *Measurement in Physical Education and Exercise Science*, *14*(3), 202–209. http://doi.org/http://dx.doi.org.libproxy.csun.edu/10.1080/10913671003726968

Woods, A. M., Bolton, K. N., Graber, K. C., & Crull, G. S. (2007). Chapter 5: Influences of perceived motor competence and motives on children's physical activity. *Journal of Teaching in Physical Education*, *26*(4), 390–403.

Appendix A: Participant Recruitment Packet

Will you help us look at different motor skills of children using two movement skill tests?

Graduate student, Mackinsey Woolever and faculty advisor Ovande Furtado are comparing motor skills using two assessment tools for children age 5 to 10 years.

What: Your child will be asked to complete movement skills (e.g., running, jumping, hopping, catching, kicking, and throwing and others). The results from both tests will be compared to determine if the new test can be used to look at gross motor performance in children 5-10 years.
When: We plan to begin collecting data the week of March 16th, 2015.
Where: Data will be collected during recess and lunch breaks and will occur once a week for 8 weeks.

Please read the attached Parent/Guardian consent form explaining the details of the study and potential benefits. If you are interested, please sign one copy of the Parent/Guardian consent forms. In addition, please complete the attached physical activity readiness questionnaire (PAR-Q). A researcher will confirm verbal assent from your child prior to participating in the study. Please keep the second copy of the consent form for your records. If you are interested in having your child participate in this study please return the signed copy of the parent consent and the PAR-Q in the attached envelope, sealed to the main office for Mackinsey to pick up by ______________________.

Mackinsey Woolever is a graduate student at California State University Northridge, in progress of receiving her Masters degree in Kinesiology/Motor Behavior. **Phone:** (818) 641-9208
Email: Mackinsey.woolever.67@my.csun.edu

Dr. Ovande Furtado is a professor who is assisting Mackinsey on her thesis project, and who developed one of the assessment tools being used in this study. **Phone:** (818) 677-5968
Email: Ovande.furtado@csun.edu

California State University, Northridge
PARENT OR GUARDIAN CONSENT FOR CHILD PARTICIPATION IN RESEARCH

Assessing movement by comparing the results from two assessment tools.

You are being asked to consent for your child to participate in a research study. This study is being conducted by Mackinsey Woolever as part of the requirements for the M.S. degree in Kinesiology/Motor Behavior. Participation in this study is completely voluntary. Please read the information below and ask questions about anything that you do not understand before deciding if you want to allow your child to participate. A researcher listed below will be available to answer your questions.

RESEARCH TEAM

Researcher:
Mackinsey Woolever
Department of Kinesiology
18111 Nordhoff St.
Northridge, CA 91330-8287
818-641-9208
Mackinsey.woolever.67@my.csun.edu

Faculty Advisor:
Ovande Furtado Jr.
Department of Kinesiology
18111 Nordhoff St.
Northridge, CA 91330-8287
818-677-5968
ovande.furtado@csun.edu

PURPOSE OF STUDY

The purpose of this research study is to compare the scores from two assessment tools by having children complete motor skills (running, jumping, throwing, kicking, etc.).

SUBJECTS

Inclusion Requirements

Your child is eligible to participate in this study if he/she is between the ages of 5 and 10 years and is attending Calahan Elementary School. In addition, a physical activity readiness questionnaire (PAR-Q) must be completed for each child before they are eligible to participate.

Exclusion Requirements

Your child is not eligible to participate in this study if he/she a) has any known movement delays or disabilities that affect his/her ability to move , b) is participating in another study that may affect his/her performance in this study, c) has sustained an injury within the last 6 months and d) you or your child declines the option of being video recorded. In addition, if you answer "yes" on items 1, 2, 3, 4, or 5 on the PAR-Q your child will be excluded from the study. For any "yes" answers for items 6-13 on the PAR-Q, please be sure to explain in detail in the space provided on the PAR-Q form.

Time Commitment

This study will involve your child completing motor skills (e.g., running, jumping, throwing, and kicking, etc.) once a week, for approximately 15-20 minutes a day, over the span of 7 weeks for a total time of 135-180 minutes.

