Summer 8-9-2022

Development of An Online Rater Training Program, and Its Impact on Accuracy of Scoring TGMD-3 Performance of Children with Developmental Disabilities

Hyokju Maeng
Georgia State University

ACCEPTANCE

This dissertation, DEVELOPMENT OF AN ONLINE RATER TRAINING PROGRAM, AND ITS IMPACT ON ACCURACY OF SCORING TGMD-3 PERFORMANCE OF CHILDREN WITH DEVELOPMENTAL DISABILITIES, by HYOKJU MAENG, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education & Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

_____
Deborah Shapiro, Ph.D.
Committee Chair

_____
Rachel Gurvitch, ED.D.
Committee Member

_____
Hongli Li, Ph.D.
Committee Member

_____
Elizabeth Webster, Ph.D.
Committee Member

07. 20. 2022
Date

_____
Jerry Brandon, Ph.D.
Chairperson,
Department of Kinesiology and Health

_____
Paul Alberto, Ph.D.
Dean, College of Education &
Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education & Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

Hyokju Maeng

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Hyokju Maeng
Department of Kinesiology and Health
College of Education & Human Development
Georgia State University

The director of this dissertation is:

Deborah R. Shapiro
Department of Kinesiology and Health
College of Education & Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Hyokju Maeng

EDUCATION:

Ph.D.	2016 – present	Georgia State University Department of Kinesiology and Health (Concentration in Physical Education Teacher Education)
Master's degree	2016	University of Michigan Sport Science (Movement Science)
	2006	Seoul National University Sport Science (Adapted Physical Education)
Bachelor's degree	2004	Seoul National University Physical Education

PROFESSIONAL EXPERIENCE:

8/2016 – present	Graduate Teaching and Research Assistant Georgia State University, Atlanta, GA, U.S.
9/2014 – 7/2016	Graduate Student Research Assistant University of Michigan, Ann Arbor, MI, U.S.

PRESENTATIONS AND PUBLICATIONS:

Maeng, H., & Shapiro, D. (2022). *Impact of a TGMD-3 Rater Training Program for Novice Raters on Scoring Accuracy of Fundamental Motor Skills among Children with Developmental Disabilities.* Virtual Poster Presentation at 2022 National Consortium Physical Education for Individuals with Disabilities.

Li, C., Haegele, J., Sun, F., Alves, M., Ang, S., Lee, J., Ng, K., Alves, I., Healy, S., Huang, W., Rintala, P., Tan, J., Wu, Y., Yang, H., Kama, E., Maeng, H., Schilemann, A., & Ding, D. (2022). Meeting the 24-hour movement guidelines and health-related outcomes among youth with autism spectrum disorder: A seven-country observational study. *Child and Adolescent Psychiatry and Mental Health*, *16*(1), 1-8.

Kwon, H., & Maeng, H. (2022). Impact of a TGMD-3 training program for pre-service adapted physical education teachers on scoring accuracy of fundamental motor skills among children with developmental disabilities. *Children*, *9*(6), 881.

Kwon, H., Maeng, H., & Kim, K. (2021). The Association between Physical Activity, Sleep Duration, and Incidence Risk of Dyslipidemia in Korean Adults: Korean Genome and Epidemiology Study (2001-2012). *Journal of Korean Society of Sports Science. 30*(4).757-771.

Maeng, H., & Shapiro, D. (2021). *Development of the TGMD-3 Rater Training Program for Accurate Scoring in Children with Developmental Disabilities: A Delphi Study*. Virtual Poster Presentation at 2021 National Consortium Physical Education for Individuals with Disabilities.

Maeng, H., Shapiro, D. (2020). *Testing of Factor Structure Models of the Test of Gross Motor Development- Third edition (TGMD-3.* 2020 SHAPE America National Convention & Expo, Salt Lake City, UT.

Maeng, H., & Shapiro, D. (2020). *Antecedent-Based Interventions to Increase Physical Activity among Individuals with Developmental Disabilities: A Systematic Review.* 2020 SHAPE America National Convention & Expo, Salt Lake City, UT.

Kim, G., & Maeng, H. (2020). *Social Media Analytics of Physical Education Professions. Poster presentation*. 2020 National Association for Kinesiology and Higher Education, Palm Springs, CA.

Kim, G., Maeng, H., & Sur, M. (2019). *Analysis of HPE Professions' Use of Twitter Power ed by Twitter-API. Oral Presentation.* 2019 SHAPE America National Convention & Expo, Tampa, FL.

Maeng, H.J., Yang, H., & Shapiro, D.R. (2018). *Reliability of the Test of Gross Motor Development- Third edition in children with developmental disabilities.* 2018 National Federation of Adapted Physical Activity Conference, Corvallis, OR.

Gurvitch, R., Lima, M., Hollibaugh, C., Sur, M., Maeng, H.J., & Kim, G. (2018). *EdTPA training the educator: The PhD student's Perspective.* 2018 PETE & HETE Conference, Salt Lake City, UT.

Kwon, H.J., Maeng, H.J., & Shapiro, D.R. (2018). *The association between sport club and internet use in Korean adolescents.*2018 SHAPE America National Convention & Expo, Nashville, TN.

Maeng, H.J., Webster, E.K., Pitchford, A.E., & Ulrich, D.A. (2017). *Inter-and Intrarater Reliabilities of the Test of Gross Motor Development–Third Edition Among Experienced TGMD-2 Raters.* Adapted Physical Activity Quarterly, 34, 442-455.

Maeng, H.J., & Shapiro, D.R. (2017). *Objective measurement of physical activity for students with autism spectrum disorder.*2018 National Consortium Physical Education for Individuals with Disabilities, Washington D.C.

Kwon, H.J., Maeng, H.J., & Haegele, J.A. (2017). *The association between physical activity and vitamin D in adolescents.* 2017 SHAPE America National Convention & Expo, Boston, MA.

Maeng, H.J., Webster, K., Pitchford, A., & Ulrich, D.A. (2016). *Reliability of the Test of Gross Motor Development-3 (TGMD-3).* 2016 SHAPE America National Convention & Expo, Minneapolis, MN.

Kwon, H., Maeng, H.J., Oh, J.W., & Kim, Y.S. (2014). *The effects of dancesport activity on age-related hormones in older adults with intellectual disability.* 2014 North American Federation of Adapted Physical Activity Symposium, Ann Arbor, MI.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS:

2017 – present	The National Consortium for Physical Education for Individuals with disabilities (NCPEID)
2015 – present	Korean Alliance for Health, Physical Education, Recreation, and Dance
2014 – present	The Society of Health and Physical Educators (SHAPE America)
2014 – present	The North American Federation of Adapted Physical Activity (NAFAPA)
2006 – present	Korean Society for Adapted Physical Activity and Exercise

DEVELOPMENT OF AN ONLINE RATER TRAINING PROGRAM AND ITS IMPACT ON THE ACCURACY OF SCORING TGMD-3 PERFORMANCE OF CHILDREN WITH DEVELOPMENTAL DISABILITIES

by

HYOKJU MAENG

Under the Direction of Deborah Shapiro

ABSTRACT

Based on the importance of fundamental movement skills (FMS) and common performance traits of persons with developmental disabilities (DD) that impact the performance of FMS, a rater training is necessary to evaluate FMS accurately among this population. The first purpose of this study was to develop a rater training protocol on the Test of Gross Motor Development-3 (TGMD-3) for novice raters. The second purpose was to validate this rater training protocol using a modified Delphi method. The third purpose was to examine novice raters' accuracy in scoring the motor skills of children with DD after completing an online rater training protocol. A total of eight experts completed two rounds of a modified Delphi method with intraclass coefficient statistic (ICC) of .75 or above providing evidence of consensus on content and presentation of training material. A total of 41 novice raters completed three rounds of training and scoring separated by 5 days each. Data analysis compared the change of rating

accuracy of novice raters with that of experts on the run and two-hand strike skills on the TGMD-3 across three different occasions. There was a significant impact on scoring accuracy of novice raters to score the run ($F(1, 39)= 56.431$, $p < .001$), two-hand strike skills ($F(1, 39)= 35.549$, $p < .001$), and the total skill score ($F(1, 39)= 64.323$, $p < .001$). The TGMD-3 online rater training program for novices in the present study provides a model training program to improve the accuracy of scoring FMS among children with DD.

