Cronfa - Swansea University Open Access Repository

Paper:

Tyler, R., Foweather, L., Mackintosh, K. & Stratton, G. (2018). A Dynamic Assessment of Children's Physical Competence. *Medicine & Science in Sports & Exercise*, 1
http://dx.doi.org/10.1249/MSS.0000000000001739

A dynamic assessment of children's physical competence: The Dragon Challenge

Richard Tyler[1*], Lawrence Foweather[2], Kelly A. Mackintosh[1], Gareth Stratton[1]

[1.] Applied Sports Technology Exercise and Medicine (A-STEM) Research Centre, College of Engineering, Swansea University; Bay Campus, Fabian Way, Swansea, SA1 8EN, UK.

[2.] Physical Activity Exchange, Research Institute for Sport and Exercise Science, Liverpool John Moores University; 62 Great Crosshall Street, Liverpool, L3 2AT, UK.

[*]Corresponding author: Richard Tyler; Applied Sports Technology Exercise and Medicine (A-STEM) Research Centre, College of Engineering, Swansea University; Bay Campus, Fabian Way, Swansea, SA1 8EN, UK; (+44)07983629791; R.P.O.Tyler.837039@swansea.ac.uk.

Abstract:

Purpose: The first aim was to develop a dynamic measure of physical competence that requires a participant to demonstrate fundamental, combined and complex movement skills, and for assessors to score both processes and products (Dragon Challenge; DC). The second aim was to assess the psychometric properties of the DC in 10-14 year old children.

Methods: The first phase involved the development of the DC, including the review process that established face and content validity. The second phase used DC surveillance data ($n=4,355$; 10-12 years) to investigate construct validity. In the final phase, a convenience sample ($n=50$; 10-14 years) performed the DC twice (one-week interval), the Test of Gross Motor Development-2 (TGMD-2), and the Stability Skills Assessment (SSA). This data was used to investigate concurrent validity, and test-retest, inter-rater and intra-rater reliability.

Results: In support of construct validity, boys ($p < 0.001$) and secondary school children ($p < 0.001$) obtained higher DC total scores than girls and primary school children, respectively. A principal component analysis revealed a nine-component solution, with the three criteria scores for each individual DC task loading onto their own distinct component. This nine-factor structure was confirmed using a confirmatory factor analysis. Results for concurrent validity showed that there was a high positive correlation between DC total score and TGMD-2 and SSA overall score ($r(43) = .86$, $p < 0.001$). DC total score showed good test-retest reliability (ICC = 0.80; 95%CI: 0.63, 0.90; $p < 0.001$). Inter- and intra-rater reliability on all comparison levels was good (all ICCs > .85).

Conclusion: The DC is a valid and reliable tool to measure elements of physical competence physical competence in children aged 10-14 years.

Key words: Physical competence, Motor competence, Assessment, Measurement, Children, Reliability, Validity.

Introduction:

The International Physical Literacy Association defines physical literacy as the motivation, confidence, physical competence, knowledge and understanding to value and take responsibility for engagement in physical activities for life (1). Such a definition describes the multidimensional and complex nature of physical literacy, highlighting the purported importance of physical literacy as a precursor to physical activity (2). Therefore, given that physical activity has been shown to result in numerous health benefits (3), the promotion of physical literacy is fundamental for physical activity-associated health benefits. According to Lundvall (4), accurate assessment of physical literacy is essential, and there is a need to develop valid tools that effectively and efficiently assess each of the affective, cognitive, and psychomotor domains in order to evaluate whether programmes are successful (5).

One of the key elements of physical literacy is physical competence, which, even within itself, is a multidimensional concept. Whitehead (p204; 6), describes physical competence as "the sufficiency in movement vocabulary, movement capacities and developed movement patterns plus the deployment of these in a range of movement forms." Specifically, movement vocabulary refers to the repertoire of movements that one can perform, and can be expanded through experience and progressive challenge in the deployment of a wide range of movement capacities/skills and movement patterns (6).

Movement capacities are the integral abilities that make it possible to improve and develop physical competence (6). These capacities or skills consist of three interrelated constructs: fundamental or simple movement skills (FMS) (balance, core stability, coordination, speed variation, flexibility, control, proprioception, and power), combined movement (poise, fluency, precision, dexterity, and equilibrium), and complex movement (bilateral coordination, inter-limb coordination, hand-eye coordination, turning, twisting and rhythmic movements, and control of acceleration/deceleration; 6,7). FMS comprise locomotor skills (moving the body in any direction from one point to another), stability skills (balancing the body in one place or while in motion), and object control/manipulative skills (handling or controlling objects with the hand, foot, or an implement; 6–8). Children have the potential to master FMS by 7-8 years of age, with FMS developing rapidly between 3 and 8 years (8).

The procurement of movement capacities/skills and the ability to utilise them to produce movement patterns are essential for the development of physical competence within physical literacy capability (6). Movement patterns, described as general (e.g., sending, striking, receiving, running, jumping, rotating), refined (e.g., throwing, dribbling, catching, sprinting, hopping, turning) and specific (i.e. sport-specific movement patterns), are amalgamations of movement that stem from the selection and application of movement skills (6). More refined and specific movement patterns are achieved when fundamental, combined and complex movement skills are utilised (5–7). There is therefore much need to develop combined and complex movement skills, to take part in more advanced physical activities in a variety of settings (i.e., land, water, air, ice; 3,6) and movement forms (i.e., adventure, aesthetic, athletic, competitive, fitness and health, interactional/relational; 6), and thus this development is posited to be a foundation stone in developing physical literacy in maturing children (5,7).

Whilst many existing land-based movement skill assessments measure physical competence (7,9), the majority involve the performance of discrete skills in isolation (e.g., the Test of Gross Motor Development (TGMD-2/3; 10), the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2; 11), the Movement Assessment Battery for Children-2 (MABC-2; 12), CS4L: Physical Literacy Assessment for Youth Fun (PLAYfun;

13), Passport for Life: Movement Skills Assessment; 14). This static testing environment limits transferability and applicability to multi-skill and sport environments and does not assess combined and complex movement skills (7). Moreover, it has been suggested that considering skills in isolation ignores a constraints-based approach (15), in which environmental constraints are taken into account, and by doing so this approach is not 'authentic'. An authentic environment is one that is developmentally-appropriate and considers the interaction of the individual and the environment, as well as the specified movement skill (15,16). Performance of movement skills in isolation does not incorporate the measurement of an individuals' ability to alter and combine movement skills according to the task at hand and the environment, both of which are important traits to advance physical competence and progress one's physical literacy (6). Finally, assessments that measure skills in isolation have also been criticised for being time- and resource-intense (7,17). Thus, tools that measure physical competence in children aged over 8 years should assess fundamental, combined and complex movement skills in a dynamic and more authentic environment, in an efficient manner. The assessment of refined and specific movement patterns in a variety of novel combinations and complexities will more accurately reflect one's physical competence.

Physical competence can be evaluated by process- or product-based assessments (10–14). Primarily process-based assessments (e.g., TGMD-2, CS4L: PLAYfun, Passport for Life: Movement Skills Assessment) measure how children move and provide qualitative information on the technique of the movement patterns (18). This type of assessment can be sensitive to assessor experience and subjectivity (19). On the other hand, assessments that are primarily product-based (e.g., MABC-2, BOT-2) are usually quantitative and focus on the outcome of the movement (20), but potentially lack the sensitivity needed to identify individual differences in movement abilities (7). The equivocal relationship between process- and product-based assessments of physical competence has resulted in the use of combined assessments for measuring physical competence (20–22). Therefore, a single assessment that aims to equally assess both the process/technique and the product/outcome aspects of physical competence is warranted.

The assessment of physical competence can be formative or summative. Specifically, formative assessments measure current levels of performance to identify a baseline and the individual needs of children, enabling the development of an educational programme catered to those children, whereas summative assessments are used to measure progress of a child at the end of a period of education (23). Therefore, a physical competence assessment tool developed within the context of education, should aim to be both formative and summative, so that it can be used as a self-referenced assessment, which is able to compare a child's pre- and post-educational programme performance.

Recently, the Canadian Agility Movement Skill Assessment (CAMSA) was developed and validated to assess physical competence in 8-12-year-old children for surveillance, as well as examining movement skills over time (24). This assessment requires a series of seven movement tasks (two-footed jump, side slide, catch, throw, skip, hop, and kick) to be completed in a continuous dynamic obstacle course to create a more authentic environment and to assess combined and complex movement skills. Performances are assessed using the time taken to complete an obstacle course consisting of 14 process/technique- and product/outcome-based criteria (24). Whilst this assessment has shaped the way towards assessing movement skills in a dynamic fashion, there are noteworthy design limitations of the CAMSA. For example, the course does not include any specific stability movement skill tasks and there are a greater number of locomotor movement skill tasks than object control movement skill tasks. In addition, the scoring is unbalanced between locomotor and object control criterion, as well as between product- and process-based criterion. As such, an assessment targeting older aged children and adolescents (10-14 years), with a more balanced design and specific to children in the UK, is warranted.

Therefore, the first aim of this study was to develop a dynamic assessment to measure elements of physical competence (Dragon Challenge; DC), that requires the demonstration of fundamental (e.g., balance), combined (e.g., poise) and complex (e.g., rhythmic movements) movement skills through refined (complex) and specific movement patterns (e.g., hopping, turning, jumping patterns), measured by both product/outcome- and process/technique-based evaluations. The study sought to produce an assessment that would be feasible for national surveillance, and could be used as both a formative and summative assessment in the educational context. The second aim of the study was to assess the psychometric properties of the DC in measuring physical competence in children, including construct and concurrent validity and test-re-test and inter- and intra-rater reliability, as per American Educational Research Association, American Psychological Association, and National Council on Measurement in Education guidelines.