PROCEDURES

The following procedures will occur:

1. Your child will be asked to complete a total of 13 motor skills over a 7-week period. The order is as follows:
   Week 1: Running, Horizontal Jumping, and Skipping
   Week 2: Galloping and Hopping
   Week 3: Sliding and Leaping
   Week 4: Throwing and Dribbling
   Week 5: Kicking
   Week 6: Catching and Underhand Rolling
   Week 7: Striking a moving/stationary ball and Make up day for any missed skills

2. Your child will complete these skills during scheduled recess and lunch breaks to avoid missing class time. Your child will complete these skills inside a gym/auditorium using a wood surface to reduce the risk of injury. If, a gym/auditorium is not available, the testing will take place on the playground. Undergraduate Kinesiology students from California State University Northridge (CSUN) will assess your child's motor performance using the (Test of Gross Motor Development-2 (TGMD-2) and the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System (FG-COMPASS). During each session your child will be video-recorded for the purpose of accurately assessing motor performance. Your child will be given an identification number to ensure confidentiality between your child and the CSUN students. This identification number is assigned by the research assistants and is only for the purpose of this study.

3. Prior to performing the skill(s) your child will state and show their ID number to the CSUN students and the camera. During the session your child will receive instructions and demonstrations if needed from the CSUN students for each skill. Your child will complete multiple trials (practice and testing) for each skill for both tests. Once your child has completed both the TGMD-2 and FG-COMPASS trials they may return to their recess or lunch break.

RISKS AND DISCOMFORTS

The possible risks and/or discomforts associated with the procedures described in this study include: any injury that could occur during regular physical activity participation or a physical education class. To minimize the possible risks of injury your child will complete each skill in a gym area if accessible on a wooden surface. Your child will be performing skills in a large open area free from any object or obstructions that may get in their way. Your child will also only be allowed to complete the assessment each week if he/she is wearing proper shoe attire (tennis shoes). Your child may sustain an injury due to performing these movement skills (i.e., falling, bruises, abrasions muscle strains, sprains, and in rare cases broken bones). Potential risks may require further treatment from medical professional at your own cost. School protocol(s) will be followed in case of an injury or emergency. The school nurse will be notified immediately, referral to additional medical treatment is at your own cost. This study involves no more than minimal risks. There are no known harms or discomforts associated with this study beyond those encountered in normal daily life.

BENEFITS

Subject Benefits

Your child may not directly benefit from participating in this study.

Benefits to Others or Society

Assessing motor development is critical in school-aged children. If physical education teachers were able to use an assessment tool during their lessons, they could apply the information into planning their activities. It is important to know if children are struggling in certain areas so that the physical education teacher(s) can design activities to advance their development. The FG-COMPASS is designed to be used in a live school setting; therefore, it is a more practical alternative to other assessment tools. The FG-COMPASS can be used by any physical education teacher in any school district if it is found to be a valid and reliable assessment tool.

ALTERNATIVES TO PARTICIPATION

The only alternative to participation in this study is not to participate.

COMPENSATION, COSTS AND REIMBURSEMENT

Compensation for Participation

Your child will not be paid for his/her participation in this research study.

Costs

There is no cost to you for your child's participation in this study.

Reimbursement

You will not be reimbursed for any out of pocket expenses, such as parking or transportation fees.

WITHDRAWAL OR TERMINATION FROM THE STUDY AND CONSEQUENCES

You are free to withdraw your child from this study at any time. **If you decide to withdraw your child from this study you should notify the research team immediately.** The research team may also end your child's participation in this study if he/she does not follow instructions, misses scheduled visits, or if his/her safety and welfare are at risk.

CONFIDENTIALITY

Subject Identifiable Data

All identifiable information (your child's name) will be collected and stored in a locked file cabinet in Dr. Furtado's office. In addition, a list linking your child's name and their ID number will also be kept in the locked file cabinet. A spreadsheet that includes: your child's ID #, gender, birthdate, and classroom teacher will be kept in an encrypted folder on a password-protected computer in the Motor Development lab at CSUN (Redwood Hall Room 178). Access to this classroom is restricted by omni code. The ID number and identifiable information will be kept in separate locations to ensure confidentiality.