INDEX WORDS: fundamental movement skills, children with disabilities, TGMD-3, online training, frame-of-reference training, Delphi method

DEVELOPMENT OF AN ONLINE RATER TRAINING PROGRAM AND ITS IMPACT ON THE ACCURACY OF SCORING TGMD-3 PERFORMANCE OF CHILDREN WITH DEVELOPMENTAL DISABILITIES

by

Hyokju Maeng

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Health and Physical Education

in

Department of Health and Kinesiology

in

the College of Education & Human Development

Georgia State University

Atlanta, GA
2022

Table of Contents

LIST OF TABLES ..... **iii**

LIST OF FIGURES ..... **v**

1 THE PROBLEM **1**

Purpose **15**

Research Hypotheses **15**

Significance of the Study **15**

Assumptions and Limitations **16**

Overview of the Study **17**

2 REVIEW OF THE LITERATURE **27**

3 METHODOLOGY **66**

Instruments **66**

Study 1: Methodology of developing for the TGMD-3 online frame-of-reference training program module development procedure **66**

Study 2: Methodology of validation for an online frame-of-reference training program Delphi study procedure **71**

Study 3: Methodology of evaluating rating accuracy **74**

participants **75**

Data Collection Procedures **77**

Data Analysis **79**

4 RESULT **81**

Delphi Study Validation of the TGMD-3 Online Training Modules **81**

Effect of a TGMD-3 Online Training Program on Rater Accuracy **100**

5 Discussion of the TGMD-3 Online Training Program on Rater Accuracy **106**

Conclusion **109**

Limitations **109**

Further Research Directions **110**

REFERENCES **112**

APPENDICES **143**

LIST OF TABLES

Table 1. Demographic Information of Expert Panels in the Modified Delphi Method……..…...72

Table 2. Foundation Scores of Two Skill Performance on the TGMD-3 by Experts…….....74

Table 3. Demographic Information of Participants……..…….....76

Table 4. Participants' Completion Period of Three Rounds……..…79

Table 5. Statistic Results from Round 1 of Frame-of-Reference Training Module Content…....87

Table 6. Result of Alignment of Content with Learning Outcome from Round 1…….…...88

Table 7. Round 1 Question Conformance & Alignment with Learning outcome of the Test of Module 1.. ……..….89

Table 8. Round 1 Question Conformance & Alignment with Learning outcome of the Test of Module 2…….....90

Table 9. Round 1 Result of Online Training Module Content…….……..……...90

Table 10. Statistic Results from Round 2 Frame-of-reference Training Module Content……96

Table 11. Result of Alignment of Content with Learning Outcome from Round 2….……..…97

Table 12. Round 2 Question Conformance & Alignment with Learning outcome of Test of Module 1……..……..……..……..……..……..……..…….....98

Table 13. Round 2 Question Conformance & Alignment with Learning outcome of Test of Module 2……..……..……..……..……..……..……..…….....99

Table 14. Round 2 Result of Online Training Module Content ……..……..……....99

Table 15. Descriptive result of the run skill ……..……..……..…….....100

Table 16. Pairwise comparison result of the run skill ……..……..…..101

Table 17. Descriptive result of the two-hand strike skill ……..…..……102

Table 18. Pairwise comparison result of the two-hand strike skill……..……...…....103

Table 19. Descriptive result of the total skill score……..……...…...104

Table 20. Pairwise comparison result of the total skill score……..…….....105

LIST OF FIGURES

Figure 1. A conceptualization of Newell's dynamic systems theory ……..…..……....31

Figure 2. Phase of motor development …….……..……..36

Figure 3. Study procedures …….…….…..……..77

Figure 4. Scoring Difference of the run skill between the expert and novice raters among 3 rounds…….……100

Figure 5. Scoring Difference of the two-hand strike skill between the expert and novice raters among 3 rounds…….……..102

Figure 6. Scoring Difference of the total skill score between the expert and novice raters among 3 rounds…….……..……...104

1 THE PROBLEM

Defining Fundamental Movement Skills

Fundamental movement skills (FMS) are universally used skills that are major components of physical development, and foundational forms for specialized movements in advanced games or sports (Burton & Miller, 1998; Barnett et al., 2014; Goodwin et al 2019; Hardy et al., 2013; Lubans et al., 2010). FMS forms a necessary base or core of the human movement. Children utilize this base to facilitate more possibilities for developing movement responses in their environments. The development of FMS generally involves movements that transfer the body from one location to another and/or launch or receive objects. These movements can be divided into three subgroups: locomotor skills (e.g., running), object control skills (e.g., throwing a ball), and stability (e.g., balancing; Goodway et al., 2019). Movement stability is defined as the ability to continue the balanced movement by coordinating actions of bones, muscles, joints, tissues, and neuromuscular systems against the force of gravity (Crockford, 2015). This involves static and dynamic balance, bending, curling, turning, twisting, stretching, and transferring weight. Locomotor movement is associated with traveling of the body from point to point through various performance patterns that primarily use the feet. The means are walking, running, jumping, hopping, leaping, skipping, galloping, and sliding from one location to another (Goodway et al., 2019; Haywood & Getchell, 2019). Manipulative movement involves moving or using an object with the hands, feet, or even the body to complete a task. This type of movement consists of two types of skills that are fine motor and gross motor skills. Fine motor skills pertain to the complex use of the muscles of the hand and wrist, such as typing, writing, and cutting. Gross motor skills involve the manipulation of an object with

accuracy and control, such as striking, kicking, dribbling, throwing, and catching (Goodway, Ozmun, & Gallahue, 2019). Gross motor skills in manipulative movements are often described as object-control skills.

Importance of Fundamental Movement Skills

FMS are important developmentally as they play a significant role in the development of children's physical, cognitive, and social growth. Children learn, use, refine, and apply various movement skill performance during childhood (Haywood & Getchell, 2019). For instance, advanced movements of throwing or striking can be transferred to various sports (e.g., baseball, tennis, badminton). This context can be adapted or varied at different levels across the lifespan depending on personal experiences (Goodway et al., 2019; Langendorfer, Roberton, & Stodden, 2011). Children, who develop movement skill competence in a variety of situations, tend to be more audacious and favorable to trying their skills at the next level of movement activities and sports. In addition, Stodden and colleagues (2008, 2013) emphasized that the development of FMS is important to consistently engage in physical activity across the lifespan to promote health related fitness. Systematic review studies (Holfelder & Schott, 2014; Lubans et al., 2010) similarly reported a significantly positive association between FMS competence and physical activity.

Learning FMS in childhood also impacts a child's cognitive development (Jaakkola et al., 2015). According to Rosenbaum et al. (2001), intellectual learning (i.e., cognitive skill) and motor skill acquisition are processed similarly. The authors reported that learning rates, training effects, and learning stages of intellectual skills were remarkably similar to acquiring movement skills. Lopes et al. (2013) presented that children with low FMS showed a higher probability of

low academic achievement compared to those with average or good FMS. Providing more opportunities to promote FMS for children may have a significant impact on academic performance.

The effect of motor skill competence on psychological development (i.e., social skills) in children has also been recognized in the literature (Stodden et al., 2008). Children's movement skills acted as determinants or correlates to promote physical activity as well as to influence perceived competence and self-efficacy. Several studies have been conducted to investigate the association between level of motor competence and self-perception (Cantell et al., 2003; Piek et al., 2006; Skinner & Piek, 2001). They presented that children's levels of motor competence were significantly related to their self-perception in many aspects, such as social acceptance, self-evaluation, and behavioral conduct. Given the importance of FMS development related to physical activity, cognitive skills, and social skills, they are frequently taught in preschool, elementary physical education, and adapted physical education programs (Kelly & Melograno, 2014; SHAPE America, 2014).

Fundamental Motor Skills and Instruction

Many studies regarding obtaining FMS in early childhood showed that significant improvement of FMS in young children stems from educational environments as opposed to free play time (Logan et al., 2012; Lubans et al., 2010; Morgan et al., 2013; Robinson et al., 2009). Instructional program interventions play a significant role in movement skill improvement among students. Given this evidence-based practice for the inclusion of FMS in educational settings, Standard 1 of the SHAPE National Standards for K-12 Physical education incorporates FMS content into the curricula for elementary school students from kindergarten to grade 5

(SHAPE America, 2014). Similarly, the Individuals with Disabilities Education Act (IDEA, 2004), defines physical education for students with disabilities to be inclusive of fundamental motor skills and patterns. Under IDEA, general and adapted physical education teachers must contribute to the process of individualized education plan (IEP) and implement specialized instructions of FMS and patterns for student with disabilities. The FMS learning objectives of students with special needs can be achieved by properly designed program content and instructional methods for those with disabilities. Winnick and Porretta (2016) stated that understanding the learning process and designing appropriate interventions are essential for teachings FMS to students with developmental disabilities.

Assessing Fundamental Movement Skills

Educators and researchers can assess the development of motor skill in children applying a quantitative and/or qualitative approach (Liu et al., 2017). A quantitative method uses numeric measurements of the outcome or product (e.g., running time, distance, accuracy, velocity) (Davids et al., 2003; McMorris, 2014), whereas a qualitative or process approach examines technique and describes the critical movement elements needed to achieve the task such as how the individual strikes a ball (Roberton et al., 2017; Seefeldt & Haubenstricker, 1982). Different types of assessment instruments have been developed using a quantitative or qualitative approach to evaluate FMS performance in children and adolescents (e.g., Test of Gross Motor Development [TGMD], Movement Assessment Battery for Children [MABC], Bruininks-Oseretsky Test of Motor Proficiency [BOMP]; Burton & Miller, 1998; Cools, De Martelaer, Samaey, & Andries, 2009; Goodway, Brian, Change, & Park, 2014). For purposes of the present study, only the TGMD will be reviewed and discussed.