Methods:

This study involved three phases. Phase one included the development of the DC, including the review process to establish face and content validity. Phase two included gathering surveillance data and establishing construct validity and phase three involved investigating concurrent validity, test-retest, inter-rater and intra-rater reliability. The COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) framework was used to guide the design and evaluate the methodological quality (25). This study would achieve a quality level of good to excellent on the COSMIN rating system. The protocol, validation and reliability study of the DC were approved by the institutional

Research Ethics Committee (PG/2014/37 & PG/2014/39). Informed parental consent and participant assent were obtained prior to participation.

Phase 1. Development of the DC:

Programme of Research to Develop the DC:

Paediatric exercise science academics, practitioners, and professionals from schools and community sport (n>30) co-designed a land-based measure of elements of physical competence in children (10-14 years of age) that was aligned to physical education and sport coaching school and community programmes that aimed to promote physical literacy. The circuit of tasks were collectively named the 'Dragon Challenge' to align with the Sport Wales' Dragon multi-skills and sport initiative (http://sport.wales/community-sport/education/dragon-multi-skills--sport.aspx). The DC assessment tool underwent several stages of development. The first stage involved desk research, where an initial review was conducted on existing movement skill assessment tools that inform physical competence (8,10–12,26). From this, each of the 10 tasks/skills in the first protocol of the DC were examined for initial content validity. Subsequently, the second stage involved an iterative process of designing and testing the DC, whereby each task and its subsequent process- and product-based criteria were defined, with significant input from expert practitioners in physical education and community sport from across Wales (n>30). This stage included six iterations of protocol development, with the overall aim being to refine and assess the suitability of tasks, and to establish whether each individual task, and the overall assessment tool, could be used as an appropriate measure of children's physical competence. The initial tasks selected were therefore modified to incorporate refined and specific movement patterns that would adequately challenge children's fundamental, combined and complex movement skills, developed during physical education curriculum and the Dragon Sport multi-skill and sport initiative. The protocol development process was completed over a 12-month period

(July 2013 to 2014). Two hundred and eighty-eight children aged 10-12 years took part in the DC pilot testing days. The final DC protocol included nine tasks ordered to create continuity of movement and allow assessors to accurately observe children's performances *(see Dragon Challenge Circuit Video, Supplemental Digital Content 1, which presents the nine tasks being completed).* Process/technique and product/outcome indicators for the assessment criteria were continuously developed and refined by discussion and consensus until the DC was finalised.

Establishing Face and Content validity:

Face and content validity refers to how well a specific assessment measures what it intends to measure. The group of University paediatric exercise specialists, with expertise in physical education, physical competence and physical literacy research were involved in reviewing the DC. Face and content validity was qualitatively reviewed by a trained researcher (LF) with over 10 years' experience of physical competence and movement skill assessment. In addition, internationally recognised experts (n=5) in childhood movement skill, fitness, and physical literacy assessment within the personal networks of this researcher, advised LF and provided comments (in confidence) to inform the review process.

The review process comprised of in situ observations of children's performances, and a subjective analysis of the assessment protocol. Checks were made for the inclusion of critical movement tasks in accordance with a developmentally-appropriate assessment of physical competence through comparisons with existing assessment tools (8,10–12,26). Further checks were made to ensure that the DC circuit of tasks were in line with physical education curriculum content for children in this age range (10-14 years old), in that it required the utilisation of fundamental, combined, and complex movement capacities/skills to perform refined and specific movement patterns. Finally, clarity in behavioural definition (descriptions of the movement characteristics associated with the performance of each task) used in the assessment criteria was ensured.

Face validity:

Children complete the nine DC tasks in a set sequence; Table 1 shows the primary and secondary skill types necessary for each component. Several tasks (five out of nine) require children to perform a combination of skills and movement patterns, to demonstrate competence. Components of motor fitness such as agility, balance, coordination, strength, power, speed and reaction time are all widely utilised within the DC. The DC challenges children to demonstrate movement skills and motor fitness in combinations of different movement patterns and in continuous fashion as opposed to discrete skills in assessments such as the TGMD-2 or MABC-2. Further, children are required to demonstrate movement concepts and attributes expected of a physically competent person, (i.e., “movement with poise, economy and confidence in a wide variety of challenging situations” and “sensitive perception in ‘reading’ all aspects of the physical environment, anticipating movement needs or possibilities and responding appropriately to these, with intelligence and imagination”; 6). Thus, the DC tasks were representative of multiple elements of physical competence.

[INSERT ‘Table 1. Description of Dragon Challenge tasks’ ABOUT HERE]

Content Validity:

Internationally recognised experts (n=5) in childhood movement skill, fitness, and physical literacy assessment, confirmed that the DC was a valid and practical measure of physical competence, and that each task was challenging, achievable, and age-appropriate. Further, the tool was praised for its feasibility and efficiency.

DC task design: Balancing, running, hopping, jumping, throwing, dribbling, catching and sprinting are common skills that are assessed in isolation within existing movement skill assessment tools (8,10–12,26). Whilst the DC incorporates these skills and others, it is conducted in a continuous fashion within a timed trial, thus tasks are dynamic, sequential and include additional layers of complexity. The order of the tasks is standardised (as displayed in Table 1) but children perform the challenge under the illusion that the order is random, except for the final task, which is always the sprint (note, the full demonstration is in a different order to the standardised protocol). Each subsequent task is displayed on an iPad/tablet. Thus, the DC also explores perception-action coupling, as participants must coordinate recognising environmental information and the associated movement responses to such information, in order to complete the goal of each task.

Children observed a demonstration of each DC task and then the full DC. An introduction and demonstration video *(see Dragon Challenge Video Resources, Supplemental Digital Content 2, which displays the video material hyperlinks to support delivery of the DC)* of the DC was produced to ensure consistent administration and adequate demonstrations of the tasks were provided to the children in line with those outlined in the DC manual. In addition, the full video of the completion of the DC *(see Dragon Challenge Circuit Video, Supplemental Digital Content 1, which presents the nine tasks being completed)* could be shown. Children were given two practice attempts at each challenge task but they did not practise the challenge in full.

Children typically took between 90 and 240 seconds to complete the DC. An assessor used a stopwatch to record completion time (to nearest 0.1s). Each assessment required at least one trained assessor and one administrator. An additional assistant was required to supervise the non-participating children. The space requirement was designed to fit within the dimensions of a full-sized badminton court (13.4m x 6.1m), which most school gymnasiums and community sports centres are likely to have. Taken together, including set-up (15 minutes), the viewing of the videos and questions (26 minutes for a full group), and practice and completion of DC (approximately 10 to 12 children in 60 minutes), the total assessment time per child was approximately 10 minutes. For further information on the DC assessment including equipment list and descriptions of the assessment, *see Dragon Challenge v1.0 Manual, Supplemental Digital Content 3, which provides information on the administration of the DC assessment, as well as, the set-up schematic*.

DC assessment criteria: The DC indicators included both technical (process) and outcome (product) characteristics of movement performance (Table 2). Due to the challenges of real-time observation, the number of criterion to be assessed was limited to three per task (i.e., two technical/process criteria and one outcome/product criteria). Given that there were several technical characteristics that could be examined for each task, it was important that assessment criteria represented critical features of movement. Existing assessment tools and reference to developmental sequences were used to inform these decisions (8,10–12,26). A global review of the criteria (Table 2) suggested that the majority assess important characteristics of each task.

The DC was scored in three ways in accordance with the instructions specified within the DC manual (*see Dragon Challenge v1.0 Manual, Supplemental Digital Content 3, which provides information on the administration of the DC assessment, as well as, the set-up schematic*): (1) TECHNIQUE - 1 point was given for each of the technical/process criterion (n=18) successfully demonstrated by the child (2) OUTCOME - 2 points were awarded for each outcome/product criterion (n=9) successfully demonstrated by the child, and (3) TIME - time taken to complete the DC was recorded and converted to a score (higher scores for faster time). Each of these constructs (technique, outcome, and time) were scored out of 18 in order to be equally weighted, and then summed to give a total score (DC total score=54). Cut-points were also produced for the DC total score using the 33rd, 66th and 95th percentiles based on pilot data collected across Wales in 2015. These percentile thresholds were selected to categorise typically-developing 10-12 year old children into Bronze, Silver, Gold and

Platinum bands, thus making results easier to interpret by children, coaches, teachers, and parents.

[INSERT 'Table 2. Dragon Challenge Assessment Criteria' ABOUT HERE]

Phase 2. Surveillance Data and Construct Validity:

Participants and Procedures:

During the development process, a workforce of physical educators, coaches and other professionals in related areas, were trained to implement the DC assessments across four regions of Wales: South East, Mid & West, Central, and North. At least two assessors from each region received >20 hours of training led by LF, and were only permitted to do assessments once reaching an 85% level of agreement (3 errors per child) with LF. This workforce acted as 'gold standard assessors' within their respective region, and rolled out training to their constituents, with use of a gold standard training package for other professionals to be assessed against. In total, circa 200 assessors were trained across the four regions. Trained regional teams then conducted DC assessments in schools between January 2015 and November 2016.

