Technological approaches to fundamental movement skill assessment in children

Brodie James Ward
Bachelor of Science (Honours)

This thesis is presented for the degree of Doctor of Philosophy at
The University of Western Australia

School of Human Sciences (Sport and Exercise Science)
2018
Abstract

Fundamental movement skill (FMS) competency in early childhood is recognised as an important facilitator of lifelong engagement in physical activity. Teachers within the early childhood and primary education system bear substantial responsibility for the development of children’s FMS competency, with their role including ongoing assessment and monitoring of FMS. Commonly, process-oriented tools are used in education settings to assess and monitor FMS, however the reliability of these tools is reduced, primarily due to time and resource constraints, as well as low levels of domain expertise within primary and early childhood teachers. To overcome these limitations, consumer-level motion capture technology could be employed to develop computational assessment strategies that mimic expert human assessment, minimising the time and expertise required from teachers to administer FMS assessments in education settings. The mimicry of expert human assessment by computational means requires a greater understanding of assessment processes, including identifying the kinematic information assessors employ when making FMS proficiency judgments.

This thesis presents a series of manuscripts employing consumer-level motion capture technology to extend the understanding of human FMS assessment strategies and to investigate the kinematic information required for computational assessment approaches. The first manuscript presents a peer reviewed publication of the proposed protocols for the project. Being the first project of its nature, it was important that the proposed protocols underwent critical review and were accepted as a valid means to inform the development of computational FMS assessment strategies.

The second manuscript within the thesis aimed to establish the use of a consumer-level depth sensor for the capture of FMS performances by comparing FMS assessment scoring from point-light (PL) displays and traditional video presentations. PL displays produced using the kinematic data extracted from the depth sensor provide a visual representation of the data that can be computationally analysed, therefore, it was important to establish whether the data was appropriate for FMS assessment. Fifty-three sports science students in the final year of their degree assessed sixteen performances of four FMS in each presentation style using
proficiency criteria derived from the Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000) process-oriented FMS assessment tool. Moderate agreement between proficiency criteria scoring was observed across presentation styles suggesting that assessors were inconsistent in their assessments. Accuracy scores for the scoring of proficiency criteria were also moderate in both styles, and significantly lower in PL than video. Assessors reported high reliability in both presentation styles when performance scores (summed number of present criteria for each performance) were analysed, as is typically the case in studies of assessment reliability. The results of our initial investigation revealed that the identification of proficiency criteria during real-time FMS assessment can be inaccurate regardless of whether skills are viewed in PL or video displays. Results also suggested that traditional analyses of assessment reliability may not capture inaccuracies in identifying proficiency criteria. It was difficult to ascertain whether the inconsistencies between presentation styles were a result of assessor experience or limitations of the depth sensor, so it was important that the subsequent investigation considered assessment scoring in each display across levels of assessor experience and also sought to understand the observed discrepancy between criterion and performance level scoring accuracy.

Manuscript Three provides an in-depth investigation of accuracy across levels of FMS assessment scoring interpretation and also assessor experience. Seven paediatric professionals and ten primary teachers represented experienced and inexperienced assessor groups respectively. Assessors were required to assess ten performances of four FMS in PL and video displays using proficiency criteria derived from the Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000). Moderate agreement of criterion level scoring across presentation styles was observed in both assessor groups. Accuracy at the criterion level was also moderate for both assessor groups in both presentation styles. No significant differences were observed between assessor groups or presentation styles. Both assessor groups reported high reliability at the performance score level in both presentation styles, further establishing the disparity between criterion level accuracy and performance score reliability. The disparity was further explained when the average accuracy in observing each individual criterion was calculated across assessor groups. Average accuracy was below 50% for some criteria and many others fell below the 70% acceptability threshold. Our results suggested that the inaccuracies in identifying proficiency criterion during real-time FMS
assessment may be washed out when performance scores are calculated. This presents an important consideration for assessment administrators given the reported strength of process-oriented assessment tools is their ability to yield accurate information about specific skill deficiencies. Results also suggested that even highly experienced assessors had difficulties in observing proficiency criteria and that moderate agreement between presentation styles is therefore more likely due to challenges in assessment rather than limitations of consumer-level depth sensors.

Manuscript Four aimed to identify and compare the observation strategies of experienced and inexperienced assessors to further understand the patterns of assessment scoring identified in Manuscript Two and to highlight the kinematic information required for computational assessment approaches. Whilst undertaking the process-oriented FMS assessment task outlined in Manuscript Three, participants gaze behaviours were recorded using an eye-tracker. Fixations (gaze behaviours that allow conscious processing of information) were analysed in reference to anatomical locations and the temporal phases of each FMS, and compared to the spatio-temporal information outlined in proficiency criteria. Results established an independence of observation strategies employed by assessors and the spatio-temporal sequencing of proficiency criteria. Gaze behaviours that would suggest conscious search for the presence of proficiency criteria were not observed, evidenced by many cases in which assessors were not attending to the kinematic features outlined in proficiency criteria in the relevant temporal phase. We suggest that assessors do not have the attentional capacity to attend to all criteria during real-time observation of FMS performances and therefore a more efficient search strategy that prioritises key kinematic features is adopted. These results went some way to explain the scoring patterns observed in Manuscripts Two and Three. Assessors are still able to make an overall proficiency judgement by employing an observational strategy that does not require attendance to all proficiency criteria, which results in high reliability at the performance score level. However, in not having the attentional capacity to consciously attend to all criteria during real-time FMS observation, assessors may rely upon inference for the scoring of some criteria, resulting in reduced accuracy at the criterion level. This is likely further exacerbated by increased subjectivity or difficulty in interpreting some criteria when compared to others and further reinforces the considerations assessors must make when interpreting criterion level scoring.
In recognising the inconsistencies in human assessment processes during the earlier investigations, we concurrently explored a data driven approach to the assessment of FMS proficiency. The final investigation presented within this thesis aimed to explore the relationship between inter-trial variability and horizontal jump proficiency as a means of computationally classifying FMS performances. Three horizontal jump performances of thirty-one children were captured with two Microsoft® Kinect cameras. Children were also administered the TGMD-2 (Ulrich, 2000) to determine FMS proficiency levels. Resultant joint velocities for 20 joint centres across each of the three trials were calculated before being temporally aligned and time normalised for each participant. The variability of velocity was then calculated across the three trials for each child to yield an inter-trial variability value. Inter-trial variability was then compared with TGMD-2 standard locomotor scores for each child. Results showed a positive relationship between variability and proficiency in horizontal jump performances. When the relationship between variability and proficiency was considered for proximal, mid, and distal joints, the strength increased across the kinematic chain, likely evidencing the release of degrees of freedom in more proficient performances. Results also showed that distal joint data alone was the strongest predictor of jump proficiency. The investigation revealed that inter-trial variability calculated from consumer-level depth sensor data may be an appropriate means to computationally assess FMS proficiency, but research needs to extend our work to include other FMS.

The culmination of this research highlights significant considerations for those administering and developing real-time FMS assessments. This thesis emphasises the challenges of real-time process-oriented assessment, and presents novel investigations of assessment scoring and observational strategies of FMS assessors across experience levels. This thesis initially aimed to investigate human FMS assessment as a means to inform computational assessment. However, our results provide evidence that human-modelled templates may not appropriately inform the development of computational FMS assessment techniques; primarily due to the limitations of human assessment, even in samples of experienced assessors.
# Table of Contents

Abstract ................................................................................................................................. ii

Table of Contents .................................................................................................................. vi

List of Tables ......................................................................................................................... x

List of Figures ......................................................................................................................... xii

List of Key Abbreviations .................................................................................................... xv

Acknowledgements ............................................................................................................... xvi

Declaration ........................................................................................................................... xx

Statement of Candidature Contribution ............................................................................... xxi

Introduction ......................................................................................................................... 1

Background ............................................................................................................................ 2

Statement of the problem ....................................................................................................... 2

Thesis aims and hypotheses ................................................................................................. 3

Thesis Outline ....................................................................................................................... 3

Thesis Overview ................................................................................................................... 4

Significance of this research ............................................................................................... 6

Delimitations ......................................................................................................................... 7

Limitations ............................................................................................................................ 7

The Microsoft® Kinect® ...................................................................................................... 8

References ............................................................................................................................ 12

Literature Review .................................................................................................................. 13

Introduction ......................................................................................................................... 14

Fundamental movement skill acquisition and development ............................................. 15

Fundamental movement skill assessment ......................................................................... 16

Use of technology for fundamental movement skill assessments .................................. 22

Conclusion ........................................................................................................................... 26

Appendix F – Proficiency criteria for FMS assessment (sourced from the TGMD-2 (Ulrich, 2000)) ..........................................................................................................................................................190
Chapter 2 – The Microsoft® Kinect®

Table 1. Detailed specifications of the Microsoft® Kinect® marker-less 3-D motion capture system 10

Chapter 5 - Investigating the use of a consumer-level depth sensor for FMS capture and assessment

Table 1. Mean absolute agreement percentage between PL and video assessments ............... 67
Table 2. Mean agreement within presentation styles based on assessments of duplicated performances .................................................................................................................. 67
Table 3. Mean proficiency criteria recognition accuracy for PL and video assessments, accuracy differences between presentation styles, and Wilcoxon signed-rank of accuracy differences between presentation styles........................................................................................................ 69
Table 4. Average criterion presence/absence error distribution including Wilcoxon signed-rank tests of differences between presentation styles ........................................................................................................ 71
Table 5. Intra-class correlations and 95% confidence intervals for performance score level comparisons for all, and individual skill types, as well as Spearman correlations between mean performance scores and expert ratings across all assessments (*p<0.05) ........................................................................................................ 72
Table 6. Summed absolute distance from true performance scores of video and PL performance scores and Wilcoxon signed-rank tests of differences between presentation styles. ............... 73

Chapter 6 - Can proficiency criteria be accurately identified during real-time fundamental movement skill assessment?

Table 1. Descriptive statistics of criterion level agreement figures for paediatric professionals and primary teachers across video and point-light FMS assessments ........................................ 89
Table 2. Mean criterion level accuracy figures for paediatric professionals and primary teachers across video and point-light FMS assessments ........................................................................................................ 90
Table 3. Intraclass correlation coefficients (ICC) and 95% confidence intervals (CI) of performance level scoring across assessor groups, and Spearman correlations (r_s) of median performance scores with true performance scores ........................................................................................................ 90
Table 4. Mean accuracy across assessor groups for identification of individual proficiency criteria in video and point-light presentation styles ........................................................................................................ 92
Chapter 7 - Observation strategies employed during process-oriented FMS assessment:

Comparisons between experienced and inexperienced assessors

Table 1. Kinematic markers used to define phases of skill performances for spatio-temporal analyses of visual search and relevant proficiency criteria for each skill phase .................................. 112

Table 2. Average criterion level assessment accuracy for paediatric professionals and primary teachers, as well as high and low accuracy assessor groups ................................................................. 114

Table 3. Average fixation count and average fixation duration per assessment for paediatric professionals and primary teachers across skills and presentation styles. ...................... 115

Table 4 Average fixation count and average fixation duration per assessment for high and low accuracy assessors across skills and presentation styles.......................................................... 120

Chapter 8 - Using inter-trial variability to classify horizontal jump proficiency in children with the Microsoft® Kinect® sensor

Table 1. Proximal, mid and distal joint groups based on proximity to the trunk ................. 141

Table 2. Descriptive statistics of Test of Gross Motor Development-2 (Ulrich, 2000) results ... 142

Table 3. Pearson Correlations between TGMD-2 Standardised Locomotor Scores and inter-trial variability for individual joints............................................................................................................. 143

Table 4. Descriptive statistics of mean variability and mean velocity (m·s⁻¹) for distal, mid, and proximal joint groups.......................................................................................................................... 144

Table 5. Correlations between Standard Locomotor Scores and Distal, Mid, and Proximal joint group mean inter-trial variability............................................................................................................. 144

Table 6. Correlations between Standard Locomotor Scores and Distal, Mid, and Proximal joint group mean velocity (m·s⁻¹) ............................................................................................................. 145

Table 7. Results of backward stepwise regression analyses of Distal, Mid, and Proximal joint group mean inter-trial variability on Standard Locomotor Scores .............................................. 147

Table 8. Results of backward stepwise regression analyses of Distal, Mid, and Proximal joint group mean velocities (m·s⁻¹) on Standard Locomotor Scores.................................................. 147
List of Figures

Chapter 2 – The Microsoft® Kinect®

Figure 1. Simplified overview of the Microsoft® Kinect® pose estimation process (Shotten et. al, 2011) .........................................................................................................................10

Chapter 4 – Protocols for the investigation of information processing in human assessment of fundamental movement skills

Figure 1. Study Overview........................................................................................................................................43
Figure 2. Kinect camera setup for FMS performance collection........................................................................45
Figure 3. Fixation location coding, adapted from Lee M.J.C., Lloyd D.G., Lay B.S., Bourke P.D., & Alderson J.A. (2013). ........................................................................................................50

Chapter 5 - Investigating the use of a consumer-level depth sensor for FMS capture and assessment

Figure 1. Bland-Altman plot of PL and video assessment accuracy across all skill assessments (n=47)............................................................................................................................................70

Chapter 7 - Observation strategies employed during process-oriented FMS assessment: Comparisons between experienced and inexperienced assessors

Figure 1. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria........................................................................................................................................116
Figure 2. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria........................................................................................................................................116
Figure 3. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria........................................................................................................................................116
Figure 4. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria........................................................................................................................................117
Figure 5. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................118

Figure 6. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................118

Figure 7. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................119

Figure 8. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................119

Figure 9. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................122

Figure 10. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase High Accuracy (HA) and Low accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................122

Figure 11. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................123

Figure 12. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.........................................................................................................123

Chapter 8 - Using inter-trial variability to classify horizontal jump proficiency in children with the Microsoft® Kinect® sensor

Figure 1. Scatterplots of the relationship between mean variability and mean velocity for distal (a), mid (b), and proximal (c) joint groups including Pearson’s r and significance values for respective comparisons.................................................146

Figure 2. Evidence of degree of freedom releasing in proficient jumping (A) in comparison to non-proficient jumping (B) captured by the Microsoft® Kinect®.................................................................150
Chapter 9 – Summary and Conclusions

Figure 1. Overview of investigations presented within this thesis including the key findings and contributions of each work…………………………………………………………………………156

Figure 2. Number of new publications containing the keyword 'Kinect' from 2010 to 2018 on Clarivate Analytics Web of Science (as at 17/07/2018)………………………………………………………162
### List of Key Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOTMP-2</td>
<td>Bruininks-Oseretsky Test of Motor Proficiency – 2</td>
</tr>
<tr>
<td>FMS</td>
<td>Fundamental Movement Skills</td>
</tr>
<tr>
<td>DOF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>GSGA</td>
<td>Get Skilled, Get Active</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>PE</td>
<td>Physical Education</td>
</tr>
<tr>
<td>PL</td>
<td>Point-light (Presented as PLD (point-light display) in Chapter 4)</td>
</tr>
<tr>
<td>RGB</td>
<td>Red, Green, Blue</td>
</tr>
<tr>
<td>SDK</td>
<td>Software Development Kit</td>
</tr>
<tr>
<td>TGMD-2</td>
<td>Test of Gross Motor Development – 2</td>
</tr>
</tbody>
</table>
This thesis not only represents the culmination of four years of hard work, but a substantial portion of my life, that would not have been possible without the significant support and contribution of so many people. To list them all would result in a document as long as this one, but here are a few special mentions:

To my supervisors:

It was late 2013, Rosey and myself had just got out of the surf and made the natural pilgrimage to the nearest café. Over coffee we discussed possible PhD opportunities. I will never forget Rosey telling me that he “had done all that he wanted in his research career, and just wanted to focus on helping people”. I knew straight away that it would be remiss of me to give up the opportunity to work under someone who would become both an incredible academic mentor, and friend. Rosey, I will forever appreciate the work you put into this project, the sleepless nights you had on my behalf, and your constant strive to better me as an academic under your guidance. Your ability to think of alternate solutions to current problems and tenacity to pursue the path others may not consider is something I will always admire.

Ashleigh, I find it hard to think that I will ever encounter an someone that inspires me in the way that you do. Your unwaveringly positive approach to your work and life, the way you treat those around you, and your selflessness is something that I will always look up to. You have such a profound and positive influence on each and every person lucky enough to work or spend time with you, myself included, and for that I will always be grateful.

Brendo, thank you for your devotion to the project, managing to juggle everything within your life and still find time for my work. I will always appreciate the opportunities you provided to sit down and muse about motor control theory, which constantly stimulated my interest in the field. Your knowledge and ideas throughout the project were invaluable.

To my family and friends:
Mum and Dad, even though you may still not fully understand what it is I do, your constant support over the last 26 years has allowed me to achieve so many things including the completion of this thesis. The work ethic you both demonstrate and instilled within me is
something I am constantly grateful for and strive to replicate. Just don’t ask me what I’m going to do next.

Britt, words cannot describe my appreciation for everything you have done for me over the past 5 years. A PhD is in many ways a selfish pursuit, and you have always remained supportive and understanding, even though sometimes you felt you came second. You continue to inspire me each and every day, and your persistence and determination is something I will always look up to. You always picked me up when I was down and made me keep going, even when you were struggling as well. I am so lucky to have had you join me on this journey and look forward to many more together.

To all of my mates outside of uni, those who finished and got real jobs, and real incomes (very smart), life has gone in different directions for us all. However, I can always count on you to be there for beers and yarns, even if I’m not offering much.

To those in Sport Science:
Marcus, your support as an office mate and life mate has significantly influenced the completion of this thesis. Through some low lows and dizzying highs, you never let your eyes waiver from the end goal, and showed me what it truly meant to work hard (and play harder). Steve, you were always there to bounce ideas off, help me with coding, and just be a great mate. The opportunity you presented me during honours spurred my interest in motor control and I wish you all the best with the completion of your PhD. To the admin staff, thank you for dealing with all those mundane tasks and naïve questions I made you deal with. Thank you to all the others that provide the little things; a listening ear, a little bit of a break to get the mind off things, help with stats, a teaching opportunity, or just words of encouragement.

Finally, to everyone else that contributed to the research within this thesis:
Du, Winston and Dave in computer science that provided much needed assistance with coding matters, Nigel in psychology that assisted at the drop of a hat with eye-tracking, and all the participants and families that willingly provided your time and effort to make this research possible.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship and I am grateful for the financial support during my candidature.
Publications Arising from this Thesis


Peer-Reviewed Conference Proceedings


I, **Brodie Ward**, certify that:

This thesis has been substantially accomplished during enrolment in the degree.

This thesis does not contain material which has been submitted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text and, where relevant, in the Declaration that follows.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

The research involving human data reported in this thesis was assessed and approved by The University of Western Australia Human Research Ethics Committee. Approval #: RA/4/1/7445.

This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Signature

Date: 04/
Statement of Candidature Contribution

This thesis contains published work and/or work prepared for publication, some of which has been co-authored. The bibliographical details of each work and the location in which they appear within the thesis are outlined below. A statement for each publication or manuscript that clarifies the contribution of the student to the work is also provided.

Chapter 4

*Planning*

The candidate contributed to the conception of the ideas, development of experimental designs, and sourcing of relevant literature to the proposed study.

*Manuscript*

The candidate prepared the first version of the manuscript. In addition, the candidate was responsible for revising manuscripts following review by co-authors and led the preparation of journal review rebuttals and revisions.

Chapter 5

*Planning*

The candidate contributed to the conception of the ideas, decisions on assessment tools and capture technologies utilised, as well as the development of the experimental design.

*Data Collection*

The candidate was the primary individual responsible for the collection, analysis, and interpretation of data.
**Manuscript**

The candidate prepared the first version of the manuscript. In addition, the candidate was responsible for revising manuscripts following review by co-authors and led the submission of the manuscript for journal review.

**Chapter 6**


**Planning**

The candidate contributed to the conception of the ideas, decisions on assessment tools, decisions on appropriate participants, as well as the development of the experimental design.

**Data Collection**

The candidate was the primary individual responsible for the collection, analysis, and interpretation of data.

**Manuscript**

The candidate prepared the first version of the manuscript. In addition, the candidate was responsible for revising manuscripts following review by co-authors and led the submission of the manuscript for journal review.

**Chapter 7**


**Planning**

The candidate contributed to the conception of the ideas, decisions on assessment tools and eye-tracking apparatus, decisions on appropriate participants, as well as the development of the experimental design.

**Data Collection**
The candidate was the primary individual responsible for the collection, analysis, and interpretation of data.

**Manuscript**

The candidate prepared the first version of the manuscript. In addition, the candidate was responsible for revising manuscripts following review by co-authors.

**Chapter 8**


**Planning**

The candidate contributed to the conception of the ideas, decisions on assessment tools, as well as the development of the experimental design.

**Data Collection**

The candidate was the primary individual responsible for the collection, analysis, and interpretation of data.

**Manuscript**

The candidate prepared the first version of the manuscript. In addition, the candidate was responsible for revising manuscripts following review by co-authors.

I, **Michael Rosenberg** certify that the student statements regarding their contribution to each of the works listed above are correct.

As all co-authors’ signatures could not be obtained, I hereby authorise inclusion of the co-authored work in the thesis.

Coordinating supervisor signature: Date: 04/12/2018
CHAPTER 1
Introduction
Background

It is important to ensure that every child is equipped with the necessary movement skills to engage in an active and healthy lifestyle throughout adolescence and adulthood. Central to childhood participation in physical activity is the acquisition of fundamental movement skills (FMS). FMS are typically considered the ‘building blocks’, or foundation movements for more specific skills that allow individuals to engage in active play, and more complex sporting and physical activities (Gallahue, Ozmun, & Goodway, 2012). Assessment and monitoring of FMS throughout childhood is important to ensure that low movement skill competency, and its negative correlates, are identified and prevented.

FMS assessment tools and techniques have been investigated at length, including comprehensive reviews by Cools and colleagues in 2009 and Bardid and colleagues in 2018; many studies within these reviews identified issues in translating assessment practices, and implementation outside of research environments. The greatest barrier to effective and reliable FMS assessment was the level of training required to administer assessments and the time required to ensure quality. This is especially common in primary education settings, where teachers’ bear the responsibility of assessing FMS. Administering effective and reliable assessment in a primary school setting is challenging, with limited time available, and limited training and experience of teachers. A lack of viable alternatives to current assessment techniques is a major barrier to effective assessment in primary schools and often results in teachers using procedures that suit their own needs.

Some of the barriers to administering reliable process-oriented FMS assessment in education settings, may be aided by recent advances in consumer-level motion capture technology. Consumer-level marker-less motion capture systems like the Microsoft® Kinect® depth sensor may provide 3-dimensional kinematic capture for computational FMS proficiency analysis. However, their suitability in this application is not yet understood.

Statement of the problem

Reliable and regular FMS assessment is challenging and costly in education settings. Advances in motion capture technology may provide the foundation for more efficient and objective
computational assessment strategies reducing the burden on primary teachers. The suitability of such technologies for the capture and assessment of FMS proficiency is yet to be established. The development of computational FMS assessment strategies also requires a far greater understanding of the observational, and information processing strategies employed by experienced assessors. Whilst a significant body of research has considered the constructs and reliability of current assessment tools, to date there has been no investigation of the perceptual strategies employed by FMS assessors during test administration.

**Thesis aims and hypotheses**

The overall aim of this thesis is to explore the suitability of consumer-level depth cameras for the capture and assessment of FMS performances and to develop a greater understanding of the observational strategies employed by human assessors during FMS assessment to inform computational FMS assessment techniques.

**Thesis Outline**

This thesis is presented as a series of manuscripts that establish new understanding of current FMS assessment techniques and explores the use of technologies to improve FMS assessment in education settings. Following an overview of the project (Chapter 1: Introduction), the Literature Review (Chapter 3) presents existing research relevant to FMS assessment, and highlights the gaps within the literature that the current project seeks to address. The literature review is followed by a peer-reviewed article (Ward, Thornton, Lay, & Rosenberg, 2017) outlining the proposed protocols for two of the original investigations presented within the thesis. Subsequently, four original investigations (Chapters 5-8) are presented. Firstly, Chapter Five aims to investigate the suitability of the Microsoft® Kinect® sensor for the capture of FMS by comparing assessment scoring from point light (PL) displays produced by the Kinect®, and typical video presentations of skill performances. Chapter Six investigates process-oriented FMS assessment scoring from primary teachers and experienced FMS assessors from video and PL displays. Chapter Seven explores the observation strategies employed by primary teachers and experienced FMS assessors through the recording of gaze behaviours during assessment. Subsequently, Chapter Eight investigates the objective classification of FMS proficiency using the Microsoft® Kinect sensor, via assessment of inter-
trial variability between horizontal jump performances. Finally, the Summary and Conclusions (Chapter 9) draws conclusions from the findings of this thesis, presents the projects developments in the understanding of FMS assessment, and provides directions for future research.

Outlined below are the specific aims and hypotheses of the four original investigations presented within this thesis:

**Thesis Overview**

1. Investigate the use of consumer-level depth capture systems for process-oriented FMS assessment
2. Understand scoring patterns of process-oriented FMS assessments administered by experienced and inexperienced assessors
3. Explore the kinematic information employed by experienced and inexperienced assessors during process-oriented FMS assessment
4. Investigate a preliminary computational approach to FMS assessment using measures of inter-trial variability from digitised kinematic data

**Development towards computational approaches to FMS assessment**
Chapter 5 Investigating the use of the Kinect® system for movement skill capture and assessment

Aim: This investigation aimed to examine the suitability of the Microsoft® Kinect® system for the capture of FMS by comparing assessment scoring from point-light displays produced by the Kinect®, and typical video presentations of skill performances.

Hypothesis: Kinect® depth capture would provide sufficient skill information, and therefore human ratings of the presence or absence of particular skill features assessed using both the RGB and depth images would be highly comparable.

Chapter 6 Can proficiency criteria be accurately identified during real-time fundamental movement skill assessment?

Aim: This investigation aimed to examine and compare patterns of process-oriented FMS assessment scoring between paediatric professionals and primary teachers under video and point-light display conditions. The investigation also sought to examine differences in scoring accuracy when assessments are considered at the criterion level in contrast to performance scores.

Hypothesis #1: No differences in assessment accuracy would be observed between video and point-light presentation styles when assessments were made by the paediatric professional group

Hypothesis #2: Paediatric professionals would be significantly more accurate in FMS assessments than primary teachers

Hypothesis #3: Lower accuracy would be observed for criteria that could be considered more subjective or that may be interpreted differently by different assessors

Chapter 7 Observation strategies employed during process-oriented FMS assessment: Comparisons of experienced and inexperienced assessors
**Aim:** This investigation sought to examine and compare the kinematic information employed by paediatric professionals and primary teachers during process-oriented FMS assessments under video and point-light conditions

**Hypothesis #1:** Paediatric professionals would employ more efficient visual search strategies when compared to primary teachers

**Hypothesis #2:** Paediatric professionals would rely on proximal sources of kinematic information more than primary teachers

**Hypothesis #3:** Both assessor groups would attend to regions outside of those outlined in proficiency criteria within relevant temporal phases

**Chapter 8 Using inter-trial variability to classify horizontal jump proficiency in children with the Microsoft® Kinect® sensor**

**Aim:** This investigation aimed to examine the objective classification of proficiency in horizontal jump performances captured by the Microsoft® Kinect® camera, through computational assessment of inter-trial variability.

**Hypothesis:** A positive relationship would be observed between proficiency and inter-trial variability in jump performances, and that more proficient children would exhibit greater average joint velocities than less proficient children.

**Significance of this research**

FMS assessment in education settings continues to be hindered by the nature of current field-based assessment techniques and the expertise of assessment administrators. This thesis seeks to establish a foundation for the use of technology to advance the efficiency and reliability of FMS assessment in education settings. In doing so, the thesis also advances the understanding of current FMS assessment techniques and explores assessment strategies at a perceptual level. The findings arising from this work will inform both, effective training strategies for test administrators, and the development of computerised assessment techniques.
Delimitations

Investigations within this thesis were designed in alignment with the manner in which FMS are assessed within education environments, importantly, that time constraints dictate that skill performances are viewed once in real-time. To maintain ecological validity the Microsoft® Kinect® was employed within the thesis as an off-the-shelf capture system that could be easily disseminated in the case of the development of a computational assessment technique. Thus, kinematic data captured by the Kinect® underwent minimal post-processing.

Limitations

The studies within the thesis concerned with human assessment of FMS employed simulated real-time assessment, whereby assessors viewed each skill performance once, in real-time, as would be the case in a field-based environment. To allow the comparison of assessment patterns across large samples of assessors, it was imperative that consistent kinematic information was available across all assessors. In a field-based environment it is problematic to have large samples of assessors observing a single performance due to differences in observation angles and therefore available kinematic information, nor can it be assumed that repeated performances of the same skill are consistent in the kinematics the child employs. It is the authors belief that the consistent attentional constraints imposed upon assessors in simulated real-time and field-based real-time assessment allows generalisation to field-based assessment. The four FMS assessed within this thesis were selected as they represented both locomotor and object-control subcategories of FMS, which could be performed within the capture space of the Microsoft® Kinect®. Whilst consistent patterns of assessment were observed across the four FMS investigated, generalisation of the assessment patterns observed within the project to assessment of other FMS, must be made with care.
Chapter 2
The Microsoft® Kinect®
The Microsoft® Kinect® consumer-level depth capture system is employed throughout the current project to capture FMS performances for computational analyses and to produce point-light (PL) displays. It is important to provide an overview of the Kinect® system and its potential use in FMS capture, including a detailed description of the motion capture system and several validation studies of Kinect® pose estimation. At the commencement of this project the Microsoft® Kinect® was the most advanced consumer-level motion capture system to produce 3-dimensional kinematic feature extraction. Kinematic feature extraction was previously only achievable using expensive lab-based motion capture systems, or via laborious coding of video. Although recently surpassed by sophisticated pose estimation algorithms that extract features from video, the Kinect® represented significant leap forward in cost-effective and user-friendly means to access 3-dimensional digital kinematic data in field-based environments.

The Kinect® is a single camera depth-based marker-less motion capture system designed for active video gaming in conjunction with the Xbox 360 gaming system (Jungong et al., 2013). Since its release, the potential for the system to be extended beyond gaming as a low-cost alternative to traditional 3-D motion capture system has been exploited (Hondori & Khademi, 2014; Jungong et al., 2013).

The Kinect® system contains an infrared (IR) projector, IR camera, and RGB camera, which simultaneously capture depth signals and RGB images (Jungong et al., 2013; Mobini, Behzadipour, & Saadat Foumani, 2014). The depth sensing component of the Kinect®, consisting of the IR projector and IR camera, operates by projecting an IR point cloud into the 3-dimensional environment and measuring deformations to the point cloud to infer the depth of objects within the projection space. The depth sensing capabilities of the Kinect® also include an underlying body tracking algorithm trained on around 500,000 depth images (Shotton et al., 2011). The body tracking algorithm implements tomographic reconstruction, enabling the Kinect® to estimate the 3-dimensional location of 20 skeletal joint centres based on position within the IR point cloud (Hondori & Khademi, 2014; Shotton et al., 2011). A simplified overview of this process is presented in Figure 1. The RGB camera captures traditional 2-dimensional colour video. Detailed specifications of the Kinect® depth and RGB sensors are outlined in Table 1 (Hondori & Khademi, 2014).
Released in conjunction with the Kinect® was a software development kit (SDK), allowing access to the depth and skeletal tracking data captured by the depth sensors. The access to such data in a low-cost, portable system released Kinect®’s potential as a real-time motion capture system in physical therapy and rehabilitation contexts (Hondori & Khademi, 2014).

Table 1. Detailed specifications of the Microsoft® Kinect® marker-less 3-D motion capture system

<table>
<thead>
<tr>
<th></th>
<th>IR Depth Camera</th>
<th>RGB Camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling frequency</td>
<td>30 frames per second</td>
<td>30 frames per second</td>
</tr>
<tr>
<td>Output video resolution</td>
<td>640x480 pixels</td>
<td>640x480 pixels</td>
</tr>
<tr>
<td>Capture Range</td>
<td>0.8 m – 4 m</td>
<td>(Optimal range: 1.2-3.5 m)</td>
</tr>
<tr>
<td>Angular field of view</td>
<td>57° horizontally</td>
<td>43° vertically</td>
</tr>
</tbody>
</table>

Validation of the Microsoft® Kinect® in capturing gross movements

Multi-camera IR marker-based motion capture systems such as Vicon (Oxford Metrics, Oxford, United Kingdom) currently represent the gold standard of kinematic analysis (Galna et al., 2014). Lack of portability, high cost, and laborious data collection protocols confines marker-based systems to laboratory environments and renders them unviable for field-based data collection with large samples (Burton & Miller, 1998; Hondori & Khademi, 2014). Due to its single camera nature, lower sampling rate, and marker-less kinematic capture, the Kinect® is less accurate in capturing kinematics than marker-based systems. However, a substantial
body of research has concluded that the system may be sufficiently accurate for kinematic capture of gross movements for general populations (Hondori & Khademi, 2014).

Galna et al. (2014) and Clark et al. (2012) both observed good agreement between Kinect® and Vicon for measuring trunk and lower limb kinematics during standing balance tests. Furthermore, Galna et al. (2014) reported excellent correlations (ICC > 9) between Kinect® and Vicon for temporal characteristics of all movements tested. Spatial features captured with the Kinect® such as range of motion were less accurate, however relative agreement to Vicon was still strong. The spatial accuracy of the Kinect® tends to be the lowest for fine aspects of movement, such as finger displacement during hand clasping and pronation and displacement during pronation and supination. Accuracy in capturing features within gross movements tends to be high (Galna et al., 2014).

The accuracy of Kinect® depth measurement is reported to be within 1-4cm within a 4m range, and when the resolution of the sensor is considered the estimation of joint displacements is considered reasonable (Mobini et al., 2014; Obdrzálek et al., 2012). Obdrzálek et al. (2012) reported greater inaccuracies in their comparison between Kinect® and the Impulse motion capture system (PhaseSpace Inc, San Leandro, CA), however, they noted that some error may have been introduced by joint position estimates of the reference motion capture system. Mobini et al. (2014) found smaller errors in joint position estimation when compared to a fabricated model, which provides an exact reference in contrast to estimated references from other motion capture systems.