Data Storage

All research data will be stored electronically on a secure computer with password protection.

The video recordings will have no identifiable information and will also be stored on a computer in the Motor Development Lab and will be erased 5 years after the completion of this study.

Data Access

The researcher and faculty advisor named on the first page of this form will have access to your child's records. Any information derived from this research project that personally identifies your child will not be voluntarily released or disclosed without your separate consent, except as specifically required by law. Publications and/or presentations that result from this study will not include identifiable information about your child. The CSUN students will have access to the video recordings; however, the video recordings of your children will not be viewed outside the Motor Development lab under any circumstances.

Data Retention

The researchers intend to keep the research data for approximately 5 years and then will be destroyed.

Mandated Reporting

Under California law, the researchers are required to report known or reasonably suspected incidents of abuse or neglect of a child, dependent adult or elder, including, but not limited to, physical, sexual, emotional, and financial abuse or neglect. If any researcher has or is given such information, she may be required to report it to the authorities.

IF YOU HAVE QUESTIONS

If you have any comments, concerns, or questions regarding the conduct of this research please contact the research team listed on the first page of this form.

CSU, Northridge
Human Subjects Committee
Approved: 3/13/15
Void After: 12/8/15 PS

If you have concerns or complaints about the research study, research team, or questions about your child's rights as a research participant, please contact Research and Sponsored Projects, 18111 Nordhoff Street, California State University, Northridge, Northridge, CA 91330-8232, or phone 818-677-2901.

VOLUNTARY PARTICIPATION STATEMENT

You should not sign this form unless you have read it and been given a copy of it to keep. **Participation in this study is voluntary.** Your child may refuse to answer any question or discontinue his/her involvement at any time without penalty or loss of benefits to which you and your child might otherwise be entitled. Your decision will not affect your relationship with California State University, Northridge or Calahan Elementary School. Your signature below indicates that you have read the information in this consent form and have had a chance to ask any questions that you have about the study.

A member of the research team will explain the study to your child and ask your child whether or not he/she wishes to participate. If your child declines to participate then your child will not be included in the study. Additionally, if your child says yes and declines later your child will be withdrawn from the study at his/her request.

I agree to allow my child to participate in the study.

___ My child may participate in this study and be video recorded.
___ My child may not be video recorded.

Child's Name: ____ Sex: M F
Last First MI (circle one)

Age: ______ years ______ months Child's Date of Birth: ______________________

____ ____________________
Parent or Guardian Signature Date

____
Printed Name of Parent or Guardian

____ ____________________
Researcher Signature Date

____
Printed Name of Researcher

CSU, Northridge
Human Subjects Committee
Approved: 3/13/15
Void After: 12/8/15 Cg

Child's PAR-Q Screening Form

Child's Name: ______________________________________

Parent/Guardian Name: ____________________________

Child's Date of Birth: ________________________________

Emergency Contact Details

Home: ____

Name and relationship to child: _____________________

Work: ____

Name and relationship to child: _____________________

Mobile: ____

Name and relationship to child: _____________________

Health Questions:

Does your child have or has he or she ever experienced any of the following?	Please circle
High or Low Blood Pressure	Y / N
Elevated blood cholesterol	Y / N
Diabetes	Y / N
Chest pains brought on by physical exertion	Y / N
Childhood epilepsy	Y / N
Dizziness or fainting	Y / N
A bone, joint or muscular problems with arthritis	Y / N
Asthma or respiratory Problems	Y / N
Any sustained injuries or illness	Y / N
Any allergies	Y / N
Is your child taking any medication	Y / N
Has your doctor ever advised your child to exercise	Y / N
Is there any reason not mentioned above why any type or physical activity may not be suitable for your child	Y / N

If answered "yes" to any of the above questions please give full details here:

In signing this form, I the parent/guardian of the aforementioned child, affirm that I have read this form in its entirety and I have answered the questions accurately and to the best of my knowledge.

Parents Signature: ________________________________Date: __________________

Verbal Assent Script (Children 5-10 years)

Hi. My name is Mackinsey Woolever. I'm a student at California State University of Northridge. Right now, I'm trying to look at two different movement tests to see if they are similar. I would like to ask you to help me by being in my study, but before I do, I want to explain what will happen if you decide to help me.