The Test of Gross Motor Development (TGMD). The TGMD (Ulrich, 1985; 2000; 2019) is designed for use with children, both with and without disabilities. The TGMD is a qualitative assessment to measure gross motor skill performance in children. This assessment instrument has two approaches to evaluate FMS performance in children, the presence of an aspect of skill performance (qualitative; i.e., criterion-referenced) and assigned numerical value (quantitative; i.e., norm-referenced). The TGMD was developed to identify the FMS development of children between 3 to 10 years of age. The TGMD can be used to plan a program to improve children's FMS and evaluate the effect of FMS instruction or intervention (Burton & Miller, 1998; Ulrich, 1984). Common uses for the TGMD include the measurement of FMS performance in experimental research (Logan, Robinson, Wilson, & Lucas, 2012), and in clinical and educational settings to assess the current level of motor skill development for children. The TGMD-3 (Ulrich, 2019) has been released through a revision to accommodate changes in the normative population, as well as to incorporate recommendations from experts in the field of motor development and practitioners alike.

Reliable measurements and valid evaluations of levels of FMS in children are important components to precisely evaluate one's performance capability. Webster and Ulrich (2017) and Allen et al. (2017) assessed test-retest reliability of the TGMD-3 among children using intraclass correlation coefficients (*ICC*). The researchers presented the assessment instrument to have strong levels of agreement (*ICC*: 0.81 – 0.97). Studies testing inter-rater reliability of the TGMD-3 (Maeng et al., 2017; Rintala et al., 2017; Valentini et al., 2017) presented overall *ICCs* (0.51– 0.99) were acceptable to excellent. Similarly, intra-rater reliability has also been shown to be good to excellent (Allen et al., 2017; Maeng et al., 2017; Rintala et al., 2017). The studies found high reliability of the TGMD-3 to use as a tool when applying to typically developing

children. Additionally, reliability of the TGMD-3 should be investigated when using it with children with disabilities. According to Maeng et al. (2017), evaluating certain groups of children who may be more or less consistent in their performance such as children with disabilities, is also important to evaluate rater reliability separately for appropriate sub-groups.

Researchers examining reliability of the TGMD-2 and 3 found that several skills on the assessment reported lower reliability coefficients (Barnett et al., 2014; Kim et al., 2014; Kim et al., 2012; Maeng et al., 2017; Rintala et al., 2017). Among locomotor subtest skills, the run (*ICC*: 0.60 – 0.74; Kim et al., 2014; Kim et al., 2012; Maeng et al., 2017), slide (*ICC*: 0.60 – 0.74; Kim et al., 2014; Kim et al., 2012), and horizontal jump (*ICC*: 0.80; Kim et al., 2012; kappa coefficient: 0.39; Rintala et al., 2017) showed lower reliability coefficients compared to other skills. Regarding ball skill subtest, researchers noted three skills had lower reliability coefficients (*ICC*: 0.40 – 0.74 as fair or good), including the two-hand strike (Kim et al., 2014; Rintala et al., 2017), two-hand catch (Barnett et al., 2014; Kim et al., 2012; Maeng et al., 2017), and kick a stationary ball (Barnett et al., 2014; Kim et al., 2012; Maeng et al., 2017). These studies pointed out the need for training raters to have good-to-excellent reliability on the TGMD.

Urbina (2014) described that validation of measurement instrument must be verified to assure the property of the test for confirming the standardized data set. This is because validity is an essential element in the developmental process of an assessment instrument (AERA, APA, & NCME, 2014). The TGMD-3 has been investigated in various studies to examine its psychometric properties since the developer furnished the record form of the TGMD-3 in 2014 (Brian et al., 2018; Estevan et al., 2017; Magistro et al., 2018; Mohammadi et al., 2019; Valentini et al., 2017; Wagner et al., 2017; Webster & Ulrich 2017). The results of those studies

presented acceptable fit indexes using the factor analysis in both the locomotor and ball skills subtests on the TGMD-3.

Accurate Evaluation of Performance

The accuracy of ratings is the primary importance in a performance rating system to evaluate human performance, especially in FMS assessment instruments. Rating accuracy is typically evaluated by comparing an individual's ratings across dimensions to ratings made by expert raters (i.e., "true" scores). The closer these ratings are to the true score, the more accurate they are believed to be (Sulsky & Balzer, 1988; Werner & Bolino, 1997).

Providing effective rater training is important to support raters (i.e., researchers, practitioners, educators) to produce accurate assessment outcomes in any field (McIntyre et al., 1984; Roch et al., 2012; Woehr & Huffcutt, 1994). Especially, it is necessary for designing curriculum and programs to address the developmental levels of motor skills of children across a variety of teaching and coaching settings. Kennedy et al. (2013) presented that rating accuracy is important to assure the ability of motor skill performance in children. According to Sulsky and Balzer (1988), an evaluation method of rating accuracy was reviewed to compare rating scores of novice raters by expert raters across task elements on a selected assessment of performance. The researchers asserted that effective rater training could improve the level of rating accuracy among raters when appraising task performance of individuals. Gorman and Rentsch (2009) examined the effect of rater training on rating accuracy and found effective rater training improved the accuracy of performance rating among undergraduate students.

A study by Palmer and Brian (2016) supported the examination of scoring differences between novice and expert raters on the TGMD-2. The study showed there was no significant

agreement about scoring the TGMD-2 between the two groups. The researchers presented that more rigorous and constant training protocols are required to score children on the TGMD-2. The raters showed different severity/leniency levels caused by the raters individual characteristics when judging gross motor skill performance, even though the raters were familiar with rating of items on all editions of the TGMD. The study suggested that a rater training protocol should be executed to reduce rater effects, which potentially introduces errors to scores, and to achieve certain levels of evaluation across the items on the assessment instrument (Y. Kim et al., 2012). Kim and colleges (2012) investigated rater effects to score FMS items on the TGMD-2 among children with developmental disabilities. The study found a set of rater effects influenced reliability across FMS assessment outcomes on the TGMD-2. The researchers suggested a rater training to reduce rater effects caused by different rating accuracy across the items on the movement assessment as well as characteristics of raters. Rater training intends not only to provide the opportunity to accumulate rating experiences on the movement assessment, but also to achieve certain level of rating accuracy across the skill items on the movement assessment (Myford & Wolfe, 2003). However, there have been few training protocols for researchers or practitioners in the field of physical education to score the TGMD-3 among children, especially children with disabilities.

Rater Training

Four types of rater training have been proposed to improve rater accuracy: (a) rater error training, (b) performance dimension training, (c) behavioral observation training and (d) frame-of-reference training (Bittner, 1948). These strategies provide a comprehensive framework of assessment as well as content of training for raters to improve rating accuracy (Smith,1986;

Spool. 1978). According to McIntyre and colleges (1984), rater training has shown two major benefits:(a) to enhance raters' knowledge and skills for carrying out evaluations, and (b) to motivate raters to use the knowledge and skills learned in the training program. Of these four strategies, only the frame of reference training emphasizes the importance of raters' awareness of the multidimensional performance criteria in comparison with the common errors often observed with actual performance (Bernardin & Buckley, 1981). Given this methods similarity to the format and scoring of the TGMD-3, it will be the sole focus of rater training for this study.

Frame-of-reference training. This rater training strategy was designed to familiarize raters with identifying correct performance dimensions. Furthermore, it was for diverse raters to evaluate ratees' different performance under natural circumstances through sharing a common framework and conceptualization of the performance (McIntyre et al., 1984; Roch et al., 2012). Thus, this strategy uses the training method that enables raters to systematically compare information regarding actual versus desired performance (Roch & O'Sullivan, 2003). Frame-of-reference training has been used to ensure more accurate rating using training content with the presentation of sample performance based on the correct performance dimension (Smith, 1986). Primarily, frame-of-reference training was developed to supplement the inaccurate results of rater error training (Rosales Sánchez et al., 2019).

Since the frame-of-reference strategy was proposed, it has been widely used in various studies (Aguinis et al., 2009; Bernardin & Pence, 1980; Cardy & Keefe, 1994; Chirico et al., 2004; Gorman & Rentsch, 2009; Keown-Gerrard & Sulsky, 2001; Lievens, 2001; Loignon et al., 2017; Schleicher et al., 2002). Those studies showed the positive effect of the frame-of-reference training on rating accuracy when scoring individual performance. Especially, Lievens and Sanchez (2007) found that trained raters with the frame-of-reference training had significantly

higher values in validity, interrater reliability, and rating accuracy compared with untrained raters. Woehr and Huffcutt (1994) examined a meta-analysis of frame-of-reference training, and they reported a large average effect size (d = .83) of frame-of-reference training studies that compared trained and untrained groups. Those studies support that frame-of-reference training positively impacts accurate rating of performance by influencing raters' memories and perceptions to score ratee's performance.

Online training

Online training is a form of instruction that is completed on the internet. It involves a variety of multimedia components including video, audio, graphics, and web-links as educational technology (Mangal & Mangal, 2009). Online training technology has been introduced to be an effective and efficient method to distribute knowledge and enhance individuals' intellectual context (Chafouleas et al., 2015). Learning through online training is more meaningful for leaners than a traditional face-to-face instruction due to its convenience and flexibility (Ginder & Stearns, 2014).