The collective research suggests that the Kinect® is an acceptable and affordable depth sensor for capture of gross movement such as FMS (Hondori & Khademi, 2014). However, the system is not without limitations. Due to its single camera nature the Kinect® can suffer from occlusion and can also be affected by objects within the scene such as chair legs in the capture of seated poses (Hondori & Khademi, 2014; Obdrzálek et al., 2012). Joint position estimation can also be affected by unusual poses that differ highly from those within the algorithms training set, however the movements performed in FMS assessment align closely with those in active video gaming so this is unlikely to be an issue (Xu & McGorry, 2015).
References


**Introduction**

Fundamental movement skills (FMS) are typically considered the basic components of movement, and form the foundation for the development of sport-specific skills (Hardy, Barnett, Espinel, & Okely, 2013; Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012). FMS can be classified as locomotor (e.g. running, jumping, hopping, galloping), object control (e.g. throwing, catching, striking), or stability (e.g. balancing) skills (Gallahue, Ozmun, & Goodway, 2012; Logan, Robinson, Wilson, & Lucas, 2012; Lubans, Morgan, Cliff, Barnett, & Okely, 2010). The importance of developing FMS throughout childhood and early adolescence is their transferability to a wide range of sports and recreational activities that promote the establishment of a physically active lifestyle (Barnett et al., 2016). For example, the overarm throw establishes foundation movement patterns that can be developed into more specialised skills like pitching, serving, or bowling, allowing children to engage in sports, or active social pursuits such as baseball, tennis and cricket respectively. The benefits of FMS competence are not limited to childhood; FMS competency enables engagement in active pursuits and the foundation for a physically active lifestyle throughout adolescence and adulthood. Engagement in regular physical activity facilitated by the development of FMS proficiency throughout the early years can result in many health-related benefits including improved cardio-respiratory fitness, improved weight status, and increased perceived physical competence (Bardid, Vannozzi, Logan, Hardy, & Barnett, 2018; Hardy et al., 2012; Logan, Robinson, Wilson, et al., 2012; Lubans et al., 2010; Stodden et al., 2008). Without identification and intervention, the negative correlates of low FMS competence such as physical inactivity and obesity can persist throughout adulthood, making the early years a key period for FMS development (Holfelder & Schott, 2014; Logan, Robinson, Wilson, et al., 2012; Veldman et al., 2018).

Intervention for children identified as having low FMS competency can be beneficial when appropriate interventions are provided (Logan, Robinson, Wilson, et al., 2012; Piek, Hands, & Licari, 2012). Successful interventions have been shown to rely upon an understanding of motor development and FMS acquisition, and the ability to accurately measure FMS development and competence (Holfelder & Schott, 2014; Logan, Robinson, Wilson, et al., 2012).
**Fundamental movement skill acquisition and development**

The acquisition and development of FMS does not occur through natural maturation and relies upon exposure, instruction, feedback and intervention during early childhood (Gallahue et al., 2012; Lubans et al., 2010). FMS acquisition and development is longitudinal in nature as performers acquire basic skill processes before progressing through stages of skill learning towards proficiency; and in the case of FMS, transferability to sport-specific skills (Clark, 2005; Davids, Bennett, & Newell, 2006; Fitts & Posner, 1967; Gallahue et al., 2012; Utley, 2008).

Although considered basic movements, FMS can be complex for those in the early stages of learning. All FMS require performers to co-ordinate individual components of their motor system in a way that produces efficient and adaptable movement patterns (Button, Macleod, Sanders, & Coleman, 2003). The individual components of the motor system, such as limb segments, are termed ‘degrees of freedom’ (DOF), which must be controlled during skill performance to achieve an optimum outcome (Bernstein, 1967; Button et al., 2003). In the early stages of FMS acquisition, it is typical for performers to ‘freeze’ DOF in order to reduce the complexity, or variability of an unfamiliar task. This results in performances appearing rigid as the performer tightens joint couplings and restricts the range of motion of segments in an effort to reduce the complexity of the movement. In later stages of FMS acquisition performers ‘release’ DOF, loosening joint couplings and moving through larger ranges of motion, resulting in adaptable and efficient skill performances (Bernstein, 1967; Davids et al., 2006; Gallahue et al., 2012; Seifert, Button, & Davids, 2013; Vereijken, Emmerik, Whiting, & Newell, 1992). Children therefore need to be taught FMS and given appropriate opportunities to learn and practice skills to ensure adaptable and proficient performances result (Gallahue et al., 2012; Hardy et al., 2012; Logan, Barnett, Goodway, & Stodden, 2016; Lubans et al., 2010).

There are multiple environments where children can be taught FMS, however, school physical education programs have become central to structured FMS development due to the learning environment provided by schools and domain expertise of teachers (Morgan & Hansen, 2008; Whipp, Hutton, Grove, & Jackson, 2011). A shortage of specialist physical education (PE)
teachers has resulted in generalist teachers being increasingly relied upon for FMS development, particularly in early primary years, reducing the number of domain experts delivering PE (Morgan & Hansen, 2008; Whipp et al., 2011). FMS acquisition and development is a complex process, requiring a substantial level of domain knowledge from those responsible for delivering PE to early childhood students (Lander, Morgan, Salmon, & Barnett, 2016). The provision of structured environments for students to develop FMS, and ability of teachers to assess FMS proficiency is paramount to identifying skill deficiencies and ensuring effective FMS development (Lander, Eather, Morgan, Salmon, & Barnett, 2017; Longmuir et al., 2017). PE specialist teachers are provided with the tools to be effective in movement observation and FMS proficiency assessment throughout their pre-service education, which is often not the case for generalist teachers (Haynes & Miller, 2014; Logan, Robinson, Wilson, et al., 2012; Morgan & Hansen, 2008). Meaning some children do not receive the benefit of specialised FMS instruction throughout the early years, when it would be most beneficial.

The complexities of FMS assessment mirror those of FMS development and in many cases require teachers to interpret kinematic features of FMS in comparison to proficient examples of the respective skills (Lander et al., 2016). There is an increasing amount of evidence that suggests that early childhood teachers are not equipped with the required domain knowledge to accurately and reliably assess FMS. A lack of domain expertise, time, and resources, in addition to large class sizes often dictate that the ideal FMS assessment tools aren’t viable in the school environment, or can’t be administered in the way they were designed (Barnett, Minto, Lander, & Hardy, 2014; Cools, De Martelaer, Samaey, & Andries, 2009; Haynes & Miller, 2014; Lander et al., 2017; Lander et al., 2016; Lander, Barnett, Brown, & Telford, 2015). It is suggested that more appropriate FMS assessment tools and techniques for educational settings are required to ensure accurate and effective skill monitoring and intervention can take place (Furtado & Gallagher, 2012; Giblin, Collins, & Button, 2014; Longmuir et al., 2017).

**Fundamental movement skill assessment**

Many valid FMS assessment tools exist; however, no single tool has been accepted as the gold standard (Bardid et al., 2018; Barnett & Peters, 2004; Cools et al., 2009; Logan, Robinson, & Getchell, 2011; Piek et al., 2012). All assessment tools are designed with the overarching goal
of evaluating movement proficiency, although their constructs differ depending on the target assessment group or implementation (Bardid et al., 2018; Cools et al., 2009). FMS assessment tools should provide an efficient and reliable means to capture the complex process of FMS acquisition and development into a simple interpretable outcome measure.

Motor skill assessments more broadly have been administered in some manner since the early 20th century, initially through developmental scales, whereby motor achievements were recorded according to chronological age. Further refinements to include qualitative aspects of movement patterns were established within assessments during the 1920’s, through observation checklists outlining components of proficient skill performances (Branta, Haubenstricker, & Seefeldt, 1984). These early assessment techniques established the foundations for contemporary FMS assessment tools which can be broadly categorised as either product-oriented (also: quantitative or norm-referenced), or process-oriented (also: qualitative or criterion-referenced), with some tools employing a combination of both (Bardid et al., 2018; Barnett et al., 2009; Cools et al., 2009).

**Product-oriented FMS assessments**

Product-oriented FMS assessments classify proficiency based on skill outcomes such as distance, speed, height, and accuracy (Barnett et al., 2014; Burton & Miller, 1998; Hands, 2002; Logan et al., 2016). Few assessments of FMS rely on product-oriented assessment in isolation, as the technique doesn’t capture any information about how skills are performed, or specific performance deficiencies (Burton & Miller, 1998; Hands, 2002). Product-oriented assessments do little to inform interventions beyond identifying children who may need intervention, however can provide valuable information for proficiency screening (Cools et al., 2009; Hands, 2002). A benefit of product-oriented assessment can be ease of administration and scoring which requires little training, or understanding of motor development. The quantitative nature of product-oriented assessment ensures high test-retest and inter-rater reliability which allows normative referencing of proficiency scores across populations (Cools et al., 2009).
Most product-oriented assessment tools such as the Movement Assessment Battery for Children-2 (MABC-2) (Henderson, Sugden, & Barnett, 2007) and the McCarron Assessment of Neurological Development (MAND) (McCarron, 1982) have been developed to include additional qualitative components which provide information about skill performances. Although the two assessments have strengths including high reliability, consistency of administration and score interpretation, and norm-referencing, the protocols can be time consuming and are designed to screen for motor impairment rather than assessing proficiency of gross movements such as FMS, rendering them inappropriate for PE environments (Cools et al., 2009; Logan, Robinson, Rudisill, Wadsworth, & Morera, 2012).

**Process-oriented FMS assessments**

Process-oriented assessments tend to be preferred for the assessment of FMS in education settings (Barnett et al., 2014; Hands, 2002). Developed from early qualitative assessment approaches, process-oriented assessments measure how skills are performed rather than skill outcomes (Barnett et al., 2014; Hands, 2002; Logan et al., 2016). Assessments that are qualitative in nature are defined as ‘the systematic observation and introspective judgement of the quality of human movement for the purpose of providing the most appropriate intervention to improve performance’ (Knudson, 2000). The strength of process-oriented assessment tools lies within their tendency to yield valuable information about the way skills were performed and identification of specific deficiencies to inform intervention strategies.

In Australia, the most popular and widely administered process-oriented FMS assessment tools designed for educational environments are the Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000), Get Skilled Get Active (GSGA) (NSW Department of Education and Training, 2000), and observation records presented in the fundamental movement skill resources produced for the Western Australian (Western Australian Minister for Education, 2004) and Victorian (Department of Education Victoria, 1996) primary education systems. All four tools employ checklists outlining criteria of proficient skill performance to provide an observational and scoring structure for assessors (Hands, 2002). Process-oriented assessment structures used in PE contrast performances with proficient examples of the skill, rather than assessing stages of skill development. This approach to FMS assessment is in part due to the
complexity of skill learning and the challenges involved in quantifying developmental stages (Hands, 2002; Stodden et al., 2008). Administrators assess FMS performances by identifying the presence or absence of proficiency criteria outlining key kinematic patterns that would be performed in the final stage of skill development (Hands, 2002). Assessing FMS in this manner allows administrators to identify specific skill deficits, which provide valuable information to plan and monitor intervention strategies (Barnett et al., 2009; Barnett et al., 2014; Hands, 2002).

High assessment reliability is important to ensure accurate proficiency information is captured and assessment scoring can be compared across children. However, limited time and expertise are considered barriers to reliable process-oriented assessment in education environments (Barnett et al., 2009; Barnett et al., 2014; Haynes & Miller, 2014; Lander et al., 2015). Teachers tend to rely on real-time assessments in the field, whereby each performance is viewed once only due to time constraints, large class sizes, and the ability to offer immediate feedback. Both the GSGA (NSW Department of Education and Training) and TGMD-2 (Ulrich, 2000) have both demonstrated fair and excellent reliability respectively when implemented in field-based settings (Barnett et al., 2009; Barnett et al., 2014; Valentini, 2012; van Beurden, Zask, Barnett, & Dietrich, 2002). Barnett et al. (2014) reported that reliability figures varied for individual skills and assessment components for each skill in the TGMD-2 (Ulrich, 2000), proposing that some skills and skill components are more difficult to assess in the field than others. An important consideration when interpreting assessment reliability figures reported in research studies is the comprehensiveness of assessor training, and the comparably little training provided to teachers.

Field-based, process-oriented assessment requires assessors to undergo rigorous training in administration and skill observation, as well as understanding motor development to ensure reliable data is captured (Barnett et al., 2009; Barnett et al., 2014; van Beurden et al., 2002). Most studies reporting field-based assessment reliability employ expert assessors and substantial training protocols prior to testing (Barnett et al., 2009; Barnett et al., 2014; Valentini, 2012). In contrast, the level of training undertaken by teachers in both pre-service programs, and in-service professional development (PD) is likely to be insufficient for reliable real-time skill assessment in PE environments (Hands, 2002; Lander et al., 2017; Lander et al.,
Lander and colleagues (2014) found that of 168 physical education (PE) teachers less than half had received more than several FMS lectures during their pre-service education and almost 70% had undertaken no more than 4 hours of FMS PD since completing their degree. These statistics refer to PE specialist teachers, less training is likely to be undertaken throughout pre-service courses for general primary and early childhood specialist teachers.

With sufficient training, assessors can reliably administer process-oriented assessment tools, such as the TGMD-2 (Ulrich, 2000) and GSGA (NSW Department of Education and Training), in PE environments where performances are viewed once, in real-time (Barnett et al., 2009; Barnett et al., 2014). The TGMD-2 (Ulrich, 2000) and GSGA (NSW Department of Education and Training, 2000) take around 20 minutes per child to administer, even when assessment are made in real-time (Cools et al., 2009). Children should also be tested individually, which can make assessments impractical for implementation in a standard PE class (Lander et al., 2016). It is often the case that assessments are not implemented in school settings in the way they are validated and tested in research settings and appropriate tools for teachers assessing PE settings should be developed (Lander et al., 2017; Lander et al., 2015).

Techniques that reduce the observational difficulties of process-oriented assessment include the three-stage approach (Gallahue & Donnelly, 2007) and the observation plan approach (Haywood & Getchell, 2014). The three-stage approach (Fundamental Movement Pattern Assessment Instrument) (McClanaghan & Gallahue, 1978) simplified classification by initially limiting options to three-stages; initial, elementary and mature. The test provided a general classification of FMS development and was later accompanied by a follow-up assessment (Gallahue & Donnelly, 2007) employing criteria to identify specific deficiencies. However, the number of criteria assessors were required to observe (4-8) was challenging and impractical. The observation plan approach, first presented in Haywood and Getchell (2009), employs decision trees or flow charts to simplify classification of developmental stages where the observer makes ‘yes/no’ decisions across 3 branches of the tree that relate to components outlined in developmental sequences. Subsequently, Furtado and Gallagher (2012) combined the two approaches to produce the FG-COMPASS; an assessment that employs decision trees that weight skill components based on the strength of their ability to discriminate skill
proficiency. The process of assessment in the FG-COMPASS (Furtado & Gallagher, 2012) requires initial identification of the presence of a skill feature considered a strong discriminator of high and low proficiency, followed by the identification of the presence of one of two confirmatory skill features, which further discriminate the performance into initial, elementary or advanced proficiency classifications. The assessment effectively provides a weighting system to proficiency criteria. Whilst the reduction of criteria did provide greater ease of administration, the developers concluded that the strength of the assessment was limited by subjectivity of criteria, and the challenge of choosing appropriate criteria to accurately reflect skill development and discriminate performers in to proficiency groups (Furtado & Gallagher, 2012).

Assessment checklists such as the FG-COMPASS (Furtado & Gallagher, 2012) have been developed to overcome issues of expertise in process-oriented assessment, however, the subjectivity of observations continue to affect assessment reliability and make it difficult to compare scoring from multiple assessors (Barnett et al., 2009; Barnett et al., 2014). Another criticism of the process-oriented assessment structures is the time cost of requiring each child to perform multiple trials of individual FMS in isolation making it difficult to complete during a typical PE lesson (Longmuir et al., 2017). The ecological validity of children performing skills in isolation has also been questioned from a physical literacy perspective, due to the disparity between assessment environments and sporting or active play environments (Giblin et al., 2014; Longmuir et al., 2017).

One approach to improve FMS ecological validity and time efficiency is to develop assessments that can be administered within a typical PE class and mimic the way children may be required to perform skills in sport or active play. Two examples of this are the Athletic Skills Track (AST) outlined in (Hoeboer et al., 2016) and the Canadian Agility and Movement Skills Assessment (CAMSA) (Longmuir et al., 2017). Both assessments employ ‘obstacle course’ structures whereby children are timed whilst they move through a course performing multiple movement skills in succession. The primary measure of both assessments is the time a child takes to complete the course, and the CAMSA (Longmuir et al., 2017) additionally scores up to two process-oriented components for each skill. Both assessments provide ecological validity and time efficiency which makes them a likely choice for teachers.
However, the minimal process-oriented components included make the assessments more appropriate as proficiency screens rather than tools that identify specific deficiencies.

Considerations for test developers include whether skills should be assessed in isolation, whether specific skill deficiencies should be identified, and whether it is sufficient to only classify proficiency. Ecologically valid assessments, such as Skills Tracks may be more appropriate for PE environments and effective as proficiency screening tools, but lack the insight into skill deficiencies provided by tools where skills are assessed in isolation. In contrast, assessing skills in isolation provides valuable process-oriented information, but is time consuming and can be less reliable than product-oriented measures. The identification of individual skill components that can be used to inform intervention continues to be highly valued amongst assessors. Thus, assessors must consider the trade-off between time efficiency and expertise when required to assess proficiency criteria. A recent consideration is to employ advancements in technology to minimise the time burden and improve the reliability of process-oriented assessment techniques in education environments.

**Use of technology for fundamental movement skill assessments**

Employing technologies to assist FMS assessment is not new. Video capture has been utilised for decades and offers many advantages in improving the reliability of process-oriented assessment. Video capture allows observation of performances multiple times with variable play speed. This can minimise the perceptual demand of observing performances in real-time and improve assessment accuracy (Haynes & Miller, 2014; Lander et al., 2016; Logan, Robinson, Rudisill, et al., 2012; Sorsdahl, Moe-Nilssen, & Strand, 2008). Video capture also allows multiple assessors to view performances, and reach consensus, as well as scoring by expert assessors, which is often used for assessor training and can provide a baseline for ensuring assessment reliability is high (Barnett et al., 2009; Barnett et al., 2014; Bisi, Pacini Panebianco, Polman, & Stagni, 2017). The time and expertise available in research environments allows the effective implementation of video capture, however the process is time consuming; performances must be captured, uploaded and processed prior to eventual assessment (Burton & Miller, 1998; Sorsdahl et al., 2008). The time costs of video assessment
and limited expertise of primary teachers renders video capture largely impractical in PE settings (Bisi et al., 2017; Burton & Miller, 1998; Sorsdahl et al., 2008).

**Computational approaches to automated FMS assessment**

Reducing the burden on teachers to implement FMS assessment may be possible through computational analyses of FMS performances (Bisi et al., 2017). Ideally, a computational FMS assessment would produce reliable, quantitative analyses of FMS proficiency that mimics expert assessment (Pirsiavash, Vondrick, & Torralba, 2014). Computational assessment could minimise the understanding of motor development and skill observation required of teachers, and allow rapid capture and analyses of FMS performances within education environments.

There are two main approaches proposed in the development of computational FMS assessment that provide equivalent skill information to traditional process-oriented assessments. One method is to mimic process-oriented assessment tools by manually fitting algorithms to kinematic thresholds that capture performance of proficiency criteria, allowing automatic quantification of criteria presence or absence (Bisi et al., 2017). The second approach is, to employ machine learning methods to mimic human assessment processes by automatically determining the kinematic information that represents proficient performance of FMS.

Recently, Bisi et al. (2017) employed the earlier method and developed 23 algorithms to automatically assess proficiency criteria within the locomotor subset of the TGMD-2 (Ulrich, 2000) using inertial measurement unit (IMU) data collected from the wrists, ankles and lower back of participants. Bisi et al. (2017) showed good generalisability across their sample in the assessment of proficiency criteria presence, which can be challenging when strict thresholds are employed to assess highly variable movement patterns. Their approach allowed accurate automated assessment of the locomotor subset of TGMD-2 (Ulrich, 2000) without the need for extensive assessor training or understanding of motor development. Analysis time was reduced to 2 minutes per child for automated analyses of IMU data, not including capture and data upload time. The study was an important step forward in the computational
assessment of FMS and the quantification of traditionally qualitative assessments. Bisi et al. (2017) noted that future approaches should look to reduce the number of sensors required, presumably to minimise the set-up and data upload time. Computational analyses of skill performances rely upon the accurate extraction of spatial and temporal kinematic features of poses and movements (Hondori & Khademi, 2014; Pirsiavash et al., 2014). Reducing the number of sensors may reduce the amount of important data captured. For example, five sensors positioned on the trunk and extremities as employed by Bisi et al. (2017) effectively capture movement of all limbs, while a single sensor on the trunk is unlikely to capture the amount of information required for sophisticated assessment. Pose estimation techniques may provide a means to extract and capture kinematic features from a large number of anatomical segments using a single sensor (such as videos or depth sensors) (Pirsiavash et al., 2014; Shotton et al., 2011).

Extracting the kinematic data required for computational analyses from video is complex and time consuming, requiring substantial computational processing power (Hondori & Khademi, 2014; Pirsiavash et al., 2014). Videos are traditionally processed manually, extracting kinematic features using software such as Siliconcoach (The Tarn Group, Dunedin, NZ) or Dartfish (Dartfish, Fribourg, Sweden). However, this can be laborious and time-consuming (Logan et al., 2016). Over the last decade, more efficient and accessible depth sensors that estimate 3-dimensional pose characteristics using tomographic reconstruction have become available. These systems offer efficiencies previously unavailable in FMS capture and assessment.

Consumer-level marker-less motion capture technologies, such as the Microsoft® Kinect® depth sensor, allow kinematic feature extraction in real time and have been widely implemented in movement analyses (Hondori & Khademi, 2014; Jungong, Ling, Dong, & Shotton, 2013). Initially developed for active video gaming, the Kinect® pose estimation was subsequently validated for the capture of gross movements for rehabilitation and skills training, and implemented for the capture of FMS performances (Clark et al., 2012; Galna et al., 2014; Hondori & Khademi, 2014; Obdržálek et al., 2012; Rosenberg et al., 2016; Sgro, Nicolosi, Schembri, Pavone, & Lipoma, 2015; Xu & McGorry, 2015). The Kinect® captures 3-dimensional kinematic data of 22 joint centres in real-time, and has been considered a cost-
effective, consumer-level alternative to gold standard infrared systems such as Vicon (Clark et al., 2012; Galna et al., 2014). The Kinect® provides many benefits for cost-effective motion capture, especially efficient kinematic analyses. Unfortunately, a lack of popularity amongst gamers led to the discontinuation of the system in 2014. However, the Kinect® continues to be implemented outside of gaming. The field of consumer kinematic data extraction is rapidly evolving from depth cameras, through video based systems including neural network approaches such as DensePose (Güler, Neverova, & Kokkinos, 2018). DensePose produces accurate pose estimation and 3-Dimensional body tracking from traditional 2-dimensional video, representing a significant leap forward in computer vision capabilities. A limitation of depth camera’s, such as the Kinect® is the small capture range and environmental constraints which reduces freedom in the way skills can be captured (Hondori & Khademi, 2014). These limitations are overcome with networks such as DensePose (Güler et al., 2018) due to the freedom in capture environments and filming range allowed by traditional video. The opportunities to develop accurate and reliable computational FMS assessments that mimic human assessment are increasing, as systems for kinematic data extraction become more sophisticated and accessible (Güler et al., 2018; Pirsiavash et al., 2014). The major challenge to mimicking human FMS assessment process with machine learning is the limited understanding of the human process to be mimicked.

An important component of any machine learning approach is reducing data to key variables or features to minimise error introduced by unnecessary features (Roweis & Saul, 2000). Likewise, accurate human FMS assessment relies upon assessors’ ability to visually detect key kinematic features that represent proficient performance (Giblin, Farrow, Reid, Ball, & Abernethy, 2015; Hernández, Romero, Vailllo, & Campo, 2006). It is known that assessors with sufficient domain knowledge and training can reliably determine overall FMS proficiency levels, however, no studies have explored observational or information processing strategies expert assessors employ to make FMS proficiency judgements (Barnett et al., 2014). Thus, the identification of features employed by expert FMS assessor when making proficiency judgements is an important step in the development of machine learning approaches that mimic expert FMS assessment. Not only will a greater understanding of perceptual processes employed by expert assessors during FMS assessment inform machine learning approaches, but also more effective training for novice assessors including teachers.
Conclusion

FMS assessment remains a critical component of identifying movement proficiency during childhood and adolescence. This review has highlighted the challenges involved in assessing and monitoring FMS development in education environments, and the need to develop appropriate assessment tools that are fast and reliable to administer with minimal training. A lack of training is recognised as a significant barrier to effective FMS assessment in education environments, however, it is unlikely that teachers have the time or resources to engage in sufficient professional development to reliably implement current assessment tools. Minimising the human input of FMS assessment via the employment of emerging technologies that capture FMS and analyse proficiency computationally may provide an alternate solution.

Computational assessments could analyse FMS performances by replicating current tools, although we propose that a more robust computational assessment may be generated with a data driven approach employing machine learning. For machine learning to be accurate, clear ‘learning’ outcome needs to be established and data needs to be reduced to ensure key variables are used for ‘training’. The detection of critical kinematic features underpins accurate and reliable skill observation, however, there is little understanding of the way FMS assessors make proficiency judgements from an observational perspective. To develop a computational approach to FMS proficiency assessment that mimics the process of expert human assessors, a greater understanding of the human process of assessment is required. The following series of studies aim to investigate assessment patterns and perceptual processes of expert and novice FMS assessors employing recent advancements in computer vision technologies; with a clear goal of identifying the information necessary to develop computational assessment strategies.
References


Furtado, O., Jr., & Gallagher, J. D. (2012). The reliability of classification decisions for the Furtado-Gallagher Computerized Observational Movement Pattern Assessment


CHAPTER 4
Protocols for the investigation of information processing in human assessment of fundamental movement skills

This chapter presents the original manuscript of an article published in The Journal of Motor Behavior by Taylor & Francis on 23rd December 2016, available online at http://dx.doi.org/10.1080/00222895.2016.1247033:

Foreword

The first chapter within this thesis presents the proposed protocols for the investigation of information processing during FMS assessment including the employment of consumer-level depth sensors for the capture of FMS performances. This was the first project to propose the use of consumer-level depth sensors for FMS capture and the investigation of information processing during FMS assessment with the goal of informing computational FMS assessment techniques. Therefore, it was important that the proposed protocols underwent critical review and were accepted as a valid approach to answering the research questions within the thesis.
Abstract

Fundamental movement skill (FMS) assessment remains an important tool in classifying individuals’ level of FMS proficiency. The collection of FMS performances for assessment and monitoring has remained unchanged over the last few decades, but new motion capture technologies offer opportunities to automate this process. To achieve this, a greater understanding of the human process of movement skill assessment is required. This manuscript presents the rationale and protocols of a project which aims to investigate the visual search patterns and information extraction employed by human assessors during FMS assessment, as well as the implementation of the Kinect® system for FMS capture.
Background

Fundamental movement skills

Fundamental movement skills (FMS) are widely accepted as core components of physical development and form the foundations for more specialised and sport specific movement skills (Hardy, Barnett, Espinel, & Okely, 2013; Lubans, Morgan, Cliff, Barnett, & Okely, 2010). Amongst Australian children FMS competency is low, with greater than 90% of grade 2 boys exhibiting low competency across all FMS (Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012). Many issues can arise from low FMS competency including poorer health outcomes, as well as academic, social, and emotional difficulties (Hardy et al., 2012; Piek, Hands, & Licari, 2012).

Children of primary school age represent a particularly key developmental period in regards to FMS acquisition, and early intervention for children identified as having low FMS competency can be beneficial (Hands, 2002; Piek et al., 2012). Without intervention, the problems associated with low competency can persist throughout adolescence and into adulthood, causing long-term health concerns (Lubans et al., 2010; Piek et al., 2012; Slater, Hillier, & Civetta, 2010). Intervention programs rely on accurate information surrounding FMS performances (Hands, 2002; Piek et al., 2012). One of the key challenges in identifying children with low FMS competency is the development of appropriate FMS assessments that can be easily and rapidly implemented, and yield the information necessary for intervention.

Fundamental movement skill assessment

No single movement skill assessment has been accepted as a ‘gold standard’. However, many valid instruments for the assessment of movement skills exist. Most instruments have a similar purpose, but differ in their constructs and assessment techniques (Barnett & Peters, 2004; Cools, De Martelaer, Samaey, & Andries, 2009; Piek et al., 2012).

Movement skill assessments usually employ either process-oriented, or outcome-based measures, or a combination of both (Burton & Miller, 1998; Hands, 2002). Outcome-based assessments measure the product or outcome of a skill performance, for example, distance
jumped, or time to run (Burton & Miller, 1998). These quick and easy to administer, objective assessments allow easy comparison of children with similar characteristics, with little understanding of human movement required by the assessor (Hands, 2002). However, outcome-based assessment strategies provide little information about movement patterns or processes, making it difficult to inform specific intervention strategies using the assessment results (Hands, 2002).

Process-oriented movement assessments assess the features, or processes within a particular skill performance, and are the most commonly used FMS assessment construct for children of primary school age (Hands, 2002). Many of the process-oriented assessments used for FMS compare an individual’s performance against a set of established skill criteria, often in the form of observation records or checklists (Burton & Miller, 1998; Cools et al., 2009; Hands, 2002). Checklists offer a more standardised means of process-oriented assessment, although human variability makes it difficult to compare performances evaluated by different assessors and can undermine assessment results (Barnett et al., 2009; Hands, 2002). Many checklists are derived from theoretical approaches to motor development, although most assessment manuals neglect to justify the basis for performance criteria selection (Hands, 2002). The number of performance criteria for each particular skill is also highly varied between assessment tools, suggesting there is a wide range of movement patterns that can define proficient performance (Ulrich, 2000; Western Australian Minister for Education, 2004). Due to their subjective nature, process-oriented assessments require extensive assessor training, which can render them costly and time consuming (Hands, 2002). In addition to this, in a school education context, many teachers required to administer FMS assessments have not undergone sufficient training to achieve proficiency. Recent developments in motion capture technology may allow the development of objective process-oriented assessments, that do not require assessors to have a deep understanding of movement pattern analysis.

Capturing fundamental movement skill performances

The collection and recording of movement skill data has remained relatively unchanged for several decades. Most assessments are carried out in a live setting and employ traditional pen and paper recording. This is mostly due to the inapplicability of motion capture technologies
for population level analyses, and the time efficiency that live assessment offers in comparison to recorded capture techniques. Live assessment is the most feasible method for those working with large populations. However, some FMS may be difficult to assess live (Barnett et al., 2009), due to difficulties in observing multiple skill components during such discrete skill performances.

Video cameras have been widely employed for the collection of movement skill performances, allowing the storage of footage for post processing. Video capture allows assessment by multiple assessors, as well as repeated viewing of performance at reduced speeds (Haynes & Miller, 2014; Logan, Robinson, Rudisill, Wadsworth, & Morera, 2012). Sorsdahl, Moe-Nilssen, and Strand (2008) found video observation to offer many advantages in their comparison of two movement measures, but found the procedure time consuming. Haynes and Miller (2014) found significant differences in the observation of some FMS criteria between live and video assessments, although their analyses did not allow them to conclude whether one observation method was considered superior to the other. Video analyses, whilst providing more objective and accurate assessments, are time inefficient and not viable for those dealing with larger populations. A more recent consideration is to develop automated computer based assessments, that may ultimately lead to FMS assessments with real-time proficiency feedback (Pirsiavash, Vondrick, & Torralba, 2014).

Using current video technology with an automated computer based assessment of movement is problematic as the extraction of 3-dimensional kinematic information from the 2-dimensional video is difficult (Pirsiavash et al., 2014). The visual depth of kinematic assessment required to assess movement skill proficiency currently appears too great for a computer system to achieve with 2-dimensional analysis (Pirsiavash et al., 2014). When it comes to automating movement assessments using a computer, a system able to capture 3-dimensional kinematic data is required. These systems, such as Vicon (Oxford Metrics, Oxford, UK) exist, employing highly accurate motion capture technologies, that provide 3-dimensional kinematic information using infrared multi-camera systems and markers placed on specific points on the body. These systems are the current gold standard of motion capture although their cost, laboriousness, and lack of portability confines them to research laboratories and limits their use at a population level (Burton & Miller, 1998; Harris & Heriza, 1987). Whilst
many capture technologies allow post-hoc assessment approaches, increasing the reliability and objectivity of FMS assessments, live assessment continues to be primarily used due to the time efficiency and real-time feedback of this approach. Ideally, a tool would be developed that has built in validity and reliability of post-hoc assessment platforms, with the advantage of real-time feedback. With the release of the Microsoft® Kinect® camera in 2009 developed for active video game play, it is now possible to use the in-built feature extraction capabilities for the assessment of FMS.

The Microsoft® Kinect® for Windows is a marker-less, single camera motion capture system increasingly implemented in the field of kinematic analysis (Jungong, Ling, Dong, & Shotton, 2013). Using a machine learning algorithm with millions of training samples, the Kinect® system estimates joint centres from tomographic reconstruction. The tomographic reconstruction and joint centre estimation occurs in real time, frame-by-frame, as the body identified in the depth image moves through the 3-dimensional infrared map projected by the camera (MacCormick). The system simultaneously captures 3-dimensional skeletal tracking data and RGB video (Yeung, Cheng, Fong, Lee, & Tong, 2014). The Kinect® motion capture system is not as accurate, nor sensitive as the gold standard Vicon 3-dimensional motion capture system due to its lower capture rate and single camera nature. Validation studies suggest that the depth data produced by the Kinect® is sensitive enough for assessing gross movements similar to those performed in FMS (Galna et al., 2014; Mousavi Hondori & Khademi, 2014; Yeung et al., 2014). The accuracy with which the Kinect® system is able to capture the kinematics of a performer has been studied comprehensively with varying degrees of accuracy found. The error in depth mapping within a static scene ranges from 2mm to 7cm as the object increases from 1m from the sensor to 5m respectively (Khoshelham & Elberink, 2012). Similar levels of accuracy (28-36mm) are reported for joint centre location estimates of a static upper body model across a distance of 0.15-1.6m from the sensor (Mobini, Behzadipour, & Saadat Foumani, 2014). A comprehensive study of full body pose estimation by Obdrzálek et al. (2012) found slightly higher levels of error when comparing to the PhaseSpace motion capture system, but noted that the Kinect® was significantly affected by objects within the collection space such as chairs, and the occlusion of limbs during movements that involve crossing over. It should also be noted that inaccuracies tend to arise when movements captured are largely different to those within an active video gaming
context, for which the Kinect® was developed (Xu & McGorry, 2015). The Kinect® system is unlikely to be accurate enough for clinical kinematic assessments but could present a reliable and cost-effective tool for population level movement assessment, however the specific implementation of the system in FMS assessment is yet to be explored.