The DC was scored in accordance with the instructions specified within the DC manual. For comparison purposes, technique and outcome scores were also summed to give sub-category scores for tasks primarily utilising stability (sum of technique and outcome criteria in tasks 1-3), object control (sum of technique and outcome criteria in tasks 4-6), and locomotor skills (sum of technique and outcome criteria in tasks 7-9; Table 1). Overall, data were successfully collected for analysis on 4,355 participants from 66 schools, aged 10-12 years from Central South Wales (n=875), South East Wales (n=1,238), Mid and West Wales (n=1,336) and North Wales (n=906). Within this overall sample, 49.9% of participants were boys, 7.2% were black and minority ethnic, 20.7% classified as special educational needs/additional learning needs status and 13.2% received free school meals (a proxy measure used in Wales for social economic status).

Construct Validity:

To ascertain whether the DC behaves according to motor development theory (8), total, technique/process, outcome/product, and time scores, as well as successful demonstration of each criterion, were examined by sex (boys expected to have higher scores than girls) and age/school level differences (older children expected to achieve higher scores than younger children). The factor structure of the DC was also examined. As each of the nine DC tasks required combinations of movement skills (Table 2), it was hypothesised that the outcome may not produce a 3-factor structure (namely, stability, object control and locomotor), but instead produce a structure with a greater number of factors, each representing a distinct combination of skills. It was also hypothesised that these factors would load on to a higher order factor, namely physical competence.

Phase 3. Concurrent Validity and Reliability:

Participants and Procedures:

A convenience sample of 50 participants (52% boys) aged 12.66±1.51 years from two schools performed the DC twice with a one-week interval between the two DC data collection days. Participants were from school year 5 (n=8; 10.32±0.31 years), year 6 (n=8; 11.28±0.32 years), year 7 (n=10; 12.42±0.23 years), year 8 (n=12; 13.48±0.25 years), and year 9 (n=12; 14.51±0.26 years) and had a mixture of abilities according to their physical education teacher. Each attempt at the DC was video recorded using two tripod-mounted video cameras [Sony Handycam, Model HDR-PJ410, Sony Corporation, Tokyo, Japan]. Scoring was completed by an expert assessor (>50 hours of DC training and in situ experience), trained assessor (20 hours of DC training and in situ experience), and/or newly trained assessor (5 hours of DC training), in accordance with the instructions specified within the DC manual. For comparison purposes, sub-category scores were also calculated for tasks primarily utilising stability skills, object control skills, and locomotor skills.

On a separate day, participants performed two trials of the Test of Gross Motor Development-2 (TGMD-2; 10) and the Stability Skills Assessment (SSA; 27), previously validated movement skills assessments, which required the completion of six locomotor (run, gallop, hop, leap, horizontal jump, and slide) and six object control (striking a stationary ball, stationary dribble, catch, kick, overhand throw, and underhand roll) subtest skills, and three gymnastics training stability skills (rock, log- roll, and back support), respectively. Participants were video recorded using two tripod-mounted video cameras [Sony Handycam, Model HDR-PJ410, Sony Corporation, Tokyo, Japan]. A trained assessor scored the video footage based on the presence (1) or absence (0) of three to five component (process) criteria for each of the skills in both trials of the TGMD-2 and SSA (10,27). 'Overall skill scores', the cumulative criteria scores for each skill across both trials, were calculated for each of the TGMD-2 and SSA tasks. 'Overall skill scores' for each of the TGMD-2 (0-96) and SSA (0-24) tasks were summed to give a 'combined TGMD-2 and SSA overall skill score' (0-120). Lastly, subcategory skill scores were also calculated for stability, object control and locomotor skill tasks (e.g., 'overall skill scores' for each of the stability tasks were summed to give a stability skill score).

Concurrent validity:

Concurrent validity refers to the extent to which the DC relates to a previously validated movement skills assessment. This was first investigated at an overall level by examining the extent to which the week 1 DC scores related to the TGMD-2 and SSA scores. Further, the relationship between week 1 DC score and TGMD-2 skill score was investigated.

The TGMD-2 and SSA were used as the comparison measures for concurrent validity for the following reasons: (i) the validity and reliability for both assessments have been established (10,27); (ii) the TGMD-2 has been extensively used as an assessment for movement skill performance; (iii) the SSA provides additional stability tasks that are missing in the TGMD-2, and tasks have been validated to add to the measurement model (27); (iv) the TGMD-2 and SSA have been used in movement skill research in school settings; (v) the TGMD-2 has been validated for children/adolescents of similar age (28); (vi) although the skills in both the TGMD-2 and SSA are completed in isolation by children, the skills assessed within these batteries more closely align with those included in the DC than those used in other movement skill assessments available at the time of study development (no comparative dynamic movement assessments were available); (vii) while TGMD-2 and SSA are considered primarily process-based assessments, there are a selection of product-based criteria (e.g., hop three consecutive times, dribble ball for four consecutive bounces (10), log roll for four complete rotations, and back support held for 30 seconds; 27), thus aligning scoring more closely with the DC.

Reliability:

Test-re-test reliability was examined by the stability of participants' DC results over the repeated rounds of assessment. The same expert assessor scored each participant on both time-points, and the level of agreement was evaluated.

Inter-rater reliability was explored by investigating how consistent two or more assessors' scores were when observing the same performance. Inter-rater reliability was first assessed at an overall level using the scores given by 3 separate expert assessors on video footage from 12 participants of mixed ability completing the DC. In order to investigate whether amount of training and experience received by assessors influenced reliability, additional analyses examined consistency between expert and newly trained assessor and between expert and trained assessor when scoring DC for 12 and 15 participants, respectively.

Intra-rater reliability was examined by investigating the consistency between scores, when the same trial was scored by the same rater on two separate occasions. Three expert assessors each scored video footage of 12 participants of mixed ability completing the DC on two occasions, with a one-week interval between viewings, and levels of agreement between the scores for each assessor was evaluated.

Statistical Analysis:

Descriptive statistics are presented as mean ± SD. All statistical tests, with the exception of the confirmatory factor analysis (CFA), were completed using SPSS, version 24 [IBM SPSS Statistics Inc., Chicago, IL, USA]. The CFA was completed using lavaan version 0.6-1 (29), in R version 3.5.0 [R Core Team, Vienna, Austria]. In all analyses, data were assessed for violation of the assumptions of normality and statistical significance was set at $p < 0.05$. Participant results were included in each respective analysis if they had sufficient data for the variable concerned.

Surveillance Data and Construct Validity: The proportion of participants successfully demonstrating each DC criterion for the surveillance data was calculated. Two-way ANOVA tests and Chi-squared tests were used to explore the effects of sex and school level on DC scores and on each individual DC task assessment criterion, respectively.

A Principal Component Analysis (PCA) was performed on all DC binary criteria scores. The suitability of each PCA was assessed prior to analysis by inspection of the correlation matrix (each variable required to have at least one correlation with another variable above $r = 0.3$), further the Kaiser-Meyer-Olkin (KMO) measure needed to be at least 0.6, for sampling adequacy (30). In addition, Bartlett's test of sphericity had to achieve statistical significance ($p < 0.05$). To establish DC components, the eigenvalue-one criterion was used (31), as well as visual inspection of the scree plot. A Varimax orthogonal rotation was used to aid interpretation, where applicable. A loading of .40 or greater was used to align items onto factors.

Based on the results of the PCA, a CFA was performed to cross-validate the factor structure of the DC. As binary criteria scores were used as indicator variables, weighted least square mean and variance adjusted estimator was used to fit the model. By default, the factor loading of the first indicator of a latent factor was fixed to 1, thereby fixing the scale of the latent factor. Error terms from the indicator variables were allowed to co-vary within the same factor. Comparative fit index (CFI), Tucker-Lewis Index (TLI), and root mean square error of approximation (RMSEA) were used to assess model fit, with CFI and TLI of >0.95 and RMSEA of <0.05, indicating a good fit (32).

Concurrent Validity: A Pearson's product-moment correlation was used to investigate the strength of relationships between DC, and TGMD-2 and SSA scores and sub-category scores. An *r* value of, 0–0.19, 0.2–0.39, 0.4–0.59, 0.6–0.79, >0.8 were interpreted to demonstrate no, low, moderate, moderately-high and high correlation coefficients, respectively (33).

Reliability: To ascertain evidence for test–retest, inter-rater and intra-rater reliability, intraclass correlation coefficients (ICC), two-way random single measures for absolute agreement (ICC 2,1), with 95% confidence intervals (95%CI), were used to evaluate the level of agreement of week 1 and week 2 scores and of rater scores. A reflect and square root transformation was used where data was non-parametric. For presentation purposes, these variables were transformed for analysis and back transformed. Intraclass correlation coefficients below .50 indicate poor reliability, those between .50 and .75 indicate moderate reliability, and those above .75 indicate good reliability (34). ICC results that indicated moderate reliability (<.75) were further examined using a t-test to investigate if there was a statistically significant mean difference between scores.

Results:

Table 3 provides age and sex characteristics of participants that took part in the DC for phase 2 and 3 of the study. On the basis of missing demographic characteristics, 95 participants from the surveillance data were excluded from all construct validity analyses (n=4260), except for the PCA and CFA (n=4,355).

[INSERT 'Table 3. Study participants' ABOUT HERE]

Construct Validity:

Mean scores and standard deviations for DC surveillance data, broken down by sex and school level, are presented in Table 4. There were no statistically significant interactions between sex and school level on DC scores. Therefore, analyses of main effects for each variable were performed. Boys scored higher than girls for all score categories, except stability skills, and secondary school level children scored higher than primary school level children on all score categories apart from time score. The proportion of children who successfully demonstrated each DC criterion, as well as statistically significant sex and school level differences, highlighted by the Chi-squared test, are shown in Table 5.