I will ask you to complete skills such as running, jumping, kicking and more each week during your recess or lunch breaks. You will only be asked to do these skills for 15-20 minutes each week. We will film you as you complete each skill so my helpers can observe how you move. This study has a small risk of you being hurt, such as falling, or receiving bruises and/or cuts. The chances of you getting hurt during my study are no greater than you participating in regular physical activities during lunch or recess.

When I tell other people about my study, I will not use your name, and no one will be able to tell who I'm talking about.

Your mom/dad/guardian says it's okay for you to be in my study. But if you don't want to be in the study, you don't have to be. What you decide won't make any difference with your school and friends. I won't be upset, and no one else will be upset, if you don't want to be in my study. If you want to be in my study now, but change your mind later, that's okay. You can stop at any time. If there is anything you don't understand you should tell me so I can explain it to you

You can ask me questions about the study. If you have a question later that you don't think of now, you can call me or ask your parents/teacher to call me or send me an email.

Do you have any questions for me now?

Would you like to be in my study?

Is it okay if I videotape you during my study?

Name of Child: ____________________ **Parental Permission on File:** ☐ Yes ☐ No
(If "No," do not proceed with assent or research procedures.)

Child's Voluntary Response to Participation: ☐ Yes ☐ No

Child's Voluntary Response to being Video Recorded: ☐ Yes ☐ No

Signature of Researcher:____________________ **Date:** ____________

(Optional) Signature of Child: ____________________

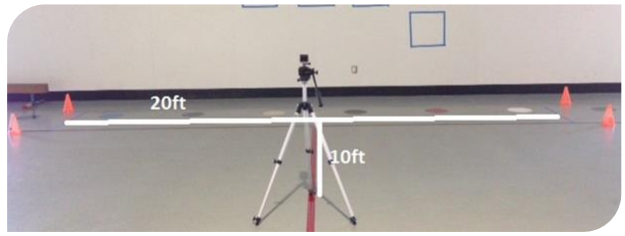
Appendix B: FG-COMPASS Selected Testing Protocols

HOPPING

→ Test Setup and Directions

<table>
<tr>
<td>
<u>Set up</u>
▪ Construct a 15-feet long traveling lane using cones.
▪ Tape starting and ending lines on the floor.
▪ Stand perpendicular to the traveling lane so you can see both the starting and ending points.
▪ Place 2 cones (each end) 1 foot before the starting and ending lines.

<u>Directions for Students</u>
▪ I want to see you hopping.
▪ Choose your preferred leg to hop.
▪ Start from the starting line and do not stop until you pass the ending line; then, come back using the same leg.
▪ This is not a race. Show your best form.
▪ Watch as I demonstrate.
▪ Ready, set, go!

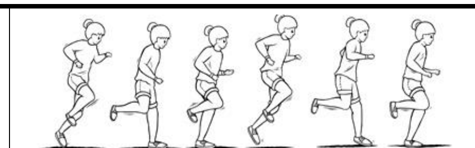
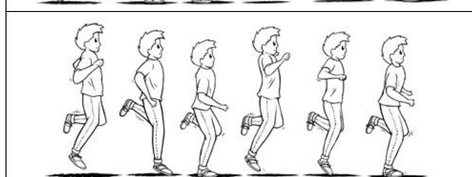
<u>Observations for Assessors</u>
▪ Give examinee 4 trials (1st trial is for practice)
</td>
<td>


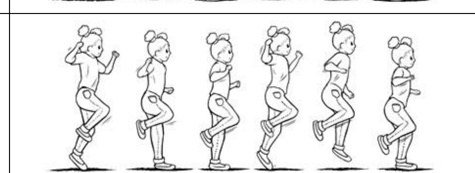
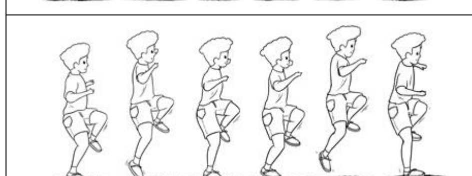
</td>
</tr>
</table>