The objective and quantitative nature of the Kinect® system offers a rapid, portable and reliable capture system for FMS assessment. The 3-dimensional skeletal tracking data also requires much less digital storage space than Vicon, or video, and is more rapidly transferrable than traditional video data, allowing easy storage and sharing of performance information. The skeletal data provides the opportunity for software to automatically assess movement proficiency, including movement processes and skill components, eliminating the human variability in assessment. The system has been successfully implemented to capture FMS performances during active video game (AVG) play, but the analysis of these performances has not extended beyond FMS frequency counts (Thornton, Lay, Rosenberg, Granich, & Braham, 2014).

A machine learning approach to fundamental movement skill assessment

The assessment of movement quality can be a relatively simple task for highly trained humans. However, automating this assessment using a computer is incredibly complex. One approach to automating kinematic analysis is machine learning. Over the last decade machine learning approaches have become mainstream, as a means to automatically classify or predict outcomes from data (Domingos, 2012). The overarching goal of machine learning in an FMS assessment context would be to produce a program (algorithm) able to classify proficiency of unseen FMS after being trained based upon a known FMS dataset (Mohri, Talwalkar, & Rostamizadeh, 2012).

A key component of any machine learning approach is defining a clear outcome, which in the case of FMS assessment might be as simple as a proficient/not proficient classification, or a more complex gradation based upon a spectrum of proficiency. Regardless of the selected outcome, the machine-learning approach is dependent on firstly, providing accurate information and secondly, on reducing the data to key points when training the computer. In
part, we have tried to provide all features to the machine and let it determine what is required. However, the computer does not know what forms the basis of a good movement and therefore at some point you have to provide this information. This approach is similar to training human assessors to judge the quality of movement using kinematic information. However, humans find it relatively easy to learn how to focus on the key aspects of FMS in their assessment, while computers lack the innate ability to filter irrelevant data. Perhaps as there has not been a need until now, little is really known about how humans process relevant information when making FMS decisions. A better understanding is now required to assist in training computers to replicate the human assessment process.

**Perception of biological motion**

Computer visual representation of 3-dimensional human movement is commonly represented as a point light display (PLD). These displays are visual representations of movement, which depict the kinematic features of motion using single light points, most frequently placed on the joint centres (Blake & Shiffrar, 2007). PLD’s can be produced using computer animations, and also using markered motion capture technology (Blake & Shiffrar, 2007). The PLDs produced from the Kinect® capture represents raw digitised kinematic data (3-dimensional joint positions in time), allowing a direct link between the information a human observer sees, and the data available for an automated assessment.

It is well established that humans are able to identify and name a range of gross movement actions being performed in PLD’s (Blake & Shiffrar, 2007; Dittrich, 1993; Norman, Payton, Long, & Hawkes, 2004). Evidence suggests that humans are able to yield information from PLD’s that is sufficient not only classify actions, but to also to perceive the kinematic features of the actions to determine gender, identity, and emotion of a PL-defined performer on most occasions (Blake & Shiffrar, 2007; Clarke, Bradshaw, Field, Hampson, & Rose, 2005). The threshold at which humans can no longer identify what is being performed within a PLD seems to be less related to the action itself, and more concerned with how the PLD is presented, for example, whether all joint centres are represented, whether the joint centres appear within an array of unrelated light points, or if the PLD is only presented for a short time. Even within these scenarios, perception of motion appears to be remarkably robust (Blake & Shiffrar,
The use of PLD’s has also been extended to sporting contexts, particularly in the study of expert-novice differences in cue utilisation (Abernethy & Zawi, 2007; Sparrow, Shemmel, & Shinkfield, 2001).

PLD’s provide the opportunity to explore the kinematic information employed by assessors in their assessment of FMS proficiency. To date, the study of biological motion perception using PLD’s is largely biased towards tasks involving movement recognition, and although evidence suggests that the kinematic information provided by PLD’s could be sufficient, research is yet to extend to movement proficiency assessment (Blake & Shiffrar, 2007). To automate the movement proficiency assessment using motion capture systems a better understanding is required of the processes involved in movement recognition and assessment.

There are direct and indirect techniques for understanding the information being used during skill observation (Magill R.A., 2007). The most direct approach is the recording of gaze behaviours, a direct observation of where the eyes are focused within the visual field (Magill R.A., 2007). This technique has been used within the sporting context to investigate visual search since the late 1980’s (Abernethy, 1988; Savelsbergh G.J.P., van Gastel P.J., & van Kampen P.M.; Vaeyens, Lenoir, Williams, & Philippaerts, 2007). Eye movement recordings have been predominantly implemented to study the way athletes use visual information (Janelle & Hillman, 2003). The extension of eye movement recording to investigate coaches and judges visual search is limited, with the few studies available reporting conflicting results (Bard, Fleury, Carriere, & Halle, 1980; Hernández, Romero, Vaillo, & Campo, 2006; Petrakis, 1987).

It can be the case that fixations do not relate directly to information extraction, which brought about the notion of a difference between ‘looking’ and ‘seeing’ (Williams A.M., Davids K., & Williams J.G., 1999). To overcome this, spatial occlusion can be used to identify the specific visual information assessors are using (Magill R.A., 2007). By selectively withholding kinematic information from the assessors and investigating how it affects their judgment, spatial occlusion can be used to determine the information assessors are using, both consciously and unconsciously. The ability to detect information that is being used unconsciously allows spatial occlusion to overcome some of the limitations of eye movement recordings. The
method has also been used successfully since the 1980’s to identify the information athletes are using during sport specific tasks (Abernethy B. & Russell D.G., 1987; Lyle & Cook, 1984; Magill R.A., 2007). The combination of both eye movement recording, and spatial occlusion in this project will result in a targeted approach to investigating the information assessors are employing whilst observing FMS.

Summary

The capture and analysis of fundamental movement skills at a population level remains an important tool in identifying individuals at risk of low engagement in lifelong physical activity. However, little has changed over the last few decades in how FMS is captured. New technologies, like the Kinect® motion capture system provide new opportunities for remote and rapid collection of FMS, as well as streamlined FMS data sharing and automated assessment previously unavailable.

Progression towards automating FMS assessment using the Kinect® system would greatly increase the efficiency and accuracy of assessment but requires a much deeper understanding of the human process of assessment than is currently understood. Through investigating the visual search and information extraction of human FMS assessors, we can identify the FMS components that have the greatest bearing on human assessment of movement proficiency. This information can then be incorporated into an automated assessment mimicking the human process of assessment.

In addition to extending the literature surrounding movement skill assessments, the study will also allow us to develop a greater understanding of the perception of biological motion and the information required for proficiency assessment. This information is not only valuable for the progression towards automated assessments, but also highly valuable for the development of theory driven process-oriented skill assessments.

The primary aim of this project is to determine the perceptual information and visual search strategies assessors use while assessing fundamental movement skills. A secondary aim is the
implementation of the Kinect® system for FMS capture and subsequent assessment of FMS using point-light display,

**Methods**

**Study 1**
FMS assessment reliability between RGB video, and point light displays to explore the use of PLD capture in FMS assessment

**Study 2:**
Eye tracking to explore the visual search behaviour and information extraction of FMS assessors scoring PLD skill performances.

**Study 3:**
Spatial occlusion techniques based on the findings of part 1 to explore the information assessors may be extracting outside of focal vision

**Figure 1. Study Overview**

Through three independent studies, this project will investigate the efficacy of PLD’s for FMS assessment, and the information human assessors require when evaluating FMS proficiency. The first study compares video assessment and PLD assessment using the Test of Gross Motor Development-2 (Ulrich, 2000), to measure the PLD’s produced by the Kinect® in FMS assessment. Study 2 uses eye-tracking to investigate the visual search behaviour and information extraction of assessors while they watch children performing FMS through PLD’s. The third study will use spatial occlusion of the PLD (removing segments from the display) and
eye tracking methods to isolate the minimum information human assessors require for accurate assessment of FMS.

**Fundamental movement skill Capture**

An initial bank of children’s FMS performances captured using the Kinect® system, will be gathered prior to study 1. The Kinect®’s ability to simultaneously capture both RGB video and 3-dimensional depth data of estimated joint centres (used to produce PLD’s) simultaneously allows the direct comparison between traditional video assessment techniques and the point light assessment.

**Sample**

An initial FMS collection (those who will be performing the FMS) will occur amongst 100 primary school aged (5-12 y/o) children recruited through their school.

**Procedure**

FMS capture will take place at various primary schools in the Perth metropolitan area after School principal, parent and child informed consent. Capture will occur in an indoor space suitable for the use of the Kinect® infrared receptors. Two Kinect® cameras placed at approximately 44 Degrees apart to minimise joint occlusion will simultaneously capture RGB video and skeletal tracking data for all performances using software designed in-house. This software translates and merges the 3-dimensional skeletal tracking data from both Kinect® cameras into a single 2-dimensional PL video file, which is an exact representation of the whole body RGB video file. Cameras will be placed 2 metres apart, at a height of 50 centimetres, with lenses angled to converge at 2.5 metres, the point from which participants will perform the skills (see Figure 2).

Participants will perform five FMS encompassing both locomotor and object control skills, as well as a static balance task that can be executed within the visual capture space of the Kinect® system. These skills are the horizontal jump, hop, overarm throw, catch, and kick. Following a calibration check of the Kinect® system, participants will perform three valid trials of each FMS and a single static balance trial on each leg. All skills, with the exception of the
kick will be performed 2.5 metres from the cameras. The kick will be performed 2 metres from the camera with a step-up allowed from 2.5 metres.

Performances will undergo post-hoc assessment using the Test of Gross Motor Development-2 (Ulrich, 2000) to ensure a representative distribution of proficiency levels is implemented in the succeeding studies.

![Figure 2. Kinect camera setup for FMS performance collection](image)

**Study 1: Comparison of RGB video and PLD presentation in FMS assessment**

**Study overview**

The use of PLD's from the Kinect® sensor as an appropriate alternative to traditional RGB video assessment mediums requires validation. As there is no gold standard for the medium through which FMS are assessed, this study will compare PLD analysis, with video assessment using a within subjects’ design. The study is designed to explore the use of depth cameras in FMS capture and assessment using a connected PL display as an alternative to traditional video presentation methods. We hypothesise that there will be no significant difference in performance criteria recognition between PLD and RGB video conditions.

**Sample**

The sample of assessors for the first study will be 130 undergraduate students in the field of sports science. Assessors will be approached through their university course. The sample is representative of a population that would conduct FMS assessments within their line of work and will be familiar with motor development and FMS performances. Due to their
undergraduate status, experience levels will be reasonably consistent, allowing us to control for the effect of experience on assessment and isolate the effects of skill presentation style (PLD or RGB video). Involvement in the study will be voluntary. Interested participants will be provided with information and written consent forms, to be returned to the researchers prior to participating in the testing session.

Procedure.

Aim

To determine the intra-rater reliability of movement skill performance criteria when skills are presented in connected PLD’s, in comparison to RGB video.

Procedure

Assessors will be invited to attend a single, 1-hour testing session at the School of Sport Science, Exercise and Health, University of Western Australia. Within each session, assessors will be required to evaluate performances of 4 different FMS (horizontal jump, hop, throw, kick) presented in either video or PLD format, using Ulrich’s (2000) TGMD-2. The TGMD-2 is a previously validated FMS assessment tool and is implemented within the study to provide structure for the analysis of intra-rater agreement between video and PLD assessment. The use of video capture allows us to draw intra-rater comparisons surrounding the way skill performances are assessed, whilst controlling for any inter-trial variability between live FMS performances.

Testing will be conducted in groups of up to 25 assessors. Assessors will view FMS performances presented within a video on individual LCD computer monitors and score the performances on an iPad® using the Qualtrics® survey platform.

Upon arrival participants will be familiarised with the manner in which skills will be presented throughout the protocol, and also the TGMD-2 performance criteria for each of the 4 skills. The familiarisation will also include 2 practice assessments of each skill in both video and PLD formats. Whilst there may be differing individual interpretations of the TGMD-2 criteria, the within-subjects nature of the study negates the need for more comprehensive familiarisation and training with the assessment tool. Within the familiarisation protocol participants will
also rate on a 5-point scale, their level of experience assessing FMS, and also their confidence in their ability to assess FMS.

The study protocol will consist of one video and one PLD block for each of the four skills, totalling 8 assessment blocks. Participants will assess both the video and PLD block of each skill successively, with the order counterbalanced between participants. Each block will contain 16 performances of the skill, with performance order consistent between video and PLD blocks, but randomised between participants. The 16 performances within each block will be selected from 8 different children representing a normal distribution of FMS proficiency, who will contribute 2 trials each. In 10 cases, the performances will consist of 2 different trials, whilst the remaining 6 cases will consist of a single trial duplicated. This will allow analysis of inter-rater agreement within, as well as between, each of the presentation styles. Duplicated performances will be separated by a minimum of two other performances to ensure participants remain blinded to the duplications.

Within the assessment video assessors will be presented with the skill name, followed by 5 seconds to review the performance criteria checklist for the skill. The 5 second revision time will be accompanied by a 3 second visual and auditory countdown prior to the playing of the skill performance. Following each performance, assessors will have 15 seconds to complete the TGMD-2 (Ulrich, 2000) performance criteria checklist for the skill. A 3-second frame presented at the start of each block will notify participants if the succeeding performances will be in video or PLD format. A 3-minute break will be granted to participants half way through the collection protocol (following block 4).

Analysis

To explore differences in how assessors rate movement when presented with either video or PLD descriptive statistics will be used in the first instance to describe patterns of responses. Intra rater reliability will comprise Kappa correlations in the rating of movements between presentation methods. This will provide statistical evidence if differences between methods exists.

Study 2: Investigation of visual search and information extraction during FMS assessment
The second study aims to investigate the visual search behaviours and information extraction of movement assessors through the use of eye movement recording. The study is designed to identify the visual cues being used consciously by assessors.

**Sample**

Participants for studies 2 and 3 will be recruited simultaneously and will comprise 30 individuals that administer movement skill assessments on a regular basis (at least every 6 months) within their professional roles, who will all partake in both studies. The sample will include, but not be limited to, primary teachers, exercise physiologists, occupational therapists, and physiotherapists. Participant experience will be collected via an initial questionnaire and treated as a covariate during data analysis to make comparisons across experience levels. No standardised accreditation for movement skill assessment exists within Australia. For this reason, it is difficult to make expert-novice comparisons, or to classify specific levels of ‘assessor proficiency’. This is a potential limitation in a study that aims to identify the information used by high-level movement assessors.

**Aim**

To investigate where observer’s gaze is directed when assessing connected PL representations of movement skills.

**Procedure**

Participants will be required to attend a single testing session at the University of Western Australia. Upon arrival participants will be fitted with the Dikablis Essential (Ergoneers GmbH, Germany) eye-tracking system which will be used to quantify gaze behaviour throughout the assessment protocol. The eye-tracking system will be calibrated using a standardised calibration screen prior to data collection.

Participants will view 20 performances of each FMS presented as complete PLD’s, on an LCD computer monitor. FMS performances will be sourced from the same bank of performances as study 1, however may not be the exact same performances. Each block of 20 skill performances will contain an even distribution of proficiency levels. Block presentations will be counterbalanced, and skills within each block will be randomised to minimise order effects.
Participants will view each skill performance once in real time, prior to rating the performance out of 10, with 10 representing an expert performance of the skill and 1 representing the worst possible performance of the skill. Implementing a more structured assessment of movement quality such as the TGMD-2, is likely to bias the visual search of assessors, as there are pre-determined observational criteria the assessors are required to attend to. A 1-10 scale allows assessors to rate the quality of performances, without introducing any visual search bias that may contaminate results.

Following the scoring of each skill performance participants will be required to provide a verbal response to the following questions regarding the performance they have just viewed:

1. What was/were the key movement feature/s that you observed to generate your performance score?
2. What information did this/these feature/s give to enable you to make a judgement on the entire skill performance?

Participants will record assessment scores within the presentation software immediately after each performance viewing, and document their responses

**Eye movement recordings**

For each skill assessment, the quantity, duration, and location of fixations will be collected using a Dikablis Essential (Ergoneers GmbH, Germany) eye-tracking system. A fixation will be classified as gaze fixed on a single location for 100ms or greater. For expert and non-expert comparison, fixation data will be classified for analysis in the following variables; average number of fixations, average durations of fixations, most common location of first fixation, and most common fixation location (determined by greatest overall fixation duration). Fixation locations will be categorised using a model adapted from Lee M.J.C. et al. (2013), displayed in Figure 3.
Self-report protocol

Responses to question one, regarding key features, will be assessed for references to spatial locations, which will then be coded using the same model employed for gaze analysis adapted from Lee M.J.C. et al. (2013). Responses to question two will be assessed for spatial linkages and indicators of proficiency following a protocol adapted from McPherson (1993), including references to movement concepts such as ‘momentum’, ‘stability’, or ‘propulsion’. These responses will not undergo any further analysis, although may be used in future to inform inexperienced assessors about the information that can be yielded from key movement features by experienced assessors.

Analysis

Descriptive statistics will be used to explore visual search patterns and assessments of movement proficiency across all participants. To explore whether experience impacts upon assessment a series of ANCOVA’s will be conducted to compare experience levels with assessment strategies and skill scoring. Age, gender and other relevant characteristics will be included in the model as covariates.
Study 3: Investigating unconscious visual processing during FMS assessment

Aims
Through the application of spatial occlusion techniques to PLD’s, study 3 aims to better understand the information filtration processes employed during the assessment of FMS. It may also establish a theoretical base for performance criteria selection in future assessments.

Procedure
Following the recognition of key features in study 2, this study will employ spatial occlusion techniques to reduce the information presented to participants to only those identified features. The digital nature of the capture technique allows spatial occlusion through the removal of all data points (estimated joint centres) that are not recognised as integral to the assessment of each skill.

Participants will view 20 performances of each FMS presented as PLD’s, on an LCD computer monitor. Each block of 20 skill performances will contain an even distribution of proficiency levels. Block presentations will be counterbalanced, and skills within each block will be randomised to minimise order effects.

Participants will view each skill performance twice, in real time, prior to rating the performance out of 10, with 10 representing an expert performance of the skill and 1 representing the worst possible performance of the skill.

The set of skill performances will remain consistent across the studies 2 and 3, however skill performances in study 3 will be spatially occluded versions of the PLD performances presented in study 2. The spatial occlusions will be strategic in nature, based upon the key skill components identified in study 2. Features that are identified as integral to the skill assessment will remain, whilst all other features are occluded within the PLD’s.

Analysis
Descriptive statistics will be used to explore visual search patterns and assessments of movement proficiency across all participants. To explore whether experience impacts upon assessment a series of ANCOVA’s will be conducted to compare experience levels with
assessment strategies and skill scoring. Age, gender and other relevant characteristics will be included in the model as covariates.

**Discussion**

FMS assessment and monitoring has remained unchanged for many years and is yet to take advantage of recent advancements in portable motion capture technologies. These technologies could provide opportunities to automate assessment, removing human error and greatly decreasing the time, and financial costs of current assessment techniques. The implementation of such technologies requires a much greater understanding of the human process of FMS assessment than is currently available. Through the investigation of gaze behaviours and information extraction during FMS assessment we can gain a greater insight into the human process that would need to be duplicated in an automated approach to assessment. This information will also provide a theoretical basis for the inclusion of performance criteria in manual assessments through the identification of the specific features required for proficiency judgments.

**Implementation**

This project will be the first to identify the kinematic information required by human observers to assess fundamental movement skills. The knowledge gained from this study will contribute to a greater understanding of biological motion perception. It will also provide a theoretical basis for the inclusion or exclusion of skill performance criteria in process-oriented FMS assessments. Finally, the information identified will provide an insight into the information required by a computer to automatically assess FMS performances.
References


CHAPTER 5

Investigating the use of a consumer-level depth camera for movement skill capture and assessment

This manuscript was submitted for publication in the *Journal of Motor Learning and Development* in August, 2018 and is currently under review.

Foreword

The first investigation within this thesis was concerned with establishing the suitability of consumer-level depth sensors for the capture and assessment of FMS. At the time of the project’s inception, consumer-level depth sensors provided an efficient means to extract the digital kinematic data that would be required for computational FMS assessment, whilst also being ecologically valid for education environments. Consumer-level depth sensors such as the Microsoft® Kinect® were being established as suitable capture systems for gross movements, however their suitability for FMS capture and subsequent assessment was not well understood. To ascertain whether Kinect® captured kinematic data is appropriate for FMS assessment this investigation compared assessment scoring between traditional video displays and point-light displays produced using Kinect® data.

Note: The FMS collection protocol for this following investigation was modified from that in the proposed protocols presented in Chapter 4. FMS capture took place at The University of Western Australia rather than in primary schools and the catch task was not included.
Abstract

Introduction

Current fundamental movement skill (FMS) assessment practices have inherent limitations that can impact the administration of assessments in early childhood settings. Recent advancements in cost-effective, consumer-level kinematic depth capture may overcome barriers to administration. This study aimed to investigate the use of the Microsoft® Kinect® system for the capture of FMSs by comparing assessment scoring from point light (PL) displays produced by the Kinect®, and video presentations of FMS performances.

Methods

Fifty-three assessors scored 16 performances of four FMS presented in videos and PL displays using process-oriented FMS assessment techniques.

Results. Moderate agreement between video and PL criteria identification was reported. PL assessments were found to be significantly less accurate than video assessments. A disparity was found between performance level reliability and criterion level accuracy suggesting that the kinematic information assessors use to judge proficiency may not align with assessment criteria.

Conclusion

PL displays produced by the Kinect® may not be appropriate for FMS assessment by moderately trained assessors and require further investigation with experienced assessors. It is also necessary to further explore criterion level scoring patterns in live process-oriented FMS assessment when conducted by moderately trained assessors.
Introduction

Developing proficient movement patterns in the early years of life plays an important role in establishing healthy physical activity behaviours throughout the lifespan. As such, the assessment and monitoring of fundamental movement skills (FMS) is a priority during childhood (Barnett, van Beurden, Morgan, Brooks, & Beard, 2009; Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012; Holfelder & Schott, 2014; Logan, Robinson, Rudisill, Wadsworth, & Morera, 2012; Lubans, Morgan, Cliff, Barnett, & Okely, 2010). In kindergarten and pre-primary, a key period of motor development, it is often the responsibility of early childhood specialist teachers to assess and develop FMS (Whipp, Hutton, Grove, & Jackson, 2011). Despite the important role they play in the assessment and teaching of FMS, primary and early childhood specialist teachers may undergo very little training around motor development and FMS assessment throughout their pre-service education (Haynes & Miller, 2014; Lander, Barnett, Brown, & Telford, 2014). This has a significant impact on teacher’s willingness and confidence to undertake FMS monitoring and development (Lander, Morgan, Salmon, & Barnett, 2016).

FMS assessment practices continue to rely predominantly on human assessment in a field-based setting, or video capture and post hoc human assessment (Barnett, Van Beurden, Morgan, Lincoln, et al., 2009; Hands, 2002). Process-oriented FMS assessment tools tend to be preferred as they can be used to identify specific skill deficiencies requiring intervention, by interpreting the presence or absence of criteria pertaining to proficient performances skills (Barnett, Minto, Lander, & Hardy, 2014; Burton & Miller, 1998; Cools, De Martelaer, Samaey, & Andries, 2009; Hands, 2002; Hulteen et al., 2015). Process-oriented assessments yield valuable skill information but can be hindered by their subjectivity, and large financial and time burdens (Logan, Barnett, Goodway, & Stodden, 2016). Low levels of confidence in teachers to conduct movement assessments, combined with the large time and financial burdens associated with the assessment of FMS, requires the investigation of more effective assessment and monitoring techniques to be prioritised. A recent consideration is to implement advancements in motion capture technologies, to develop more efficient, objective assessments of FMS, including computational approaches such as machine learning (Ward, Thornton, Lay, & Rosenberg, 2017)
Depth sensors such as the Microsoft® Kinect® enable the extraction of the kinematic data required for computational analyses with greater efficiency than traditional video (Pirsiavash, Vondrick, & Torralba, 2014; Rosenberg et al., 2016). The Kinect® is a low-cost, single camera, marker-less motion capture system able to simultaneously capture both RGB-video and 3-Dimensional skeletal tracking data. Initially developed for video gaming, the system has since been used in the collection and analysis of kinematic data (Jungong, Ling, Dong, & Shotton, 2013). Whilst video capture allows more freedom in terms of capture environment than depth cameras, the extraction of kinematic data from 2-dimensional video remains computationally expensive (Hondori & Khademi, 2014; Hu et al., 2015; Pirsiavash et al., 2014). As an off-the-shelf system, readily available to consumers, the Kinect® allows efficient capture and extraction of kinematic information required for computational analyses, but is yet to be implemented for process-oriented FMS assessment (Hu et al., 2015).

The Kinect® advanced skeletal tracking has been implemented to analyse human activity including pose estimation, rehabilitation monitoring, and movement coaching (Hondori & Khademi, 2014; Hu et al., 2015; Mobini, Behzadipour, & Saadat Foumani, 2014; Obdrzálek et al., 2012). The system has also been used to detect FMS during active video gaming, but analysis did not extend to proficiency assessments (Rosenberg et al., 2016). The reliability of Kinect® captured depth data may be reduced due to its single-camera capture and relatively low sampling rate, especially when compared to multi-camera, marker-based systems, so it is important to initially understand its suitability for proficiency analyses (Hondori & Khademi, 2014; Hu et al., 2015). If, however, the Kinect® is established as a feasible capture tool for FMS, a move towards computational assessments could result, improving the efficiency and effectiveness of FMS assessment. Computational approaches employing machine learning are concerned with the development of algorithms that that mimic human processes. A major consideration with machine learning for FMS assessment is initially ensuring that the data fed into the algorithm contains the relevant information for the human processes to be mimicked. It was therefore important in the first instance to ensure that the digital kinematic data captured by the Kinect® was suitable for human FMS assessment (Ward et al., 2017).
Because the Kinect®’s depth image and RGB video capture occur simultaneously, direct comparisons can be made from assessment outcomes using the two presentation styles, to ensure Kinect® data contains the information required for FMS assessment. Kinect® depth data is commonly visually reconstructed as a ‘connected’ point-light (PL) display (Blake & Shiffrar, 2007). Unlike videos, PL displays produced from Kinect® data present kinematic information that can be utilised by both human observers and computational assessment approaches. PL displays typically present the performer as an array of points positioned at the locations of joint centres, and have been used extensively to investigate human perception of biological motion (Blake & Shiffrar, 2007). The perception of movements in PL displays is robust and humans can readily and accurately recognise PL actions and additionally interpret features of the performer such as gender, weight, and mood. This suggests that observers are able to make in-depth judgements of kinematic features, even though environmental context has been removed (Blake & Shiffrar, 2007; Gold, Tadin, Cook, & Blake, 2008; Norman, Payton, Long, & Hawkes, 2004). If the Kinect® system is capable of capturing sufficient kinematic data for FMS assessment, we would expect little difference in assessment accuracy using Kinect®-produced PL displays when compared to accuracy from full-figure presentations within videos.

This study investigated the suitability of kinematic data captured by the Microsoft® Kinect® for FMS assessment by comparing human assessments made from Kinect®-produced PL displays to full-figure video presentations typically used for FMS assessment. We hypothesised that Kinect® depth capture would provide sufficient skill information, and therefore human ratings of the presence or absence of particular skill components assessed using both the RGB and depth images would be highly comparable. Support for this hypothesis would suggest that kinematic data captured by consumer-level depth cameras is appropriate for FMS proficiency assessments by humans and could subsequently be used in the development of computational assessment strategies.

**Methods**

Data collection for this study comprised three parts; (1) The capture of FMS performances using the Microsoft® Kinect®, (2) The assessment of FMS performances by a sample of trained
assessors, (3) The development of expert ratings to use as a basis for comparison and analysis. The assessment protocols throughout the study were derived from Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000), however some adaptations were made due to capture limitations of the Kinect® and the primary aim of the study. The TGMD-2 (Ulrich, 2000) is a previously validated and reliable process-oriented FMS assessment tool that is widely used in field-based, real-time assessments. All components of the study had ethics approval from the University of Western Australia (UWA, RA/4/1/7445).

**FMS Collection**

**Sample**
The sample of performers for the FMS capture was 16 primary school aged children (age 8.2 ± 2.2 years) recruited through the paediatric movement programs available at the School of Sports Science, Exercise & Health, at UWA. Participants were instructed to wear clothes and footwear they would typically use for school sport for the collection protocol.

**Procedure**
FMS capture took place in an indoor movement lab devoid of natural light to minimise any possible interference with the infrared sensors of the Kinect®. Within the space a Kinect® camera was placed 50cm above the ground, 2.5 metres in front of the participant at a 22-degree angle to the sagittal plane, a viewing angle common in FMS assessment. The camera position ensured that performers remained within both the frame and the capture space for the Kinect® depth sensor for the entirety of performances (Jungong et al., 2013). The camera simultaneously captured RGB video and 3-dimensional joint trajectory data at 30Hz using in-house software designed with the Microsoft® Kinect® SDK. As the study sought to test the feasibility of the Kinect® as an off-the-shelf motion capture tool, PL displays were constructed using raw Kinect® joint trajectory data, with no post-hoc data cleaning or filtering.

Two locomotor and two object control skills were selected from the 12 FMS within the TGMD-2 Participants performed ten trials of each FMS (horizontal jump, hop, overarm throw, kick) in front of the camera. Ten trials were captured as a requirement for subsequent
investigations, however, only the first two trials were employed in the current study, as is the case within the TGMD-2 (Ulrich, 2000). Due to the Kinect® depth sensor’s maximum capture distance of 3.5m, the kick and hop tasks were modified from the standardised TGMD-2 instructions (Jungong et al., 2013). Instead of 3 hops on each leg, participants performed a single hop for distance on their preferred leg. The run-up distance for the kick was reduced to 1.5 metres. As the study sought to compare the two capture techniques for live process-oriented FMS assessment, rather than to validate a particular assessment tool, we expected FMS task changes to have little impact on the study results.

Following the Kinect® collection of the FMS, participants were administered the full TGMD-2 (Ulrich, 2000) to determine each child’s level of motor proficiency, to ensure a range of proficiency levels were employed for comparisons between presentation styles.

PL and video assessment

Of the initial 16 FMS participants, performances from 8 participants were selected for the assessment stage. These participants reflected a range of FMS proficiency levels with 2 in the 20th percentile, 4 in the 20-80th percentile, and 2 in the 80-100th percentile of the sample based on their standardised TGMD-2 (Ulrich, 2000) scores.

Sample

The sample of assessors comprised 47 university students in the final year of their sports science degree. Participants were all familiar with motor development and skill observation, yet had no prior experience with the TGMD-2 (Ulrich, 2000). Prior to undertaking the formal assessments participants were familiarised with the manner in which skills would be presented throughout the protocol, and also the proficiency criteria for each of the 4 skills. The familiarisation also included 2 practice assessments of each skill in both video and PL formats, to ensure assessors met the recommendations for reliable assessment in the TGMD-2 Assessor Manual (Ulrich, 2000). Within the familiarisation protocol participants also rated on a 5-point scale, their level of experience assessing FMS, and also their confidence in their ability to assess FMS. Experience levels were expected to have little impact on comparisons between presentation styles. Participants were provided with information and written consent forms, to be returned to the researchers prior to participating in the testing session.
**Procedure**

Assessors attended a single, 1-hour testing session at the School of Sport Science, Exercise and Health, UWA. Within each session, assessors were required to evaluate performances of 4 different FMS (horizontal jump, hop, throw, kick) presented in both video and PL format, using proficiency criteria derived from the TGMD-2 Ulrich (2000). In the case of the hop, the TGMD-2 (Ulrich, 2000) proficiency criteria assessing whether the performer ‘Takes off and lands three consecutive times on preferred foot’, was adjusted to ‘Takes off and lands on same foot’, to reflect the change in the task. Whilst the minor changes made to the skill performances may impact the generalisation of results to process-oriented assessment techniques, the consistency in changes for both PL and video presentations should not affect comparisons between the two. Participants viewed video and PL performances as recorded clips, however each performance was only observed once in real-time to simulate field-based assessment.

FMS performances were presented on individual LCD computer monitors and assessors scored the presence of skill components on an iPad® using the Qualtrics® survey platform. Assessors were instructed to select the criteria that were present in the performance, and leave absent criteria unselected.

The assessment protocol consisted of one video and one PL block for each of the four skills, totalling 8 assessment blocks. Participants assessed both the video and PL block of each skill successively, with the order counter-balanced between participants. Each block contained 16 independent performances of the respective skill, with performance order consistent between video and PL blocks, but randomised between participants. Each of the 8 final FMS participants contributed 2 trials each. In 10 cases, the performances consisted of 2 different trials, whilst the remaining 6 cases consisted of a single trial duplicated. This allowed analysis of agreement within, as well as between each of the presentation styles. Duplicated performances were separated by a minimum of two other performances to ensure participants remained blinded to the duplications.

Prior to each assessment, assessors were presented with the skill name, followed by 5 seconds to review the proficiency criteria checklist for the skill. The 5 second revision time
was accompanied by a 3 second visual and auditory countdown prior to the playing of the skill performance. Following each performance, assessors were given 15 seconds to complete the proficiency criteria checklist for the skill on the iPad®. A 3-second frame presented at the start of each block notified participants of the presentation style for the subsequent performances. A 3-minute break was granted to participants halfway through the collection protocol (following block 4) to prevent any effects of fatigue.

**Expert ratings development.**

Following unexpected results from initial data analyses, a set of expert ratings for FMS performances were produced to analyse accuracy of PL and video FMS assessments, via a two-expert consensus.

**Sample**

The sample for the expert ratings included three paediatric movement researchers from the School of Sports Science, Exercise & Health at UWA. Each assessor had over 3 years’ experience in paediatric movement analysis and assessment, including extensive experience with the process-oriented assessment tools.

**Procedure**

Two of the assessors completed the provided expert ratings, with the third available to assist in case consensus could not be reached between the two assessors, however, this was not required. To ensure complete understanding of the test, both experts re-familiarised themselves with the proficiency criteria for each of the four skills. Both experts then completed three practice assessments of each skill prior to undertaking the formal assessment protocol.