[INSERT 'Table 4. Descriptive statistics for Dragon Challenge' ABOUT HERE]

[INSERT 'Table 5. Proportion (%) of children successfully…' ABOUT HERE]

PCA on DC criteria scores:

PCA was found to be suitable according to the correlation matrix, overall KMO (.76) and Bartlett's test of sphericity ($p < 0.001$). The PCA revealed nine components that had eigenvalues greater than one, 5.11, 2.53, 2.01, 1.83, 1.54, 1.42, 1.37, 1.19, 1.15, and which explained 18.94%, 9.39%, 7.46%, 6.76%, 5.71%, 5.24%, 5.09%, 4.40%, 4.26%, respectively. Visual inspection of the scree plot also indicated that nine factors should be retained. This nine-component solution explained 67.24% of the total variance and the rotated solution exhibited a simple structure. The interpretation of the data was consistent with the skill combinations the DC was designed to measure, with strong loadings of balance bench criteria scores on component one, core agility criteria scores on component two, wobble spot criteria scores on component three, overarm throw criteria scores on component four, basketball dribble criteria scores on component five, catch criteria scores on component six, t-agility criteria scores on component seven, jumping patterns criteria scores on component eight, sprint criteria scores on component nine. Component loadings of the rotated solution *(see Table 6, Supplemental Digital Content 4, which presents the rotated component matrix from the principal component analysis on Dragon Challenge criteria scores)* were all >0.4.

CFA on the DC criteria scores:

Based on the PCA results, CFA was conducted to confirm the nine-factor structure, as well as to examine whether the nine latent factors loaded onto a higher order factor (physical competence). Following the addition of three correlations between error terms within the same factor, the fit for the hypothesised model (Figure 1), was good (CFI, 1.00; TLI, 1.00; RMSEA, 0.038; 90% confidence interval 0.037 – 0.040). Factor loadings ranged from 0.45 – 0.99, showing that the factor validity was acceptable to excellent.

[INSERT 'Figure 1. Factor Structure of DC' ABOUT HERE]

Concurrent Validity:

Results for concurrent validity show that there was a significant high positive correlation between DC total score (35.9 ± 8.5) and 'combined TGMD-2 and SSA overall skill score' (72.5 ± 10.9) ($r(43) = .86$, $r^2 = .74$, $p < 0.001$). Relationships for sub-category scores between DC and TGMD-2 and SSA skills scores, across stability tasks (7.2 ± 3.2, 7.8 ± 3.7; $r(43) = .46$, $p = 0.001$), object control tasks (8.0 ± 3.4, 32.5 ± 6.9; $r(43) = .83$, $p < 0.001$) and locomotor tasks (8.5 ± 2.5, 32.2 ± 3.4; $r(43) = .60$, $p < 0.001$), showed significant moderate to high positive correlations. Finally, there was a significant high positive correlation between DC score (35.93 ± 8.54) and TGMD-2 'overall skill score' (64.71 ± 8.66) ($r(43) = .81$, $r^2 = .66$, $p < 0.001$).

<u>Reliability:</u>

Test-Retest Reliability:

The DC total score showed good test-retest reliability across the one-week interval (ICC = 0.80; 95%CI: 0.63, 0.90; $p < 0.001$). Evidence for test-retest reliability was good for technique scores (ICC = 0.77; 95%CI: 0.58, 0.88; $p < 0.001$), and high-moderate for time scores (ICC = 0.74; 95%CI: 0.57, 0.85; $p < 0.001$) and for outcome scores (ICC = 0.71; 95%CI: 0.52, 0.83; $p < 0.001$). Follow up t-tests revealed no significant mean difference in time score between test (12.18 points) and retest (12.93 points) scores ($t = .837$, $p = 0.41$) and no statistically significant mean difference in outcome score between the test (11.95 points) and retest (12.00 points) scores ($t = .103$, $p = 0.92$).

Further, test-retest reliability for skill sub-categories was good for object control skills score (ICC = 0.80; 95%CI: 0.67, 0.89; $p < 0.001$), high-moderate for locomotor skills score (ICC = 0.68; 95%CI: 0.49, 0.81; $p < 0.001$), and moderate for stability skills score (ICC = 0.60; 95%CI: 0.38, 0.76; $p < 0.001$). No significant mean difference was found in locomotor skills score between test (8.43 points) and retest (8.59 points) scores ($t = .525$, $p = 0.60$), nor in stability skills score between test (7.14 points) and retest (6.61 points) scores ($t = -1.25$, $p = 0.22$).

Inter-Rater and Intra-Rater Reliability:

Inter-rater and intra-rater reliability on all comparison levels (*see Table 7, Supplemental Digital Content 5, which reports the inter- and intra-rater reliability results for Dragon Challenge scores and sub-category scores*) showed significant relationships and were classed as good (all ICCs > .85).

Discussion:

Many current measures that inform physical competency as part of physical literacy assessments (7,9), in children and adolescents (10–14), use isolated movement skills. Assessing discrete movement skills in isolation fails to account for the utilisation of combined and complex movement skills observed during physical activity and play, and needed to demonstrate physical competence and physical literacy (6). This study therefore aimed to develop the DC, a land-based dynamic measure of movement capacities/skills and movement patterns to assess elements of physical competence for 10-14 year olds.

The DC consists of nine tasks completed in a timed circuit, incorporating the utilisation of fundamental, combined and complex movement skills/capacities, to produce refined/complex and specific movement patterns. The DC can be used for assessment for learning (summative and/or formative), and as a national surveillance tool, that can be aligned to physical literacy programmes and physical education curriculum. The assessment criteria for the DC includes both technique (process) and outcome (product) indicators of movement performance, to provide a more complete picture of physical competence levels than currently used assessments that include primarily product- or process-based criteria (10–14). Given that the DC is completed in a continuous circuit, tasks are dynamic, sequential and include additional layers of complexity in a more open 'authentic' environment than many existing measures that assess skills in isolation (10–14). The DC is internally paced by the participants, whom are required to perform the tasks competently as fast as they can, thereby requiring a speed-accuracy trade-off. Although not directly measured, children also need to apply awareness of spaces, effort, and relationships to objects, goals, and boundaries to complete the challenge. Thus, within the DC, children are required to demonstrate movement concepts and attributes expected to be displayed by a physically competent child, for example, "movement with poise, economy and confidence in a wide variety of challenging situations” and “sensitive perception in ‘reading’ all aspects of the physical environment, anticipating movement needs or possibilities and responding appropriately to these, with intelligence and imagination” (6). Therefore, given the paucity of dynamic measures of movement skills/capacities and varying complexities of movement patterns to inform physical competence in children aged 10-14 years, this study fills a critical gap in the current literature in this field.

Construct Validity:

Boys obtained significantly higher DC total, time, technique and outcome scores (Table 4). When broken down into sub-categories for comparison purposes, boys scored significantly higher than girls for tasks primarily utilising object control skills, with more detailed analysis (Table 5) showing that significantly more boys demonstrated proficiency at each of the assessment criteria for the overarm throw, basketball dribble and catch. These sex differences seem to be in line with numerous studies that have shown that boys outperform girls at object control skills (13,35,36). On the other hand, girls scored significantly higher than boys for tasks primarily utilising stability skills, with significantly more girls demonstrating proficiency at each of the assessment criteria for core agility, as well as two of the assessment criteria for balance bench (criterion 1.1, 1.2; Table 5). While literature regarding sex differences in stability skills is less prevalent, young girls have been shown to display greater aptitude in process-oriented balancing skills (37). In line with many studies that report no gender difference in locomotor skills (13,35,36), no significant difference was found in score between boys and girls for the locomotor skills sub-category. Moreover, girls typically excel at hopping and skipping in comparison to boys (38), supporting our findings that significantly more girls were proficient in two of the jumping patterns criteria (criteria 8.2 and 8.3). Considering these findings within the context of sex differences, the DC data are aligned to current literature on physical competence and movement skill competence.

Not only did secondary school level children obtain significantly higher DC total, technique and outcome scores compared to primary school level children, but they also scored significantly higher for object control skills, locomotor skills, and stability skills. Given that gross motor skill is developmental by age and stage, these results are standard within the literature (8). It is worth noting, however, that there was no significant difference in time score between primary and secondary school children. This was unexpected as previous studies have shown that running speed increases with age in children (38), although this discrepancy may be explained by the speed-accuracy trade-off made by children when completing the DC. Thus, the higher accuracy of the secondary school level children at the DC tasks would have resulted in them taking longer to complete the tasks than the less accurate primary school level children. In summary, the findings in relation to sex and age differences are consistent with the literature.

Since the factor structure showed good model fit (Figure 1), it is reasonable to conclude that, unlike existing measures of physical competence (10,26,27), the DC does not measure movement skills in isolated skill categories (i.e., stability, object control, locomotor skills; 8), but rather requires the application of different combinations of movement skills for each task. Thus, the good fit of the model adds support to the design of the DC, as each task was selected to include the utilisation of skills from multiple movement categories to produce a series of movement patterns, and to the contention that the DC includes combined and complex movement skills. Additionally, the adequate factor loadings of each criterion scores onto its respective latent factor suggests that each criterion score is a good indicator, giving strength to the choice of criteria in the DC scoring system. Finally, as each of the nine first order latent factors (skill combinations) loaded onto a higher order latent factor (physical competence), it suggests that the combination of skills required by each DC task is needed for children to be physically competent. It must be noted, however, that physical competence is a multidimensional concept, therefore there are additional aspects of physical competence that are not represented in this model, for example, combinations of movement skills in different settings (water, air, ice), or movement forms (3,6).