Frame of reference training procedures have been used with online training protocols. Chafouleas and colleges (2015) investigated the impact of frame-of-reference training on the accuracy of performance rating using an online approach method to random rater samples. Their findings supported the effect of online frame-of-reference training on increased rating accuracy regardless of rater discrepancy and varied settings. The researchers concluded online rater training could be an important application method for accurate rating on different targets and performance levels. According to Aguinis et al. (2009), online frame-of-reference training helped to reduce the impact of rater biases in performance appraisal. They presented frame-of-reference

training as an effective intervention to minimize the scoring discrepancy compared to experts. Therefore, the present study will develop an online frame-of-reference training for the TGMD-3 to help ensure accurate rating on FMS.

Delphi Method

The Delphi method is a structured communication technique, originally developed as a systematic and interactive forecasting method which cooperates with a group of experts (Linstone & Turoff, 1975; Rowe & Wright, 2001; Sackman, 1974). The experts are asked to answer questionnaire survey or scale in two or more rounds. During each round, a facilitator or researcher provides a summary of the experts' responses from the previous round as well as the reasons or evidence for their decisions. The experts are encouraged to reevaluate their answers through considering answers of other experts in the group. Throughout this process, the range of answers will be narrowed down and converge towards the ideal outcomes. Finally, the process is stopped and results are determined after deriving a refined criterion with eligible mean or median scores in the final round (Rowe & Wright, 1999).

Several researchers have applied the Delphi method to make a consensus in assessment tool development (Brian et al., 2016; Columna et al., 2014), educational effectiveness (Bulger & Housner, 2007; Ross et al., 2014; Taliaferro & Bulger, 2020), and implementation features (Dyer et al., 2011) among informed experts. These studies showed that the Delphi method provided a justifiable and practical means for developing measurement or survey tools as well as curriculum content. The present study utilized the Delphi method to develop and validate the online frame-of-reference training program for the TGMD-3 when scoring FMS among children with DD.

Fundamental Movement Skills in Children with Developmental Disabilities

Students eligible for special education services are a diverse group with varying disability classifications such as physical, intellectual, developmental disabilities or sensory impairments. Interest in the present study is students with developmental disabilities (DD). Developmental disability is an assorted group of conditions that affects the trajectory of the individual's physical, intellectual, or emotional development, influencing personal, social, academic, or occupational function (Center for Disease Control and Prevention [CDC], 2013; American Psychiatric Association, 2013).

There are multiple types of developmental disabilities, such as intellectual disability (ID), autism spectrum disorder (ASD), Down syndrome, and cerebral palsy. Children with developmental disabilities (DD) show unique characteristics due to a limitation in physical, learning, language, or behavior areas (Center for Disease Control and Prevention [CDC], 2013). Children with DD may have limitations in reciprocal social interaction, verbal and nonverbal communication and imaginative activity, repertoire of activities and interests, cognition, fundamental movement skills (FMS; i.e., gross motor skills), physical fitness, and behavior patterns (American Psychiatric Association, 2013). Children with DD, tend to be significantly delayed in their FMS compared with peers without disabilities (MacDonald et al., 2013; Rintala & Loovis, 2013; Staples & Reid, 2010; Westendorp et al., 2011). A number of studies have shown that children with ID (Simons et al., 2008; Simons & Eyitayo, 2016; Westendorp et al., 2011) and ASD (Edwards et al., 2017; Liu, Breslin, & Elgarhy, 2017; Lloyd, MacDonald, & Lord, 2013; Pan, Tasi, & Chu, 2009; Staples & Reid, 2010) tended to demonstrate delayed FMS development. According to Westendorp and colleges (2011), motor delays among children with ID appeared in motor performance requiring greater speed and movement control. By all

accounts, characteristic of children with ASD in motor development was similar to students with ID. Common among them is a delay of FMS (Allen et al., 2017; Block, 2016; Provost et al., 2007; Staples & Reid, 2010). In some cases, children with DD may show completely different performance on the movement assessments (e.g., walking or running across the room instead of required performance of galloping or skipping). In other situations, students with DD may move their body in atypical ways (e.g., rocking or self-stimulatory behavior) while performing an overhand throw. Behavioral challenges such as those just described among children with DD make it difficult to not only perform motor skills on the TGMD properly but also evaluate motor skill performance by raters accurately.

Skill selection for the online frame-of-reference training.

For developing the online frame-of-reference training for the TGMD-3, the present study used 2 skills which consists of one locomotor skill (run) and one ball skill (two-hand strike). There are three reasons for the selection of these skills for the training module in this study: (a) low reliability in preceding research, (b) lower levels of fundamental movement skill development among children with DD, and (c) high frequency of use in game or sports activity in elementary physical education programs.

According to studies examining reliability of the TGMD-2 and 3, lower reliability coefficients (*ICC*: 0.40 – 0.74) among locomotor and ball skill were found in the run (Kim et al., 2014; Kim et al., 2012; Maeng et al., 2017) and two-hand strike (Kim et al., 2014; Rintala et al., 2017). Those coefficient values were fair to good (Fleiss, 2011). ICC values rated fair to good have been considered insufficient about meeting acceptable levels of reliability of the TGMD-2 and 3 even though other skills on the assessment were rated as having excellent ICC values.

Thus, it was inferred there might need to be procedures or strategies developed (such as rater training protocols) to prevent or minimize the rater effects in performance ratings (Myford & Wolfe, 2003) and ensure higher ICC ratings for these specific skills.

Delayed development of FMS among children with DD was the second reason for the selection of the run and strike. For children with DD the run showed not only lower scores (Pan et al., 2009; Westendorp et al., 2011) but also lower percentages of mastery of performance criteria for each skill (Capio et al., 2018; Pan et al., 2009; Rintala & Loovis, 2013) compared to typically developing children. Delayed developmental tendency had also been document in children with DD in run and two-hand strike (Capio et al., 2018; Edwards et al., 2017; Pan et al., 2009a; Rintala & Loovis, 2013; Westendorp et al., 2011).

The third reason that these skills on the TGMD-3 were selected for the rater training module is related to physical education programs in elementary school. Physical education curricula for children from Kindergarten to grade 5 are established to help their physical, psychological, and social developments according to Standard 1 of the SHAPE National Standards for K-12 Physical education (SHAPE America, 2014). This standard recommends a variety of FMS and physical activity. For children with disabilities, the Individuals with Disabilities Education Act (IDEA, 2004) states that physical education must include instruction on FMS and patterns. The run is a fundamental skill required for participation in activities requiring running, chasing, dodging, and fleeing and is used across sports like basketball, football, and soccer. The two-hand strike is a practical skill for participation and inclusion in activities like baseball, tennis, and badminton for example. Therefore, these FMS play a significant role in engagement in PE and sport for children (Clark & Metcalfe, n.d.; Goodway et al., 2019; Kalaja, 2012).

Purpose

The purpose of this study was threefold: first, to develop a pilot online rater training program for scoring the Test of Gross Motor Development-3 (TGMD-3), second, validate the online training program and third, investigate the impact of the rater training program on accuracy of scoring fundamental movement skill performance on the TGMD-3 of children with developmental disabilities.

Research Question and Hypotheses

The following research question guided this study: Is there a difference in rater accuracy for the run, two-hand strike, total score across three scoring occasions?

Hypothesis #1: Rater accuracy for the run, two-hand strike and total score will improve from the 1st to 2nd scoring

Hypothesis #2: Rater accuracy for the run, two-hand strike and total score will improve from 2nd to 3rd scoring

Hypothesis #3: Rater accuracy for the run, two-hand strike and total score will improve from 1st to 3rd scoring

Significance of the Study

Raters (i.e., educators, students in higher education) learn crucial points of movement skill assessment when they score or evaluate FMS among children with and without disabilities. Acquiring a skill of accurate scoring of movement skills among children, especially those with disabilities, was beneficial to a practitioner not only to teach FMS as well as games, sports, and dance, but also to provide activity services (Barnett et al., 2013; Goodway et al., 2019; Haywood

& Getchell, 2019). Based on the importance of FMS and common performance traits of persons with DD that impact performance of FMS, a rater training was necessary to evaluate FMS accurately among this population. This study documented whether scoring accuracy of raters through a rater training program on the TGMD-3 among children with DD could be more like expert raters. The results of this study may contribute to the body of evidence for the effectiveness and necessity of a rater training to improve rating accuracy of novices. Second, there were not enough opportunities to learn accurate scoring of FMS among children with and without disabilities even though some physical education programs in higher education offer curriculum on fundamental motor skills. The rater training modules developed for this study were intended to serve as a pilot curriculum to train rater for greater reliability in scoring. The modules for training raters in this study might serve as a comprehensive and practical curriculum for use in physical education courses in which the instruction and assessment of FMS are taught.

Assumptions

This study supposes the following assumptions:

- The Test of Gross Motor Development-3 (TGMD-3) provides the standardized performance criteria to evaluate the performance proficiency among children with developmental disabilities (DD).
- Data subjects are children with DD diagnosed by a psychiatrist or psychologist.
- The rater training module is an appropriate method for all novice raters in the study.
- The rater training program provides understandable information to score accurately motor skill performance of children with DD.

- Primarily, novice raters' rating on the TGMD-3 is significantly different compared to expert raters.