Only video performances were used for expert ratings, as the study aimed to determine how closely Kinect® captured movements represented the skill information that would typically be presented and assessed in videos. Initially, each expert independently rated proficiency criteria presence in the video performances. Performances were separated into four skill blocks, with all performances of each skill assessed consecutively. Presentation order within each block was counterbalanced to negate any order effects, and assessors were required to
have a 1-hour break between assessment blocks to minimise fatigue. To ensure maximum accuracy of assessment, expert assessors were able to view each performance as many times as they felt was required.

Following the independent ratings from each expert, the data were analysed to highlight any disagreements. The experts reviewed any disagreements and reached consensus. Once consensus had been reached a final dataset was constructed representing the expert rating of criterion presence for each of the performances.

*Data analysis and Results*

Following the collection protocol, assessment data for each participant were entered with proficiency criteria coded as 1, present, or 0, absent. Coded data were then de-randomised and PL and video assessments aligned for comparison.

For clarity throughout the remainder of the paper, data analysed at the ‘criterion level’ refers to analyses which were performed on the binary present/absent scoring of proficiency criteria for each assessment. Analyses at the ‘performance score level’ refer to analyses of skill performance scores, calculated by summing the number of present criteria recorded by assessors for a given performance, to give a score out of four.

To understand whether the information presented in PL displays was representative of video performances, it was imperative to assess agreement at the individual criterion level for each performance. This prevented misrepresenting the actual agreement between the two presentation styles.

*Analyses of agreement*

Initially, data were analysed for absolute agreement at the criterion level between PL and video assessments for each individual assessor. Absolute agreement for each assessor was calculated as the percentage of time the same presence/absence judgement was recorded for a skill criterion in both PL and video. Mean absolute agreement was then calculated across all assessors, giving an indication of how similar scoring was between the two presentation styles. Kappa statistics were not used to analyse agreement due to the binary nature of
criterion level data (present/absent), and the tendency for kappa statistics to be affected by uneven prevalence distributions of binary data (Byrt, Bishop, & Carlin, 1993).

Table 1 presents percentage agreement between point-light and video assessments, and percentage agreement between repeat assessments within each medium, averaged across all assessors. The mean agreement between PL and video assessments was 67.98%. When agreement was considered for individual skills, agreement was the greatest for the kick, followed by the throw, the jump, and lastly the hop.

Table 1. Mean absolute agreement percentage between PL and video assessments

<table>
<thead>
<tr>
<th>Direct PL vs Video Comparison</th>
<th>Point-light vs. Video % (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mean agreement:</td>
<td>67.98 (5.81)</td>
</tr>
<tr>
<td>Kick:</td>
<td>70.58 (10.91)</td>
</tr>
<tr>
<td>Hop:</td>
<td>66.51 (10.82)</td>
</tr>
<tr>
<td>Throw:</td>
<td>68.51 (9.62)</td>
</tr>
<tr>
<td>Jump:</td>
<td>68.72 (10.90)</td>
</tr>
</tbody>
</table>

Secondly, data were analysed to determine the level of agreement between the duplicated performances presented to participants. This was implemented to determine whether assessors were being consistent in their assessments and whether one medium resulted in greater assessment consistency. The same protocol for the analysis of absolute agreement was followed however agreement was analysed within each presentation style rather than between.

Table 2. Mean agreement within presentation styles based on assessments of duplicated performances

<table>
<thead>
<tr>
<th>Within Style Agreement</th>
<th>Video % (SD)</th>
<th>Point-Light % (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mean agreement:</td>
<td>77.27 (8.76)</td>
<td>69.94 (7.61)</td>
</tr>
<tr>
<td>Kick:</td>
<td>80.08 (14.39)</td>
<td>73.06 (11.54)</td>
</tr>
<tr>
<td>Hop:</td>
<td>75.32 (17.04)</td>
<td>68.90 (12.16)</td>
</tr>
<tr>
<td>Throw:</td>
<td>77.69 (15.03)</td>
<td>70.28 (13.15)</td>
</tr>
<tr>
<td>Jump:</td>
<td>76.06 (13.65)</td>
<td>68.11 (16.86)</td>
</tr>
</tbody>
</table>
Across duplicated performances in each style assessors made the same assessment 69.94% of the time in PL and 77.27% of the time in video (Table 2). The kick again showed the highest level of agreement, and the hop the lowest. All individual skills with the exception of the kick fell below the 80% threshold for an acceptable level of agreement (McHugh, 2012).

**Analyses of assessment accuracy**

The low agreement figures observed were surprising given that the Kinect® had previously been used to capture gross movements, humans are robust at observing movements presented in PL displays, and assessors were trained to the level recommended for reliable assessment. Agreement figures did not allow us to determine why there were inconsistencies across styles or whether low agreement was due to one presentation being substantially different to the other, so expert ratings were developed to provide a basis for accuracy analyses. Data were subsequently analysed for percentage accuracy at the criterion level in comparison to the expert ratings for each performance. To determine patterns of response and accuracy at the criterion level the percentage of correct judgements were calculated across the 47 assessors for each individual criterion, in both presentation styles. Wilcoxon signed-rank tests were carried out for overall accuracy and accuracy for each skill, to determine if statistically significant differences were observed between PL and video assessment accuracy and Bland-Altman plots were produced to explore the patterns of bias between PL and video accuracy scores.

Table 3 presents the accuracy of PL and video performances assessments at the criterion level when compared to expert ratings. This comparison found that assessors averaged 61.02% overall accuracy when making assessments from PL displays and 69.92% overall accuracy when making assessments from videos. Table 3 also presents assessment accuracy for each of the four skills analysed individually. Wilcoxon signed-rank tests revealed that PL assessments were significantly less accurate than video assessments across all skills and when considered for each individual skill. Mean assessment accuracy was greatest for the Jump in both presentation styles. The greatest disparity in accuracy between the two presentation styles was found for hop assessments as evidence by the largest z score. Large standard deviations were observed for assessment accuracy for each of the four skills across the 47
assessors, showing that variation in assessment accuracy across assessors was large. Bland-Altman plots (Figure 1.) supported lower accuracy figures in PL assessments, however, revealed no systematic trends in PL and video accuracy.

Table 3. Mean proficiency criteria recognition accuracy for PL and video assessments, accuracy differences between presentation styles, and Wilcoxon signed-rank of accuracy differences between presentation styles

<table>
<thead>
<tr>
<th>Accuracy percentages</th>
<th>Video % (SD)</th>
<th>Point-light % (SD)</th>
<th>Difference %</th>
<th>Wilcoxon Signed-rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Assessments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>69.92 (6.08)</td>
<td>61.02 (5.50)</td>
<td>-8.90</td>
<td>Z = -5.91, p &lt; 0.01</td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick assessment</td>
<td>69.91 (8.89)</td>
<td>62.28 (9.82)</td>
<td>-7.63</td>
<td>Z = -4.34, p &lt; 0.01</td>
</tr>
<tr>
<td>Throw assessment</td>
<td>64.47 (10.17)</td>
<td>59.20 (8.57)</td>
<td>-5.27</td>
<td>Z = -3.78, p &lt; 0.01</td>
</tr>
<tr>
<td>Hop assessment</td>
<td>69.76 (10.55)</td>
<td>56.03 (11.98)</td>
<td>-13.73</td>
<td>Z = -6.00, p &lt; 0.01</td>
</tr>
<tr>
<td>Jump assessment</td>
<td>76.07 (10.14)</td>
<td>66.15 (9.03)</td>
<td>-9.92</td>
<td>Z = -5.03, p &lt; 0.01</td>
</tr>
</tbody>
</table>
Differences in assessment accuracy between styles was also analysed for groups separated by low (<3) and high (=>3) self-reported FMS assessment experience, the difference between styles for high experience assessors was non-significant ($z = -1.40, p = 0.16$) whereas the difference for low experienced assessors was significant ($z = -5.20, p = 0.00$). These figures suggest that those who were more experienced in FMS assessment, or movement observation may be more robust to the different presentation styles.

*Analyses of error distribution within presentation styles*

Following results reporting low assessor accuracy in each style, it was also important to determine the distribution of errors. For example, if assessors tended to report greater false negatives, or false positives in scoring criteria presence. A greater number of false negatives in PL assessment may indicate that information was missing from those displays, or assessors were having greater difficulty in observing skill features. Each present/absent score was initially subtracted from the true present/absent score. A score of -1 would indicate that the assessor had marked the criterion as present when it was absent (overscored), a score of 1...
would indicate the opposite, and a score of 0 would indicate that the assessor was accurate. The distribution scores were summed across all criteria assessed by each assessor. A summed distribution close to zero would suggest that the assessor was neither tending towards overscoring nor underscoring criterion presence. Finally, summed distribution scores were compared for differences between presentation styles using Wilcoxon signed-rank tests.

The distribution of errors suggested that assessors overscored criterion presence significantly more in videos than PL displays across all assessments and for individual skills (Table 4). The distribution of inaccuracies across PL assessments was close to zero, suggesting that participants were equally overscoring and underscoring criterion presence/absence. It important to note that distributions of error do not indicate the number of inaccuracies, rather the direction of error in the case that assessors were inaccurate.

Table 4. Average criterion presence/absence error distribution including Wilcoxon signed-rank tests of differences between presentation styles.

<table>
<thead>
<tr>
<th></th>
<th>Mean criterion error distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video (SD)</td>
<td>Point-light (SD)</td>
</tr>
<tr>
<td>All Assessments:</td>
<td>-16.94 (18.23)</td>
<td>0.57 (22.60)</td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick</td>
<td>-5.53 (7.64)</td>
<td>1.06 (7.99)</td>
</tr>
<tr>
<td>Throw</td>
<td>-8.66 (7.59)</td>
<td>-10.55 (7.15)</td>
</tr>
<tr>
<td>Hop</td>
<td>-1.72 (6.31)</td>
<td>4.57 (7.65)</td>
</tr>
<tr>
<td>Jump</td>
<td>-1.02 (6.76)</td>
<td>5.49 (5.60)</td>
</tr>
</tbody>
</table>

Analyses of performance level scoring

Finally, present criteria (given a value of 1) were summed for each individual performance to give an overall score out of four, as would typically be the case in reliability analyses of process-oriented assessments (Cools et al., 2009). To determine inter-rater reliability of performance level scoring, intra-class correlations (2, k, consistency) of performance scores were calculated for PL and video assessments across all assessors. ICC’s were calculated
across all 47 assessors for all performances assessed (64 performances) and also for each skill type independently (16 performances of each skill type). Significant, strong positive ICC’s allowed the use of mean scores for each performance across all assessors to determine the validity of assessments in each presentation medium using Spearman correlations with true performance scores. An interesting corollary of performance score analyses was the ability to examine performance score level reliability and how it differed from accuracy at the criterion level.

Table 5. Intra-class correlations and 95% confidence intervals for performance score level comparisons for all, and individual skill types, as well as Spearman correlations between mean performance scores and expert ratings across all assessments (*p<0.05)

<table>
<thead>
<tr>
<th></th>
<th>Point-light</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N ICC</td>
<td>95% CI</td>
</tr>
<tr>
<td>All Assessments:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall ICC (64 performances)</td>
<td>47 0.93* 0.90-0.96</td>
<td>0.89* 0.84 - 0.93</td>
</tr>
<tr>
<td>Correlation with expert-rating (64 performances)</td>
<td>47 0.55*</td>
<td>0.83*</td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick ICC (16 performances)</td>
<td>47 0.87* 0.81 - 0.91</td>
<td>0.86* 0.79 - 0.91</td>
</tr>
<tr>
<td>Throw ICC (16 performances)</td>
<td>47 0.86* 0.80 - 0.91</td>
<td>0.78* 0.68 - 0.86</td>
</tr>
<tr>
<td>Hop ICC (16 performances)</td>
<td>47 0.80* 0.72 - 0.87</td>
<td>0.85* 0.78 - 0.90</td>
</tr>
<tr>
<td>Jump ICC (16 performances)</td>
<td>47 0.65* 0.48 - 0.77</td>
<td>0.80* 0.71 - 0.87</td>
</tr>
</tbody>
</table>

ICC’s for performance scoring reliability across assessors were found to be 0.933 (p<0.001) and 0.892 (p<0.001) for PL and video assessments respectively (Table 5). Bivariate correlations of the average score assessors gave to each performance and respective expert score reported Spearman’s Rho to be 0.554 (p<0.001) and 0.826 (p<0.001) for PL displays and video respectively. When considered for each skill, ICC’s were highest for the kick in both presentation styles. Strong positive correlations were found across all skills in both presentation styles with the exception of PL assessments of jumps. The lower reliability in assessments of jump performances was likely due to performances being skewed towards the upper end of proficiency resulting in less variation within the assessment set which can lower ICC’s.
The absolute distance of each performance score from expert ratings was also calculated as a secondary performance level accuracy measure. Wilcoxon signed-rank tests were used to determine accuracy of the performance scores was significantly different across presentation styles. Wilcoxon signed-rank tests showed that performance scores in PL assessments were significantly further from true performance scores than those from video assessments (Table 6), which supports the weaker Spearman correlations of PL performance scores with expert ratings. This was observed across all assessments and for individual skills.

Table 6. Summed absolute distance from true performance scores of video and PL performance scores and Wilcoxon signed-rank tests of differences between presentation styles.

<table>
<thead>
<tr>
<th></th>
<th>Mean distance from expert ratings</th>
<th>Wilcoxon Signed-rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video (SD)</td>
<td>Point-light (SD)</td>
</tr>
<tr>
<td>All Assessments:</td>
<td>55.57 (13.60)</td>
<td>73.26 (13.56)</td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick</td>
<td>11.93 (3.26)</td>
<td>13.84 (3.84)</td>
</tr>
<tr>
<td>Throw</td>
<td>13.64 (4.91)</td>
<td>16.09 (3.94)</td>
</tr>
<tr>
<td>Hop</td>
<td>19.04 (6.40)</td>
<td>26.91 (7.10)</td>
</tr>
<tr>
<td>Jump</td>
<td>12.82 (5.81)</td>
<td>18.96 (4.87)</td>
</tr>
</tbody>
</table>

Discussion

The current study sought to investigate the use of PL displays generated by the Microsoft® Kinect® as an alternative to traditional video presentation for FMS assessment, to establish whether data captured by the Kinect® is suitable for FMS assessment. Contrary to our primary hypothesis, proficiency criteria identification was inconsistent across presentation styles evidenced by moderate absolute agreement (67.97%) between scores from PL and video assessments. Subsequent analyses of scoring accuracy for each presentation style independently, identified scoring patterns that may explain the lower than expected agreement figures. Scoring accuracy in video and PL was also moderate (69.92% and 61.02% respectively) suggesting that assessors had difficulty in accurately perceiving proficiency.
criteria in both display styles. Wilcoxon signed-rank tests revealed that assessors were significantly less accurate at the criterion level when making assessments from PL displays than video displays.

Moderate observed accuracy levels observed in both styles may be attributed to the moderate experience level of assessors. Interestingly, when scoring reliability was analysed in terms of performance scores, as is typically the case in process-oriented assessments, assessors exhibited high reliability. This suggests that high inter-rater reliability in FMS assessments can be observed in the presence of low criterion scoring accuracy. It is likely that the low number of criteria within process-oriented assessments prevent a sufficient range of scoring possibilities to elicit low inter-rater reliability. Many studies of FMS assessment inter-rater reliability neglect to report scoring accuracy, and instead employ highly trained assessors, assuming that assessors remain accurate throughout assessments. Whilst this observation was an interesting corollary of our analyses and should be the focus of future research, comparisons to other studies of assessment reliability must be made with caution given the experience of assessors and modification of assessment tasks within the study.

The observed impact of presentation style may be explained by perceptual processing during performance assessment. It has been established that during performance assessment observers interpret performances by comparing what they observe with a set of internally stored templates of what they expect to observe, based on historical observations of similar skill performances (Eckstein, Ahumada, & Watson, 1997; Gold et al., 2008; Lange & Lappe, 2006). Given that FMS performances are typically observed live or from videos, where the performer is presented as full anatomical figures, assessors’ internally stored observation templates likely depict full-figure representations of skill performances. If assessors were discriminating PL performances based on more familiar full-figure templates, there would be a disparity between what assessors are observing and the templates they are drawing comparisons from. We could therefore expect lower accuracy in PL assessments, especially from assessors with less experience. Whilst we did not expect assessors to elicit a significant accuracy differences in assessments from both presentation styles, FMS assessment is a perceptually complex task, and it may be possible that unfamiliar PL presentations introduced further complexity for inexperienced assessors. The distribution of error in PL assessments at
the criterion level showed that there was no systematic bias towards overscoring or underscoring criterion presence. If PL displays were presenting less information than video displays, we may expect the distribution of errors to tend towards underscoring, which was not observed. Therefore, lower accuracy is likely to be due to difficulties in assessing rather than inconsistencies in the kinematic information presented across displays.

The current study sought to investigate assessment scoring of FMS performances from Kinect® produced PL displays in comparison to video presentations. Agreement between the two styles was lower than expected and significant differences in assessment accuracy were observed across presentation styles. Distributions of error in PL displays at the criterion level suggests that experience levels of assessors may have been a larger factor in the observed accuracy levels rather than discrepancies in the kinematic information presented, however this is difficult to ascertain. The non-significant differences in accuracy figures across styles observed from assessors with high self-reported experience also suggests that assessors familiar with observation of FMS are robust to PL presentations of FMS performances. Results therefore show that digitised kinematic data as provided by depth cameras such as the Kinect® sensor may be appropriate for FMS assessment when assessors are experienced in FMS observation, however further research is required to establish this. Although accuracy levels were comparable between video and PL when assessments were made by assessors of greater self-reported experience, accuracy levels were concerning in both conditions. Accuracy of criterion identification during real-time FMS assessment should be a focus of future research to identify training thresholds required for FMS assessment tools. Low agreement is likely to be contributed to by variability and inaccuracies in assessment in combination with possible discrepancies in the information provided by each display. As computer vision capabilities advance towards sophisticated and reliable extraction of 3-dimensional kinematic data from 2-dimensional videos such as DensePose (Güler, Neverova, & Kokkinos, 2018) some of the limitations that impact depth sensors may be overcome allowing greater freedom in FMS capture for computational assessment. In light of the emergence of sophisticated extraction of kinematic features from video, it is less important that future investigations address the limitations of depth-capture systems for FMS assessment, and instead, develop a greater understanding of human assessment processes to inform computational approaches.
References


Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46*(5), 423-429. doi:[http://dx.doi.org/10.1016/0895-4356(93)90018-V](http://dx.doi.org/10.1016/0895-4356(93)90018-V)


Holfelder, B., & Schott, N. (2014). Relationship of fundamental movement skills and physical activity in children and adolescents: A systematic review. *Psychology of Sport and Exercise, 15*(4), 382-391. doi:[http://dx.doi.org/10.1016/j.psychsport.2014.03.005](http://dx.doi.org/10.1016/j.psychsport.2014.03.005)


Can proficiency criteria be accurately identified during real-time fundamental movement skill assessment?

This manuscript was submitted for publication in Research Quarterly for Exercise and Sport in November, 2018 and is currently under review

Foreword

The results presented in Chapter Five were somewhat surprising, particularly with regards to the accuracy of FMS assessment across criterion and performance level scoring interpretation and inconsistencies in scoring between presentation styles. It was difficult to ascertain whether the inconsistencies in assessment were due to assessor inexperience or the impact of presentation styles. Thus, it was important to further investigate the disparity between criterion and performance score accuracy and whether low accuracy at the criterion level persisted within assessors of high experience. It was also important to investigate whether presentation styles had a significant impact on FMS assessments to establish digitised kinematic data from depth cameras as an alternative to video capture. This paper provides a comprehensive investigation of FMS assessment accuracy across levels of scoring interpretation, and compares real-time assessment accuracy between paediatric professionals and primary teachers in both video and PL displays.
Abstract

Introduction

Fundamental movement skill (FMS) assessors in education environments rely upon real-time FMS assessment due to the time and resource efficiency the technique offers. It has recently been suggested that the recognition of proficiency criteria during real-time process-oriented FMS assessment may be problematic. Studies into the reliability of process-oriented FMS assessment commonly neglect to consider the accuracy of identifying proficiency criteria. Given that a key strength of process-oriented FMS assessment techniques is the identification of skill deficiencies based on proficiency criteria scores, the current study aimed to further understand assessors’ ability to recognise proficiency criteria during real-time FMS assessment and the impact of assessor experience on assessment accuracy.

Methods

10 primary teachers, and 7 paediatric professionals assessed 10 performances of four FMS presented in videos and point-light displays using process-oriented FMS assessment techniques.

Results

Moderate accuracy in identifying proficiency criteria was reported in both assessor groups. In contrast, reliability of performance scoring was high in both groups. Some proficiency criteria were more difficult to assess than others, with reported accuracy as low as 36%.

Conclusion

The study reinforces the difficulty of observing proficiency criteria during real-time FMS assessment regardless of assessor experience. Results also suggest that performance level reliability may overstate assessor’s ability to accurately score proficiency criteria, which is an important consideration for test administrators. A greater understanding of perceptual processing during FMS assessment may shed light on the information employed by assessors during proficiency assessments and provide a foundation for assessment developments.
Introduction
Identifying and intervening upon movement deficiencies throughout childhood continues to be valuable in helping establish positive lifelong health behaviours. Thus, fundamental movement skill (FMS) assessment has an important role to play in both understanding movement proficiency and informing movement intervention (Barnett et al., 2009; Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012; Holfelder & Schott, 2014; Logan, Robinson, Rudisill, Wadsworth, & Morera, 2012). The responsibility for FMS assessment and monitoring within Australia falls mostly upon early childhood and primary teachers as a curricular requirement (Lander, Barnett, Brown, & Telford, 2015). There are significant barriers to good quality assessment administration in primary educational environments including a lack of teacher training, time, and resources (Haynes & Miller, 2014; Lander, Barnett, Brown, & Telford, 2014; Netelenbos, 2005; Whipp, Hutton, Grove, & Jackson, 2011). A lack of training, resulting in low domain knowledge and confidence, impacts the frequency of assessment administration in schools and the likelihood of teachers using ideal assessment tools and techniques (Lander, Eather, Morgan, Salmon, & Barnett, 2017). Time constraints and large class sizes also dictate that teachers typically assess FMS performances in real-time in the field, which is rarely replicated in research studies (Lander et al., 2014; Lander et al., 2015). Whilst primary teachers are important proponents of FMS assessment and monitoring (Lander et al., 2015), research studies have neglected to investigate the accuracy and reliability of FMS assessments made by inexperienced assessors such as primary teachers, in contrast to experienced assessors.

The reliability of field-based FMS assessment structures is commonly addressed by providing experienced assessors with a significant amount of further training prior to the administration of assessment items (Barnett et al., 2009; Barnett, Minto, Lander, & Hardy, 2014; Valentini, 2012). Training typically involves assessors undertaking a high number of observations and assessments of video performances, which are compared with expert ratings until an acceptable accuracy threshold is reached (Barnett et al., 2009; Barnett et al., 2014). It is well established that teachers are not afforded a comparative level of FMS assessment training as assessors employed within research studies (Lander et al., 2015). Teachers also rarely have the same time allowances or personnel resources to administer FMS assessments in
alignment with standardised protocols. The background of assessors in education environments is likely to impact the reliability and accuracy of FMS assessment and warrants further investigation.

In research and educational environments, people seeking to determine FMS proficiency tend to employ process-oriented assessment tools rather than simpler and faster product-oriented assessments (Lander et al., 2015). Process-oriented assessment techniques assess skill performances by judging the presence or absence of criteria pertaining to proficient movement patterns (Barnett et al., 2014; Burton & Miller, 1998; Cools, De Martelaer, Samaey, & Andries, 2009; Hands, 2002). In this context, a performance describes a single repetition of an FMS. Process-oriented assessment tends to be preferred as it allows administrators to use assessment results to identify specific proficiency criteria or skill components that need development or monitoring (Barnett et al., 2014; Hands, 2002; Lander et al., 2015). Scoring patterns within process-oriented assessment tools are typically analysed using performance scoring (summed number of present criteria for each performance), or standardised scores derived from performance scores of multiple skills (Cools et al., 2009). Reliability is rarely assessed in terms of the recognition of individual proficiency criteria and therefore specific criteria that may require attention are rarely examined (Barnett et al., 2014).

Barnett et al. (2014) recently identified that reliability for some individual components of process-oriented FMS assessments was much lower than the excellent reliability of performance level scores. A reported strength of process-oriented FMS assessments is the ability to identify specific skill deficiencies from criterion level scores (Barnett et al., 2014; Hands, 2002; Lander et al., 2015). However, if the identification of proficiency criteria is unreliable, especially when assessments are made in real-time, it may be problematic to assume that criterion level scoring is reflective of actual skill deficiencies (Lander et al., 2015). Barnett et al. (2014) stressed the importance of understanding the reliability of criterion recognition in process-oriented assessments, so that test administrators can accurately identify skill features that require intervention. The study suggested that the identification and scoring of some skill components may be problematic for assessors due to differences in criterion interpretation. Therefore, further investigations of criterion level scoring patterns are warranted. Barnett et al. (2014) also employed highly trained assessors that reached a
high accuracy benchmark prior to testing and it could be expected that difficulties in criterion identification may be exacerbated when FMS assessments are administered by inexperienced assessors such as teachers. It is therefore important to understand assessment scoring of inexperienced assessors when using process-oriented FMS assessment techniques and how they compare with experienced assessors.

The current study was undertaken within a wider research project investigating the use of consumer-level depth cameras and digitised kinematic data in FMS assessment processes. If appropriate capture systems are established, digitised kinematic data may allow computational assessment of FMS, reducing the training and time burden on primary teachers (Ward, Thornton, Lay, & Rosenberg, 2017). The development of process-oriented FMS assessments towards the use of digitised kinematic data requires further understanding of process-oriented FMS assessment scoring by inexperienced and experienced assessors to assist with data reduction in computational approaches (Ward et al., 2017). Additionally, if visual representations of digital kinematic data such as point-light (PL) displays are appropriate for human assessment, digital kinematic data could allow more efficient data storage and sharing, as well as anonymity which can overcome ethical concerns of video capture in educational environments (Barnett et al., 2014; Ward et al., 2017). Therefore, it was also important to establish whether digitised kinematic data presented as point-light (PL) displays is appropriate for FMS assessment across levels of assessor expertise.

The current study aimed to compare criterion and performance level scoring patterns of primary teachers and paediatric professionals assessing video and PL presentations of FMS using a process-oriented FMS assessment. Due their high level of training and experience, it was hypothesised that paediatric professionals would report comparable video and PL scoring accuracy, however, primary teachers would be less accurate in PL than video. We also hypothesised that paediatric professionals would have greater accuracy in identifying proficiency criteria than primary teachers. We further hypothesised that lower accuracy would be observed for criteria that are more subjective in nature or may be interpreted differently by different assessors.
Methods

Participants
A total of 17 participants were recruited to the current study; seven paediatric movement professionals (age $28.7 \pm 6.55$ years) and ten primary school teachers (age $34.5 \pm 13.5$ years). The paediatric movement professionals were considered experienced assessors due to their high level of training, and regular observation and assessment of paediatric movement within structured environments, such as intervention programs. All paediatric movement professionals had completed bachelor’s degrees in sport science and administered 2 years of paediatric movement intervention programs as minimum. Three were also accredited exercise physiologists, and one holds a PhD in the field of paediatric motor development. All paediatric professionals had experience with process-oriented FMS assessment tools. Primary teachers were included in the study as inexperienced assessors. Although relied upon for FMS assessment within primary schools, teachers receive little training and professional development in the area (Lander et al., 2014). Teachers had an average of $9.7 \pm 12.1$ years of early childhood teaching experience, and only two had partaken in further FMS assessment training since completing their initial teaching degree.

Procedure
The FMS performances to be included in the assessment task were collected using the Microsoft® Kinect® system, which at the time of data collection provided the most efficient extraction of kinematic data in a cost effective and portable system, which is important for the ecological validity of FMS capture. Sixteen primary school aged children (age $8.2 \pm 2.2$ years) performed ten trials of four FMS (horizontal jump, hop, kick, throw), which were recorded with Microsoft® Kinect® camera using in-house software. FMS capture took place in an indoor movement lab devoid of natural light to minimise any possible interference with the infrared sensors of the Kinect®. Within the space a Kinect® camera was placed 50cm above the ground, 2.5 metres in front of the participant at a 22-degree angle to the sagittal plane, a viewing angle common in FMS assessment. The camera position ensured that performers remained within both the frame and the capture space for the Kinect® depth sensor for the entirety of performances (Jungong, Ling, Dong, & Shotton, 2013). Participants
were instructed to perform each skill following the protocols outlined in Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000), as the proficiency criteria within the TGMD-2 would be employed to provide a structure for scoring comparisons throughout the study. Due to the Kinect® depth sensor’s maximum capture distance of 3.5m, the kick and hop tasks were modified from the standardised TGMD-2 instructions (Ulrich, 2000). Instead of 3 hops on each leg, participants performed a single hop for distance on their preferred leg. The run-up distance for the kick was reduced to 1.5 metres. Following the Kinect® collection of the FMS, participants were administered the Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000) by an experienced movement assessor to determine each child’s level of motor proficiency and ensure a range of proficiency levels were employed throughout the study. RGB video and digitised kinematic data captured by the Kinect® were extracted and digitised kinematic data for each performance was visually represented as a PL figure, which was spatially and temporally aligned with the respective video performance. Ten trials of each FMS were then selected for the FMS assessment task.

The subsequent FMS assessment task required primary teachers and paediatric professionals to assess the proficiency of FMS performances presented in PL and video displays using process-oriented FMS assessment techniques. Assessors were required to evaluate performances of four FMS (jump, hop, throw, kick), using proficiency criteria derived from the TGMD-2 Ulrich (2000), which was employed to provide a structure for scoring comparisons. The proficiency criteria for the hop was adapted to reflect the change in task due to the Kinect® capture environment, however, we expected the adaptation would have little impact on the study aim of comparing assessment scoring accuracy across levels of analyses. All other criteria remained consistent with those presented in the TGMD-2 (Ulrich, 2000). Gaze behaviours during FMS assessment were also collected for subsequent analyses, thus, participants maintained a stable head position within a chin rest during data collection to ensure accurate gaze behaviour recording.

Participants were initially familiarised in the presentation of skills, by observing two performances of each skill in each presentation style. Participants were then given as much time as necessary to acquaint themselves with the proficiency criteria for each skill and were able to seek clarification if required. Following this, participants practiced viewing and
assessing three performances of each skill in each presentation style. This concluded the familiarisation procedure. No additional FMS assessment training was provided to participants to ensure that assessment scoring was reflective of assessor historical training and experience.

The formal FMS assessment task required participants to assess ten performances of each FMS (jump, hop, throw, kick). Prior to the presentation of each performance, the relevant proficiency criteria were displayed for ten seconds for assessors to review. Assessors viewed each performance once in real-time, as would be the case in a field-based environment. Following the presentation of the skill performance, proficiency criteria were again displayed accompanied by a yes (performed) and no (not-performed) check-box for each criterion. Assessors were given 20 seconds to select the relevant check-boxes based on their observation. Performances were grouped by skill and presentation style, resulting in 8 assessment blocks. The video and PL block of each skill were paired, and participants viewed them successively, with the order counter-balanced between participants. Presentation order of clips within each block remained consistent between paired video and PL blocks, but was randomised between participants and skill types. Following each block participants were given the opportunity to have a break and remove their head from the chin rest if required.

Data Analysis

Prior to data analysis all criterion level scores were coded as 1 (present) or 0 (absent) and de-randomised for comparisons. For clarity throughout the remainder of the paper, data treated at the ‘criterion level’ refers to analyses of the binary present/absent (1/0) scoring of individual proficiency criteria during assessment. Data treated at the ‘performance level’ refers to analyses of performance scores, calculated by summing the number of individual criteria recorded as present for each performance to give a score out of four.

Percentage agreement of criterion level scoring across presentation styles was calculated as the number of times assessors made the same criterion presence/absence judgement across both video and PLD assessments of the same performance, divided by the total number of criteria assessed. Percentage accuracy for each presentation style was calculated as the
number of times assessors made the same criterion presence/absence judgement as the respective true criterion score, divided by the total number of criteria assessed. True assessment scores for accuracy analyses were determined by a two-expert assessor consensus. The protocol required two experienced assessors to independently score criterion presence across repeated viewings of each performance to ensure maximum accuracy, before reviewing disagreements until a consensus was reached for each performance. Both percentage agreement and percentage accuracy at the criterion level were averaged across each assessor group within each presentation style for all assessments, and also for each skill type. Analyses of differences between assessor groups for both percentage agreement and percentage accuracy were conducted using Mann-Whitney U tests. Accuracy differences between presentation styles within each assessor group were analysed using Wilcoxon signed-ranks tests.

The number of present criteria for each performance were then summed to yield performance scores. Reliability of performance level scoring was analysed for all assessments and for the four FMS across each assessor group and presentation style using intra-class correlations (2,k). To determine the level to which performance scores made by assessors correlated with true scores, Spearman correlations were calculated between true performance scores and the median score given to each performance across assessor groups in each presentation style.

To analyse assessment patterns for individual proficiency criteria, the mean accuracy across assessor groups was determined for each individual criterion (i.e. average accuracy in assessing the first criterion in jump performances). Accuracy differences between assessor groups for individual criteria were analysed using Mann-Whitney U tests. Accuracy differences for individual criteria between presentation styles within each assessor group were analysed using Wilcoxon signed-ranks tests. These analyses allowed us to determine whether the accuracy of the underlying constructs of the assessment were representative of the final assessment score and also whether specific components were more or less difficult to assess in each presentation style.
Results

Both paediatric professionals and primary teachers revealed moderate agreement between video and PL scores at the criterion level (Table 1). Acceptable agreement is typically considered to be 80% and both groups fell below this figure (McHugh, 2012). Paediatric professionals had higher agreement between presentation styles than primary teachers, however Mann-Whitney U tests revealed that the difference was not significant. Paediatric professionals reported the lowest agreement for hop assessment, whilst primary teachers revealed the lowest agreement for throw assessments.