Concurrent Validity:

The DC total score showed a significant high positive relationship with the 'combined TGMD-2 and SSA overall skill score' and with 'TGMD-2 overall skill score', demonstrating strong concurrent validity between the assessments. When broken down across sub-categories, there was a significant high relationship between object control task scores in the DC and TGMD-2, whereas the DC stability and locomotor task scores showed only significant moderate relationships with those included in the TGMD-2 and SSA. While the stability and locomotor skills in the two assessments were matched for comparison purposes, the tasks required by the TGMD-2 and SSA compared to the DC were not identical. Moreover, as evidenced by the CFA on DC criteria scores, each of the DC tasks require a unique combination of movement skills/capacities to perform the refined/complex and specific movement patterns. Therefore, the differences in stability and locomotor tasks in the TGMD-2 and SSA compared to the DC probably contributed to lowering the correlation of these subcategories. Nevertheless, all relationships, both in total scores and in subcategory scores, between the tools were significant moderate to strong, indicating that the DC ranks children in similar order to previously-validated tools.

Reliability:

Test-retest reliability for both DC total and technique scores, was good across a one-week interval. However, time and outcome scores only showed high-moderate and moderate test-retest reliability, respectively. This may also be reflective of the speed-accuracy trade-off associated with the DC assessment tasks, with children perhaps making different decisions as to which to prioritise when performing the DC on multiple occasions. Upon further investigation, there was no significant difference in mean outcome or time scores between the test and retest, providing support that no learning effect was present. Since DC total score showed good test-retest reliability over a one-week interval, and all other scores showed moderate-to-good test-retest reliability, with statistics at least as strong as those for other measurement tools (10–12,17,24), then the tenet that the DC is a stable measure is supported.

Inter-rater reliability was good for each of the DC total, time, technique, outcome, and sub-category scores when comparing three separate expert standard assessors' score. These levels of inter-rater reliability are similar to those of the TGMD-2, BOT-2 and MABC-2, but stronger than those of the CAMSA measurement tool (10–12,17,24). Inter-rater reliability was also good for the DC total, technique, outcome, and subcategory scores, both when comparing the level of agreement of expert assessor's scores and newly-trained assessor's scores and when comparing the level of agreement of expert assessor's scores and trained assessor's scores. There was stronger reliability between the expert assessor and trained assessor than between the expert assessor and newly-trained assessor in all scores. This suggests that the additional field time that the trained assessor undertook compared to the newly-trained assessor may have resulted in more reliable assessments. Taken together, the inter-rater reliability results may imply that only one skilled assessor is needed to achieve a reliable assessment of participants taking part in the DC. Moreover, each of the three expert assessors' DC total, time, technique, outcome, and sub-category scores showed good intra-rater reliability, consistent with the levels of intra-rater reliability of other measurement tools (10–12,17,24), suggesting that the current DC assessment criteria are sufficiently clear to allow an accurate assessment of a participant in one viewing.

Feasibility:

There are currently no guidelines for determining the optimal duration of an assessment tool, therefore the purpose, information yielded, and time for completion should all be considered when examining assessment feasibility. Assessing a child in the DC required at least one assessor and administrator, with an additional assistant to supervise the non-participating children, in line with most other assessments (10–12,24). While this may seem burdensome, the balance between developing sufficient data for surveillance and adequate detailed insight to provide feedback and promote learning was achieved using this approach.

Children typically only took between 90 and 240 seconds to complete the DC, and an overall estimated assessment time of 10 minutes per child. Large sports halls can facilitate multiple concurrent DC circuits, thus decreasing time to assess larger numbers of children. However, the trade-off is that more assessors and administrators are required with multiple set-ups. In many previously-validated movement skill assessments (10–14,17), an average of 15-60 minute assessment time per child was required. Although some of these assessments were initially created for differing circumstances (e.g., developmental coordination disorder), they have all been used to assess the physical aspects of physical literacy, in an educational setting (7,9). In comparison to these assessments, the DC assessment time per child is considerably less, providing evidence that the DC is a time-efficient measure. Conversely, the CAMSA (24), requires less time to complete (set up time = 5–7 mins; assessment time = 25 min for 20 children) than the DC. This is due, at least in part, to the incorporation of more tasks and indeed performance criteria in the DC. It is therefore postulated that longer assessment times to yield more information are reasonable.

The DC produced important information on a child’s movement skills/capacities and varying complexities of movement patterns to inform physical competence and physical literacy, and so, as in other assessments within schools (English, Maths and Science exams), time and effort needs to be applied for progressive learning. The decreased assessment time associated with the DC compared to the many previously-validated assessments (10–14,17), increases its feasibility as a population-level surveillance tool. Furthermore, in this study we have demonstrated that we can collect data on a national sample of children (n= 4,355), supporting our premise that DC can be used as an assessment for learning and a national surveillance tool.

Limitations and Future Directions:

It is important to note that whilst, in comparison to many other existing assessments, the DC is more inclusive of the constructs of Whitehead’s interpretations of physical competence (6), it does not provide a complete assessment of physical competence. Specifically, the DC does not reflect physical competence in terms of different varieties of contexts and durations of activities, activity settings (i.e., water, air, ice; 3,6), or different movement forms (i.e., adventure, aesthetic, athletic, competitive, fitness and health, interactional/relational; 6). However, many land-based measures assume the transferability of movement capacities/skills and movement patterns assessed in the measures, to other contexts (7,9). This may also be the case for the DC, but future studies may wish to investigate the use of the DC to predict the participation is differing movement forms and activity settings. The authors of this study also acknowledge that although the DC generally showed good concurrent validity with the TGMD-2 and SSA, a gold standard measure that is more dynamic and includes more aspects of combined and complex movement skills, rather than individual skills in isolation, may have be more appropriate for comparisons. However, at the time of study design there was no gold standard assessment that assessed such movement skills. Furthermore, as a compromise for being able to assess on a population-level, some criterion that were typically considered critical movement features (e.g., hip then shoulder rotation for the overarm throw), were not incorporated into the DC assessment criteria due to the difficulty of observation in real-time during protocol development.

Although discriminant and clinical use of the DC was not a planned outcome in the current study, further analysis of the surveillance data (n=4,355), reported in a separate DC surveillance report, found that the DC was able to significantly differentiate between children with and without an additional or special learning needs, across all DC scores (39). However, additional investigations are required to develop the DC so that is fully inclusive, irrespective of disability. Moreover, the high percentage of success for both boys and girls on criterion 9.3 (Table 5), suggests that a ceiling effect may be present for this product criterion. Therefore, an adjustment of this criterion, perhaps with the use of Rasch analysis (40), may be warranted. Finally, since the tasks included in the DC were selected to be a developmentally-appropriate assessment of physical competence for children in developed countries with similar physical education curricular and sport programmes, future studies should examine cultural differences to evaluate whether the tasks chosen are also valid in jurisdictions with different physical education and sport programmes.

Conclusion:

The DC was designed as a tool to measure elements of physical competence, representing a more ecological measurement of fundamental, combined, and complex movement skills in one assessment. These skills are combined in the DC to form complex movement patterns in a more authentic environment, and can be measured in a time-efficient manner. The DC is novel in that it offers a dynamic land-based measure to inform physical competence for formative and summative assessment purposes, as well as for national surveillance, with accurate data collected from a national sample of over 4,300 children in Wales. Our results demonstrate that the DC is a valid and reliable measure in children aged 10-14 years. Further investigation into the potential of the DC to reflect physical competence in terms of different contexts, durations, and activity settings, as well as the development of measures of the remaining physical literacy domains, should be of focus to construct a full physical literacy measurement model.

Acknowledgements:

This work was supported by a grant from Sport Wales, and postgraduate support from the Swansea University Scholarship Fund. The authors would like to thank numerous researchers, including A-STEM PGRs, and practitioners from across Wales, including Julie Rotchell, Jan English, Helen Hughes, Kirsty Edwards, Jonathan Moody, Beverley Symonds, Jane Antony, and Luke Williams. This study would not have been possible without their significant contributions. We would also like to pay tribute to the many physical educators, coaches, Young Ambassadors, and other professionals who helped with the administration and assessment of the Dragon Challenge. Finally, we would like to express our gratitude to the children and schools across Wales who participated in the Dragon Challenge.