Limitations

- Novice raters may have different experiences and knowledge about fundamental movement skills in children.
- The recruitment of novice raters was restricted to individuals majoring in Kinesiology. Potential raters in other majors, such as special education, will likely show different results.
- The training program used an online format. The researcher could not control individual training and scoring conditions optimally. The results of this study may be different with a face-to-face training with novice raters.
- The data subject was one boy with DD. Different age, gender, and the ability of FMS performance caused by individual characteristics of children with DD may show different rating accuracy by raters.
- Due to the small sample size, the results may have potential limitations with generalizability of the findings.

Overview of the Study

This study developed and validated a pilot online frame-of-reference rater training program for scoring fundamental motor skills (FMS) on the TGMD-3 and examined the impact of the training program on novice raters' rating accuracy when scoring FMS performance of children with developmental disabilities (DD). This study was constructed in three parts: (a)

developing online frame-of reference training modules, (b) validating the modules, and (c) examining the effect of the rater training program on the rating accuracy to score FMS among children with DD.

Part 1: Development of online frame-of-reference training modules

Identifying frame of reference training criteria. Frame-of-reference training criteria were identified through a literature review and applied to establish effective training strategies for novice raters to score FMS of children with DD. The frame of reference criteria included: (a) identify correct performance dimension under natural circumstances, (b) provide information to score the variables in question, (c) describe correct performance on variables in question, (d) systematically compare information regarding actual versus desired performance, (e) provide information according to characteristics of performer, (f) provide correct scoring feedback on actual performance, and (g) offer practical questions with answers.

Identifying online training criteria. This study included components of online training in terms of a methodological perspective necessary to meaningfully train novice raters to score FMS correctly among children with DD. This study used the following criteria to guide the development of the online training modules: (a) training program was easy to access, (b) utilizes visual resources (angle of camera, angle of performance front and side view), (c) voice narration is clear and understandable, (d) voice narration is at an appropriate pace, (e) tests are available online, and (f) written material in video is easy to read.

Outlining learning outcomes for training modules. This study created learning outcomes for training module 1 and 2, respectively. For module 1 in which the general information of the TGMD-3 was introduced, four learning outcomes were selected: (a) the rater understands the TGMD-3 and its components, (b) the raters can identify the skills in the

locomotor subtest, (c) the raters can identify the skills in the ball skills subtest, and (d) the raters understand how to score a skill.

The second module includes information about developmental disability (DD) and behavioral and psychomotor characteristics of children with DD as well as the correct performance cues for 2 skills on the TGMD-3 one locomotor skill (i.e., run) and one ball skill (i.e., two-hand strike). The learning outcomes for module 2 were the following: (a) the rater can explain about DD, (b) the rater can list behavior and movement characteristics of children with DD, (c) the rater can recognize the performance criteria of the run skill in the locomotor subtest, (d) the raters can recognize the performance criteria of the two-hand strike in the ball skills subtest, and (e) the rater can score FMS among children with DD according to each criterion on the TGMD-3.

Developing tests to check for novice rater understanding. Tests for checking understanding of each module among novice raters were required to encourage their engagement in the training as well as maximize effectiveness of the rater training. Thus, the tests for modules 1 and 2 were made up of the most important concepts from each module aligned with the learning outcomes of the respective model (see Appendix G & I). Each test has 10 questions consisting of true or false and multiple-choice questions. A final version of each test was developed using expert panel feedback (see Appendix F & H).

Developing evaluation forms for expert validation. An evaluation survey was developed for expert raters to evaluate the effectiveness of the online frame-of-reference training modules. The evaluation examined the degree of alignment between the components of a frame of reference training and online training and the content in each of the modules. Second experts

evaluated the effectiveness of training content as well as the tests for modules 1 and 2. All ratings used a 5-point Likert type scale (see Appendix C).

Development of Module #1 Content.

Content and format of Module #1. Using multimedia online platform, a brief description presentation of the TGMD-3 using presentation format slides and narration was provided to help raters understand the general information of this movement instrument (approx.3 mins). Using the administration of the TGMD-3 on YouTube (Webster, 2014), this module then demonstrated thirteen fundamental movement skills on the TGMD-3 separated into locomotor and ball skills (approx.4 mins). The developer of the TGMD-3 demonstrated 6 locomotor and 7 ball skills. Lastly, the TGMD-3 record form and scoring methods were presented (approx. 8 mins). Presentation slides explained how to score FMS on the TGMD-3 in Module #1 with a sample illustrated example of the gallop skill. It took approximately15 minutes to complete Module #1. An evaluation test with 10 questions was required to ensure novice rater's understanding of content in Module #1 (Appendix F & G).

Development of Module #2 Content

Content and format of Module #2. Module #2 had a similar structure as Module #1. Module 2 began with information about developmental disability (DD; 4 mins) and behavioral and psychomotor characteristics of children with DD (5 mins; see Appendix B). Presentation slides with example videos were used to introduce the definition of DD, general characteristics and unique behavioral and psychomotor characteristics (e.g., challenging behavior, low motor function) among children with DD. The correct performance and cues for 2 skills on the TGMD-3 one locomotor skill (i.e., run; 4 mins) and one ball skill (i.e., two-hand strike; 6 mins) were provided to score actual performance among children with DD. This module explained how to

correctly score each performance criterion of the run and two-hand strike skills on the TGMD-3 among children with DD. For instance, a child with DD may come out of the path when s/he is required to perform the run skill on the TGMD-3. Another example could be a child with DD who might push a ball using a bat instead of striking the ball with two-hands when performing the two-hand strike on the TGMD-3. In addition, training content in this module described how to score specific components of each performance criterion either 1 or 0 on a trial. Acceptable reasons were given to help understand correct scoring on the performance criteria. It took about 19 minutes to complete Module #2. An evaluation test with 10 questions was required to ensure novice rater's understanding of content in module 2 (see Appendix F & G).

Part 2. Validating the Online Frame-of-Reference Training Modules and Tests: A Modified Delphi Method.

The present study examined content validity of the rater training program using a modified Delphi method. The modified Delphi method was primarily developed by Ziglio (1996) to effectively generalize research protocols. Two rounds of administration were used in the current study.

Selection of expert panel. Eight professionals were involved in the administrative procedures to evaluate and support the validation of the rater training protocol in this study. Panel members have expertise in teaching and studying motor development using the TGMD as well as teaching experience related to the behavior and FMS among individuals with DD. The expert panel consisted of 8 members: (a) university faculty with expertise motor development (n = 3), (b) university faculty with expertise teaching FMS among children with DD (n = 3), and (c) APE teachers with at least 5 years of teaching experience of FMS to children with DD (n = 2).

Prospective panel members were personally contacted through e-mail or telephone, provided a description of the study, its procedures, and their contribution to the study. Experts were given an honorarium for their participation in the validation phase of this study.

Round 1. An email was sent to panel expert containing a description of the study, the modified Delphi method process, the module evaluation questionnaire, and a timeline for completion. Reminders to complete round 1 of the review was sent via telephone call or through online video chat. A copy of the forms completed by the panel experts can be found in Appendix C. Panel experts responded to validation questions using a 5-point Likert type scale ranging from 1=very poor to 5 =very strong alignment or content representation. The expert panel members provided independent ratings of subject-matter content of the TGMD-3 and how well the training protocol provided information to score performance on the TGMD-3 among children with DD with respective feedback to enhance the rater training program (Lawshe, 1975; see Appendix C)

Round 2. The alignment with learning outcomes of modules 1 and 2 were evaluated for content validation (see Appendix D and G). Moreover, every module and question requested feedback if the rating score by experts was 3 or below which indicated poor or moderate alignment with learning outcomes. Individual or group discussions with expert panel members using a telephone call, an e-mail, and a video chat were implemented to make improvements to the modules and tests. The response list outlining the revisions made to the content in the modules and tests was provided to the expert panel members explaining what parts the author modified and why it was adapted. Expert panel members provided a second round of reviews with comments to guide any additional changes to the modules and tests.

The intraclass coefficient statistic was used to analyze the agreement between expert raters for calculating their agreement of content in each module and test. The above procedures

were repeated until ICC coefficient reaches at least 80% or higher agreement between individual experts on each module for the final version of modules. The ICC analysis was conducted using SPSS version 27.0 for Windows (IBM Corp., Armonk, NY, USA).

Part 3. Examining the Effects of a TGMD-3 Online Frame-of-Reference Training Program on Rating Accuracy

The video of a child with DD performing the run and two hand strike were shared with three expert raters who were asked to score the FMS using the TGMD-3 scoring form. There were three eligibility criteria to recruit the expert raters: (a) a graduate degree in motor development or adapted physical education/activity, (b) experience administering and scoring the TGMD-3, and (c) a minimum of 5-years of experience teaching FMS curricular content to children with DD in physical education, adapted physical education, or physical activity programs. Expert raters independently scored the vide. Any disagreements were addressed through discussion until reaching a 100% agreement. The scores by the expert raters on the TGMD-3 served as the foundation on which to compare the accuracy of scoring between novice versus expert raters.

Participants

Novice raters. A total of 41 novice raters completed all three rounds of the study. Novice raters were undergraduate students from a university in the U.S. and recruited according to the following inclusion criteria: (a) majoring in health and physical education or kinesiology, (b) a maximum of one course in motor development, motor behavior, motor control or movement assessment, and (c) no experience using the TGMD to score fundamental movement skills (FMS) among children with and without disabilities. The demographic characteristics of the participants can be found in the participant section of Chapter 3.