Table 1. Descriptive statistics of criterion level agreement figures for paediatric professionals and primary teachers across video and point-light FMS assessments

<table>
<thead>
<tr>
<th></th>
<th>Primary Teachers</th>
<th>Paediatric Professionals</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Assessments:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Agreement (%)</td>
<td>76.16 (6.19)</td>
<td>71.45 (8.41)</td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick assessments</td>
<td>79.64 (12.45)</td>
<td>76.67 (7.50)</td>
</tr>
<tr>
<td>Hop assessments</td>
<td>67.14 (7.28)</td>
<td>70.28 (6.67)</td>
</tr>
<tr>
<td>Throw assessments</td>
<td>80.71 (4.94)*</td>
<td>65.28 (8.41)*</td>
</tr>
<tr>
<td>Jump assessments</td>
<td>77.14 (7.56)</td>
<td>73.61 (10.98)</td>
</tr>
</tbody>
</table>

* Denotes significant difference in mean agreement between groups (p < 0.05)

Accuracy scores were moderate for both groups with paediatric professionals correctly scoring proficiency criterion presence 74.73% and 73.48% of the time for video and PL respectively, and primary teachers correctly scoring proficiency criterion presence 69.58% and 69.31% of the time for video and PL respectively (Table 2). Both groups reported the highest criterion level accuracy for jump assessments across video and PL displays. Mann-Whitney U tests revealed no significant differences in assessment accuracy between assessor groups for all assessments, or for individual FMS. Wilcoxon signed-rank tests revealed no significant differences in accuracy between video and PL assessments within assessor groups.

Intraclass correlation coefficients revealed high reliability across both assessor groups in each of the presentation styles when performance scores were considered (Table 3). When the
median scores for individual performances were correlated with the respective true score, analyses revealed significant strong correlations between assessor scores and true scores across both assessor groups and presentation styles (Table 3).

Table 3. Mean criterion level accuracy figures for paediatric professionals and primary teachers across video and point-light FMS assessments

<table>
<thead>
<tr>
<th></th>
<th>Mean assessment accuracy</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video</td>
<td>Point-light</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% (SD)</td>
<td>% (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paediatric Professionals</td>
<td>74.73 (7.52)</td>
<td>69.37 (10.00)</td>
<td>73.39 (4.58)</td>
<td>69.24 (7.14)</td>
<td></td>
</tr>
<tr>
<td>Primary Teachers</td>
<td>69.37 (10.00)</td>
<td>73.39 (4.58)</td>
<td>69.24 (7.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Assessments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick assessments</td>
<td>72.86 (8.83)</td>
<td>65.83 (13.17)</td>
<td>71.07 (7.48)</td>
<td>61.39 (14.31)</td>
<td></td>
</tr>
<tr>
<td>Throw assessments</td>
<td>73.21 (11.34)</td>
<td>63.61 (19.93)</td>
<td>73.21 (13.36)</td>
<td>62.50 (11.46)</td>
<td></td>
</tr>
<tr>
<td>Hop assessments</td>
<td>74.29 (13.28)</td>
<td>72.22 (9.97)</td>
<td>68.57 (9.88)</td>
<td>73.61 (9.85)</td>
<td></td>
</tr>
<tr>
<td>Jump assessments</td>
<td>78.57 (11.18)</td>
<td>76.67 (14.09)</td>
<td>81.07 (9.98)</td>
<td>79.72 (9.47)</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes significant difference in mean accuracy between groups (p < 0.05)
* Denotes significant difference in mean accuracy between presentation styles, within an assessor group (p < 0.05)

Table 2. Intraclass correlation coefficients (ICC) and 95% confidence intervals (CI) of performance level scoring across assessor groups, and Spearman correlations (r_s) of median performance scores with true performance scores

<table>
<thead>
<tr>
<th></th>
<th>Video</th>
<th></th>
<th>Point Light</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pediatr.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professionals</td>
<td>Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Assessments:</td>
<td></td>
<td>Overall ICC</td>
<td>0.88**</td>
<td>0.84**</td>
<td>0.87**</td>
<td>0.85**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(CI 0.80-0.93)</td>
<td>(CI 0.75-0.90)</td>
<td>(CI 0.79-0.93)</td>
<td>(CI 0.77-0.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r_s</td>
<td>0.75**</td>
<td>0.73**</td>
<td>0.79**</td>
<td>0.67**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Skills:</td>
<td></td>
<td>Kick ICC</td>
<td>0.92**</td>
<td>0.87**</td>
<td>0.92**</td>
<td>0.86**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(CI 0.80-0.98)</td>
<td>(CI 0.71-0.96)</td>
<td>(CI 0.80-0.98)</td>
<td>(CI 0.69-0.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Throw ICC</td>
<td>0.92**</td>
<td>0.89**</td>
<td>0.94**</td>
<td>0.90**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(CI 0.80-0.98)</td>
<td>(CI 0.74-0.97)</td>
<td>(CI 0.86-0.98)</td>
<td>(CI 0.78-0.97)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hop ICC</td>
<td>0.86**</td>
<td>0.80**</td>
<td>0.74**</td>
<td>0.72**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(CI 0.67-0.96)</td>
<td>(CI 0.53-0.91)</td>
<td>(CI 0.39-0.93)</td>
<td>(CI 0.35-0.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jump ICC</td>
<td>0.75**</td>
<td>0.59*</td>
<td>0.86**</td>
<td>0.71**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(CI 0.42-0.93)</td>
<td>(CI 0.05-0.88)</td>
<td>(CI 0.68-0.96)</td>
<td>(CI 0.34-0.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05
** p < 0.01
When the average accuracy of scoring individual proficiency criteria was calculated across groups, large variability in accuracy was observed in both paediatric professional and primary teachers (Table 4). The greatest criterion accuracy range within a skill for paediatric professionals was 37% observed in video kick assessments. The greatest range observed for primary teachers was 46%, which occurred in PLD throw assessments. The lowest range for paediatric professionals and primary teachers was 20% and 17% respectively. The lowest range for paediatric professionals and primary teachers were observed in PLD kick and video hop assessments respectively. Paediatric professionals were significantly more accurate than primary teachers in both presentation styles for assessments of Kick Criterion 3 (Non-kicking foot placed even with or slightly in back of ball). Primary teachers were significantly more accurate than paediatric professionals in PLD assessments for Hop Criterion 3 (Foot of non-support leg remains behind body). No other significant differences between groups were observed for accuracy in identifying individual criteria.
Table 4. Mean accuracy across assessor groups for identification of individual proficiency criteria in video and point-light presentation styles

<table>
<thead>
<tr>
<th></th>
<th>Mean Accuracy</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video % (SD)</td>
<td>Primary Teachers</td>
<td>Point Light % (SD)</td>
<td>Primary Teachers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paediatric Professionals</td>
<td>Primary Teachers</td>
<td>Paediatric Professionals</td>
<td>Primary Teachers</td>
<td></td>
</tr>
<tr>
<td><strong>Kick</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Rapid and continuous approach to the ball</td>
<td>58.57 (24.78)</td>
<td>68.89 (20.28)</td>
<td>64.29 (22.25)</td>
<td>61.11 (23.69)</td>
<td></td>
</tr>
<tr>
<td>2) An elongated stride or leap immediately prior to ball contact</td>
<td>70.00 (8.16)</td>
<td>65.56 (15.09)</td>
<td>65.71 (5.35)</td>
<td>68.89 (20.88)</td>
<td></td>
</tr>
<tr>
<td>3) Non-kicking foot placed even with or slightly in back of ball</td>
<td>67.14 (17.99)*</td>
<td>45.56 13.33)**^</td>
<td>70.00 (15.28)*</td>
<td>36.67 (12.25)**^</td>
<td></td>
</tr>
<tr>
<td>4) Kicks ball with instep of preferred foot (shoelaces) or toe</td>
<td>95.71 (7.87)</td>
<td>83.33 (20.00)</td>
<td>84.29 (37.35)</td>
<td>78.89 (20.88)</td>
<td></td>
</tr>
<tr>
<td><strong>Throw</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Windup is initiated by downward movement of hand/arm</td>
<td>57.14 (33.52)</td>
<td>48.89 36.55)</td>
<td>52.86 (39.04)</td>
<td>34.44 (37.45)</td>
<td></td>
</tr>
<tr>
<td>2) Rotates hip and shoulders to a point where the non-throwing side faces the direction of the throw</td>
<td>72.86 (14.96)</td>
<td>65.56 (20.07)</td>
<td>74.29 (13.97)</td>
<td>80.00 (16.58)</td>
<td></td>
</tr>
<tr>
<td>3) Weight is transferred by stepping with the foot opposite the throwing hand</td>
<td>82.86 (14.96)^</td>
<td>75.56 (25.55)</td>
<td>88.57 (17.73)^</td>
<td>75.56 (18.78)</td>
<td></td>
</tr>
<tr>
<td>4) Follow-through beyond ball release diagonally across the body towards non-preferred side</td>
<td>80.00 (14.14)</td>
<td>64.44 (16.67)</td>
<td>77.14 (17.04)</td>
<td>60.00 (14.14)</td>
<td></td>
</tr>
</tbody>
</table>
### Hop

<table>
<thead>
<tr>
<th>Criterion Description</th>
<th>Assessor Group 1</th>
<th>Assessor Group 2</th>
<th>Assessor Group 3</th>
<th>Assessor Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Non-support leg swings forward in pendular fashion to produce force</td>
<td>55.71 (16.18)</td>
<td>61.11 (16.91)</td>
<td>52.86 (26.28)</td>
<td>60.00 (15.81)</td>
</tr>
<tr>
<td>2) Foot of non-support leg remains behind body</td>
<td>81.43 (15.74)</td>
<td>76.67 (10.00)</td>
<td>62.86 (9.51)*</td>
<td>76.67 (13.23)*</td>
</tr>
<tr>
<td>3) Arms flexed and swing forward to produce force</td>
<td>78.57 (31.32)</td>
<td>72.22 (21.08)</td>
<td>80.00 (20.00)</td>
<td>80.00 (22.36)</td>
</tr>
<tr>
<td>4) Takes off and lands on same foot (adjusted)</td>
<td>81.43 (16.76)</td>
<td>78.89 (16.16)</td>
<td>78.57 (18.64)</td>
<td>77.78 (17.16)</td>
</tr>
</tbody>
</table>

### Jump

<table>
<thead>
<tr>
<th>Criterion Description</th>
<th>Assessor Group 1</th>
<th>Assessor Group 2</th>
<th>Assessor Group 3</th>
<th>Assessor Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Preparatory movement includes flexion of both knees with arms extended behind the body</td>
<td>87.14 (17.04)</td>
<td>81.11 (22.05)</td>
<td>81.43 (15.74)</td>
<td>82.22 (16.41)</td>
</tr>
<tr>
<td>2) Arms extend forcefully forward and upward reaching full extension above the head</td>
<td>64.29 (16.18)</td>
<td>65.56 (21.28)</td>
<td>64.29 (17.18)</td>
<td>66.67 (15.81)</td>
</tr>
<tr>
<td>3) Take off and land on both feet simultaneously</td>
<td>98.57 (3.78)</td>
<td>92.22 (10.93)</td>
<td>100.00 (0.00)</td>
<td>88.89 (19.65)</td>
</tr>
<tr>
<td>4) Arms are thrust downward during landing</td>
<td>64.29 (17.18)</td>
<td>67.78 (19.22)</td>
<td>78.57 (15.74)</td>
<td>81.11 (11.67)</td>
</tr>
</tbody>
</table>

*Denotes significant difference in accuracy between assessor groups ($p < 0.05$)

^Denotes significant difference in accuracy between presentation styles, within an assessor group ($p < 0.05$)
Discussion

The current study sought to compare scoring patterns of paediatric professionals and primary teachers during process-oriented assessments of FMS observed under video and PL conditions. Agreement between video and PL scoring for both assessor groups was moderate, suggesting that inconsistencies in criterion recognition existed across styles. In partial support of our primary hypothesis, no significant differences in assessment accuracy were observed between presentation styles across all assessments, or across the four FMS for either assessor group. Whilst agreement between styles was low, comparable accuracy figures across presentation styles provides evidence that experienced assessors are equally accurate in assessing FMS from digital kinematic data captured with depth cameras as from traditional videos. If moderate agreement was a result of a particular presentation style impacting assessment scoring, we would expect large differences in accuracy figures across styles, however, this was not observed. It is likely that moderate agreement figures are a result of moderate assessor accuracy rather than impact of presentation styles. Thus, further discussion of assessment scoring will be in relation to patterns within assessor groups rather than patterns within individual presentation styles.

Contrary to our hypotheses, paediatric professionals did not report significantly greater mean accuracy than primary teachers. Accuracy in correctly identifying proficiency criteria was moderate for both groups, which was unexpected due to the high level of experience in the paediatric professional group. Paediatric professionals and primary teachers reported good to excellent levels of reliability at the performance score level in both presentation styles. The differences between criterion and performance level scoring observed within the current study reinforce the findings from Barnett et al. (2014) that performance score reliability may not be reflective of assessors ability to correctly identify proficiency criteria presence or absence, and found the disparity to exist across levels of assessor experience. High performance score reliability and strong correlations with true scores suggest that assessors can accurately and reliably classify proficiency levels during real-time process-oriented FMS assessment, however, moderate criterion level scores suggest that the identification of individual criteria may be problematic.
In part, the disparity between performance score reliability and criterion level accuracy was explained when accuracy was analysed for each proficiency criterion individually rather than across all criteria. Average scoring accuracy for individual criteria ranged between 100% and 35%, suggesting that assessors were consistently accurate in the identification of some criteria, but highly inconsistent for others. Large accuracy ranges were observed across both paediatric professionals and primary teachers suggesting that experience had little influence on assessors’ ability to accurately identify all criterion during a single skill observation.

It is possible that inaccurate scoring of individual criteria may be washed out when scores are summed to produce overall scores. In the scenario that a child performed two out of four criteria in a performance and all assessors gave a score of two out of four, assessors were highly accurate and reliable at the performance score level (2/4). However, accuracy at the criterion level may have been as low as 50% if assessors were identifying two different criteria in each case. If accuracy for each criterion was also analysed in this scenario, the accuracy ranges between individual criterion could be as large as 100% (0% for one criterion, 50% for two, and 100% for the last, with an average accuracy of 50%). It is therefore possible that summing criteria scoring and analysing performance scores may mask inaccuracies in criterion level scoring, which provides a plausible explanation for the variation between assessment accuracy at the different levels of analysis observed in the current study. The observed variation between performance scoring accuracy and criterion level accuracy suggests that analysing accuracy at the performance score level may not be indicative of assessor’s ability to accurately perceive proficiency criteria.

The large accuracy ranges observed across individual criteria in the current study go some way to support our final hypothesis and the work of Barnett et al. (2014); who also analysed reliability for individual skill criteria and identified that some skill features may be harder to identify than others. Barnett et al. (2014) suggested that inconsistencies may be due to differences in criterion interpretation between assessors. Lower accuracy was observed in the current study for criteria that could be considered more open to interpretation, such as Kick criteria 1 and 3 (non-kicking foot placed even with or slightly in back of ball), Throw criterion 1, and Hop criterion 1; which are more subjective in comparison to the likes of Jump criterion 3 (take off and land on both feet simultaneously), which exhibited high accuracy.
Additionally, we suggest that large accuracy ranges may be due to limitations in assessor’s ability to visually attend to all proficiency criteria within the time the child performs the movement.

Accuracy figures for the recognition of some skill components were below chance levels (50%), suggesting that assessors may not have been able to consciously attend to some criteria, but relied on ‘filling-in’ information for others (Chong, Familiar, & Shim, 2016). The visual perceptual system often relies on the ‘filling-in’ or interpolation of perceptual information in the case that stimuli fall outside of attended regions of the visual field (Chong et al., 2016; Komatsu, 2006). It may be the case that assessors prioritise attention toward particular proficiency criteria or important kinematic features during the skill performance and rely on overall impression to retrospectively interpolate the presence of other features they could not attend to. If assessors are relying on the filling-in of visual information, an assessment can be reliable and valid for overall proficiency classification, even though assessor’s identification of individual proficiency criteria is inconsistent.

Further development of process-oriented assessment techniques needs to be undertaken to improve their efficiency and the value of assessment outcomes (Furtado & Gallagher, 2012; Giblin, Collins, & Button, 2014; Longmuir et al., 2017). Barnett et al. (2014) suggested that components within process-oriented assessment need to be further specified for assessors to distinguish between proficiency levels within proficiency criteria. A recent paper sought to address this through a hierarchical approach employing decision trees for assessment of skill process-oriented components (Furtado & Gallagher, 2012). Process-oriented assessment techniques must find a balance between having enough assessment components to comprehensively describe proficient movement patterns, yet not so many that assessment in real-time becomes perceptually difficult or unreliable. We believe that a beneficial approach would be to gather a greater understanding of current assessment practices from a visual-perceptual perspective, employing eye-tracking methodologies to identify the observation strategies from spatial and temporal perspectives. If specific skill components that assessors use to assess FMS proficiency are identified, an assessment structure that provides more weighting to these skill components could be developed.
Digitised kinematic data from depth cameras such as the Microsoft® Kinect® are a viable alternative to video presentations when assessments are made by experienced assessors, or assessors with high exposure to paediatric movement. Results from this study also highlighted important factors to consider when interpreting process-oriented assessment scores. The current study reinforces some of the challenges and issues involved in current process oriented assessment strategies (Barnett et al., 2014), and provides additional evidence that difficulties in accurately scoring proficiency criteria may not be represented by performance level reliability figures. Our results suggest that both experienced and inexperienced assessors may not use all criteria presented in process-oriented assessment structures to assess proficiency, suggesting more efficient strategies requiring fewer skill features may be employed. As the extraction of digital kinematic data from video becomes more sophisticated in algorithms such as DensePose (Güler, Neverova, & Kokkinos, 2018), the limitations of depth capture systems will be overcome allowing more freedom in FMS capture. Therefore, in developing computational FMS assessment techniques that mimic human proficiency assessment, it is important to prioritise understanding human information processing during FMS assessment rather than specific capture devices. Further investigation into the visual information people use to assess FMS proficiency may identify key skill features required for assessment, informing improved approaches to current assessment strategies.
References


Holfelder, B., & Schott, N. (2014). Relationship of fundamental movement skills and physical activity in children and adolescents: A systematic review. *Psychology of Sport and Exercise, 15*(4), 382-391. doi: [http://dx.doi.org/10.1016/j.psychsport.2014.03.005](http://dx.doi.org/10.1016/j.psychsport.2014.03.005)


CHAPTER 7

Observation strategies employed during process-oriented FMS assessment:
Comparisons between experienced and inexperienced assessors

This manuscript was submitted for publication in Psychology of Sport and Exercise in December, 2018

Foreword

Given the unexpected results observed in Chapter Five, further investigation was required into process-oriented FMS assessment techniques from an observational perspective, including whether assessor adopted an observational strategy independent of proficiency criteria. In Chapter Five assessors were found to experience substantial challenges in accurately identifying all proficiency criteria during real-time FMS assessment. In Chapter Six these challenges were found to persist across both inexperienced and experienced assessors. To inform computational assessment that mimic human experts and targeted assessor training, greater understanding is required of the kinematic information experienced assessors employ during FMS assessments and whether they differ from inexperienced assessors. Given that no accuracy differences between experienced and experienced assessors were observed in Chapter Six, separation of assessors by their accuracy was adopted to determine whether highly accurate assessors adopted different visual search strategies from less accurate assessors. This investigation employed eye-movement recordings during real-time process-oriented FMS assessment to investigate and compare the visual search strategies employed by experienced and inexperienced, and high and low-accuracy assessors.
Abstract

Introduction

There is currently little understanding of the perceptual processes involved in real-time FMS assessment, particularly across levels of experience or assessor accuracy. Greater knowledge in how assessors judge FMS proficiency may assist assessor training and targeted computational approaches to FMS assessment. This study employed eye-movement recording during real-time FMS assessment to investigate the kinematic information employed by FMS assessors.

Methods

10 primary teachers, and 7 paediatric professionals each assessed 10 performances of four FMSs presented in randomly ordered videos and PL displays using process-oriented FMS assessment techniques. Gaze behaviours during FMS assessments were recorded using an EyeLink 1000 (SR Research) eye-tracker.

Results

Visual search strategies observed show that assessors do not consciously attend to all anatomical locations outlined proficiency criteria. No differences in visual search behaviours were observed between inexperienced and experienced assessors or high and low accuracy assessors.

Conclusion

Assessors tended to be limited in their ability to attend to all proficiency criteria during real-time FMS assessment. Instead, assessors adopted a more efficient search strategy that allowed proficiency assessment by prioritising information-rich anatomical locations. Due to the limitations of human information processing during real-time assessment, more work is required to understand human assessment strategies and the ideal foundation for computational FMS assessment techniques.
Introduction

The responsibility for FMS assessment and monitoring within Australia falls mostly upon early childhood and primary teachers (Lander, Barnett, Brown, & Telford, 2014; Lander, Morgan, Salmon, & Barnett, 2016). Typically, teachers rely on real-time process-oriented assessments (concerned with motor skill processes rather than outcomes), allowing them to identify proficient movement patterns within the limited class time they are provided to conduct movement assessments (Barnett, Minto, Lander, & Hardy, 2014; Cools, De Martelaer, Samaey, & Andries, 2009; Lander, Barnett, Brown, & Telford, 2015). In addition to limited time, lack of training and domain knowledge are further barriers teachers experience when conducting quality movement assessment (Haynes & Miller, 2014; Lander et al., 2014; Netelenbos, 2005; Whipp, Hutton, Grove, & Jackson, 2011). These barriers to administering process-oriented assessment tools, have resulted in calls to provide more effective training for teachers to reliably administer FMS assessments. Alternatively, rather than investing time to assessor training, there may benefits in developing computational assessment strategies employing machine learning to mimic expert FMS assessment processes, minimising the time and expertise demands on teachers and enabling standardised and reliable FMS assessment within education environments (Ward, Thornton, Lay, & Rosenberg, 2017).

Machine learning refers to a process whereby a computer learns to identify patterns in data and make predictions with minimal human intervention (Domingos, 2012). For FMS assessment this would encompass an algorithm that could mimic expert human assessment, analysing kinematic features of FMS performances to make predictions of proficiency. Initially however, the algorithm must be ‘trained’ to make accurate predictions based upon a large sample of input data, in a similar way that a human must be taught what a proficient jump looks like by watching many jump performances and being told which ones are proficient and which ones are not. To ensure accuracy in the final prediction, machine learning also relies upon only relevant information being supplied during the ‘training’ stage (Lewandowski, Makris, & Nebel, 2009; Roweis & Saul, 2000), much like accurate and efficient FMS assessment relies upon assessors’ ability to visually detect key kinematic features that represent proficient performance and filter out unnecessary features (Giblin, Farrow, Reid, Ball, & Abernethy, 2015; Hernández, Romero, Vaíllo, & Campo, 2006). Investigating visual
search strategies employed by experienced and inexperienced FMS assessors can assist in identifying kinematic features that can be built into ‘training’ machine learning algorithms for FMS assessments as well as informing targeted assessor training.

Experienced assessors are able to reliably assess the overall proficiency of FMS performances when skills are viewed in real-time, as in educational environments (Barnett et al., 2009; Barnett et al., 2014; Cools et al., 2009; Valentini, 2012). However, there is little understanding of the kinematic information experienced assessors employ to make real-time proficiency judgements, nor the observational strategies they use to extract relevant information. The ability of experienced movement observers to visually detect kinematics pertaining to proficient performance is essential to optimise assessment accuracy and the efficacy of subsequent intervention (Giblin et al., 2015; Hernández et al., 2006). There are no published studies exploring observational expertise or visual search strategies of FMS assessors. It may be assumed that assessors search for the proficiency criteria outlined in process-oriented assessment tools, however, recent studies have suggested that the identification of proficiency criteria may be challenging during real-time FMS assessment (Ward, Thornton, Lay, Chen, & Rosenberg, 2018; Ward, Thornton, Lay, & Rosenberg, 2018) (Chapters 5 & 6) (Barnett et al., 2014). This suggests that assessors may not rely on kinematic information outlined in assessment criteria to make overall proficiency judgements and therefore, assessment criteria may not provide an appropriate foundation for machine learning approaches.

A large number of studies have examined differences in perceptual expertise and information pick-up in experienced and inexperienced coaches and judges, who share similar roles to FMS assessors in regards to skill observation (Imwold & Hoffman, 1983; Leas & Chi, 1993; Sparrow & Sherman, 2001; Ste-Marie, 1999) . These studies however, tend to explore ‘what’ experienced observers can do, rather than ‘how’ they are doing it. Experienced movement observers are able to yield and recall more kinematic information relevant to the task, are more critical in analysis (rather than descriptive) and are able to more accurately anticipate movement outcomes or recall errors than novice observers (Giblin et al., 2015; Woorons, 2001). Through an expert’s superior ability to understand and interpret relevant kinematic information, information processing demands are reported to be reduced, allowing experts
to allocate more attentional resources to performance analysis (Ste-Marie, 1999). Research is yet to consider the observational strategies employed by experienced FMS assessors in yielding the relevant kinematic information during skill observation and how they differ from inexperienced assessors.

The few studies that have explored visual search behaviour of movement observers have employed eye-tracking methodologies to quantify the allocation of conscious attention (Giblin et al., 2015; Hernández et al., 2006; Petrakis, 1987; Waters, Lay, Tidman, & Benjanuvatra, 2014). Using visual search behaviour as a proxy for attentional focus has historically been considered problematic due to the inability to capture allocation of attention to peripheral vision (Abernethy & Zawi, 2007; Giblin et al., 2015). A recent study by Ryu, Abernethy, Mann, Poolton, and Gorman (2013) investigating the use of information within both central and peripheral visual fields was the first to show that the gaze-behaviours tended to be consistent regardless of whether the entire visual field was available or whether observers were restricted to foveal sources of information. Thus, gaze characteristics such as fixation count, duration, location, and temporal sequencing can be reliably used to ascertain visual search strategies and information extraction from static and dynamic environments.

The development of computational approaches to process-oriented FMS assessment requires a greater understanding of human FMS assessment processes including the identification kinematic features critical to proficiency assessment (Ward et al., 2017). Therefore, identifying the kinematic information employed by assessors when observing visual representations of digitised kinematic data, such as point-light (PL) displays becomes important also, as the kinematic information presented in PL displays is consistent with that which can be computationally analysed. Through measurement of visual search behaviours including spatial and temporal gaze characteristics, the current study aimed to investigate the visual search strategies of experienced and inexperienced movement assessors to identify the kinematic information employed during process-oriented FMS assessments. Within the study paediatric professionals were employed as experienced assessors and primary teachers were employed as inexperienced assessors. We hypothesised that paediatric professionals would exhibit more efficient visual search strategies than primary teachers evidenced by fewer fixations of longer duration. We also hypothesised that paediatric professionals would have
more fixations upon proximal anatomical locations than less experienced primary teachers. It was hypothesised that both groups would attend to locations outside of those referred to by proficiency criteria within the relevant temporal phase of each FMS. Finally, we hypothesised that no differences would be observed between the use of video or PL in the assessment of FMS across the entire sample.

**Methods**

**Participants**

A total of 17 participants were recruited to the current study; seven paediatric movement professionals (age 28.7 ± 6.55 years) and ten primary school teachers (age 34.5 ± 13.5 years). The paediatric movement professionals were considered experienced assessors due to their high level of training, and regular observation and assessment of paediatric movement within structured environments, such as intervention programs. All paediatric movement professionals had completed bachelor’s degrees in sport science and had minimum 2 years’ experience in paediatric skill development programs. Three were also accredited exercise physiologists, and one holds a PhD in the field of paediatric motor development. Primary teachers were included in the study as inexperienced assessors. Although FMS assessment in Australia is predominantly administered within primary schools, teachers receive little training and professional development in the area (Lander, Eather, Morgan, Salmon, & Barnett, 2017). Primary teachers had an average of 9.7 ± 12.1 years of early childhood teaching experience, and only two had partaken in further FMS assessment training since completing their initial teaching degree. All components of the study had ethics approval from the University of Western Australia.

**Procedure**

Participant’s eye-movements were recorded by an EyeLink 1000 (SR Research) eye-tracker in a desk mounted configuration. Using pupil centre corneal reflection, and nine calibration points, the EyeLink 1000 recorded monocular gaze at 1000 Hz, with up to .25° accuracy and .01° spatial resolution. Participants were seated at a viewing distance of 70cm with their head firmly secured in a chin rest to prevent movement. Participant information regarding FMS assessment experience and formal assessment training was initially gathered, in addition to
informed consent. Participants were then seated in front of the EyeLink 1000 eye-tracker and the chin rest was adjusted to suit. The eye-tracker was calibrated prior to the FMS assessment task using a nine-point calibration procedure. Calibration was monitored throughout the experiment on a secondary control computer, and re-administered if necessary. Following calibration, participants undertook the FMS assessment task.

The FMS assessment task was developed using Experiment Builder v1.10.1241 (SR Research Ltd, Mississauga, Canada) and presented on a 24” widescreen monitor at a resolution of 1920px x 1080px. Skill performances in the FMS assessment task were sourced from the bank of FMS performances collected using the Microsoft® Kinect® motion capture protocol presented in (Ward, Thornton, Lay, & Rosenberg, 2018). Ten performances of four FMS (Hop, Kick, Jump, Throw) were presented in both video and PL displays within the current study. Video and PL clips of each performance were spatially and temporally aligned to ensure visual stimuli were consistent.

The assessment task required participants to assess the proficiency of PL and video FMS performances using proficiency criteria adapted from the Test of Gross Motor Development-2 (Ulrich, 2000). The TGMD-2 is a validated and widely administered process-oriented FMS assessment tool whereby assessors score skill performances by observing the presence or absence of criteria pertaining to proficient performance of the skill.

Participants were required to assess ten performances of each FMS in each presentation style. Prior to the presentation of each performance, the relevant proficiency criteria were displayed for ten seconds for assessors to review. Following the presentation of the skill performance, proficiency criteria were again displayed accompanied by yes (present) and no (absent) check-boxes for each criterion. Assessors were given 20 seconds to select the relevant check-boxes based on their observation. Performances were grouped by skill and presentation style, resulting in 8 assessment blocks. The video and PL block of each skill was paired, and participants viewed them successively, with the order counter-balanced between participants. Presentation order of clips within each block remained consistent between paired video and PL blocks but was randomised between participants and skill types. Following each block participants were given the opportunity to have a break and remove
Data Analysis

FMS assessment data
FMS assessment was initially analysed for agreement between PL and video criterion scoring. Subsequent analyses were concerned with how accurately assessors identified the presence or absence of assessment criteria in each stimulus. Mann-Whitney U tests were used to determine if significant differences in assessment accuracy existed between paediatric professionals and primary teachers. Wilcoxon signed rank tests were administered to determine whether significant differences existed between video and PL assessment accuracy within each assessor group.

Relative to each presentation style, assessors were also ranked by their assessment accuracy and split at the 50th percentile to yield high and low-accuracy assessment groups for further analyses. Mann-Whitney U tests were used to determine if significant differences in assessment accuracy existed between high and low-accuracy assessment groups.

Eye-tracking data
The number, duration, and location of fixations was extracted from the eye-tracking data. A fixation was recorded when gaze remained within 3 degrees of visual angle for a period of greater than 100ms as this is considered the minimum duration required for conscious processing of visual information (Vickers, 2007).

Mann-Whitney U tests were used to determine if there were significant differences in the number of fixations and the duration of fixations between paediatric professionals and primary teachers. Wilcoxon signed-rank tests were also administered to determine if significant differences in gaze behaviours existed across presentations styles, within each assessor group. For spatio-temporal analyses of visual search, the average and median number of fixations were calculated across the ten assessments of each skill and across assessors, for the spatial location and skill phase in which each fixation was made. Average fixation counts across spatial locations and skill phases provided a description of spatial and
temporal trends in visual search. Median fixation counts were also included to indicate the consistency of spatial and temporal trends within visual search across assessors. A median for fixation count of 1 or more on a location within a skill phase represents participants fixating on the respective location within the same skill phase in the majority of trials. It is important to note that median fixation counts were calculated for individual spatial locations rather than anatomical regions (i.e. torso, upper-body, lower-body) and therefore may underestimate consistency.

Spatial aspects of visual search were categorised in relation to nine spatial locations (Areas of Interest- AOI’s) on the performer. Spatial locations for each video stimulus were developed using the dynamic AOI feature of DataViewer (SR Research). Nine AOI’s (head, hips, torso, shoulders (left and right), knees (left and right), and feet (left and right)) were manually coded for each video stimulus and saved as an AOI template for the respective video. Arms were excluded from the spatial location coding due the large number of frames in which the arms were not clearly visible, caused by the capture rate of the Kinect® and the high velocity of the arms during FMS performances. AOI templates developed based on video stimuli were also used for the corresponding PL stimuli, which was possible due to the spatial and temporal alignment of stimulus pairs during the development of the FMS assessment task.

MATLAB r2017a (MathWorks, Inc) was employed for further treatment of gaze behaviour data under the following conditions:

I. When consecutive fixations were found to remain within a single AOI and separated by less than 150 milliseconds (average blink duration), the two fixations were merged. This can occur when the observer fixates within the AOI before saccading away and back without making a fixation elsewhere, or when the participant blinks.

II. When the centre of a fixation fell outside of all AOI’s, but the centre of that fixation was within 1.5 degrees of visual angle of the nearest AOI, the fixation was coded as being within the nearest AOI. Because of the proximity of AOI’s to one another, it was not possible to increase the size of AOI’s to automatically buffer the abovementioned circumstance.
To investigate temporal sequencing of visual search, four temporal phases were determined for each skill type outlined in Table 1. When gaze data was considered in respect to the four phases of each skill, it was found that in many cases a fixation may start in one phase but continue across multiple phases. As a result, if fixations were only considered relative to the phase in which they started, it could be the case that the spatio-temporal gaze data in subsequent phases may be missed. To overcome this, a fixation was coded for a respective phase if it remained within that phase for greater than 100ms, even if it had already been coded for a previous phase. A corollary of this is that the sum of fixations across the four temporal phases may be greater than the number of fixations across the trial, however, we believed this coding to be important to ensure spatio-temporal visual search strategies were captured across skill observations.

Due to the large number of pairs for comparison (288) when presentation style, skill, skill phase, and spatial location are considered, statistical analyses of differences in spatio-temporal visual search strategies were not conducted. Instead, the data was visually represented and observed for trends, as well as similarities and differences between assessor groups and presentation styles.

Initial analyses revealed consistent gaze behaviours across paediatric professionals and primary teachers in addition to established consistencies in assessment accuracy (Ward, Thornton, Lay, Chen, et al., 2018) (Chapter 6), so we determined it important to investigate whether visual-search behaviours differed between high and low accuracy assessors. Assessors were divided at the 50th percentile into high and low-accuracy groups based on their criterion level accuracy. Mann-Whitney U tests were used to determine if significant differences in assessment accuracy existed between high and low-level assessment groups.