Conflict of Interest:

The authors declare there are no known conflicts of interest in the present study. The results of the study do not constitute endorsement by ACSM, and are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

References:

1. International Physical Literacy Association. Definition of physical literacy. [Internet]. 2016 [cited 2017 Oct 19]. Available from: https://www.physical-literacy.org.uk/
2. Edwards LC, Bryant AS, Keegan RJ, Morgan K, Jones AM. Definitions, Foundations and Associations of Physical Literacy: A Systematic Review. Sport Med. 2016;47(1):1–14.
3. Jurbala P. What Is Physical Literacy, Really ? Quest. 2015;67(4):367–83.
4. Lundvall S. Physical literacy in the field of physical education - A challenge and a possibility. J Sport Heal Sci. 2015;4(2):113–8.
5. Corbin CB. Implications of Physical Literacy for Research and Practice: A Commentary. Res Q Exerc Sport. 2016;87(1):14–27.
6. Whitehead M. Physical Literacy: throughout the lifecourse. London: Routledge; 2010. 256 p.
7. Giblin S, Collins D, Button C. Physical literacy: importance, assessment and future directions. Sports Med. 2014;44(9):1177–84.
8. Gallahue DL, Ozmun JC, Goodway JD. Understanding motor development: Infants, children, adolescents, adults. 7th ed. New York, NY: McGraw-Hill; 2012. 480 p.
9. Robinson DB, Randall L. Marking Physical Literacy or Missing the Mark on Physical Literacy ? A Conceptual Critique of Canada's Physical Literacy Assessment Instruments. Meas Phys Educ Exerc Sci. 2017;21(1):40–55.
10. Ulrich DA. TGMD-2: Test of Gross Motor Development. 2nd ed. Austin, Texas: PRO-ED; 2000.
11. Bruininks R, Bruininks B. Bruininks-Oseretsky Test of Motor Proficiency, second edition (BOT-2). Minneapolis, MN: Pearson Assessment; 2005.
12. Henderson SE, Sugden DA, Barnett AL. Movement Assessment Battery for Children –

second edition (Movement ABC-2); examiner's manual. London: Harcourt Assessment; 2007.

13. Cairney J, Veldhuizen S, Graham JD, et al. A Construct Validation Study of PLAYfun. Med Sci Sport Exerc. 2018;50(4):855–62.

14. Physical and Health Education Canada (PHE Canada). Development of Passport for Life. Phys Heal Educ J. 2014;80(2):18–21.

15. Newell K. Constraints on the development of coordination. In: Wade MG, Whiting HT, editors. Motor Development in Children: Aspects of Coordination and Control. Dordrecht, Netherlands: Martinus Nijhoff; 1986. p. 341–60.

16. Barnett LM, Stodden D, Miller AD, et al. Fundamental Movement Skills : An Important Focus. J Teach Phys Educ. 2016;35:219–25.

17. Wiart L, Darrah J. Review of four tests of gross motor development. Dev Med Child Neurol. 2001;43:279–85.

18. Hardy LL, King L, Farrell L, Macniven R, Howlett S. Fundamental movement skills among Australian preschool children. J Sci Med Sport. 2010;13(5):503–8.

19. Schoemaker MM, Niemeijer AS, Flapper BCT, Smits-Engelsman BCM. Validity and reliability of the Movement Assessment Battery for Children-2 Checklist for children with and without motor impairments. Dev Med Child Neurol. 2012;54(4):368–75.

20. Logan SW, Barnett LM, Goodway JD, Stodden DF. Comparison of performance on process- and product-oriented assessments of fundamental motor skills across childhood. J Sports Sci. 2017;35(7):634–41.

21. Robinson LE, Stodden DF, Barnett LM, et al. Motor Competence and its Effect on Positive Developmental Trajectories of Health. Sport Med. 2015;45(9):1273–84.

22. Rudd J, Butson ML, Barnett L, et al. A holistic measurement model of movement competency in children. J Sports Sci. 2016;34(5):477–85.

23. Gallahue DL, Donnelly FC. Developmental Physical Education for All Children. 4th ed. Champaign, IL: Human Kinetics; 2007. 744 p.

24. Longmuir PE, Boyer C, Lloyd M, et al. Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age. J Sport Heal Sci. 2017;6(2):231–40.

25. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. Qual Life Res. 2010;19(4):539–49.

26. NSW Department of Education and Training. Get skilled: Get active. A K-6 resource to support the teaching of fundamental movement skills. Ryde, NSW: NSW Department of Education and Training; 2000.

27. Rudd JR, Barnett LM, Butson ML, Farrow D, Berry J, Polman RCJ. Fundamental movement skills are more than run, throw and catch: The role of stability skills. PLoS One. 2015;10(10):1–15.

28. Issartel J, McGrane B, Fletcher R, O'Brien W, Powell D, Belton S. A cross-validation study of the TGMD-2: The case of an adolescent population. J Sci Med Sport. 2017;20(5):475–9.

29. Rosseel Y. lavaan: An R Package for Structural Equation Modelling. J Stat Softw [Internet]. 2012;48(2):1–36. Available from: http://www.jstatsoft.org/v48/i02/

30. Kaiser HF. An index of factorial simplicity. Psychometrika. 1974;39(1):31–6.

31. Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas. 1960;20:141–51.

32. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives. Struct Equ Model A Multidiscip J.

1999;6(1):1–55.

33. Cohen J. A Power Primer. Psychol Bull. 1992;112(1):155–9.

34. Portney LG, Watkins MP. Foundations of clinical research: Applications to practice. Norwalk, CT: Appleton & Lange; 1993. 722 p.

35. Hume C, Okely A, Bagley S, et al. Does Weight Status Influence Associations Between Children's Fundamental Movement Skills and Physical Activity? Res Q Exerc Sport. 2008;79(2):158–65.

36. Barnett LM, van Beurden E, Morgan PJ, Brooks LO, Beard JR. Gender Differences in Motor Skill Proficiency From Childhood to Adolescence : A Longitudinal Study. Res Q Exerc Sport. 2010;81(2):162–70.

37. Pienaar AE, Reenen I van, Weber AM. Sex differences in fundamental movement skills of a selected group of 6-year-old South African children. Early Child Dev Care. 2017;186(12):1994–2008.

38. Duger T, Bumin G, Uyanik M, Aki E, Kayihan H. The assessment of Bruininks-Oseretsky test of motor proficiency in children. Pediatr Rehabil. 1999;3:125–31.

39. Stratton G, Foweather L, Hughes H. Dragon Challenge: A National Indicator for Children's Physical Literacy in Wales. Surveillance Report January 2017. Cardiff: Sport Wales; 2017. 31 p.

40. Zhu W, Cole EL. Many-faceted Rasch calibration of a gross motor instrument. Res Q Exerc Sport. 1996;67(1):24–34.

- Supplemental Digital Content 1. Dragon Challenge Circuit, which presents the nine tasks being completed. mp4
- Supplemental Digital Content 2. Dragon Challenge Video Resources, which displays

the video material hyperlinks to support delivery of the DC. doc

- Supplemental Digital Content 3. Dragon Challenge v1.0 Manual, which provides information on the administration of the DC assessment, as well as, the set-up schematic. pdf
- Supplemental Digital Content 4. Table 6, which presents the rotated component matrix from the principal component analysis on Dragon Challenge criteria scores. doc
- Supplemental Digital Content 5. Table 7, which reports the inter- and intra-rater reliability results for Dragon Challenge scores and sub-category scores. doc

Table 1. Description of Dragon Challenge protocol and tasks, and types of skills utilised during each task

DC Task		Description	Stability	Locomotor	Object Control
1.	**Balance Bench**	Runs to bench. Walks length of narrow side of bench beam, completing a 360° turn at mark before dismounting at the end of the bench and returning to iPad. If the child falls off then task is ended and the child returns to iPad immediately.	●	○	
2.	**Core Agility**	Runs to gym mat. Completes 4 positions (dish on back - arch on front - dish on back - arch on front), rotating both ways. Returns to iPad.	●		
3.	**Wobble Spot**	Runs to wobble spot and picks up bean bag on floor. Completes 5 bean bag passes around body whilst balancing on wobble spot on one leg. Returns to iPad. If child falls off after starting, the task is ended and the child returns to the iPad immediately.	●		○
4.	**Overarm Throw**	Collects tennis ball from hoop. Overarm throw at target from badminton court service box line approx. 10m from target. The child does not collect ball and returns to iPad.			●
5.	**Basketball Dribble**	Collects basketball from hoop. Dribbles around coloured spots on floor in *z* formation (body <u>and</u> ball move around outside of spots) with either hand. After dribbling around last spot, finishes with a dribble down the middle, returning ball to hoop/iPad.		○	●
6.	**Catch**	Runs forward and collects tennis ball from floor. Underarm throws ball against rebound net to catch from any distance without a bounce. Does not collect ball if dropped. Returns to iPad.		○	●
7.	**T-Agility**	Completes t-agility run, facing forwards throughout. Returns to iPad.	○	●	
8.	**Jumping Patterns**	Runs to coloured foot markers and hurdles. Follows jumping pattern sequence to finish (2-footed jump over hurdle → 2-footed landing → 2 left hops → 2 right hops → 2-footed jump over hurdle → 2-footed landing. Returns to iPad.	○	●	
9.	**Sprint**	Runs through start gate and then 10m sprint acceleration to finish line.		●	

Note: ● = primary skill category involved in task; ○ = secondary skill category involved in task