Data Collection Procedures

The video of one male subject with DD between 9 years old, was used to assess the accuracy of scoring between novice and expert raters. The male subject with DD had an average level of performance in both run and two-hand strike skills. According to Ulrich (2019), the '*Average*' performance level was defined as the descriptive term of performance between the 21th to 73th percentile. Present study designated that 5 to 7-points out of 8 for the run skill and 7 to 9-points out of 10 on the two-hand strike skill scores (see Appendix G) were needed to meet the criteria for '*Average*' performance level. Content and videos for the modules were housed on the University Qualtrics system. Novice raters were given an individual access code to the Qualtrics content. Novice raters completed a total of three rounds of scoring. Each training is described below.

Round 1 – Module 1 (intro TGMD-3) with demographic questionnaires and 1stscoring. The novice rater completed module 1 (see Appendix A) and demographic questionnaires (see Appendix E). Module 1 introduced the TGMD-3 and described the performance components of each skill item and how to complete the TGMD-3 examiner record form using the YouTube resource of the TGMD-3 (Webster, 2014). This module included the standardized performance of 13 skills based on the performance criteria of the TGMD-3. Those 13 skills consist of 6 locomotor skill and 7 ball skills. It took about 15 minutes to watch module 1. For checking understanding of module 1, raters took a test consisting of 10 questions about the content of the module 1 (see Appendix F). Raters passed module 1 when they scored above 80 percent on the test. Novice raters had an unlimited number of trials to pass the test for module 1. Immediately following completion of module 1, the researcher provided raters with the video of the data subject and ask raters to

independently score the performance videos within five days (i.e., 1st scoring; see Appendix G).

Round 2 – Module 2 (rater training) and 2ndscoring (intervention). Within 5 days of finishing round 1, the novice raters completed module 2 (see Appendix B). Module 2 presented information on how to correctly score the selected skills when evaluating the performance of a child with DD. This module described information on the behavioral and FMS characteristics of children with DD (e.g., challenging behavior, low motor function; see Appendix B), as well as how to score their performance of the 2 skills on the TGMD-3 (i.e., run, two-hand strike). This module provided sample performance videos of different children with DD and how to accurately score their performance on the TGMD-3. It took about 20 minutes to complete module 2. For checking understanding of module 2, raters completed a 10-question test about the content of the module 2 (see Appendix I). Raters passed module 2 when they scored above 80 percent on the test. Raters had unlimited number of opportunities to pass the test. Upon successful completion of module 2, the researcher resent a video link of the data subjects and asked novice raters to score the same performance videos within five days (i.e., 2nd scoring; see Appendix I).

Round 3 – Module 2 (same rater training) and 3rd scoring (intervention). Within 5 days of finishing round 2, novice raters completed the same process as round 2 then scored the same video again within five days of completing module 2 a second time (i.e., 3rd scoring; see Appendix I).

Data Analysis

The descriptive statistics of the TGMD-3 scores were calculated for total gross motor scores (i.e., selected two skills), one locomotor (i.e., run) and one ball skills (i.e., two-hand strike) for each of the 1st, 2nd, and 3rd scoring, respectively. A repeated measure analysis of variance (RM-ANOVA) was used to analyze the changes in scoring of the novice raters among the three different occasions (i.e., 1st, 2nd, and 3rd scoring) compared to expert raters. The RM-ANOVA was used to compare means across different occasions. Additionally, *t*-test was conducted to investigate the score difference between expert and novices after each round. Finally, effect sizes were calculated for each of the run, two-hand strike, and total score. Partial eta squared (η_p^2) was interpreted using the following recommendations: .01 = small, .25 = medium, and .4 = large effect (Cohen's *f* ; Cohen, 1992). All analyses were conducted using SPSS version 27.0 for Windows (IBM Corp., Armonk, NY, USA).

2 REVIEW OF THE LITERATURE

Understanding Motor Development

Motor behavior is an umbrella term comprised of three sub-disciplines: motor control, motor learning, and motor development used to understand the nature and causes of goal-directed movement (Ives, 2013). Movement is the interaction of physiology (motor control) in which the focus is on the sensory neurophysiological systems and musculoskeletal system, and psychology (behavior or motor learning) emphasizing the role of the mind in acquiring, planning, initiating, and modifying movements (Ives, 2013). According to (Schmidt et al., 2018), motor learning is the relatively permanent shift in movement ability resulting from training and experience. Motor development refers to the development of an individual's physical condition and ability to perform a variety of movements throughout the lifespan (Haywood & Getchell, 2019). Motor development is directly or indirectly related to psychology (i.e., affective domain) and neuroscience discipline (i.e., cognitive, and psychomotor domains). The focus of this paper is on motor development. The following sections provide a detailed explanation of the field of motor development.

Defining motor development. Motor development is best understood by examining three key concepts: (a) continuous and cumulative process, (b) age association, and (c) sequential event (Haywood & Getchell, 2019). First, motor development is regarded as a continuous process of change in one's functional ability to live, move, and work. Second, aging influences growth and development of individuals. Even though individual growth and maturation related to aging influence the acquisition of movement abilities, motor development of individuals shows different acquisition levels, paces, and rates throughout life. The last characteristic of motor

development is the notion of development as a sequential process. Development occurs through a series of steps to reach the ideal movement outcome.

Motor development theories. Throughout childhood, movement patterns are developed concurrently with cognitive and affective development, that coexist with environmental changes (Goodway et al., 2019); Haywood & Getchell, 2019).There are six major theories to explain motor development : (a) Maturational theory (Gesell, 1928), (b) Cognitive developmental theory (Piaget, 1936; 1954), (c) Ecological theory- Dynamic systems theory (Newell, 1986; Newell & Jordan, 2007), (d) Ecological theory- Behavior setting theory (Barker, 1969; Bronfenbrenner, 1979; 1992), and (e) Information processing theory (Schmidt et al., 2018; Schmidt & Wrisberg, 2008).

Maturational theory. Maturational theory (Gesell, 1928) was the first developmental theory to identify universal characteristics of human behaviors. According to Gesell, the sequence of individual physiological growth and developmental characteristics is controlled by personnel biological factors (Crain, 2015). With maturation comes a sequential emergence of cognitive, moral, personality, and motor development. Generally, typical phases of behaviors advance one or more stages each year. Studying infant behavior, Gesell (1928) developed a set of behavioral norms that explained the sequential and predictable stages of patterns according to the primary maturational process. Thus, the theory has an assumption that motor development follows an internal or inherent process stemming from the biological development of humans. Behavior patterns in the natural process of development of children eventually result in organized movement performance. Individuals' environment may not change genetic and biological development process of individuals, but only influence the speed of the developmental process. In the 1930s, this maturational perspective regarding motor development became a

popular theory to explain each stage of the developmental sequence of patterns. Pediatric physical therapists use the concepts of Gesell's maturational theory to help determine what babies should be able to do at various ages (Goldstein & Naglieri, 2010). However, this theory failed to specify how much variation could be expected at each age (Wolraich, 2008).

Cognitive developmental theory Piaget (1936; 1954) argued that cognitive development of individuals was a progressive reorganization of mental processes resulting from individuals' biological maturation and environmental experience. This theory claimed four major stages of cognitive development: (1) sensorimotor (birth to 2 years), (2) preoperational (2 to 7 years), (3) concrete operational (7 to 11 years), and (4) formal operational (11 to after). Using Piaget's stages of development, researchers were able to verify the association between cognitive development and motor skill ability. Einspieler et al. (2016) reviewed studies that examined the developmental association between cognitive outcome and general movements in children from 3 - 5 months to 7 – 10 years of age in which he found infant's motor repertoire during infant period to be predictive of cognitive outcomes in childhood. Leonard (2016) presented evidence for the impact of motor skill on social and cognitive development in children with Developmental Coordination Disorder (DCD). This study concluded that the development of motor skills could be interactive with the developmental process of cognitive ability. These studies strongly supported that the development of motor skills was predictive of the development of cognitive and perceptual ability. However, this theory does not account for the influence of individual psychological factors, such as motivation, self-efficacy, and anxiety on motor development.