Mann-Whitney U tests were used to determine if significant differences in fixation counts and fixation durations existed between high and low-accuracy assessors. Wilcoxon signed-rank tests were also administered to determine if significant differences in gaze behaviours existed across presentations styles, within each assessor group. For spatio-temporal analyses of visual search, the mean and median number, and respective location of fixations within each skill phase was also calculated across assessor groups. Spatio-temporal aspects of visual
search were also treated for high and low-accuracy assessor groups in the same manner as paediatric professionals and primary teachers.
Table 1. Kinematic markers used to define phases of skill performances for spatio-temporal analyses of visual search and relevant proficiency criteria for each skill phase

<table>
<thead>
<tr>
<th>Skill</th>
<th>Phase</th>
<th>Phase Start</th>
<th>Phase End</th>
<th>Relevant Proficiency Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kick</td>
<td>Static Phase</td>
<td>Start of clip</td>
<td>First forward movement of run up</td>
<td>- Rapid and continuous approach to the ball</td>
</tr>
<tr>
<td></td>
<td>Run-up</td>
<td>First forward movement of run up</td>
<td>Last grounding of kicking foot</td>
<td>- An elongated stride or leap immediately prior to ball contact</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Non-kicking foot placed even with or slightly in back of the ball</td>
</tr>
<tr>
<td></td>
<td>Preparation</td>
<td>Last grounding of kicking foot</td>
<td>Grounding of support-leg or first forward movement of kicking leg</td>
<td>- Kicks ball with instep of preferred foot (shoe-laces) or toe</td>
</tr>
<tr>
<td></td>
<td>Execution</td>
<td>First forward movement of kicking leg</td>
<td>Grounding of kicking foot following ball contact or end of clip</td>
<td>- Windup is initiated with downward movement of hand/arm</td>
</tr>
<tr>
<td>Throw</td>
<td>Static</td>
<td>Start of clip</td>
<td>First movement of throwing arm</td>
<td>- Rotates hip and shoulders to a point where the non-throwing side faces the direction of the throw</td>
</tr>
<tr>
<td></td>
<td>Preparation</td>
<td>First movement of throwing or aiming arm</td>
<td>Peak lift of throwing arm (behind body)</td>
<td>- Weight is transferred by stepping forward with the foot opposite the throwing hand</td>
</tr>
<tr>
<td></td>
<td>Execution</td>
<td>First forward movement of throwing arm or opposite foot</td>
<td>Ball release</td>
<td>- Follow-through beyond ball release diagonally across the body toward the non-preferred side</td>
</tr>
<tr>
<td></td>
<td>Follow-through</td>
<td>Ball release</td>
<td>End of clip</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td>Static</td>
<td>Start of Clip</td>
<td>First backward movement of arms or flexion of knees</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------------</td>
<td>---------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td>First backward movement of arms or flexion of knees</td>
<td>First forward movement of arms or extension of knees</td>
<td>Preparatory movement includes flexion of both knees with arms extended behind the body</td>
<td></td>
</tr>
<tr>
<td>Propulsion</td>
<td>First forward movement of arms or extension of knees</td>
<td>Peak arm extension or peak jump height</td>
<td>Arms extend forcefully forward and upward reaching full extension above the head</td>
<td></td>
</tr>
<tr>
<td>Landing</td>
<td>Peak arm extension or peak jump height</td>
<td>Full leg extension (recovery) or end of clip</td>
<td>Arms are thrust downward during landing</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hop</th>
<th>Static</th>
<th>Start of Clip</th>
<th>First backward movement of arms or non-support leg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>First backward movement of arms or non-support leg</td>
<td>First forward movement of arms or non-support leg</td>
<td>Foot of non-support leg remains behind body</td>
</tr>
<tr>
<td>Propulsion</td>
<td>First forward movement of arms or non-support leg</td>
<td>Peak arm extension or peak hop height</td>
<td>Non-support leg swings forward in pendular fashion to produce force</td>
</tr>
<tr>
<td>Landing</td>
<td>Peak arm extension or peak hop height</td>
<td>Full leg extension (recovery) or end of clip</td>
<td>Foot of non-support leg remains behind body</td>
</tr>
</tbody>
</table>

- Takes off and lands on the same foot
Results

Comparisons between paediatric professionals and primary teachers

Table 2 presents accuracy figures for paediatric professionals and primary teachers, as well as high and low accuracy assessor groups. Accuracy and visual search results will be initially be presented for paediatric professionals and primary teachers, followed by high and low accuracy assessors. As established in (Ward, Thornton, Lay, Chen, et al., 2018) (Chapter 6), there were no significant differences in assessment accuracy between paediatric movement professionals and primary teachers in either PL or video assessments and moderate assessment accuracy was observed across both groups and presentation styles (Table 2). Wilcoxon signed-rank tests also revealed no significant differences between video and PL accuracy within each assessor group.

Table 2. Average criterion level assessment accuracy for paediatric professionals and primary teachers, as well as high and low accuracy assessor groups

<table>
<thead>
<tr>
<th></th>
<th>Video assessment accuracy (%)</th>
<th>PL assessment accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paediatric Professionals</td>
<td>74.73 (7.52)</td>
<td>73.39 (4.58)</td>
</tr>
<tr>
<td>Primary Teachers</td>
<td>69.38 (10)</td>
<td>69.24 (7.14)</td>
</tr>
<tr>
<td>High Accuracy Assessors</td>
<td>79.30* (5.16)</td>
<td>76.25* (2.59)</td>
</tr>
<tr>
<td>Low Accuracy Assessors</td>
<td>64.14* (4.58)</td>
<td>65.86* (4.19)</td>
</tr>
</tbody>
</table>

*denotes significant difference in accuracy between high and low-accuracy assessor groups (p < 0.05)

Mann-Whitney U tests revealed significant differences between groups for average fixation count for PL jump assessments, however no differences in average fixation count or average fixation duration between groups were observed for other skills (Table 3). Wilcoxon signed-rank tests revealed paediatric professionals made significantly less fixations of significantly longer duration when making PL assessments in comparison to video assessments. Wilcoxon signed-rank tests also revealed that primary teachers made significantly less fixations of significantly longer duration when making PL assessments of kick performances in comparison to video assessments. Significantly longer fixation duration was also observed for PL jumps assessments by primary teachers when compared to video jump assessments.
Figures 1 through 8 present the average number of fixations on anatomical locations across skill phases for paediatric professionals and primary teachers. Visual assessment of spatio-temporal visual search data revealed that similar strategies were employed by both paediatric professionals and primary teachers across skills and presentation styles. There was only one case in which the median number of fixations across groups on a particular location within a phase was 1 for video assessments of the four skills. For PL observations there were seven cases in which the median number of fixations across assessors was 1. In all other cases the median number of fixations was 0.

Table 3. Average fixation count and average fixation duration per assessment for paediatric professionals and primary teachers across skills and presentation styles.

<table>
<thead>
<tr>
<th></th>
<th>Average Fixation Count</th>
<th>Average Fixation Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paediatric Professionals</td>
<td>Primary Teachers</td>
</tr>
<tr>
<td></td>
<td>Paediatric Professionals</td>
<td>Primary Teachers</td>
</tr>
<tr>
<td>All Skills</td>
<td>Video</td>
<td>7.46 (1.22)</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td>6.41 (1.60)</td>
</tr>
<tr>
<td>Hops</td>
<td>Video</td>
<td>7.54 (0.99)^</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td>6.60 (1.51)^</td>
</tr>
<tr>
<td>Kicks</td>
<td>Video</td>
<td>7.94 (1.04)^</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td>6.50 (1.12)^</td>
</tr>
<tr>
<td>Jumps</td>
<td>Video</td>
<td>7.41 (1.66)^</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td>6.77 (2.06)^*</td>
</tr>
<tr>
<td>Throws</td>
<td>Video</td>
<td>6.93 (1.12)^</td>
</tr>
<tr>
<td></td>
<td>PL</td>
<td>5.77 (1.76)^</td>
</tr>
</tbody>
</table>

*denotes significant differences between assessor groups (p < 0.05)
^denotes significant differences between presentation styles (p < 0.05)
Figure 1. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Figure 2. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Figure 3. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Figure 4. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Figure 5. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Figure 6. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Average fixation count per trial across spatial locations and skill phases: PL Throws

Figure 7. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Average fixation count per trial across spatial locations and skill phases: PL Kicks

Figure 8. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase for Paediatric Professionals (PP) and Primary Teachers (PT). Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Comparisons between high and low accuracy assessor groups

Initial results revealed no differences in accuracy or visual search strategies between paediatric professionals and primary teachers. Given that process-oriented FMS assessment relies upon accurate criterion identification, it was important to investigate whether different visual search strategies were adopted by assessors that were highly accurate when compared to those that were less accurate. High-accuracy assessor groups were found to be significantly more accurate than low-accuracy assessor groups in both presentation styles (Table 2). However, Mann-Whitney U tests revealed no significant differences between groups when average fixation count and average fixation duration were considered across all skill assessments (Table 4). When average fixation counts were considered for individual skills, high accuracy assessors made significantly less fixations than low accuracy assessors when assessing PL hop performances. Significant differences between groups were also found when average fixation duration was considered for PL hop observations, as well as for PL jump observations. High accuracy assessors had significantly longer average fixation durations in both cases (Table 4).

Table 4. Average fixation count and average fixation duration per assessment for high and low accuracy assessors across skills and presentation styles.

<table>
<thead>
<tr>
<th></th>
<th>Average Fixation Count</th>
<th>Average Fixation Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Accuracy Assessors</td>
<td>Low Accuracy Assessors</td>
</tr>
<tr>
<td>All Skills Video</td>
<td>7.66 (1.16)</td>
<td>7.72 (1.036)</td>
</tr>
<tr>
<td>PL</td>
<td>6.33 (1.57)</td>
<td>7.76 (1.13)</td>
</tr>
<tr>
<td>Hops Video</td>
<td>7.29 (1.01)</td>
<td>7.59 (0.78)</td>
</tr>
<tr>
<td>PL</td>
<td>6.43 (1.29)*</td>
<td>7.76 (1.22)*</td>
</tr>
<tr>
<td>Kicks Video</td>
<td>8.04 (0.90)</td>
<td>8.34 (0.82)</td>
</tr>
<tr>
<td>PL</td>
<td>6.53 (1.09)</td>
<td>7.76 (1.15)</td>
</tr>
<tr>
<td>Jumps Video</td>
<td>7.91 (1.75)</td>
<td>8.03 (1.03)</td>
</tr>
<tr>
<td>PL</td>
<td>6.79 (2.07)</td>
<td>8.53 (0.94)</td>
</tr>
<tr>
<td>Throws Video</td>
<td>7.41 (0.79)</td>
<td>6.94 (1.07)</td>
</tr>
<tr>
<td>PL</td>
<td>5.59 (1.70)</td>
<td>7.00 (0.79)</td>
</tr>
</tbody>
</table>

*denotes significant differences between assessor groups (p < 0.05)

^denotes significant differences between presentation styles (p < 0.05)
Observation of spatio-temporal visual search data revealed similar observation strategies for high-accuracy and low-accuracy assessors across all skills in both presentation styles. Again, due to the large number of pairs for comparison (288) when presentation style, skill, skill phase, and spatial location are considered, statistical analyses of differences in search behaviour were not conducted. When spatio-temporal visual search data was observed relative to the spatio-temporal sequencing of proficiency criteria, fixations often fell upon locations specified by proficiency criteria, but also fell on a range of other locations (Figures 9-12). Spatio-temporal data was also assessed in terms of the median number of fixations on each location within assessor groups as a means to identify whether assessor groups were consistently making fixations on the same location. A median of one was recorded once for both high and low-accuracy assessors across video assessments of the four skills, and twice for high-accuracy and 8 times for low-accuracy assessors across PL assessments of the four skills.
Figure 9. Average number of fixations per trial calculated across ten hop assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Figure 10. Average number of fixations per trial calculated across ten kick assessments; presented for each spatial location and skill phase High Accuracy (HA) and Low accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Figure 11. Average number of fixations per trial calculated across ten jump assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.

Figure 12. Average number of fixations per trial calculated across ten throw assessments; presented for each spatial location and skill phase for High Accuracy (HA) and Low Accuracy (LA) assessors. Shaded regions represent locations outlined in TGMD-2 assessment criteria.
Discussion

The current study sought to compare the visual search behaviours of paediatric professionals and primary teachers whilst conducting process-oriented FMS assessments from both video and PL displays. We predicted that paediatric professionals would exhibit more efficient gaze behaviours due to their increased training and experience in comparison to primary teachers, however this was not supported. We also predicted that paediatric professionals would adopt a more proximal search strategy than primary teachers, which was also not supported. The results are important in helping develop machine learning approaches to FMS assessment, and also understanding the administration of process-oriented FMS assessments by experienced and inexperienced assessors.

Somewhat unexpectedly, visual search strategies were consistent across all assessor regardless of training or experience. Visual search strategies were also consistent regardless of the accuracy of proficiency criteria identification. Proficiency criteria within process-oriented FMS assessments provide an observational structure to assessors, and therefore it could be expected that assessment calls for a search strategy that closely follows proficiency criteria. If, however, the attentional and time constraints imposed by the task prevent such a strategy during real-time FMS assessment, in which assessors get one opportunity to rate proficiency, assessors must employ an alternate strategy that still affords a proficiency assessment to be made. Our results suggest that assessors may not be using the proficiency criteria to guide their visual search, or consciously ‘looking for’ the kinematics pertaining to individual proficiency criteria, as would be the case in a ‘top-down’ model of visual processing (van Zoest & Donk, 2004). In both video and PL conditions spatio-temporal visual search strategies were often inconsistent with anatomical locations in proficiency criteria.

One strategy to circumvent attentional constraints when required to attend to multiple stimuli simultaneously is to anchor gaze in one location and use peripheral cues to infer information in surrounding locations (Abernethy & Zawi, 2007). In the circumstance that assessors cannot attend to all proficiency criteria, it is likely that gaze is directed to anatomical locations that provide the most proficiency information and allow the most efficient information processing (Abernethy & Zawi, 2007). The visual search strategies observed
within the current study appeared to be closely linked to the nature of the FMS being assessed and provide support that the strategies employed by assessors are highly task dependent (Waters et al., 2014). For object control skills, assessors prioritised anatomical locations that interacted with the object during skill execution, and likely deemed these locations to provide the most proficiency information. When observing jumps and hops assessors tended to employ a more proximal anatomical search strategy. Most fixations during assessments of locomotor skills fell upon the hips and torso of the performer, whereas many proficiency criteria relate to more distal segments. A proximally-based anatomical visual search strategy is reported to be characteristic of observers that have the ability to use proximal spatial regions to pick up and use relevant information earlier in the kinematic chain (Abernethy & Zawi, 2007; Hernández et al., 2006; Petrakis, 1993). It is also suggested that observers may use proximal locations as a convenient anchor-point around which they can extract information, improving the efficiency of visual search (Giblin, 2014).

The consistency of visual search across levels of assessor experience in the current study and the disparity between criterion and performance level scoring observed in Ward, Thornton, Lay, and Rosenberg (2018)(Chapter 5) provides further evidence that the strategy adopted by assessors to judge FMS proficiency does not rely on accurate perception of proficiency criteria. It may be the case that proficiency criteria are independent of the information assessors use to judge proficiency, to the extent that criterion identification accuracy may not be an appropriate outcome measure of perceptual expertise in FMS assessment. The consistent patterns of visual search behaviour across high and low accuracy groups even though assessment results reported significant differences in accuracy further suggests a disparity between the information employed to judge FMS proficiency and proficiency criteria. The notion of FMS is that they are fundamental to the performance of many specialised motor skills and, perhaps, independent of structured training regular exposure to FMS performances may be sufficient to develop visual search patterns that allow proficiency assessment (Giblin, 2014). Our findings suggest that the ability to observe FMS and make accurate proficiency judgements is not limited by assessor experience or the ability to accurately perceive proficiency criteria. The large range of search locations outside of the those defined by the criteria support the notion that assessors observe the entire movement based on their own observation template rather than the information outlined in proficiency
criteria, and subsequently rate the presence or absence of proficiency criteria based on their subjective judgement of proficiency (Chong, Familiar, & Shim, 2016; Komatsu, 2006; Ward, Thornton, Lay, & Rosenberg, 2018). As suggested in (Ward, Thornton, Lay, & Rosenberg, 2018) (Chapter 5), assessors observe and rate the proficiency of FMS performances by making comparisons to their own observation templates developed over historical skill observations. Subsequently, assessors may use the assessment structure to ensure that the proficiency score they generated via their own judgement is simply reflected in the number of proficiency criteria they score, rather than ensuring they score the correct criteria.

Given that so few differences were observed between visual search behaviours across comparisons between expertise and assessment accuracy, the findings of the current study unfortunately give little information that can be used to inform assessor training. The current study provides further evidence that real-time FMS assessment is challenging regardless of assessor experience and suggests that assessors do not consistently attend to spatial regions outlined within proficiency criteria, supporting the disparity between performance score and criterion level reliability observed by Barnett et al. (2014). Accurate and reliable assessment of proficiency criteria may require human assessors to make multiple observations of captured performances, overcoming the perceptual limitations observed in the current study. Spatio-temporal search strategies observed in the current study suggest that there may be inconsistencies between performance criteria and the information employed by observers when assessing FMS proficiency in real-time. Therefore, using criterion level scoring to identify specific skill deficiencies and subsequently inform interventions may be problematic.

The results of the current study suggest that human perception of FMS during assessment is misaligned with criteria outlined in process-oriented assessment tools. The allocation of attention highlighted within the current study also suggests that it may be difficult to inform computational assessment strategies based upon spatio-temporal patterns of human visual search during real-time assessment. The complexity and variability of the human perceptual system appears to be a significant barrier to the accurate real-time administration of process-oriented FMS assessment structures and the identification of consistent perceptual processes across assessors. Additionally, the information processing capacities of computers are not limited in the same way as human attentional capacities and can simultaneously process
multiple streams of information. Humans appear to currently have a superior ability to extract and filter relevant information as a means to overcome limitations in attentional capacity. In an FMS environment, computers may have the information processing capacity to rapidly comprehend more information than human assessors, but lack the innate ability to determine which information is relevant. Whilst machine learning approaches require key information to be identified prior to ‘training’, our results suggest that a human observation template during real-time FMS assessment may not provide a suitable foundation for data filtering. We suggest that machine learning approaches to FMS assessment should capitalise on the reliability of human proficiency classification to provide the required assessment outputs, but rely upon statistical means to determine the key kinematic features that should be included in the assessment algorithm.

Finally, although this study did not provide evidence to support the hypothesis that expert assessors would exhibit fewer fixations of longer duration, it has instead shown consistency in visual search across levels of experience. The results also offer new insights into the inherent limitations of both experienced and inexperienced humans to objectively assess proficiency criteria during real-time FMS assessment, and suggests that assessors instead rely upon established observational templates to make proficiency judgements. One implication of these findings is the limitation of real-time FMS assessment in informing clinical interventions. Furthermore, in progressing the development of technology-based assessments, the results cast doubt on the use of real-time human assessment as the gold standard in the development of computational approaches.
References


Giblin, G. (2014). *What a coach can see and an athlete can feel.* (PhD), University of Victoria, Victoria.


CHAPTER 8
Using inter-trial variability to classify horizontal jump proficiency in children with the Microsoft® Kinect® sensor

This manuscript was submitted for publication in Human Movement Science in October, 2018 and is currently under review.

Foreword

Initial investigations within this thesis (see Chapters Five, Six and Seven) revealed limitations of real-time human FMS assessment. It was therefore pertinent to concurrently investigate a possible data-driven, computational approach to classifying FMS proficiency, minimising the human involvement in FMS assessment. Using kinematic data captured using the Microsoft® Kinect®, this investigation examined the relationship between inter-trial variability and FMS proficiency in horizontal jump performances as a means to rapidly classify FMS proficiency using consumer level depth cameras.
Abstract

Introduction

Fundamental movement skill (FMS) assessment is crucial throughout early childhood to monitor motor skill development. Current assessment practices require a substantial understanding of motor development and can be time-consuming, which can impact the administration of assessments on a large scale. The Microsoft® Kinect® motion capture system may overcome barriers to administration by allowing cost-effective motion capture and the adoption of computational approaches to FMS assessment. The current study investigated the objective classification of proficiency in repeated horizontal jump performances captured by Microsoft® Kinect® cameras, through computational assessment of inter-trial variability.

Methods

Thirty-one children (7.8±2.1yrs) performed 3 horizontal jumps in front of two Microsoft® Kinect® cameras. Inter-trial variability of velocity for each joint centre was compared to standardised TGMD-2 (Ulrich, 2000) locomotor scores

Results

A significant positive association was found between inter-trial variability and standardised locomotor scores. Distal joints were found to correlate more strongly with proficiency than proximal joints. Strong positive correlations were also found between joint inter-trial variability and joint velocity in distal joints.

Conclusion

The study establishes a link between inter-trial variability and proficiency in horizontal jumping. We also provide evidence for the feasible computational classification of movement proficiency using the Microsoft® Kinect® motion capture system, which could provide a basis for automated FMS proficiency classification.
Introduction

Fundamental movement skill (FMS) assessment is crucial for the early identification of skill deficiencies and monitoring of motor development throughout childhood (Hands, 2002; Piek, Hands, & Licari, 2012). Identifying and remediying skill deficiencies in children is important to prevent the negative correlates of poor movement proficiency, such as obesity and physical inactivity persisting through adolescence and adulthood (Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012; Lubans, Morgan, Cliff, Barnett, & Okely, 2010; Piek et al., 2012; Slater, Hillier, & Civetta, 2010). Many popular FMS assessments focus on skill processes (how a skill is performed), however they require substantial training and understanding of human movement, and can suffer from low reliability, due to their subjective nature (Barnett et al., 2009; Barnett, Minto, Lander, & Hardy, 2014; Burton & Miller, 1998; Hands, 2002). Comparatively, product-oriented assessments (those concerned with skill outcomes such as distance, speed, or accuracy) are simpler, faster to administer, and have greater reliability, although they provide little information about movements contributing to the performance (Barnett et al., 2009; Burton & Miller, 1998; Hands, 2002). Regardless of the method, FMS assessments are typically administered by human assessors, live in the field, or through video capture and post hoc analysis, meaning the time and personnel costs associated with such assessments are high. To overcome the time and costs associated with current assessment techniques, a more recent consideration is to develop approaches to FMS assessment which replace the human with technology and in doing this the time costs and subjectivity associated with movement assessment may be improved (Bisi, Pacini Panebianco, Polman, & Stagni, 2017; Ward, Thornton, Lay, & Rosenberg, 2017).

Recent advances in consumer-level depth capture systems could be used to execute computational assessment of kinematic patterns including inter-trial variability. One such depth capture system is the Microsoft® Kinect®; an off-the-shelf, consumer-level marker-less motion capture system that captures 3-dimensional digitised kinematic data (Jungong, Ling, Dong, & Shotton, 2013). The capture of kinematic data by the Kinect® depth camera allows far more computationally efficient extraction of kinematic features for motion analysis when compared to video data captured by traditional video cameras (Pirsiavash, Vondrick, & Torralba, 2014). The Kinect® is not as accurate at movement capture when compared with
gold standard multi-camera marker-based infrared systems (e.g. Vicon (Oxford Metrics, Oxford, United Kingdom)), as it has a reduced capture rate, and can by impacted by spatial occlusion. The Kinect® is however cost effective, easy to use and portable making it ecological viable for rapid FMS capture, and subsequent assessment of large samples (Ward et al., 2017). The suitability of consumer-level depth cameras like Kinect® for the classification of individuals’ FMS proficiency is yet to be established. One approach to objective FMS assessment that warrants consideration is the computational measurement of inter-trial variability between successive FMS performances. FMS assessments typically require the execution of a motor skills in a closed environment where the intended goal of the task and regulatory features remain invariant across successive performances. This well-structured and controlled environment affords the opportunity to assess the link between movement proficiency and inter-trial variability.

Movement pattern variability during skill acquisition was once considered noise that should be minimised (Dhawale, Smith, & Ölveczky, 2017; Stergiou, Harbourne, & Cavanaugh, 2006). More recent advances in research have identified variability during skill acquisition to represent flexibility within the motor system to explore and an essential part of the progress towards the most appropriate performance of a given task (Preatoni et al., 2013). The human body is a complex and dynamic motor system, consisting of many degrees of freedom (DOF) or components, that must be coordinated to achieve any given motor task. Successful performance of a motor skill relies on an individual’s capability to effectively coordinate and control the motor system’s DOF to achieve the desired or optimal outcome (Button, Macleod, Sanders, & Coleman, 2003). It’s important to note that variability is often studied in precision tasks where variability is considered a means by which skilled performers compensate for errors within movement patterns to achieve end-point consistency, such as consistent ball release parameters in a basketball free-throw (Button et al., 2003). The task constraints of FMS rarely require performers to achieve end-point stability and it is therefore important to consider how variability may be elicited across stages of FMS development. It has been reported that in the early stages of skill acquisition, such as learning a new FMS, performers are typically rigid and invariable in their movement because they ‘freeze’ DOF, tightening joint couplings to reduce the complexity and variability of movements (Bernstein, 1967; Davids, 2008; Davids, Bennett, & Newell, 2006; Piek, 2002; Seifert, Button, & Davids, 2013; Vereijken,
Emmerik, Whiting, & Newell, 1992). In contrast, skilled performers exhibit movement patterns in which DOF are released, resulting in efficient skill performances, a notion originally observed by Nikolai Bernstein (1967) and studied across a variety of movement contexts thereafter (Button et al., 2003; Davids et al., 2006; Gallahue, Ozmun, & Goodway, 2012; Piek, 2002; Pretoni et al., 2013; Vereijken et al., 1992).

The release of DOF is present within proficient performance of many FMS and as such, many process-oriented FMS assessments reflect the release of DOF in assessment criteria pertaining to proficient skill performances (Haywood & Getchell, 2014). Regarding the horizontal jump for example, an effective arm swing via the loosening of joint couplings in the upper-limb is a major component of proficient performance (Ashby & Heegaard, 2002; Hara, Shibayama, Arakawa, & Fukashiro, 2008; Lees, Vanreunterghem, & Clercq, 2004), and as such is represented in many process-oriented jump assessments (Bruininks, 1978; NSW Department of Education and Training, 2000; Ulrich, 2000) (Western Australian Minister for Education, 2004). Like many locomotor skills, velocity production is a key component of skilled jumping and spatial accuracy is of comparatively little importance. Optimised velocity production is achieved through effective summation of forces throughout the kinematic chain and can only be achieved once the performer’s movement patterns are mature enough to reflect the release of DOF across joints (Ashby & Heegaard, 2002). From a process-oriented perspective it is important that assessments go some way to consider how a movement is produced, rather than skill outcome alone. It could be the case that variability is increased as a result of movement pattern maturation independent of increases in velocity production, and therefore inter-trial variability in resultant joint velocity may be a better predictor of movement pattern proficiency than the magnitude of velocity alone.

The current study investigated the relationship between inter-trial variability in resultant joint velocity and movement proficiency in horizontal jump performances in typically developing children, with a clear goal of exploring the use of variability to classify the proficiency of FMS performances. We hypothesised that increased inter-trial variability in velocity would be associated with greater movement proficiency in jump performances, and that more proficient children would exhibit greater average joint velocities than less proficient children.
Methods

Participants

Participants were 31 typically developing primary school aged children (7.8±2.1, 15 Male, 16 Female) recruited from paediatric exercise programs, and research mailing lists at the University of Western Australia. Kinect® FMS motion capture took place in a carpeted indoor movement laboratory, and TGMD-2 assessment of movement proficiency took place in a grassed outdoor laboratory adjacent to the indoor laboratory at the University of Western Australia. The FMS assessment, and motion capture collection were counterbalanced between participants and a five-minute break was given in between to negate any order effects and the effects of fatigue. Participants were instructed to wear clothing they would typically wear when participating in physical activity for the testing protocol. All components of the study had ethics approval from the University of Western Australia.

Procedure

Kinect® Movement Capture

We chose to assess the horizontal jump because the movement features, such as the start and end points, are easily defined and provide efficient assessment of variability by automated means. Jump performances were captured by 2 Microsoft® Kinect® motion capture cameras placed 50 cm above the ground, 2 m apart, 2.5 m in front of the participant, with each camera converging on the skill starting position at a 21° angle to the sagittal plane. Two offset cameras were used in the study to minimise segment occlusion, an established limitation of the Kinect® system. Performances were captured at 30Hz by both cameras on a single computer using software designed in-house with the Microsoft® Kinect® software development kit.

Participants were required to perform 10 horizontal jumps for distance from a fixed starting position. Ten performances were initially captured as we were unsure of the minimum trials required to reveal associations and were concerned with collecting too few. Each child was instructed to perform a 2-footed horizontal jump as far as they could, given a demonstration of the skill by the researcher and a practice trial to ensure they understood the task, as is
common in FMS assessments. If it was clear that they had misunderstood the task, a second demonstration and practice was provided.

**Test of Gross Motor Development-2 administration**

Within the same data collection session, participants were administered the Test of Gross Motor Development-2 (Ulrich, 2000) by an experienced movement assessor. The TGMD-2 (Ulrich, 2000) is a validated and widely administered process-oriented FMS assessment tool whereby assessors score skill performances by observing the presence or absence of criteria pertaining to proficient performance of the skill. The TGMD-2 (Ulrich, 2000) is a 12 item assessment containing both locomotor (6 skills) and object control (6 skills) items to determine movement proficiency in a field-based setting.

**Data Processing**

**Treatment of TGMD-2 data**

Raw scores for locomotor and object control subtests were summed, prior to being age and gender standardised, and percentile ranked using the tables provided in the assessment manual (Ulrich, 2000). Overall tests scores were also converted to Percentiles and Gross Motor Quotients (GMQ) using the tables provided in the assessment manual (Ulrich, 2000). Standardised locomotor scores were used as the dependent variable representing skill proficiency in further analyses even though they are calculated using the entire locomotor subset, as this allowed proficiency to be standardised for both age and gender which is not available for individual skills.

**Processing of Kinect® data**

Kinect® data for jump performances was processed in MATLAB R2017a (The MathWorks, Inc). The raw joint data captured by each Kinect® camera represents x, y, and z joint positions for 20 estimated joint centres over time. The system also assigns a tracking state for each frame for each joint as, tracked, inferred, or not-tracked, depending on whether the joint position was visible, not clearly visible, or not visible to the camera. In addition to this, each frame for each camera is coded with a timestamp based on the capture computer’s internal clock, which was later used for temporal alignment of the camera streams. The data for the hands and feet were removed from further processing and analysis due to a large number of invalid data points for these joint centres, resulting in joint positions for 16 joints. The final set of
joints for processing were; Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), Right Hip (acetabulofemoral), Left Elbow, Left Wrist, Right Elbow, Right Wrist, Left Knee, Left Ankle, Right Knee, Right Ankle. The two Kinect® cameras captured performances independently during collection, and the merging of the streams to minimise data loss was done in post-processing.

**Merging camera streams.** For each performance, joint positions from both cameras were merged to minimise the amount of ‘inferred’ and ‘untracked’ data due to occlusion. For the merging of Kinect® data streams, the left camera (stage right) was treated as the primary data stream, and the right as a secondary. Prior to merging, streams were time synchronised and trimmed to the primary camera stream based on the primary camera timestamps. Because the two cameras have their own local co-ordinate systems, joint data must be transformed into a single global co-ordinate system. The primary camera co-ordinate system was treated as the global system, into which the secondary camera data would be transformed. An optimal transformation matrix was calculated for data alignment. To ensure the greatest transformation accuracy the matrix was calculated using only joints that were ‘tracked’ by both cameras. Once the streams were spatially and temporally aligned, the data was merged based on the tracking state of each frame, for each camera. When processing the data for merging, distal joints were considered ‘child’ joints of proximal joints, i.e. the ankle is the child joint of the knee, and both are child joints of the hip. We found that in many cases, if a joint was coded as ‘untracked’ by the Kinect®, the child joint position was quite inaccurate, even if it had a ‘tracked’ state. For this reason, if the primary camera had a joint that was ‘untracked’, but was ‘tracked’ by the secondary camera, data for the joint, and all child joints was sourced from the secondary stream. The resulting single stream of merged x, y, z joint positions for each performance was used for subsequent velocity processing.

**Calculating joint velocities.** The x, y and z joint positions were used to calculate the resultant velocity in metres per second (m·s⁻¹) for each of the remaining 16 joints. Joint centre velocity sequences were used for analyses of inter-trial variability. This decision was influenced by the goal of automating the analysis process for possible implementation as an FMS proficiency classification tool, and the limitations of the Kinect® system. Whilst a global
co-ordinate system was established for individual trials to merge kinematic data from multiple cameras, the Kinect® does not employ a global co-ordinate system in the same manner that Vicon (Oxford Metrics, Oxford, United Kingdom) or similar marker-based motion capture systems do. This renders it computationally expensive and difficult to analyse positional variability in 3-dimensional joint space across multiple trials.

**Time-normalising performances.** When analysing variability between multiple signals, it is important to ensure the start and points are consistent across the signals (Mueen & Keogh, 2016). Velocity profiles were automatically trimmed from the start of the preparatory squat to the peak of the landing squat using the hip joint as the reference joint due to its relatively consistent velocity profile across all children. Trimmed velocity profiles were then plotted over the full trial velocity profile and visually inspected to ensure trimming was correct. If the algorithm had misidentified the timestamps for each squat, the timestamps were manually coded. In the case that the hip velocity profile was irregular, the sequence was crosschecked with the video performance to establish if the irregularity could be explained by the child’s movement. If the irregular velocity profile could not be explained by the movement, capture error was assumed, and the trial was rejected. This was the case in 3 of the 93 trials used in the final analysis (~3%). In the event that a trial was rejected, the next valid trial was taken. Velocity sequences for the remaining 15 joints were then trimmed relative to the squat timestamps, and time-normalised to 100 data points using linear interpolation.

**Calculating inter-trial variability.** Inter-trial variability analyses were calculated for each joint across time-normalised velocity sequences of the first 3 valid trials. Whilst 10 trials of each skill were captured to ensure sufficient data was collected, only the first three trials were used for variability calculations to maintain the ecological viability of the study in regards to overcoming the time-costs of FMS assessment. Time-normalised velocity sequences were aligned for each of the 16 joint centres. To analyse variability across entire performances, variability was evaluated at each of the 100 time-normalised points for each joint, by calculating the standard deviation of velocity values across the 3 trials. The standard deviations at each time normalised point were then averaged to calculate the mean deviation, which represented the average variability between the velocity profiles of 3 trials for each
joint centre. Average variability and velocity for individual joint centres was also aggregated into 3 joint groups based on proximity to the trunk presented in Table 1.