Table 2. Dragon Challenge Assessment Criteria

DC Task	Technique Criterion	Technique Criterion	Outcome Criterion
1. Balance Bench	1.1 Moves without hesitation up to turn	1.2 Body posture stable (*head & trunk stable, minimal arm flailing*)	1.3 Walks length of beam, completes full turn at 3/4 mark without falling off, dismounts at end zone
2. Core Agility	2.1 Hands & legs extended & held with tension, with shoulders & feet off the floor	2.2 Controlled & fluent transition through shapes	2.3 Completes 4 positions in correct order (*dish on back - arch on front - dish on back - arch on front*), rotating both ways
3. Wobble Spot	3.1 Non-support foot does not touch support leg/foot/wobble spot/floor	3.2 Body & head are stable/still	3.3 Completes 5 bean bag passes around body whilst balancing on wobble spot on one leg # 'correct' passes 0 1 2 3 4 5
4. Overarm Throw	4.1 Throwing arm moves in a backward arc to initiate throw (*shoulder rotates*)	4.2 Steps with the foot opposite throwing hand towards target	4.3 Overarm throw directly hits target (*ball should not bounce prior to hitting target*)
5. Basketball Dribble	5.1 Pushes ball with fingertips (*not slapping at the ball*)	5.2 Controlled directional dribbling	5.3 Dribbles around all spots using either hand. (*body & ball must move around outside of spots*). Cannot catch ball/use two hands simultaneously
6. Catch	6.1 Feet move in line with rebound	6.2 Catches ball with hands only (*must be caught without a bounce*)	6.3 Successful catch off rebound net (*must be caught without a bounce*)
7. T-Agility	7.1 Plants & drives off outside foot (*right to left & left to right*)	7.2 Side-stepping on balls of feet (*right to left & left to right; feet don't cross)*	7.3 Moves through all points of 'T' facing forwards (*must enter both right & left court tramlines*)
8. Jumping Patterns	8.1 Arms drive over first hurdle (*elbows bent & arms swing to produce force*)	8.2 Rhythmical pattern throughout	8.3 Completes jumping pattern sequence correctly. No contact with hurdles
9. Sprint	9.1 Drives off balls of feet, leaning forwards	9.2 Arms bent, driving forward & backwards (*arms bent at approx. right angles*)	9.3 Runs through start gate & then through to finish (*must be running not walking*)

Table 3. Age (mean±SD) and gender (%) of participants who took part in the Dragon Challenge in study phase 2 and 3

	Boys	Girls	Total
<u>Construct validity</u>			
Primary School Level			
Age (years)	10.9 ± 0.5	10.9 ± 0.5	10.9 ± 0.5
n (%)	765 (51.9)	709 (48.1)	1,474 (100)
Secondary School Level			
Age (years)	11.7 ± 0.3	11.6 ± 0.3	11.7 ± 0.3
n (%)	1,362 (48.9)	1,424 (51.1)	2,786 (100)
Total			
Age (years)	11.4 ± 0.5	11.4 ± 0.5	11.4 ± 0.5
n (%)	2,127 (49.9)	2,133 (50.1)	4,260 (100) †
<u>Concurrent validity</u>			
Age (years)	12.8 ± 1.5	12.1 ± 1.6	12.5 ± 1.6
n (%)	25 (55.6)	20 (44.4)	45 (100)
<u>Test-retest reliability</u>			
Age (years)	12.7 ± 1.6	12.3 ± 1.5	12.5 ± 1.5
n (%)	22 (50.0)	22 (50.0)	44 (100)
<u>Inter-rater reliability</u>			
Expert Assessor vs Newly Trained Assessor			
Age (years)	11.6 ± 1.6	12.0 ± 2.1	11.8 ± 1.8
n (%)	7 (46.7)	8 (53.3)	15 (100)
Expert Assessor vs Trained Assessor			
Age (years)	13.9 ± 0.5	13.0 ± 0.6	13.5 ± 0.7
n (%)	6 (50)	6 (50)	12 (100)
<u>Inter-rater reliability and Intra-rater reliability (video analysis)</u>			
3 x Expert Assessors			
Age (years)	11.3 ± 1.0	11.4 ± 1.1	11.3 ± 1.0
n (%)	6 (50)	6 (50)	12 (100)

Note: † On the basis of missing gender, 95 participants from the surveillance data were excluded from all construct validity analyses, except for the PCAs. For these analyses n=4,355.
Expert assessors: >50 hours of DC training and in situ experience; Trained assessor: 20 hours of DC training and in situ experience; Newly trained assessor: 5 hours of DC training.

Table 4. Descriptive statistics (mean±SD) for Dragon Challenge (surveillance data) score categories

Score Category	**Score Range**	**Boys** (n=2,127)	**Girls** (n=2,133)	**Primary** (n=1,474)	**Secondary** (n=2,786)	**Total** (n=4,260)
DC Total Score	0-54	33.8 ± 8.6**	31.1 ± 8.3	31.7 ± 8.3	32.8 ± 8.6**	32.4 ± 8.5
Technique Score	0-18	10.9 ± 3.7**	9.6 ± 3.8	9.9 ± 3.9	10.4 ± 3.8**	10.2 ± 3.8
Outcome Score	0-18	11.0 ± 3.8**	10.5 ± 3.6	10.3 ± 3.6	11.0 ± 3.7**	10.8 ± 3.7
Time Score	0-18	11.9 ± 2.5**	11.0 ± 2.6	11.5 ± 2.4	11.4 ± 2.7	11.4 ± 2.6
Stability Score	0-12	6.2 ± 3.3	6.6 ± 3.3**	6.1 ± 3.3	6.6 ± 3.3**	6.4 ± 3.3
Object Control Score	0-12	7.6 ± 3.2**	5.5 ± 3.1	6.3 ± 3.3	6.7 ± 3.3**	6.5 ± 3.3
Locomotor Score	0-12	8.1 ± 2.9	8.0 ± 2.8	7.8 ± 2.9	8.2 ± 2.8**	8.1 ± 2.9

Note: Stability skills = sum of technique and outcome criteria in tasks 1-3; Object Control skills = sum of technique and outcome criteria in tasks 4-6; Locomotion skills = sum of technique and outcome criteria in tasks 7-9. Differences examined using two-way analysis of variance (ANOVA) test.

*= significant sex/school level difference ($p < 0.05$)

** = significant sex/school level difference ($p < 0.001$)

Primary = Primary school-aged children

Secondary = Secondary school-aged children/young people

Table 5. Proportion (%) of children successfully demonstrating each Dragon Challenge criterion (surveillance/normative data)

DC Task		All (n=4,260)	Boys (n=2,127)	Girls (n=2,133)	Primary (n=1,474)	Secondary (n=2,786)
1.Balance bench						
1.1▪	Moves without hesitation up to turn	85.5	83.8	87.1*	84.5	86.0
1.2▪	Body posture stable	39.0	37.0	40.9*	39.5	38.7
1.3◊	Walks length of beam, completes full turn at 3/4 mark without falling off, dismounts at end zone	42.7	41.7	43.6	42.7	42.6
2. Core Agility						
2.1▪	Hands & legs extended & held with tension, with shoulders & feet off the floor	37.3	31.4	43.2**	33.0	39.6**
2.2▪	Controlled & fluent transition through shapes	41.1	37.9	44.3**	41.0	41.1
2.3◊	Completes 4 positions in correct order, rotating both ways	75.4	71.8	78.9**	70.5	77.9**
3. Wobble Spot						
3.1▪	Non-support foot does not touch support leg/foot/wobble spot/floor	50.5	50.3	50.8	46.0	52.9**
3.2▪	Body & head are stable/still	48.9	48.3	49.5	44.5	51.3**
3.3◊	Completes 5 bean bag passes around body whilst balancing on wobble spot on one leg	50.8	50.5	51.1	46.3	53.2**
4. Overarm throw						
4.1▪	Throwing arm moves in a backward arc to initiate throw	57.6	73.2**	42.0	56.2	58.3
4.2▪	Steps with the foot opposite throwing hand towards target	73.1	86.6**	59.7	71.7	73.9
4.3◊	Overarm throw directly hits target	47.9	53.5**	42.3	44.2	49.8*
5. Basketball Dribble						
5.1▪	Pushes ball with fingertips	61.4	75.7**	47.1	57.7	63.3**
5.2▪	Controlled directional dribbling	71.1	77.2**	65.0	67.1	73.2**
5.3◊	Dribbles around all spots using either hand. Cannot catch ball/use two hands simultaneously	64.2	69.9**	58.5	62.4	65.1
6. Catch						
6.1▪	Feet move in line with rebound	62.9	73.4**	52.5	60.8	64.1*
6.2▪	Catches ball with hands only	32.3	40.7**	24.0	31.0	33.0
6.3◊	Successful catch off rebound net	35.6	44.4**	26.9	34.8	36.0
7. T-Agility						
7.1▪	Plants & drives off outside foot	29.6	33.3**	25.9	27.8	30.5
7.2▪	Side-stepping on balls of feet	50.0	51.2	48.9	45.8	52.3**
7.3◊	Moves through all points of 'T' facing forwards	58.6	59.5	57.8	52.9	61.7**
8. Jumping Patterns						
8.1▪	Arms drive over first hurdle	72.2	71.9	72.5	72.9	71.9
8.2▪	Rhythmical pattern throughout	64.2	62.2	66.1*	60.0	66.4**
8.3◊	Completes jumping pattern sequence correctly. No contact with hurdles	65.5	62.4	68.6**	63.7	66.5
9. Sprint						
9.1▪	Drives off balls of feet, leaning forwards	70.2	74.2**	66.3	69.8	70.5
9.2▪	Arms bent, driving forward & backwards	76.6	79.1**	74.1	78.3	75.7
9.3◊	Runs through start gate & then through to finish	97.0	97.4	96.6	97.0	97.0

Note.
◊ = product/outcome characteristic/quality indicator (outcome of movement)
▪ = process/technique characteristic/quality indicator (technique or movement form)
*= significant sex/school level difference ($p < 0.05$)
** = significant sex/school level difference ($p < 0.001$)
Primary = Primary school-aged children
Secondary = Secondary school-aged children/young people

<u>Supplemental Digital Content 4:</u> Table 6. Rotated Component Matrix from PCA on Dragon Challenge criteria score