→ Coding Grid & Pictorial Representations

				EXAMINEE ID →														
HOP	Is the thigh of the nonsupport leg in a vertical position with knee flexed at 90^0 or less?	**Y**	Does the foot of the nonsupport leg pass behind/in front the support leg?	**Y**	4	4	4	4	4	4	4	4	4	4	4	4	4	4
				N	3	3	3	3	3	3	3	3	3	3	3	3	3	3
		N	Is the nonsupport leg held in front of the body?	**N**	2	2	2	2	2	2	2	2	2	2	2	2	2	2
				Y	1	1	1	1	1	1	1	1	1	1	1	1	1	1

<table>
<tr>
<td rowspan="4">Answer YES, regardless whether the leg is moving or not. This will be confirmed in the next question.

Note that having the leg extended in front of the body does NOT mean the thigh is in a vertical position. This a pattern consistent with level 1.</td>
<td rowspan="2">Intended to verify whether the entire leg swings back and forth like a pendulum as it aids in force production.</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td rowspan="2">Intended to verify whether the nonsupport leg is flexed at 90^0 or less with the nonsupport thigh parallel to the surface.

Answer YES, even when the nonsupport leg is extended in front of the body.</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</table>

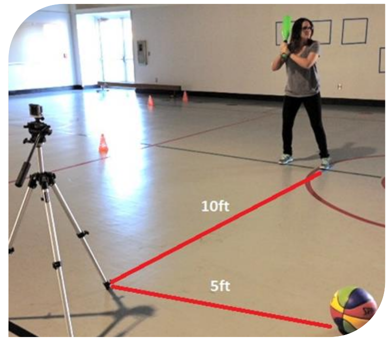
Betting

→ Test Setup and Directions

<u>Set up</u>
- Tape an "X"on the floor 20 feet from the wall.
- Stand slightly to the side (about 12 feet) facing the examinee.
- Invert the position, having the examinee facing the opposite wall/open space if he/she is left handed.

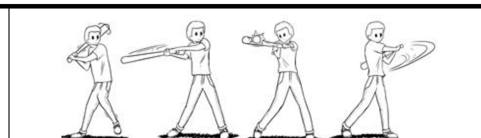
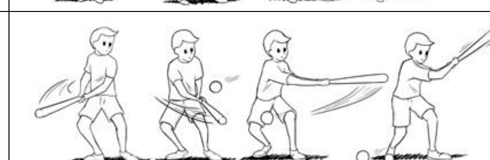
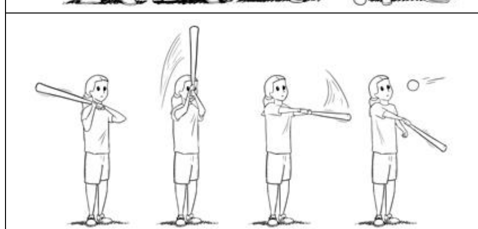
<u>Directions for Students</u>
- I want to see you striking a ball tossed in your direction.
- Try to stay near the "X", but you are free to move as the ball approaches.
- Strike the ball against the wall/to the open space.
- Watch as I demonstrate.
- Read, set, go!

<u>Observations for Assessors</u>
- Give the examinee 4 trials (1st trial is for practice).
- Use an underhand toss.
- Toss the ball just above the examinee's hip level.
- Repeat any attempt that results in a bad toss (too high/low or on the sides).
- Only assess consistency on the three "valid" tosses.



→ Coding Grid & Pictorial Representations

					EXAMINEE ID →														
BATTING	Does the strike occur in a long (full arc) AND in a horizontal plane?	**Y**	Is the transfer of weight in a contralateral pattern?	**Y**	4	4	4	4	4	4	4	4	4	4	4	4	4	4
				N	3	3	3	3	3	3	3	3	3	3	3	3	3	3
		N	Is the motion from back to front in a downward plane?	**N**	2	2	2	2	2	2	2	2	2	2	2	2	2	2
				Y	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Say YES only if both the "full arch" and "horizontal plane" patterns are visible.	Say YES if you see a shift of weight onto the forward (opposite) foot just prior to initiation of the swing.		4
			3
	Say YES if the striking movement is a vertical chopping motion.		2
			1