Table 1. Proximal, mid and distal joint groups based on proximity to the trunk

<table>
<thead>
<tr>
<th>Joint Group</th>
<th>Joints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal</td>
<td>Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), Right Hip (acetabulofemoral)</td>
</tr>
<tr>
<td>Mid</td>
<td>Left Elbow, Right Elbow, Left Knee, Right Knee</td>
</tr>
<tr>
<td>Distal</td>
<td>Left Wrist, Right Wrist, Left Ankle, Right Ankle</td>
</tr>
</tbody>
</table>

Statistical Analyses

Descriptive statistics and correlations were conducted using SPSS version 24, and regressions models were analysed using STATA version 13.1. Descriptive statistics were initially produced for TGMD-2 proficiency scores, followed by analyses of variability and velocity data. The term ‘variability’ within the results, and subsequent sections, refers to the average variability across 3 velocity sequences (trials), calculated at each of the 100 time-normalised data points. The term ‘velocity’ refers to the average joint velocity across 3 velocity sequences (trials), calculated at each of the 100 time-normalised data points represented in m·s⁻¹. Pearson product-moment correlations were initially conducted to determine the relationship between proficiency, and variability for each individual joint. Subsequently, descriptive statistics of average velocity and inter-trial variability were produced for each joint group. Pearson product-moment correlations between standard locomotor scores and joint group data were then analysed to assess the relationship between proficiency and variability, and also between proficiency and velocity. Fisher r-to-z test were implemented to assess whether inter-trial variability and velocity correlated with proficiency in the same manner. Linear regression modelling was then used to assess the association of joint group variability on proficiency, and joint group velocity on proficiency.
Results

Descriptive statistics of TGMD-2 (Ulrich, 2000) revealed a large and normally distributed range of standardised locomotor scores across the sample (Table 2.). Pearson correlations between variability of individual joint centres and standard locomotor scores revealed significant positive correlations for 11 of the 16 joint centres, with the strength of correlations ranging from weak to moderate (Table 3.).

Table 2. Descriptive statistics of Test of Gross Motor Development-2 (Ulrich, 2000) results

<table>
<thead>
<tr>
<th>TGMD-2 Component</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Locomotor Score</td>
<td>31</td>
<td>9.10</td>
<td>2.55</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Locomotor Percentile</td>
<td>31</td>
<td>41.68</td>
<td>27.27</td>
<td>1</td>
<td>84</td>
</tr>
<tr>
<td>Gross Motor Quotient</td>
<td>31</td>
<td>100.58</td>
<td>14.85</td>
<td>52</td>
<td>121</td>
</tr>
</tbody>
</table>

Descriptive statistics revealed that both variability, and mean velocity increased from proximal to distal joint groups (Table 4.). Pearson correlations revealed moderate to very strong, significant relationships between joint group variability and joint group velocity for all group comparisons suggesting that variability increases linearly with velocity. The strongest correlation between variability and velocity was observed between distal joint groups ($r = 0.844$, $p < 0.01$), followed by mid joint groups ($r = 0.750$, $p < 0.01$) and proximal joint groups ($r = 0.619$, $p < 0.01$) (Figure 1.).
When the mean variability for each joint group was correlated with standard locomotor scores, a clearer picture of the relationship between variability and proficiency was observed (Table 5). Distal joint variability had the strongest positive correlation with standard locomotor scores ($r = 0.46$, $p < 0.01$), followed by proximal joint variability ($r = 0.426$, $p < 0.05$), and mid-segment joint variability ($r = 0.407$, $p < 0.05$). A similar pattern was observed when Pearson correlations were analysed between joint group velocities and proficiency (Table 6), with distal joints again having the strongest correlation ($r = 0.405$, $p < 0.05$). The relationship between velocity and proficiency was weaker than that between variability and proficiency across all joint groups, however Fisher r-to-z tests revealed that differences were not significant.

Table 3. Pearson Correlations between TGMD-2 Standardised Locomotor Scores and inter-trial variability for individual joints

<table>
<thead>
<tr>
<th>Joint Centre</th>
<th>Correlation with Standardised Locomotor Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Hip (acetabulofemoral)</td>
<td>$0.423^*$</td>
</tr>
<tr>
<td>Left Hip (acetabulofemoral)</td>
<td>$0.417^*$</td>
</tr>
<tr>
<td>Left Shoulder (glenohumeral)</td>
<td>$0.407^*$</td>
</tr>
<tr>
<td>Spine (mid)</td>
<td>$0.398^*$</td>
</tr>
<tr>
<td>Hip Centre (sacroiliac)</td>
<td>$0.392^*$</td>
</tr>
<tr>
<td>Right Wrist</td>
<td>$0.392^*$</td>
</tr>
<tr>
<td>Head</td>
<td>$0.384^*$</td>
</tr>
<tr>
<td>Shoulder Centre (C7)</td>
<td>$0.378^*$</td>
</tr>
<tr>
<td>Left Elbow</td>
<td>$0.366^*$</td>
</tr>
<tr>
<td>Left Ankle</td>
<td>$0.362^*$</td>
</tr>
<tr>
<td>Right Ankle</td>
<td>$0.358^*$</td>
</tr>
<tr>
<td>Right Knee</td>
<td>$0.313$</td>
</tr>
<tr>
<td>Right Shoulder (glenohumeral)</td>
<td>$0.299$</td>
</tr>
<tr>
<td>Left Knee</td>
<td>$0.298$</td>
</tr>
<tr>
<td>Left Wrist</td>
<td>$0.296$</td>
</tr>
<tr>
<td>Right Elbow</td>
<td>$0.279$</td>
</tr>
</tbody>
</table>

* $p < 0.05$ (2-tailed).
** $p < 0.01$ (2-tailed).
Table 4. Descriptive statistics of mean variability and mean velocity (m·s\(^{-1}\)) for distal, mid, and proximal joint groups

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distal Joints</td>
<td>31</td>
<td>1.056</td>
<td>0.199</td>
</tr>
<tr>
<td>Mid Joints</td>
<td>31</td>
<td>0.751</td>
<td>0.145</td>
</tr>
<tr>
<td>Proximal Joints</td>
<td>31</td>
<td>0.547</td>
<td>0.118</td>
</tr>
<tr>
<td>Velocity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distal Joints</td>
<td>31</td>
<td>2.500</td>
<td>0.409</td>
</tr>
<tr>
<td>Mid Joints</td>
<td>31</td>
<td>2.123</td>
<td>0.284</td>
</tr>
<tr>
<td>Proximal Joints</td>
<td>31</td>
<td>1.814</td>
<td>0.235</td>
</tr>
</tbody>
</table>

Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle
Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee
Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)

Table 5. Correlations between Standard Locomotor Scores and Distal, Mid, and Proximal joint group mean inter-trial variability

<table>
<thead>
<tr>
<th></th>
<th>Standard Locomotor Score</th>
<th>Distal Joints Variability</th>
<th>Mid Joint Variability</th>
<th>Proximal Joint Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Locomotor Score</td>
<td>Sig. (2-tailed)</td>
<td>.468**</td>
<td>(N ) 31</td>
<td>(N ) 31</td>
</tr>
<tr>
<td>Distal Joint Variability</td>
<td>Pearson Correlation</td>
<td>.407*</td>
<td>(N ) 31</td>
<td>(N ) 31</td>
</tr>
<tr>
<td>Mid Joint Variability</td>
<td>Pearson Correlation</td>
<td>.426*</td>
<td>(N ) 31</td>
<td>(N ) 31</td>
</tr>
<tr>
<td>Proximal Joint Variability</td>
<td>Pearson Correlation</td>
<td>.757**</td>
<td>(N ) 31</td>
<td>(N ) 31</td>
</tr>
</tbody>
</table>

* \( p < 0.05 \) (2-tailed).
** \( p < 0.01 \) (2-tailed).

Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle
Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee
Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)
Table 6. Correlations between Standard Locomotor Scores and Distal, Mid, and Proximal joint group mean velocity (m·s⁻¹)

<table>
<thead>
<tr>
<th>Standard Locomotor Score</th>
<th>Distal Joint Velocity</th>
<th>Mid Joint Velocity</th>
<th>Proximal Joint Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.405*</td>
<td>.362*</td>
<td>.323</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.024</td>
<td>0.046</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

*. p < 0.05 (2-tailed).
**. p < 0.01 (2-tailed).

Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle
Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee
Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)
Figure 1. Scatterplots of the relationship between mean variability and mean velocity for distal (a), mid (b), and proximal (c) joint groups including Pearson’s r and significance values for respective comparisons.

Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle
Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee
Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)
Linear regression modelling of joint group variability and standard locomotor scores revealed that distal joint variability in isolation had the greatest association with proficiency (Table 7.). Linear regression of joint group velocity and standard locomotor scores revealed similar patterns, however, the association was weaker (Table 8.).

**Table 7. Results of backward stepwise regression analyses of Distal, Mid, and Proximal joint group mean inter-trial variability on Standard Locomotor Scores**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.55 (2.23)</td>
<td>2.52 (2.49)</td>
<td>2.78 (2.29)</td>
</tr>
<tr>
<td>Distal Joint Variability</td>
<td>9.97 (6.84)</td>
<td>4.36 (2.68)</td>
<td>5.98 (2.06)**</td>
</tr>
<tr>
<td>Mid Joint Variability</td>
<td>-9.70 (10.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximal Joint Variability</td>
<td>6.02 (4.92)</td>
<td>3.61 (4.22)</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.26*</td>
<td>0.23*</td>
<td>0.22**</td>
</tr>
<tr>
<td>No. Observations</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Replications</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

*: \( p < 0.05 \) (2-tailed).

**: \( p < 0.01 \) (2-tailed).

*Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle

*Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee

*Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)

**Table 8. Results of backward stepwise regression analyses of Distal, Mid, and Proximal joint group mean velocities (m·s\(^{-1}\)) on Standard Locomotor Scores**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4.78 (3.86)</td>
<td>2.47 (3.59)</td>
<td>2.78 (2.29)</td>
</tr>
<tr>
<td>Distal Joint Velocity</td>
<td>9.64 (5.91)</td>
<td>2.32 (1.23)</td>
<td>2.52 (0.93)**</td>
</tr>
<tr>
<td>Mid Joint Velocity</td>
<td>-14.64 (11.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximal Joint Velocity</td>
<td>6.22 (3.86)</td>
<td>0.45 (2.41)</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.22*</td>
<td>0.16*</td>
<td>0.16**</td>
</tr>
<tr>
<td>No. Observations</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Replications</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

*: \( p < 0.05 \) (2-tailed).

**: \( p < 0.01 \) (2-tailed).

*Distal joint group: Left Wrist, Right Wrist, Left Ankle, and Right Ankle

*Mid joint group: Left Elbow, Right Elbow, Left Knee, and Right Knee

*Proximal joint group: Head, Left Shoulder (glenohumeral), Shoulder Centre (C7), Right Shoulder (glenohumeral), Spine (mid), Left Hip (acetabulofemoral), Hip Centre (sacroiliac), and Right Hip (acetabulofemoral)
Discussion

The aim of this study was to explore the relationship between inter-trial variability in joint velocity and fundamental movement skill proficiency in children. The study focused on variability between repeated performances of horizontal jumps, and its association with locomotor skill proficiency, which was determined by TGMD-2 (Ulrich, 2000) scores. It was predicted that movement variability would increase with proficiency, due to the nature of skill development in children and the task constraints of jumping.

Inter-trial variability was found to increase with proficiency across all joint-groups, which supported our primary hypothesis, as well as contemporary theories of the role variability plays in skilled performance. Care must be taken when comparing findings of the current study to previous work, due to the type of skill investigated, and the measurement of variability. Traditionally studies of variability and expertise have considered precision object control tasks such as basketball free-throw shooting, handball throwing, or the golf swing. In these tasks joint-space variability is considered within performances as a means of reaching a consistent discrete end-point between trials (i.e. Consistent wrist position at ball release). Due to the nature of the task, it was difficult to define a discrete endpoint for analysing inter-trial variability between jump performances. In addition to this, the goal when performing a jump is not to maximise accuracy as in the aforementioned tasks, but to solely maximise power, or velocity. Therefore, the role or elicitation of variability needs to be interpreted differently when considering jump performances.

Variability was the greatest for distal joints, followed by mid-limb joints, and proximal joints respectively. The same pattern was observed across joint groups for velocity. We suggest that the variability observed in our results is less a functional product of ensuring that a consistent spatial end point is achieved, such as in precision tasks, but rather a product of maximal force production and the release of DOF throughout skill development. This is supported by the acknowledged association between variability and high velocity movements (Fairbrother, 2010), and also the process of skill acquisition in children from a dynamical systems theory perspective (Gallahue et al., 2012). A child that is maximising segment velocity via the release
of DOF and increased joint range of motion (ROM) is likely to be exhibiting both; more proficient movement patterns, and greater jump outcome success (distance).

It has been established that traditionally defined joint-space variability is associated with rapid and accelerative phases of movement (Button et al., 2003; Darling & Cooke, 1987; Wagner, Pfusterschmied, Klous, von Duvillard, & Muller, 2012). Our results suggest that the positive relationship observed in traditional analyses of velocity and joint-space variability persists when inter-trial variability is calculated from joint velocities. Wagner et al. (2012) found that joint-space variability increases through accelerative phases of movement, and additionally found that this was more so the case in distal joints than proximal ones. The current study also observed that increased inter-trial variability was characterised by increased mean velocity through distal joints, which likely to be evident of the release of DOF and the speed-accuracy trade-off (Schmidt, Sherwood, Zelaznik, & Leikind, 1985; Wagner et al., 2012). The speed-accuracy trade-off inherently results in a positive relationship between velocity and joint-space variability within trials which may also produce greater inter-trial variability in tasks where velocity production is important. With this in mind, it is important to consider why variability should be used in contrast to velocity. Whilst velocity production is a factor in skilled performance, a child can exhibit advanced movement patterns exhibiting the releasing of DOF independent of being able to produce large joint velocities. Therefore, it may be the case that inter-trial variability provides more meaningful information of skill development than velocity alone. To provide further evidence of this future investigations should consider minimising the influence of velocity by scaling the magnitude of inter-trial variability to the magnitude of velocity.

We also observed increasing variability in distal joints as proficiency increased. From a dynamic systems perspective, this association can be explained by the release of kinematic DOF, which is representative of jump maturation, especially in children. While it was not directly measured, distal joint variability would be characteristic of performers that are ‘releasing’ joint couplings to move through a full range of motion, rather than being ‘frozen’ to achieve greater control and reduce variability. The release of DOF, particularly in terms of an effective arm swing, is also established as a key contributor to jump success in relation to velocity production (Ashby & Heegaard, 2002; Lees et al., 2004; Vaverka et al., 2016). Visual
observation of jump performances provides evidence of clear differences in distal joint ROM between high and low proficiency jumpers (Figure 2.), in support of the developmental sequences outlined by Gallahue et al. (2012).

![Figure 2. Evidence of degree of freedom releasing in proficient jumping (A) in comparison to non-proficient jumping (B) captured by the Microsoft® Kinect®](image)

We suggest that the positive relationship between variability and proficiency in jumping observed in the current study is evidence of the release of DOF across skill development and its contribution to the capability of proficient performers to produce greater velocity through large joint ranges of motion. This study implemented the Microsoft® Kinect® as the primary motion capture tool, due to its ease of use, relative affordability, and efficient kinematic feature extraction, making it a viable and practical option for rapid FMS classification. It is also important to note that similar protocols could be replicated with the employment of other depth sensors or non-visual motion capture devices, such as inertial movement sensors depending on which technology is more appropriate for the collection environment. More clinically accurate results could be achieved in future studies through the use of multi-camera, marker-based motion capture systems, however, these systems have little practicality for large scale FMS assessment and thus provide little ecological validity. As computer vision capabilities progress towards reliable kinematic feature extraction from RGB video, greater freedom and efficiency FMS capture and computational proficiency assessment will be available. Nonetheless, further analyses of inter-trial variability for FMS assessment should consider quantifying additional structures of variability such as joint angle relationships and differences in joint range of motion, with the view to strengthen the variability-proficiency
associations observed in the current study. We would also suggest that future research explores the number of trials required for maximising the association between inter-trial variability and FMS proficiency, however, time costs and ecological viability needs to remain a strong consideration when considering the amount of trials. Research should extend the current investigation to a range of FMS including both locomotor and object control skills. The current study only explored one data driven approach to improving the objectivity of field-based FMS assessment, relying on a significant amount of ongoing human input, and future computational assessment techniques should look towards autonomous techniques employing machine learning.

**Conclusion**

This study revealed significant positive associations between movement proficiency and inter-trial variability in children’s jump performances captured with the Microsoft® Kinect® motion capture system. Whilst only one method of analysing inter-trial variability was investigated, and further research should build on this, results provided positive insights into the possible use of kinematic data from a consumer-level capture tool for computational assessments of FMS.
References


CHAPTER 9

Summary and Conclusions
Key Findings:
Inconsistent recognition of proficiency criteria was observed across PL and video assessments. PL assessments were less accurate than video assessments. Typical analyses of performance score reliability may not reflect inaccuracies in identifying proficiency criteria.

Key Contributions:
Suggests that consumer-level depth capture of FMS performances may not be suitable for assessment by inexperienced assessors. Suggests that FMS proficiency assessments from PL displays aren’t as accurate as those from video. Established that high assessment reliability can be observed even if accuracy of criterion identification is low.

Key Findings:
Analyses of performance score reliability may not capture inaccuracies in identifying proficiency criteria. Accurate identification of proficiency criteria during real-time FMS assessment is difficult, even for experienced assessors. Kinect®-produced PL displays may be suitable for FMS assessment by experienced assessors. Primary teachers can reliably assess FMS proficiency at the performance level.

Key Contributions:
Established differences in FMS assessment accuracy between overall proficiency scoring and identification of individual proficiency criteria. Established challenges of identifying proficiency criteria during real-time FMS assessment in experienced and inexperienced assessors.

Key Findings:
Experienced and inexperienced assessors have consistent gaze behaviours, as is the case with high and low accuracy assessors, making it difficult to inform assessor training. Assessors may not use proficiency criteria to make proficiency judgements. Using proficiency criteria scoring to identify skill deficiencies may be problematic.

Key Contributions:
Established the visual search strategies of FMS assessors across levels of experience. Established the visual search strategies of high and low accuracy FMS assessors. Established a discrepancy between locations visually attended to during FMS assessment and locations presented in proficiency criteria.

Key Findings:
Positive relationship between inter-trial variability and proficiency in horizontal jumps. Inter-trial variability in distal joints is best predictor of FMS proficiency. Strong positive relationship between joint velocities and inter-trial variability. Consumer-level depth cameras may be suitable for computational assessment of FMS proficiency.

Key Contributions:
Investigated the use of consumer-level depth capture for computational assessment of jump proficiency. Established a positive relationship between inter-trial variability and jump proficiency. Identified increasing inter-trial variability within distal joints as a predictor of jump proficiency.

Figure 1. Overview of investigations presented within this thesis including the key findings and contributions of each work.
Summary

Fundamental movement skill (FMS) assessment remains central to the monitoring of skill development and ensuring that children are equipped with the necessary movement tools to engage in physical activity throughout their life (Barnett et al., 2009; Hardy, Reinten-Reynolds, Espinel, Zask, & Okely, 2012; Holfelder & Schott, 2014; Logan, Robinson, Rudisill, Wadsworth, & Morera, 2012). Primary school settings provide an ideal environment for the implementation of health and physical education (PE) programs that incorporate FMS development in the early years (Lander, Morgan, Salmon, & Barnett, 2016). Australian curricular requirements dictate that those administering PE programs within primary schools are also responsible for assessing the FMS proficiency of children. Primary and early childhood teachers responsible for the teaching and assessment of FMS experience significant barriers to assessment administration (Morgan & Hansen, 2008; Whipp, Hutton, Grove, & Jackson, 2011).

The majority of Australian teachers are not provided the required training, time, or resources to reliably assess FMS proficiency in the school environment (Haynes & Miller, 2014; Lander, Eather, Morgan, Salmon, & Barnett, 2017). Process-oriented assessments, including those recommended within school curricula, require teachers to interpret kinematic features of movements in relation to proficiency criteria (Lander et al., 2016; NSW Department of Education and Training, 2000; Western Australian Minister for Education, 2004). Significant time constraints and large class sizes dictate that teachers typically assess students in real-time in field-based environments where performances can only be viewed once (Barnett, Minto, Lander, & Hardy, 2014; Lander, Barnett, Brown, & Telford, 2015). This is in direct contrast to clinical or research environments, where skill performances are typically captured with video, allowing assessors to observe a single skill performance multiple times to ensure assessments are completed reliably (Cools, De Martelaer, Samaey, & Andries, 2009). Process-oriented tools, considered the ideal means of FMS assessment because of the valuable information they yield about particular skill deficiencies, are not ecologically viable in education environments (Lander et al., 2015). Current assessment techniques may not be appropriate in education settings, therefore, further development of alternative options for
FMS assessment is required (Furtado & Gallagher, 2012; Giblin, Collins, & Button, 2014; Longmuir et al., 2017).

At the commencement of the current project, a lack of time was reported as the greatest barrier to effective FMS assessment in schools (Morgan & Hansen, 2008) and the project was structured to investigate solutions that may provide rapid process-oriented assessment. Throughout the project, it was established that a lack of training, experience, and resources were further barriers to reliable and effective FMS assessment by teachers (Lander et al., 2017; Lander et al., 2016; Lander et al., 2015). It was also established that process-oriented FMS assessment tools such as the Test of Gross Motor Development-2 (TGMD-2) (Ulrich, 2000) and the Get Skilled Get Active (GSGA) (NSW Department of Education and Training, 2000) were reported as reliable and appropriate for school environments (Barnett et al., 2009; Ulrich, 2000). However, there was no evidence to suggest that assessors may not be able to reliably assess the presence of proficiency criteria during real-time assessment, as later established in Barnett et al. (2014).

We identified that advancements in consumer-level motion capture technologies, such as depth cameras, may provide an opportunity to reduce the subjectivity, time, and personnel costs of FMS assessment in educational environments. More specifically, the current project set out to investigate the use of consumer-level depth cameras for FMS proficiency assessments, with a clear goal of identifying the kinematic information required for computational assessment strategies. At the beginning of the project, depth cameras offered the most efficient extraction of kinematic information available to consumers (Hondori & Khademi, 2014; Pirsiavash, Vondrick, & Torralba, 2014). In particular, the Microsoft® Kinect® was emerging as a suitable motion capture system for gross movements, and its implementation for rehabilitation and skill monitoring was increasing within the literature (Hondori & Khademi, 2014; Jungong, Ling, Dong, & Shotton, 2013). Digital kinematic data captured by the Kinect® allows users to visually represent the movements in many ways, but the most common are point light (PL) displays. PL displays have been historically used for the investigation of biological motion perception due to the isolation of kinematic features within the displays (Blake & Shiffrar, 2007; Giblin, Farrow, Reid, Ball, & Abernethy, 2015). Humans are remarkably robust at observing kinematic information presented in PL displays, and can
accurately recognise actions being performed, as well as more complex recognition such as the gender, stature, or identity of PL performers (Blake & Shiffrar, 2007). It was expected that humans could readily assess the proficiency of FMS performances presented in PL displays.

The PL displays produced by the Kinect® are a direct visual representation of the kinematic data available to computers. This presented an opportunity to investigate the suitability of the capture system for the collection and human assessment of FMS, and the information processing strategies of human FMS assessments. An important process within the development of computational approaches, especially those employing artificial neural networks (ANN’s) that mimic human processes, is the reduction of input data to only key variables, minimising random variables considered for interpretation (Roweis & Saul, 2000). In investigating the human processing of FMS assessment, we aimed to identify the key kinematic information employed for proficiency assessments, to contribute towards future computational approaches.

At the time, this was the first project of its nature, and publishing the proposed protocols was key to ensuring a valid and valuable contribution to the development of computational FMS assessment techniques would be made. The proposed protocols for the project (Ward, Thornton, Lay, & Rosenberg, 2017) were developed around three assumptions. Firstly, that the high reliability of process-oriented FMS assessment tools reported in test manuals and research studies suggested that humans could reliably score subcomponents of assessments (i.e. proficiency criteria). Therefore, criterion level scoring would be an appropriate point of comparison between video and PL display styles. Secondly, that humans can reliably and accurately observe, recognise, and analyse movements presented in PL displays. Finally, that the Microsoft® Kinect® is an appropriate system for the capture of gross movements such as FMS. As presented within Ward et al. (2017), it was initially important to investigate the suitability of the Kinect® system for capture and assessment of FMS. It was hypothesised that the initial investigation would establish the suitability of the Kinect® and Kinect®-produced PL displays for FMS assessment. The then-proposed subsequent investigations were heavily focused on determining the information processing strategies of human FMS assessment, to provide a targeted approach to computational assessment strategies. Evidently, substantial deviations from the proposed protocols of the project outlined in Ward et al. (2017) were
made following our initial investigation, in which none of the initial assumptions and hypotheses were supported (Ward, Thornton, Lay, & Rosenberg, 2018) (Chapter 5).

Our initial investigation presented in Chapter 5 aimed to compare the scoring of FMS performances presented in video and PL displays. University students in the final year of their Sports Science degree assessed FMS performances in video and PL format using process-oriented techniques. The scoring of proficiency criteria was used to determine agreement between the presentation styles. Agreement between video and PL was calculated to establish whether the digitised kinematic data captured by the Kinect® was consistent with information employed by humans when viewing skills in real-time from video or in the field. Surprisingly, inconsistent recognition of proficiency criteria across presentation styles was observed. Subsequent analyses of assessment accuracy and reliability attempted to identify whether inconsistencies lay within the information presented in the displays, or within assessment scoring. Accuracy figures suggested that assessors had difficulty in scoring proficiency criteria in both styles, more so in PL displays. This may result from the single-camera nature of the Microsoft® Kinect® that can be affected by occlusion resulting in data loss (Hondori & Khademi, 2014). We looked at the distribution of error in assessments to determine whether skill information was likely to be missing in the PL displays. If data loss was an issue, we would have expected the error distribution to skew towards a high number of false negatives. If this was the situation, assessors would have scored proficiency criteria as absent when they were actually performed. Error distributions suggested that data loss was not likely an issue, and the low accuracy figures in PL were likely to be assessor inexperience, potentially compounded by the unfamiliar display style. It was not predicted that assessor experience would impact comparisons. However, it is likely that greater inaccuracies in assessment scoring introduced additional variability into comparisons between styles.

Assessment accuracy for the recognition of individual proficiency criteria is rarely considered in research studies. However, it was important that scoring was evaluated at this level to ensure Kinect® data allowed the identification of skill deficiencies. Most studies tend to investigate assessment reliability using performance scores, rather than accuracy of individual criteria (Barnett et al., 2009; Barnett et al., 2014; Valentini, 2012). When we
analysed assessments in this manner, reliability in both displays was high, suggesting that proficiency scoring can be reliable even if the identification of individual criteria is poor. Most analyses of FMS assessment report reliability of performance scores and assume that the recognition of proficiency criteria contributing to the performance score is accurate. The results of the current thesis suggest this may not always be the case for real-time assessments. It is likely that the low number of possible outcomes in performance scores (1-4) can result in high reliability, even in the absence of accurate criteria recognition. Barnett et al. (2014) similarly proposed that the reliability of performance scoring may not reflect the reliability of scoring proficiency criteria, however, criterion level accuracy was not analysed. It is often assumed that due to high levels of assessor training, criterion level scoring would be accurate and provide an appropriate foundation for reliability analyses, which may not be the case. As outlined in Ward et al. (2017), the aims of our initial investigation were to compare two presentation styles rather than investigating the reliability of assessment techniques limiting the generalisability of results. The observed discrepancies between reliability and accuracy figures came about through exploring why large inconsistencies were found between presentation styles and suggested that further investigations of process-oriented FMS assessment scoring were warranted.

The inconsistencies observed within our initial investigation prevented us from determining if the Kinect® was a suitable capture device for FMS assessment. It was difficult to ascertain whether the Kinect® data was the primary reason for inconsistencies or whether inconsistencies were due to the challenges of real-time FMS assessment. It therefore became important to decide whether subsequent investigations should focus on further establishing the Kinect® as an appropriate capture system or developing a greater understanding of human FMS assessment processes. The release of the Kinect® system in 2010 was positively received and readily implemented within the computer vision community, and considered paradigm changing technology in human pose estimation. Over the period of the current project, implementations of the Kinect system declined due to the emergence of augmented and virtual reality systems, and was fastened by Microsoft’s® discontinuation of the Kinect® in 2017 (Figure 2).
Concurrently, the computer vision community was developing sophisticated and efficient kinematic feature extraction from standard RGB video such as DensePose (Güler, Neverova, & Kokkinos, 2018), overcoming many of the limitations of depth cameras such as the Kinect®. In recognising that the feature extraction capabilities of Kinect® would in time be surpassed by more accessible technologies, it became less important for the project to investigate limitations within the capture system, and more important to investigate the poorly understood human process of FMS assessment. Whilst the Kinect®’s popularity was declining throughout the project; the system still provided the most efficient means to rapidly extract the kinematic information required for computational assessment strategies. Therefore, Kinect® data continued to be employed throughout subsequent investigations. As a result of the commercial activities of the Kinect® and the wider computer-vision community, the thesis became more focused on investigating human assessment processes and identifying the shortcomings of human assessment.

Building upon the findings in Chapter 5, Chapters 6 and 7 aimed to ascertain whether scoring patterns observed in the initial investigation persisted within assessors of higher experience, and the particular observation strategies employed by experienced and inexperienced assessors. The strength of process-oriented assessment techniques lies within the identification of specific skill deficiencies. Chapters 6 and 7 investigated whether the accurate identification of proficiency criteria is possible during real-time FMS assessment, and if so,
what observational strategies or kinematic information were assessors employing. It was also valuable to continue to investigate the assessment of FMS from digitised kinematic data, as this is ultimately required to inform computational assessment approaches.

The results of Chapter 6 (Ward, Thornton, Lay, Chen, & Rosenberg, 2018a) suggested that high reliability in performance level scoring typically reported in studies of FMS assessment, may not capture inaccuracies in identifying proficiency criteria. Even though the reliability and accuracy of performance scoring was high for both inexperienced and experienced assessors, accuracy figures in scoring proficiency criteria were around 70%. This suggested that the identification of proficiency criteria during real-time FMS assessment is problematic, even for highly trained assessors. The results further establish a need for more appropriate real-time FMS assessment tools. Interestingly, inexperienced and experienced assessors were equally accurate and reliable in both video and PL assessments. The inexperienced assessor sample comprised primary teachers, and we suggest that even though few had undertaken formal assessment training or experience, the frequent exposure to FMS during PE and active play may have led to some level of domain expertise, decreasing the impact of unfamiliar PL displays that was observed in Chapter 5 (Giblin et al., 2015; Sparrow & Sherman, 2001).

The results of this investigation also showed that assessors had difficulty in observing proficiency criteria, although it was difficult to ascertain whether particular criteria were harder to judge than others. When accuracy was explored at the individual criterion level (i.e. Jump Criterion 3: Takes off and lands on both feet simultaneously), the results suggested that assessors have greater difficulty in scoring some criteria, particularly those that are more subjective. The accuracy results for recognition of individual criteria in Chapter 6 were supported by the variable reliability figures in Barnett et al. (2014). Interestingly, the average accuracy for some criteria was around 50%, which in a binary classification is the likelihood of correctly guessing. Whilst we do not suggest that assessors were guessing, we believe that the attentional demands of real-time FMS assessment may require assessors to estimate criterion presence in some circumstances. The results suggest this may be no better than guessing at the individual criterion level, a problematic situation for any clinical assessment tool.
During real-time FMS assessment, assessors are required to attend to multiple proficiency criteria simultaneously within a small time period, which is perceptually difficult (Furtado & Gallagher, 2012). The results presented within this thesis suggest that assessors make a judgement of proficiency based upon their observation of the entire performance, and subsequently score criteria according to their overall proficiency judgement (Chong, Familiar, & Shim, 2016). Whilst some criteria may be observed and therefore accurately scored, the scoring of others may be estimated based upon an overall impression of proficiency. This effect is described in perception research as ‘filling-in’ (Chong et al., 2016; Komatsu, 2006). Under these circumstances, high accuracy and reliability can be observed in performance scores, even if the recognition of individual criteria is inaccurate. Whilst this may not limit process-oriented assessments in monitoring overall skill proficiency, it does impact the reliable use of criterion level scoring to identify specific skill deficiencies, which is an important consideration for those using FMS assessments.

The belief that assessors aren’t able to observe all proficiency criteria during real-time FMS assessments was supported by the visual search strategies employed by assessors in Chapter 7 (Ward, Thornton, Lay, Chen, & Rosenberg, 2018b). When the fixation locations of assessors were analysed in relation to anatomical locations relevant to proficiency criteria, we observed that in many cases assessors were not attending to the relevant location at the relevant time. For locomotor skills, assessors anchored their gaze around proximal locations such as the hips and torso, and likely used information captured by peripheral vision to infer the presence of proficiency criteria relating to more distal segments such as arms and feet. For object control skills, assessors anchored gaze on segments which interact with the object, and therefore likely considered these locations to provide the greatest proficiency information and observational efficiency. If assessors were consciously looking for each individual criterion during skill performances, we would expect different visual search strategies to those we observed (van Zoest & Donk, 2004). The results suggested that assessors adopt a more efficient search strategy by prioritising attendance to anatomical locations that provide the greatest proficiency information, and relying upon peripheral sources of information to infer the presence of criteria that fall outside of those locations (Abernethy & Zawi, 2007). Rather than relying upon process-oriented observation structures, assessors may be implicitly attuned to observing and making judgement on what constitutes a proficient FMS
performance (Giblin et al., 2015; Sparrow & Sherman, 2001). It is likely that the observation strategies employed by humans are influenced by historical skill observations combined with limited information processing capacity. The visual search strategies employed by assessors provide some justification for high reliability being observed in overall proficiency assessment even when criterion recognition is inaccurate.