DC Task Criteria	Component								
	1	2	3	4	5	6	7	8	9
Balance Bench Technique 1	**.473**	.018	.103	.010	.094	.009	.022	.178	.136
Balance Bench Technique 2	**.951**	.075	.057	.043	.008	.031	.043	.003	.021
Balance Bench Outcome	**.952**	.061	.065	.040	.025	.042	.041	.016	.020
Core Agility Technique 1	.088	**.744**	.079	.028	.022	.002	.084	.107	.092
Core Agility Technique 2	.062	**.798**	.055	.042	.075	.037	.022	.096	.052
Core Agility Outcome	-.006	**.661**	.035	-.007	.082	.055	.089	.057	.044
Wobble Spot Technique 1	.088	.070	**.974**	.050	.056	.053	.048	.094	.062
Wobble Spot Technique 2	.087	.075	**.963**	.048	.056	.045	.043	.101	.066
Wobble Spot Outcome	.087	.068	**.977**	.051	.051	.049	.046	.095	.063
Overarm Throw Technique 1	-.023	-.015	.044	**.755**	.152	.076	.134	.044	.140
Overarm Throw Technique 2	.003	.019	.026	**.764**	.084	.101	.003	.009	.024
Overarm Throw Outcome	.074	.039	.038	**.544**	.058	.039	.012	.072	-.004
Basketball Dribble Technique 1	.021	.026	.032	.243	**.674**	.164	.058	.039	.102
Basketball Dribble Technique 2	.070	.113	.045	.102	**.857**	.109	.079	.085	.039
Basketball Dribble Outcome	.050	.074	.068	.029	**.800**	.067	.076	.087	.083
Catch Technique 1	.006	.074	.055	.142	.182	**.676**	.126	.058	.033
Catch Technique 2	.043	.018	.039	.060	.077	**.931**	.037	.051	.044
Catch Outcome	.035	.022	.040	.059	.075	**.945**	.030	.035	.044
T-Agility Technique 1	.030	.147	.044	.190	.075	.117	**.608**	.151	.160
T-Agility Technique 2	.046	.060	.042	.012	.065	.034	**.819**	.045	.060
T-Agility Outcome	.028	.036	.034	-.011	.066	.046	**.822**	.061	-.011
Jumping Patterns Technique 1	.078	.102	.059	.102	.081	.030	.125	**.430**	.350
Jumping Patterns Technique 2	.076	.127	.132	.068	.097	.075	.093	**.851**	.059
Jumping Patterns Outcome	.082	.116	.098	.038	.065	.052	.084	**.871**	.052
Sprint Technique 1	.045	.049	.073	.175	.106	.061	.116	.201	**.689**
Sprint Technique 2	.050	.053	.048	.103	.064	.032	.049	.122	**.768**
Sprint Outcome	.053	.062	.028	-.091	.026	.017	-.001	-.073	**.604**

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 6 iterations.

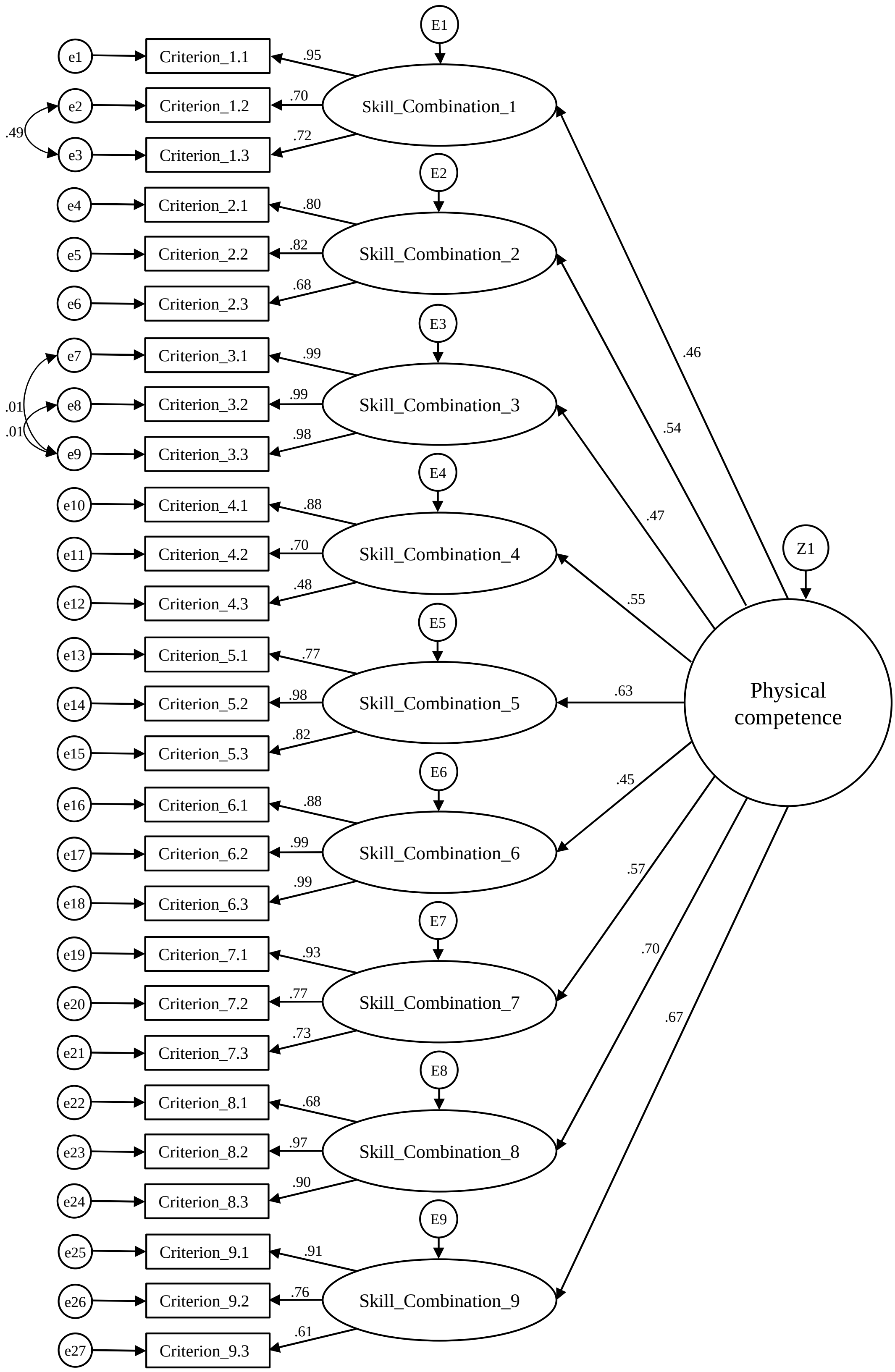

Figure 1. Factor Structure of DC

<u>Supplemental Digital Content 5:</u> Table 7. Inter- and intra-rater reliability results for Dragon Challenge scores and sub-category scores.

	Intraclass Correlation	95% CI		Strength of ICC
Reliability Test	ICC	Lower	Upper	
<u>Inter-rater Reliability</u>				
Expert Assessor vs. Newly Trained Assessor				
Dragon Challenge Total Score	.950 **	.859	.983	Good
Technique Score	.839 **	.592	.942	Good
Outcome Score	.916 **	.742	.972	Good
Time Score ♦	-	-	-	-
Stability Skills Score	.787 **	.479	.923	Good
Object Control Skills Score	.945 **	.847	.981	Good
Locomotor Skills Score	.903 **	.742	.966	Good
Expert Assessor vs. Trained Assessor				
Dragon Challenge Total Score	.986 **	.951	.996	Good
Technique Score	.987 **	.957	.996	Good
Outcome Score	.920 **	.647	.979	Good
Time Score ♦	-	-	-	-
Stability Skills Score	.941 **	.792	.983	Good
Object Control Skills Score	.972 **	.907	.992	Good
Locomotor Skills Score	.964 **	.882	.990	Good
3 x Expert Assessors				
Dragon Challenge Total Score	.942 **	.837	.982	Good
Technique Score	.889 **	.718	.964	Good
Outcome Score	.899 **	.758	.967	Good
Time Score	.990 **	.973	.997	Good
Stability Skills Score	.850 **	.666	.949	Good
Object Control Skills Score	.918 **	.803	.973	Good
Locomotor Skills Score	.904 **	.753	.969	Good
<u>Intra-rater Reliability</u>				
Dragon Challenge Total Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.990 **	.967	.997	Good
Expert Assessor 3	.999 **	.997	1.00	Good
Technique Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.977 **	.904	.994	Good
Expert Assessor 3	.995 **	.984	.999	Good
Outcome Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.947 **	.830	.984	Good
Expert Assessor 3	.989 **	.965	.997	Good
Time Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	1.000	-	-	Good
Expert Assessor 3	.991 **	.968	.997	Good
Stability Skills Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.963 **	.878	.989	Good
Expert Assessor 3	.997 **	.991	.999	Good
Object Control Skills Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.987 **	.955	.996	Good
Expert Assessor 3	.991 **	.969	.997	Good
Locomotor Skills Score				
Expert Assessor 1	1.000	-	-	Good
Expert Assessor 2	.962 **	.867	.989	Good
Expert Assessor 3	.975 **	.916	.993	Good

Note: * = $p < 0.05$; ** = $p < 0.001$. ♦ Time Score for the Expert Assessor vs. Newly Trained Assessor & Expert Assessor vs. Trained Assessor was not examined as times for each participant did not differ between assessors.