Ecological theory- dynamic systems theory. Bronfenbrenner (1979; 1992) developed the ecological systems theory that explained how an individuals' development was affected by their

social relationships and the world around them (i.e., surrounding environment). This theory (Bronfenbrenner 1979; 1992) consisted of five different levels on individuals' environment: (1) the microsystem, (2) the mesosystem, (3) the exosystem, (4) the macrosystem, and (5) the chronosystem. Bronfenbrenner proposed the microsystem to be the smallest and closest settings in children's lives. The microsystem is composed of the home, school, peer group, community service and environment of the children. Interactions within the microsystem generally involves personal relationships with parents, siblings, classmates, and teachers. These populations or groups make a direct impact on children's growth. In other words, the fact that children interact with people in the microsystem will impact a child's development. More nurturing and more supportive interactions and relationships will understandably foster the children's improved development (Lerner, Liben, & Mueller, 2015). The second level of the individuals' environment is the mesosystem. This level encompasses the interrelation among the different microsystems of other children such as between home and school, between family, groups, and community. According to Bronfenbrenner (1979; 1992), when a child's parents are engaged in activities to promote friendship and social relationships for their child, they invite the interaction of microsystems of others. The next level is the exosystem, an extension of mesosystem. This level pertains to the linkages between two or more settings at a structured local level which may indirectly influence the development of children. This level is inclusive of the neighborhood, mass media, educational policy and system, or social services. The fourth level of ecological systems theory is the macrosystem that is the largest and most distant collection of people and places to children. This level is the set of overarching beliefs, values, and norms, as reflected in the cultural, religious, and socioeconomic organization of society. The macrosystem influences development within and among all other systems and serves as a filter or lens through which an

individual interprets future experiences (Kail & Cavanaugh, 2012). Research on macrosystems provides insight into what predicts participation, why some individuals in the same activity have different experiences, and issues. The last level is the chronosystem. The chronosystem adds to the useful dimension of time, which demonstrates the influence of both change and constancy in children's environments. The chronosystem may include a change in family structure, address, parents' employment status, as well as social changes such as economic cycles and wars (Santrock, 2009). Bronfenbrenner (1992) asserted that various ecological systems contribute to the diversity of interrelated influences on the development of children.

Based on the ecological systems theory, there were two theories related to the motor development: dynamic systems theory (DST; Newell, 1986; Newell & Jordan, 2007) and behavior setting theory (Barker, 1969; Bronfenbrenner, 1992).

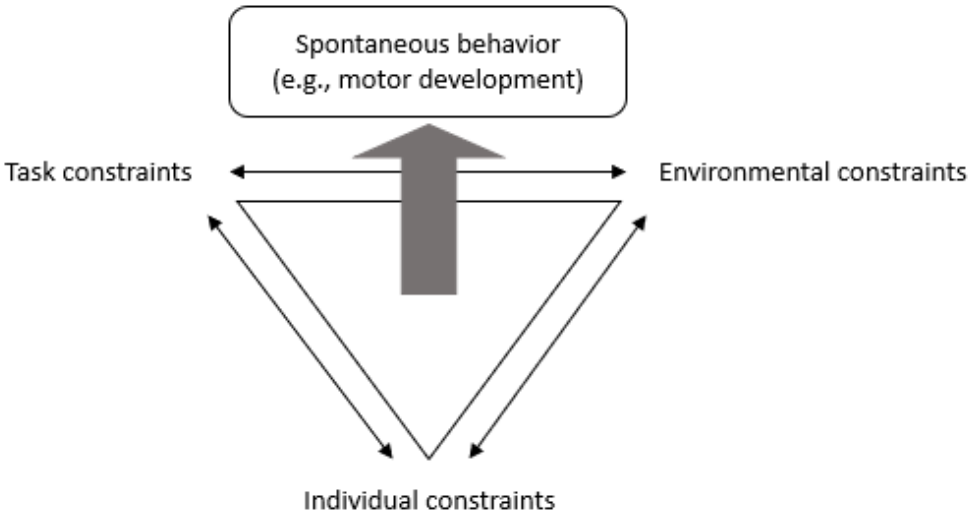

Figure 1. A conceptualization of Newell's dynamic systems theory

Dynamic systems theory (DST; Newell, 1986; Newell & Jordan, 2007) defines a behavior as a coordinated movement of the body coming from interactions between three constraints, which are individual, environment, and task constraints (see Figure 1). According to Gagen and Getchell (2006), constraints under DST refer to neither positive nor negative messages regarding motor development. It represents a neutral condition influencing behavior (i.e., movement skill, motor performance) across the life span (Colombo-Dougovito, 2017). DST primarily has been studied across multiple fields, stemming from physics, chemistry, and mathematics. The essential concepts from the natural sciences have been changed to apply to the fields, including biological, cognitive, neurological, and social sciences to explain the changes of motor development (Thelen & Smith, 1996; Thelen & Ulrich, 1991). According to Smith and Thelen (2003), the behavioral development of individuals is established on a complex dynamic system after birth that has multiple internal and external influences consistently interacting with each other. Individual constraints include the structural (e.g., height, muscle mass) and functional characteristics (e.g., attention, motivation) of individuals that are their own unique processes for behaviors (Haywood & Getchell, 2019). Environmental constraints refer to everything in the individuals' surroundings, such as the season, weather, time, space, or surface of the ground. Additionally, Hutzler (2007) included social and physical factors as environmental constraints affecting a person's behaviors. The last one is task constraints that include everything in the performance outcomes such as the directions of the task, the goals of the movement, or the devices used for the movement. DST enables one to account for the capability of motor performance based on the constraints of individuals as the coordinated movement within the characteristics of individual, environment, and task. The interaction between the three constraints could result in not just motor skills but any form of behavior. DST has been evaluated to help

understand the complicated human movements and new behaviors throughout the individuals' lifespan (Haywood & Getchell, 2019; Langley, 2001).

Researchers have studied motor development using DST to examine the effect of the constraints among various populations. Ohgi et al. (2007) evaluated motor development in infants through investigating their spontaneous movements. The authors concluded infant's instinctive movements were originated in the individual constraint which influence the performance of voluntary movement skills. In terms of the environmental constraint under DST, Renshaw and colleges (2010) applied important implications in teaching physical education to examine how environmental constraints influenced lessons. The author found that the learning environment in physical education was a significant constraint for acquisitions of stability and functional movement patterns. Another study to examine the impact of DST was Vernadakis and colleges (2015). This study aimed to compare the effects of two different interventions using a DST framework which were an exergame-based and a traditional object control skill training. The authors concluded that task modification on a fundamental movement skill (FMS) training program gave a significant effect on improving FMS of elementary school children. However, DST has a limitation that it does not provide a comprehensive understanding about how the nervous system and the body work to help people hypothesize likely control constraints. In addition, DST has difficulty to exhibit itself the exact same way twice (Renshaw et al. (2010).

Another motor developmental theory using the ecological systems theory was behavior setting theory (Barker, 1969; Bronfenbrenner, 1979). Barker (1969) developed the notion of behavioral setting including behavioral, surrounding, and structural settings. Those specific setting conditions accounted of a large portion of the individual variation among children. He stated the concept of standing patterns that was typical patterns in which people respond. This

concept was used to explain how a person could have different behavior patterns depending on the situation. For instance, a teenage girl could show active, energetic, and noisy behaviors with peers during recess time at a school. Whereas, she would have self-regulated attitudes in which she can be calm and serene in the classroom. Barker (1969) concluded that the standing patterns of behavior meant people had specifically predicted or expected behaviors in the given environmental situations.

Bronfenbrenner (1979) developed the ecology of human development and bioecological theory that was an extension of behavior setting theory by Barker (1969). He defined the ecology of human development as the progressive and mutual compromise between actively growing human beings. Human development is affected by interrelations among all embedded components in our lives. Bronfenbrenner (2005) additionally established extensive importance on the perceptions of activities, roles, and interpersonal relations in one's behavior settings. This perceived notion of proper activities, roles, and interpersonal relations were applied to the environmental contexts of the ecological theory of human development.

Information processing theory. The last major theory of motor development is information processing theory (Schmidt & Lee, 2005); Schmidt & Wrisberg, 2008). This theory is affiliated with the sensory process, such as visual, auditory, and kinesthetic. Information processing through the brain was likened to a "computer" (Haywood & Getchell, 2019) or "black box" (Schmidt & Lee, 2005). Primarily, Kephart (1960) presented the term of perceptual-motor process based on the input-output system, which was considered in every perceptual and motor learning activity. The concept of perceptual-motor resulted from perception that meant knowing or interpreting information into the brain. After that, the information contributes to distinguishing movement behavior patterns of individuals (Goodway, Ozmun, & Gallahue, 2019). This

perceptual-motor is a voluntary movement activity from the stimulated sensory information. The complementary connection between input (i.e., sensory stimulation) and output (i.e., motor performance) empowers the cooperative developments of perception and motor skills. Due to the characteristic of information processing through the brain, this perspective has been mainly used in the study of motor learning and motor control (Haywood & Getchell, 2019).

Understanding motor skill development

Theoretical approaches have strived to integrate and present motor skill development. Goodway et al. (2019) presented the phase of motor skill integrated with the ecological perspective to describe the process of motor skill development with four phases: (a) reflexive movement phase, (b) rudimentary movement phase, (c) fundamental movement phase, and (d) specialized movement phase (see Figure 2). A brief description of each phase of motor skill development follows.

Reflexive movement phase. The reflexive movement phase is the first movement phase. This phase constitutes involuntary movements of infants from birth to one year old. Reflexive movements of this phase are the infant's automatic reactions to natural stimuli, which are light, sounds, touch, and changes in pressure. These involuntary movements of infants before one year of age play a significant role in helping to gather information about their body and surroundings, obtain nourishment, and find protective responses (Goodway et al., 2019).