If assessors do not rely upon the observation and judgement of proficiency criteria to determine the proficiency of FMS performances, the reliable identification of skill deficiencies from proficiency criteria scoring becomes problematic. Assessment developers often consider the trade-off between having too many proficiency criteria that scoring becomes perceptually difficult, and too few that movement patterns for proficiency are not captured (Furtado & Gallagher, 2012). The results of this thesis suggest that proficiency criteria should ensure assessment reliability, and assessor training should ensure assessors can accurately identify proficiency criteria (Barnett et al., 2014). Interestingly, we observed no differences in criterion level accuracy, or visual search patterns between experienced and inexperienced assessors. There were also no differences in visual search patterns between high and low accuracy assessors. These results further evidence that observation strategies were independent of assessors’ ability to accurately recognise criteria. The findings in this thesis challenge the ease of designing specific training strategies to improve the accuracy of proficiency criteria identification during real-time FMS assessment. Furthermore, even with a suitable training strategy, our results suggest the amount of training required to overcome the shortcomings of real-time FMS assessment is unlikely to be viable within the time constraints of educational environments. The investigations within the thesis were designed to identify human information processing during FMS assessment to inform computational approaches, and they highlight that human assessment strategies may not provide a suitable foundation for computational assessment.

Based upon early findings in this PhD and supported by the findings of Barnett et al. (2014), we hypothesised that an alternative data-driven approach to proficiency classification might produce similar results to the current approach of human based assessments. Whilst investigating human assessment processes we also investigated a data driven approach to classifying horizontal jump proficiency. The work presented in Chapter 8 (Ward, Thornton,
Lay, Huynh, et al., 2018) was an investigation into a computational assessment of FMS proficiency using depth capture. We investigated whether inter-trial variability could classify skill proficiency in horizontal jump performances captured with the Microsoft® Kinect®. The investigation was the first to examine whether inter-trial variability calculated from digital kinematic data captured using consumer-level depth capture could be used to rapidly classify FMS proficiency. Results showed a positive relationship between inter-trial variability and jump proficiency, suggesting that kinematic depth data is appropriate for computational proficiency assessment. In contrast to the proximal sourcing of information employed by humans in assessing jump proficiency, distal joint data was shown to have the strongest statistical relationship with proficiency when considering inter-trial variability of velocity. The investigation highlighted a potential data driven FMS assessment technique, and provided further evidence that the mimicry of human assessment processes may not provide the most suitable framework for computational assessment approaches. The results are supported by Sgro, Nicolosi, Schembri, Pavone, and Lipoma (2015) who employed the Kinect® to investigate the relationship between vertical jump outcome measures and developmental stages. More recently, Bisi, Pacini Panebianco, Polman, and Stagni (2017) successfully employed inertial measurement units (IMU’s) to computationally quantify the presence of Test of Gross Motor-2 (Ulrich, 2000) proficiency criteria during locomotor skill performances, representing significant advancement in the field of computational FMS assessment. The results of this thesis confirm this area of research holds potential to provide rapid computationally driven FMS assessments suitable for teachers in an educational setting. One area for further exploration is artificial neural networks (ANN’s) that allow computational ‘perceptions’ of the movement patterns that determine proficiency, utilising a large and varied input data set to provide robust and reliable classification.

Conclusions
This thesis contributes new knowledge in the investigation of the human process of FMS assessment that inform future computational approaches. This thesis was the first to identify the challenges of assessing subcomponents of process-oriented assessment structures during real-time assessment, even by experienced assessors. This thesis also provides new evidence
that typical analyses of FMS assessment reliability may not capture inaccuracies in individual proficiency criteria. A key finding in this thesis is that assessors are limited in their ability to observe all proficiency criteria during real-time skill observation, regardless of their assessment experience. The efficacy of FMS teaching programs in primary schools relies upon accurate and reliable assessment of motor development and FMS deficiencies through teachers observations of proficiency criteria. It is therefore paramount that the development of reliable and appropriate FMS assessment tools for environments that require time and resource efficiency, such as primary education settings becomes a priority. At its inception, this thesis aimed to provide a human-modelled template for computational assessment approaches. Conversely, the results of the thesis provide evidence for future developments to focus on data driven approaches that capitalise on advancements in computer vision and data processing capabilities.

**Recommendations for Future research**

Following the line of research presented within this thesis, four key directions have been identified for future research within the field of FMS assessment, including both human and computational assessment strategies. These are:

1) Working towards agreement on ideal means for assessing motor competence
2) Understanding the capabilities of human assessors in field-based environments
3) The development of machine learning approaches using human assessment data
4) Investigating purely computational approaches to FMS assessment (unsupervised machine learning)

One of the current challenges for those assessing FMS is choosing the correct assessment tool for the given implementation. Publications such as that by Bardid, Vannozzi, Logan, Hardy, and Barnett (2018) provide some assistance in selecting appropriate assessment techniques, but unanimous agreement on ideal tools does not exist. Even within a single context such as field-based FMS assessment, there are a large number of assessment tools available, with none accepted as a gold standard. This is likely to become a limiting factor in global monitoring of FMS development and the collation of FMS competence data. There are a number of new FMS assessment tools and techniques developed every year which can
contribute to confusion in choosing the most appropriate assessment approach. As such, a focus within the field should be a move towards agreement amongst academics on the ideal means to assess motor competence, and the establishment of gold standard assessment tools and techniques.

If process-oriented assessments continue to be implemented by human assessors in field-based environments, the number of criteria that can be accurately perceived during real-time FMS assessment should be established. The impact of human attentional capacities on real-time FMS assessment has been established within this thesis, but the investigation of perceptual ability during FMS assessment was by no means exhaustive. Future research would benefit from identifying the maximum number of observable criteria as a basis for inclusion or clarification of proficiency criteria. It has been suggested that training should increase to minimise the variability in criteria interpretation (Barnett et al., 2014), however increased clarity in criteria identification may not improve reliability if the observation of criteria is not possible. Subjectivity and limited attentional capacity affect the accuracy and value of process-oriented assessment results and developments should therefore consider the maximum perceivable criteria in addition to the interpretability of criteria.

Assessment developments should also continue to make use of advancements in technology, whether it be assistive technology to aid human assessment or computational assessment that removes the need for human observation. In the first instance assessments may be translated onto portable devices such as tablets and smartphones which allow assessors to capture, store and assess performances on a single device, minimising the need for capture and post hoc uploading. Whilst this may increase the efficiency of capture and observation of performances from video, it needs to be established if this approach would realistically overcome the time constraints experienced in education environments. Any means of process-oriented assessment requiring human assessors is also still likely to be impacted by the subjectivity of human observation. An alternative approach would involve removing human subjectivity by computationally assessing proficiency. There are multiple means by which computational analyses could evolve including both traditional programming and machine learning.
A traditional programming approach would involve humans developing algorithms that reflect or mimic current process-oriented assessment tools. An example of this is the computational scoring of the TGMD-2 locomotor subset (Ulrich, 2000) developed by Bisi et al. (2017), whereby a set of algorithms were manually programmed to recognise when kinematic data represented the successful performance of a proficiency criterion. Bisi et al. (2017) noted that developers should consider the number of sensors used, to ensure efficiency of assessment protocols and post-hoc upload time. Computer vision capabilities have advanced to allow more sophisticated kinematic feature extraction from RGB video such as DensePose (Güler et al., 2017). This may allow the approach employed by Bisi et al. (2017) to be reproducible using kinematic data extracted from RGB video which employs a single sensor rather than multiple IMU’s.

An alternative approach is to allow trends in data to determine the required kinematic information to provide a reliable proficiency classification as in Ward, Thornton, Lay, Huynh, et al. (2018) (Chapter 8), or in a more sophisticated manner, by employing ANN’s. Rather than being provided with strict parameters or rules, ANN’s automatically generate rules by ‘learning’ complex trends or relationships between input data and the desired output. An example of this would be an image recognition ANN ‘learning’ to recognise an image of a dog. By analysing a large set of images correctly labelled as dog or not-dog the ANN will automatically generate identifying features based on trends in the input images, rather than being manually programmed to recognise that a dog has a tail, fur and four-legs. Similar to a human brain ANN’s change their interpretation depending on the input information and therefore develop rules to correctly distinguish between inputs. An ANN could recognise that a picture of a cat is not a picture of a dog even though both have a tail, fur and four-legs, where traditional programming may struggle.

In an FMS assessment context, an ANN may be able to develop a set of rules to classify FMS proficiency based upon a large set of labelled performances. This would reflect a supervised learning approach whereby humans assign labels to input data and the ANN develops a set of functions that match input data to their assigned labels. This approach relies upon human input throughout development, whether it is the assignment of proficiency labels to performances or the programming of rules to assess proficiency criteria. In either case, it is
important the human input provides the most accurate proficiency classifications to ensure computational assessments provide accurate and reliable outputs.

The assessment of FMS performances could be considered to increase in sophistication from binary proficiency classification, to comparisons with the final stage of skill development, to placement upon the developmental pathway. It is important that in developing ANN’s to assess FMS, the labels assigned to performances accurately represent the desired output so that the error introduced during ‘training’ is minimised. In circumstances where ANN’s are designed to perform classification tasks such as recognising an image as a dog or not, the labelling of input images is unlikely to be affected by human subjectivity. Likewise, it is likely that the labelling of FMS performances into high and low proficiency is a straightforward task for humans. In being a relatively straightforward task for untrained human assessors it is unlikely that a computational assessment that only provides binary classification of FMS performances into high or low proficiency would be sophisticated enough provide significant benefits to assessors. This is an important consideration for those developing computational assessment techniques.

Another machine learning approach is to allow computers to determine the proficiency of FMS performances without labels assigned by human assessors. Unsupervised machine learning approaches draw inferences based upon trends in input data without labels or classifications. This approach can unearth patterns or clusters based upon exposure to large datasets. Although possibly not suitable for FMS assessment, it is important that FMS researchers consider how advances in data processing may assist FMS assessment.

Objective and reliable computational FMS assessment will have a remarkable impact upon the collection of FMS proficiency data in education settings. Employing accessible capture technologies that allow real-time proficiency classification and feedback will greatly reduce the burden upon teachers, improving the likelihood of effective and regular skill assessment within education environments. Engaging computational assessment systems may improve the efficiency of identifying and monitoring population-level trends in FMS development. FMS development significantly contributes to establishing engagement in physical activity across the lifespan and negating the poor outcomes of low FMS competency. Progressions in
computer vision and data processing capabilities are providing opportunities for efficient FMS assessment which may allow monitoring at the population level and should therefore be the focus for development of objective and reliable FMS assessment techniques.
References


Our Ref: RA/4/1/7445

11 May 2015

Associate Professor Michael Rosenberg
School of Sport Science, Exercise and Health
MBDP: M408

Dear Professor Rosenberg

HUMAN RESEARCH ETHICS APPROVAL - THE UNIVERSITY OF WESTERN AUSTRALIA

Development of an Assessment for Fundamental Movement Skills Captured Using 3-Dimensional Marker Less Motion Capture Technology

Student(s): Brodie Ward - PhD - 20772149

Ethics approval for the above project has been granted in accordance with the requirements of the National Statement on Ethical Conduct in Human Research (National Statement) and the policies and procedures of The University of Western Australia. Please note that the period of ethics approval for this project is five (5) years from the date of this notification. However, ethics approval is conditional upon the submission of satisfactory progress reports by the designated renewal date. Therefore initial approval has been granted from 11 May 2015 to 01 May 2016.

You are reminded of the following requirements:

1. The application and all supporting documentation form the basis of the ethics approval and you must not depart from the research protocol that has been approved.
2. The Human Ethics office must be approached for approval in advance for any requested amendments to the approved research protocol.
3. The Chief Investigator is required to report immediately to the Human Ethics office any adverse or unexpected event or any other event that may impact on the ethics approval for the project.
4. The Chief Investigator must submit a final report upon project completion, even if a research project is discontinued before the anticipated date of completion.

Any conditions of ethics approval that have been imposed are listed below:

Special Conditions
None specified

The University of Western Australia is bound by the National Statement to monitor the progress of all approved projects until completion to ensure continued compliance with ethical principles.

The Human Ethics office will forward a request for a Progress Report approximately 30 days before the due date.

If you have any queries please contact the Human Ethics office at humanethics@uwa.edu.au.

Please ensure that you quote the file reference – RA/4/1/7445 – and the associated project title in all future correspondence.

Yours sincerely
Appendix B – Parent and Child information and consent (FMS performance collection)

---

**Investigating the relationship between movement variability and movement proficiency**

— Parent Information Sheet —

**PURPOSE**

The purpose of this study is to investigate whether a link exists between movement variability and movement proficiency in children.

**PROCEDURES**

The study requires participants to attend a single testing session at the School of Sport Science Exercise and Health at the University of Western Australia. The session will take approximately half an hour for each participant that attends. Within the session participants will be asked to perform 10 trials of 4 different movement skills (jump, hop, throw, kick). Movement skills will be recorded using 3-D motion capture cameras for analysis. Participants will also undergo a standardised movement skill assessment to determine their movement proficiency levels at the time of collection.

Given that the protocol does require some moderate physical activity, participants will need to wear comfortable clothes and enclosed footwear.

Any additional information can be obtained from the person conducting the testing session.

**RISKS**

The study testing procedure presents minimal risk to participants, however should your child have any current injuries or issues that may affect their ability to perform the movement skills, please notify the researcher prior to the testing session.

**BENEFITS**

The benefit of this study is to provide researchers with knowledge regarding the link between movement variability and movement proficiency in children. This information will be used to develop more efficient and accurate ways to assess children’s movement skills. It will also provide evidence for the use of marker less motion capture technology in the field of FMS assessment.
CONFIDENTIALITY

Your child’s confidentiality will be maintained throughout the study. All movement data will be de-identified before it is disseminated in any manner. All data obtained from testing will be stored on a secure external hard drive, accessible only to relevant research personnel and will only be released with your informed consent.

WHAT HAPPENS TO THE DATA?

Your child’s movement data will be used to determine the link between movement variability and movement proficiency in children. If further consent is given, de-identified recordings of your child’s performances may be used in other studies within the same research project.

SUBJECT RIGHTS

All children are volunteers and can choose to withdraw from the study at any time without prejudice. Parents can observe their children undertaking the testing session and can request their child be withdrawn from the study at any time without prejudice.

If you do withdraw your child we may wish to retain the data that we have recorded from you but only if you agree, otherwise your records will be destroyed.

Your child’s participation in this study does not prejudice any right to compensation that you may have under statute of common law.

If you have any questions concerning the research at any time please feel free to ask the researcher who has contacted you about your concerns. Further information regarding this study may be obtained from:

Brodie Ward
brodie.ward@uwa.edu.au

The Human Research Ethics Committee at the University of Western Australia requires that all participants are informed that, if they have any complaint regarding the manner, in which a research project is conducted, it may be given to the researcher or, alternatively to the Secretary, Human Research Ethics Committee, Registrar’s Office, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 (telephone number 6488-3783). All study participants will be provided with a copy of the Information Sheet and Consent Form for their personal records.

Mr Brodie Ward
PhD Candidate
School of Sport Science, Exercise and Health
Ph: 9449 191 783
brodie.ward@uwa.edu.au

Dr Brendan Lay
Associate Professor
School of Sport Science, Exercise and Health
brendan.lay@uwa.edu.au

Dr Michael Rosenberg
Associate Professor
School of Sport Science, Exercise and Health
michael.rosenberg@uwa.edu.au

Dr Ashleigh Thornton
Research Assistant
School of Sport Science, Exercise and Health
asleigh.thornton@uwa.edu.au
Dear Participants,

Before you become a part of our study, there are a few things we need to make sure you are aware of:

You will be asked to come to the movement assessment room at the university for 30 minutes to have your movement tested. While you are here one of our research team will ask you to perform four skills. They are:

1. Hop
2. Throw
3. Kick
4. Jump

You will be asked to do each of these skills three ten while you are being recorded. The video will be then taken back to UWA and the researchers will assess your movements.

You will also be asked to for an outside movement assessment, which includes 12 skills, but you only do each one twice.

Participant Consent Statement

- I understand why and how the study is being done
- I understand that the study may not benefit me directly.
- I understand that my involvement is my choice (voluntary) and I can withdraw at any time without affecting my relationship with the school or the research team.
- I understand that no information like my name and address (personal identifying information) will be used and that all information including videos will be securely stored for 7 years before being destroyed.
- I understand that I need to sign another consent form if video of me is to be used for presentation purposes of research findings.
- I have been allowed to ask any questions I have.

☐ I (your name) ___________________________ agree to be included in the Uni-Active study.

Your Signature: ___________________________

Date: ___/___/2016

Please fill out and sign this form and hand it back to one of the Study Team.
CONSENT FORM
For Parents

- I have read and understood the information provided about the research or have had it explained to me in language I understand.

- I have taken up the invitation to ask any questions I may have had and am satisfied with the answers.

- I understand that my child's participation in the project is entirely voluntary.

- I have discussed with my child what participation in this research project involves and they have agreed to participate.

- I understand that I am free to withdraw him/her at any time without reason and without prejudice.

- I have been advised as to what type of data will be collected.

- I understand that this research project is being conducted independently by The University of Western Australia and is not a part of school activities.

- I understand that my child will be video recorded in order to assess their movement and this footage will only be used for analysis.

- I understand that I will receive a report of the assessment results within four weeks of the testing.

- I understand that my child’s school will also receive a report on the assessment results of my child.

- I understand that all information I provide is treated as strictly confidential.

- I agree that the information gathered for the study may be published provided my child’s name or other identifying information is not used.

- I understand that I can contact the research team and discuss the assessment results if required.

☐ Yes, I (full name of parent/caregiver) ______________________ give permission for my child (full name) ______________________ to participate in the Uni-Active study.

Signature of parent/caregiver: ______________________

Date: __/___/2016

Please complete and return on your assessment day
Appendix C – FMS assessment information and consent (Chapter 5)

The effect of presentation method on performance criteria recognition during fundamental movement skill assessment

— Subject Information Sheet —

PURPOSE

The purpose of this study is to determine the effect of presentation style on the recognition of fundamental movement skill (FMS) performance criteria.

PROCEDURES

You will be required to attend one testing session of approximately one hour at the School of Sport Science Exercise and Health at the University of Western Australia. Your objective for the length of the session is to assess FMS performances using the Test of Gross Motor Development-2 (Ulrich, 2000). The FMS performances will be displayed on an LCD monitor. Over the testing period you will be required to assess 128 FMS performances in addition to 8 practise trials prior to testing.

You will be offered remuneration for your time in the form of a beverage or snack to the value of $10 from a local establishment.

Any additional information can be obtained from the person conducting the testing session.

Risks

Your exposure to FMS performance videos during the testing procedure presents minimal risk.

Benefits

The benefit of this study is to provide researchers with knowledge regarding the effect of presentation style on FMS assessment. It will also provide evidence for the use of marker less motion capture technology in the field of FMS assessment.
CONFIDENTIALITY

Your confidentiality will be maintained throughout the study. You will be randomly assigned a number to de-
indentify their data. All data obtained from testing will be stored on a secure external hard drive, accessible
only to relevant research personnel and will only be released with your informed consent.

SUBJECT RIGHTS

Participation in this research is voluntary and you are free to withdraw from the study at any time without
prejudice. You can withdraw for any reason and you do not need to justify your decision. If you do withdraw
we may wish to retain the data that we have recorded from you but only if you agree, otherwise your records
will be destroyed.

Your participation in this study does not prejudice any right to compensation that you may have under statute
of common law.

If you have any questions concerning the research at any time please feel free to ask the researcher who
has contacted you about your concerns. Further information regarding this study may be obtained from:

The Human Research Ethics Committee at the University of Western Australia requires that all participants are informed that, if they
have any complaint regarding the manner, in which a research project is conducted, it may be given to the researcher or, alternatively
to the Secretary, Human Research Ethics Committee, Registrar’s Office, University of Western Australia, 35 Stirling Highway, Crawley,
WA 6009 (telephone number 6880-3703). All study participants will be provided with a copy of the Information Sheet and Consent Form
for their personal records.

Mr Brodie Ward
PhD Candidate
School of Sport Science, Exercise and Health
Ph: 0449 191 783
brodie.ward@uwa.edu.au

Dr Brendan Lay
Associate Professor
School of Sport Science, Exercise and Health
brendan.lay@uwa.edu.au

Dr Michael Rosenberg
Associate Professor
School of Sport Science, Exercise and Health
michael.rosenberg@uwa.edu.au

Dr Ashleigh Thornton
Research Assistant
School of Sport Science, Exercise and Health
asleigh.thornton@uwa.edu.au
The effect of presentation method on performance criteria recognition during fundamental movement skill assessment.

— Consent Form —

I have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time without reason and without prejudice.

I understand that all information provided is treated as strictly confidential and will not be released by the investigator unless required to by law. I have been advised as to what data is being collected, what the purpose is, and what will be done with the data upon completion of the research.

I agree that research data gathered for the study may be published provided my name or other identifying information is not used.

☐ I agree to allow photos and/or video footage of myself to be included in academic presentations and publications under the following conditions (please list if necessary):

__________________________________________________________

Participant’s Signature  Date

The Human Research Ethics Committee at the University of Western Australia requires that all participants are informed that, if they have any complaint regarding the manner, in which a research project is conducted, it may be given to the researcher or, alternatively to the Secretary, Human Research Ethics Committee, Registrar’s Office, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 (telephone number 6488-3703). All study participants will be provided with a copy of the Information Sheet and Consent Form for their personal records.
Participant Information Form

Information processing during fundamental movement skill assessment

PURPOSE

The purpose of this study is to determine the key information used during fundamental movement skill (FMS) assessment.

PROCEDURES

You will be required to attend one testing session of approximately 1 hour at the School of Psychology Eye-tracking Lab at the University of Western Australia.

Your objective is to assess FMS performances using the Test of Gross Motor Development-2, a validated FMS assessment tool. During the session, your eye movements will be recorded using an EyeLink1000 mobile eye tracker. The nature of the eye tracker will require your head to be rested in a chin rest for the protocol. The FMS performances will be displayed on an LCD monitor.

Any additional information can be obtained from the person conducting the testing session.

Risks

Your exposure to FMS performance videos during the testing procedure presents minimal risk. The eye tracking technology is non-invasive. If the chin rest causes any irritation the testing procedures will be stopped immediately.

Benefits

The benefit of this study is to provide researchers with the key information used by assessors during FMS assessment. This information will be used to develop a new FMS assessment.
CONFIDENTIALITY

Your confidentiality will be maintained throughout the study. You will be randomly assigned a number to de-identify your data. All data obtained from testing will be stored on a secure external hard drive, accessible only to relevant research personnel and will only be released with your informed consent.

SUBJECT RIGHTS

Participation in this research is voluntary and you are free to withdraw from the study at any time without prejudice. You can withdraw for any reason and you do not need to justify your decision. If you do withdraw we may wish to retain the data that we have recorded from you but only if you agree, otherwise your records will be destroyed. Your participation in this study does not prejudice any right to compensation that you may have under statute of common law.

If you have any questions concerning the research at any time please feel free to ask the researcher who has contacted you about your concerns. Further information regarding this study may be obtained from:

Mr Brodie Ward  
PhD Candidate  
School of Sport Science, Exercise and Health  
Ph: 0449 191 783  
brodie.ward@uwa.edu.au

Dr Ashleigh Thornton  
Research Associate  
School of Sport Science, Exercise and Health  
asheigh.thornton@uwa.edu.au

Dr Michael Rosenberg  
Associate Professor  
School of Sport Science, Exercise and Health  
michael.rosenberg@uwa.edu.au

Dr Nigel Chen  
Research Associate  
School of Psychological Sciences  
nigel.chen@uwa.edu.au

Dr Brendan Lay  
Associate Professor  
School of Sport Science, Exercise and Health  
brendan.lay@uwa.edu.au

The Human Research Ethics Committee at the University of Western Australia requires that all participants are informed that, if they have any complaint regarding the manner, in which a research project is conducted, it may be given to the researcher or, alternatively to the Secretary, Human Research Ethics Committee, Registrar’s Office, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 (telephone number 6488-3703). All study participants will be provided with a copy of the Information Sheet and Consent Form for their personal records.
Participant Consent Form

Information processing during fundamental movement skill assessment

I, __________________________ have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this research project, realising that I may withdraw at any time without reason and without prejudice.

I understand that all identifiable information that I provide is treated as confidential and will not be released by the investigator in any form that may identify me unless I have consented to this. The only exception to this principle of confidentiality is if this information is required by law to be released.

☐ I agree to allow photos and/or video footage of myself to be included in academic presentations and publications under the following conditions (please list if necessary):

________________________________________________________________________

________________________________________  _______________
Participant signature                      Date

Approval to conduct this research has been provided by the University of Western Australia, in accordance with its ethics review and approval procedures. Any person considering participation in this research project, or agreeing to participate, may raise any questions or issues with the researchers at any time.

In addition, any person not satisfied with the response of researchers may raise ethics issues or concerns, and may make any complaints about this research project by contacting the Human Ethics Office at the University of Western Australia on (08) 6488 3703 or by emailing to humanethics@uwa.edu.au

All research participants are entitled to retain a copy of any Participant Information Form and/or Participant Consent Form relating to this research project.

## TGMD–2

### Test of Gross Motor Development–Second Edition

### Profile/Examiner Record Form

#### Section I: Identifying Information

Name ____________________________  
School ____________________________

Male □ Female □  Grade ____________________________

Date of Testing ____________________________  
Reason for Referral ____________________________

Date of Birth ____________________________  
Examiner ____________________________

Age ____________________________  
Examiner’s Title ____________________________

#### Section II: Record of Scores

**First Testing**

<table>
<thead>
<tr>
<th>Locomotor Object Control</th>
<th>Raw Score</th>
<th>Standard Score</th>
<th>Percentile</th>
<th>Age Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sum of Standard Scores Gross Motor Quotient** ____________________________

**Second Testing**

<table>
<thead>
<tr>
<th>Locomotor Object Control</th>
<th>Raw Score</th>
<th>Standard Score</th>
<th>Percentile</th>
<th>Age Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sum of Standard Scores Gross Motor Quotient** ____________________________

#### Section III: Testing Conditions

**A. Place Tested**

**Interfering**  

<table>
<thead>
<tr>
<th>Interfering</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Noise Level</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>C. Interruptions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>D. Distractions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>E. Light</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>F. Temperature</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Not Interfering**  

<table>
<thead>
<tr>
<th>Interfering</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Noise Level</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>C. Interruptions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>D. Distractions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>E. Light</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>F. Temperature</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**G. Notes and other considerations**

---

#### Section IV: Other Test Data

<table>
<thead>
<tr>
<th>Name of Test</th>
<th>Date</th>
<th>Standard Score</th>
<th>TGMD–2 Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Additional copies of this form (#952D) may be purchased from:  
PRO-ED, 8700 Shadel Creek Blvd., Austin, TX 78757-6869  
800/997-1202 fax 800/997-7633 www.proed.com

Copyright © 2000, 1995 by PRO-ED, Inc.  
# 9 10 11 12 13 14 15 16 17 18 19 20 21 22
### Subtest Performance Record

**Preferred Hand:**  Right  □  Left  □  Not Established  □  
**Preferred Foot:**  Right  □  Left  □  Not Established  □  

#### Locomotor Subtest

<table>
<thead>
<tr>
<th>Skill</th>
<th>Materials</th>
<th>Directions</th>
<th>Performance Criteria</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Score</th>
</tr>
</thead>
</table>
| Run     | 60 feet of clear space, and two cones | Place two cones 50 feet apart. Make sure there is at least 9 to 10 feet of space beyond the second cone for a safe stopping distance. Tell the child to run as fast as he or she can from one cone to the other when you say "Go." Repeat a second trial. | 1. Arms move in opposition to legs, elbows bent  
2. Brief period where both feet are off the ground  
3. Narrow foot placement landing on heel or toe (i.e., not flat footed)  
4. Non-support leg bent approximately 90 degrees (i.e., close to buttocks) |         |         |       |
| Gallop  | 25 feet of clear space, and tape or two cones | Mark off a distance of 25 feet with two cones or tape. Tell the child to gallop from one cone to the other. Repeat a second trial by galloping back to the original cone. | 1. Arms bent and lifted to waist level at takeoff  
2. A step forward with the lead foot followed by a step with the trailing foot to a position adjacent to or behind the lead foot  
3. Brief period when both feet are off the floor  
4. Maintains a rhythmic pattern for four consecutive gallops |         |         |       |
| 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 non-support leg remains behind body  
3. Arches 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 non-preferred foot |         |         |       |
| Leap    | A minimum of 20 feet of clear space, a beanbag, and tape | Place a beanbag on the floor. Attach a piece of tape on the floor so it is parallel to and 10 feet away from the beanbag. Have the child stand on the tape and run up and leap over the beanbag. Repeat a second trial. | 1. Take off on one foot and land on the opposite foot  
2. A period where both feet are off the ground longer than running  
3. Forward reach with the arm opposite the lead foot |         |         |       |
<table>
<thead>
<tr>
<th>Skill</th>
<th>Materials</th>
<th>Directions</th>
<th>Performance Criteria</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Horizontal Jump</td>
<td>A minimum of 10 feet of clear space and tape</td>
<td>Mark off a starting line on the floor. Have the child start behind the line. Tell the child to jump as far as he or she can. Repeat a second trial.</td>
<td>1. Preparatory movement includes flexion of both knees with arms extended behind body. 2. Arms extend forcefully forward and upward reaching full extension above the head. 3. Take off and land on both feet simultaneously. 4. Arms are thrust downward during landing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Slide</td>
<td>A minimum of 25 feet of clear space, a straight line, and two cones</td>
<td>Place the cones 25 feet apart on top of a line on the floor. Tell the child to slide from one cone to the other and back. Repeat a second trial.</td>
<td>1. Body turned sideways so shoulders are aligned with the line on the floor. 2. A step sideways with lead foot followed by a slide of the trailing foot to a point next to the lead foot. 3. A minimum of four continuous step-slide cycles to the right. 4. A minimum of four continuous step-slide cycles to the left.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Object Control Subtest**

<table>
<thead>
<tr>
<th>Skill</th>
<th>Materials</th>
<th>Directions</th>
<th>Performance Criteria</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Striking a Stationary Ball</td>
<td>A 4-inch lightweight ball, a plastic bat, and a batting tee</td>
<td>Place the ball on the batting tee at the child's belt level. Tell the child to hit the ball hard. Repeat a second trial.</td>
<td>1. Dominant hand grips bat above nondominant hand. 2. Nonpreferred side of body faces the imaginary torser with feet parallel. 3. Hip and shoulder rotation during swing. 4. Transfers body weight to front foot. 5. Bat contacts ball.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Stationary Dribble</td>
<td>An 8- to 10-inch playground ball for children ages 3 to 5; a basketball for children ages 6 to 10; and a flat, hard surface</td>
<td>Tell the child to dribble the ball four times without moving his or her feet, using one hand, and then stop by catching the ball. Repeat a second trial.</td>
<td>1. Contacts ball with one hand at about belt level. 2. Pushes ball with fingertips (not a slap). 3. Ball contacts surface in front of or to the outside of foot on the preferred side. 4. Maintains control of ball for four consecutive bounces without having to move the feet to retrieve it.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Skill Score**

Locomotor Subtest Raw Score (sum of the 6 skill scores)
<table>
<thead>
<tr>
<th>Skill</th>
<th>Materials</th>
<th>Directions</th>
<th>Performance Criteria</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Catch</td>
<td>A 4-inch plastic ball, 15 feet of clear space, and tape</td>
<td>Mark off two lines 15 feet apart. The child stands on one line and the tosser on the other. Toss the ball underhand directly to the child with a slight arm aim and hold for his or her chest. Tell the child to catch the ball with both hands. Only count those tosses that are between the child's shoulders and belt. Repeat a second trial.</td>
<td>1. Preparation phase where hands are in front of the body and elbows are flexed 2. Arms extend while reaching for the ball as it arrives 3. Ball is caught by hands only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Kick</td>
<td>An 8- to 18-inch plastic, playground, or soccer ball; a beanbag; 30 feet of clear space, and tape</td>
<td>Mark off one line 30 feet away from a wall and another line 20 feet from the wall. Place the ball on top of the beanbag on the line nearest the wall. Tell the child to stand on the other line. Tell the child to run up and kick the ball hard toward the wall. Repeat a second trial.</td>
<td>1. Rapid continuous approach to the ball 2. An elongated stride or leap immediately prior to ball contact 3. Nonkicking foot placed even with or slightly in back of the ball 4. Kicks ball with instep of preferred foot (shoelaces) or toe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Overhand Throw</td>
<td>A tennis ball, a wall, tape, and 20 feet of clear space</td>
<td>Attach a piece of tape on the floor 20 feet from a wall. Have the child stand behind the 20-foot line facing the wall. Tell the child to throw the ball hard at the wall. Repeat a second trial.</td>
<td>1. Windup is initiated with downward movement of hands and arm 2. Rotates hip and shoulders to a point where the nonthrowing side faces the wall 3. Weight is transferred by stepping with the foot opposite the throwing hand 4. Follow-through beyond ball release diagonally across the body toward the nonpreferred side</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Underhand Roll</td>
<td>A tennis ball for children ages 3 to 6; softball for children ages 7 to 10; two cones; tape, and 25 feet of clear space</td>
<td>Place the two cones against a wall so they are 4 feet apart. Attach a piece of tape on the floor 20 feet from the wall. Tell the child to roll the ball hard so that it goes between the cones. Repeat a second trial.</td>
<td>1. Preferred hand swings down and back, reaching behind the trunk while chest faces cones 2. Strides forward with foot opposite the preferred hand toward the cones 3. Bends knees to lower body 4. Releases ball close to the floor so ball does not bounce more than 4 inches high</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Object Control Subtest Raw Score (sum of the 6 skill scores)
Appendix F – Proficiency criteria for FMS assessment (sourced from the TGMD-2 (Ulrich, 2000))

*Proficiency criteria modified from original TGMD-2 assessment

| Hop | 1) Nonsupport 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 on same foot |  
| Horizontal Jump | 1) Preparatory movement includes flexion of both knees with arms extended behind the body |  
| 2) Arms extend forcefully forward and upward reaching full extension above the head |  
| 3) Take off and land on both feet simultaneously |  
| 4) Arms are thrust downward during landing |  
| Kick | 1) Rapid and continuous approach to the ball |  
| 2) An elongated stride or leap immediately prior to ball contact |  
| 3) Nonkicking foot placed even with or slightly in back of the ball |  
| 4) Kicks ball with instep of preferred foot (shoelaces) or toe |  
| Overhand Throw | 1) Windup is initiated with downward movement of hand/arm |  
| 2) Rotates hip and shoulders to a point where the nonthrowing side faces the direction of the throw |  
| 3) Weight is transferred by stepping with the foot opposite the throwing hand |  
| 4) Follow-through beyond ball release diagonally across the body towards nonpreferred side |  