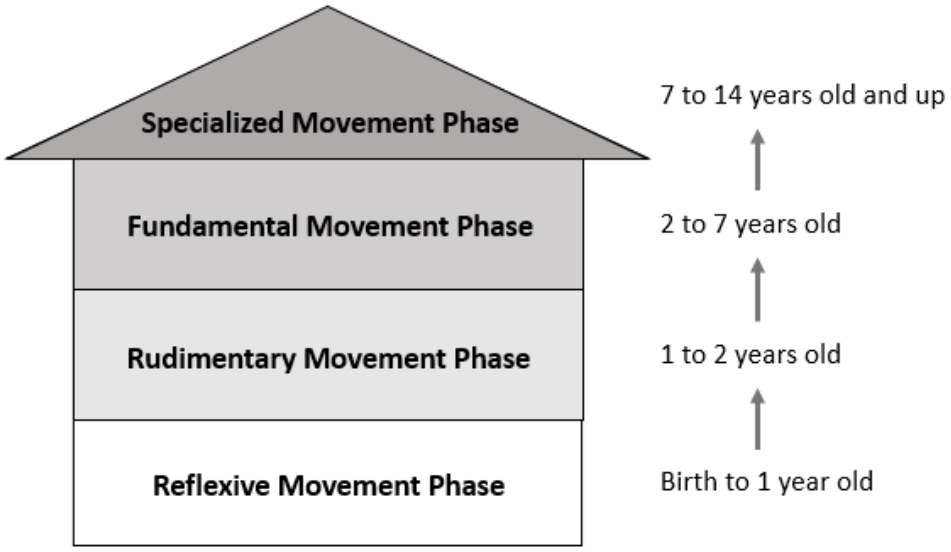

Figure 2. Phase of motor development

According to Goodway, Ozmun, and Gallahue (2019), through developing reflexive movements, infants experience two overlapping stages that are the information encoding stage and information decoding stage. The information encoding stage is identified by noticeable involuntary movements from pregnancy to about four months in infancy. The brain develops to respond involuntarily on a variety of stimuli during this stage. The information decoding stage of the reflexive development phase initiates roughly on the fourth month after birth. During this time, infants begin to develop voluntary control based on stored information as simultaneously reacting to stimuli.

Rudimentary movement phase. The next phase of motor development is rudimentary movements. Rudimentary movements are the basic forms of voluntary movement, such as stability (e.g., balancing), locomotor (e.g., crawling, walking), and manipulative movements (e.g., grasping, releasing). This phase starts at birth and further develops through approximately two years of age. The rudimentary movement phase consists of two stages, the reflex inhibition stage, and the 1stcontrol stage, that describe the highly progressive motor development of

infants. The reflex inhibition stage transpires after birth to one year in infancy. Though reflexive movements prevail in movements of this period, they begin to show various movements combined with voluntary and involuntary movements. It is caused by the development of the brain and adjustments to environmental constraints. Through this process, some reflexive movements of the infant gradually disappear. Voluntary movements are developed on differentiated and integrated forms of movement performance at the same time. However, the movement performance would likely be the low capacity of movement control. The 1stcontrol stage is characterized by toddlers who begin to perform precise and controlled movements between about one to two years old. During this stage, toddlers rapidly discover basic ways to maintain their balance, to conduct locomotor movements, and manipulate objects (Goodway et al., 2019).

Fundamental movement Phase. The third phase of motor development is fundamental movement. Children develop fundamental patterns of movement based on rudimentary movements. During this phase, children actively discover how to do a variety of stability, locomotor, and manipulative movements in their surroundings. They experience a variety of daily tasks that influence the development of fundamental movement skills, such as going up and down stairs, sustaining their balance, or lifting objects. Many factors play important roles in the development of fundamental movement patterns, characterized as task demands (e.g., jumping, kicking, the beam walk) and environmental factors (e.g., experience opportunities, instructions, the number of practices, encouragement). The fundamental movement phase is developed from two to seven years of age. Goodway et al. (2019) present that children show three stages in the fundamental movement phase according to the aging sequence, which are initial, emerging elementary, and proficient stages. The initial stage of a fundamental movement phase transpires

in children from ages two to three years (Goodway et al., 2019). This stage is defined by children attempting to perform their movements often improperly or poorly due to the lack of the spatial and time-related integration of movement. The next stage of fundamental movement phase, which can be observed from 3 to 5 years old, is the emerging elementary stage. Children attain advanced motor control and rhythmical integration of fundamental movement skills in this stage. Although children perform improved coordination of fundamental movement skills in various situations, they still display limited or overstated body movements. The proficient stage of the fundamental movement phase is described by adequate, integrated, and regulated performances. This stage occurs between five to six years old in most fundamental movement skills. Due to advances in visual and motor abilities in accordance with physical maturation in this period, children develop the ability to visually track objects and to block targets using their bodies (e.g., catching, striking; Goodway et al., 2019).

Specialized Movement Phase. The fourth phase of fundamental motor development is specialized movement. Movement skills in this phase consist of a variety of complex movements needed to perform demanding skills in sports, recreations, and daily activities. Fundamental movement skills, which are stability, locomotor, and manipulative movements, are developed to advanced levels through the process of refinement, combination, and elaboration between acquired movement skills. Rope-jump activities, sport skills, and dance activities can serve as an examples. O'keeffe, Harrison, and Smyth (2007) presented the performance level of fundamental movement skills showed significant effects in the specific sport skills. This phase develops after the age of seven and continues throughout the individual's life span. For instance, when a child has the proficiency to perform jumping and hoping, they move to the next stage of practice and instruction to execute jumping over a rope, jumping a long rope activities and single jump-rope

challenges (Hernandez et al., 2009; Roberton et al., 2017). Children learn various fundamental movement skills in this phase, then they develop the patterns of movement within the acquired skills from increasingly challenging situations such as games, sports, and physical activities. This process can be achieved by mastering fundamental movement skills to then contribute to learning complex movement skills throughout life.

According to Goodway, Ozmun, and Gallahue (2019), the specialized movement phase has three stages with children ages seven years old and older: (1) transitional (7 to 10 years), (2) application (11 to 13 years), and (3) lifelong utilization stages (after 14 years). Haubenstricker and Seefeldt (1986) presented that children aged 7 to 8 years old begin the transitional stage in the specialized movement phase. Children in this stage begin to show combinations of fundamental movement skills applied to games and sports activities, such as soccer and jump-rope activities. Also, they discover and make use of the various patterns of movement in a variety of activity settings. Children rapidly develop their capacity of motor learning and control from the accumulation of their experiences (Goodway et al., 2019).

The next stage within the specialized movement phase is the application stage, which comprises children from ages 11 to 13 years. During this stage, individuals not only have more extensive experiences and abilities, but also have enhanced complex and combined movement skills in all types of physical activities while developing their cognitive ability and furthering physical growth. The development of complex movement skills can be achieved to learn various activities by applying diverse movement strategies to tasks. This development process is significantly influenced by numerous tasks, individual, and environmental factors in the application stage. Individuals repeat the process to attain certain goals of movement skills by participating in a variety of games and sport activities (Goodway et al., 2019).

The final stage is lifelong utilization. This stage begins around the age of 14 and continues for the rest of an individual's life. This stage describes the apex of the motor development process and indicates the use of acquired movement skills throughout one's life. Thus, it comprises the longest period of motor development on daily living, sports, and physical activities. All factors, such as an individual's capability, participation opportunities, levels of activity, physical condition, quality of instruction, number of practices, learning motivation, and expenses, effect the level of mastery of movement skills. In other words, the lifelong utilization stage essentially is the pinnacle of all motor development phases and stages. The specialized skill development phase plays an important role in life to develop and make use of individuals' movement performances in various areas (Landers, Carson, & Tjeerdsma-Blankenship, 2010; Gould & Enomoto, 2009; Wiersma, 2000).

Understanding Fundamental Movement Skills

Fundamental movement skills (FMS; i.e., gross motor skills) are considered to be a critical milestone that proceed on the specialized movement phase which is comprised of a variety of complex movements. FMS are a specific set of skills that involve different body parts, such as feet, legs, trunk, head, arms, and hands, developed throughout childhood. FMS are categorized as moving through space (locomotor skill; e.g., running, jumping), controlling our body against gravity (e.g., balance, stability), and manipulating objects in the environment (object control skill; e.g., throwing, catching a ball). These components are needed to actively engage in various games, sports, and recreational activities. Balance and stability skills required for efficient FMS performance have traditionally been categorized as underlying abilities for locomotor skills as opposed to stand-alone FMS (Burton & Rodgerson, 2001; Fleishman et al.,

1984). FMS are very important to the physical development of children. Children can develop sport-specific and complex movement skills when they are confident and proficient in FMS. Clark and Metcalfe (2002) described that competency in FMS was significantly associated with future movement and physical activity. Individuals learn, use, refine, and apply various movement skill performances from childhood. According to Newell (1986; 2007), constraints affect the acquisition of various movement patterns, which are individual, task, and environmental factors. These constraints directly and indirectly influence the development of movement skills and its outcomes of children.

Importance of fundamental movement skills. The development of FMS is a critical process to have proficient movement performance in assorted sport, game, recreational and dance activities. Children have opportunities to explore their surroundings, and easily learn through the cultivation of FMS and effective and efficient combinations of FMS. For example, this notion can be illustrated through a child who has had many opportunities to practice catching, throwing, and dribbling balls, which were different sizes, weights, and shapes in their learning environment; or a child who has experienced various locomotor skills, such as running, jumping, hopping, and sliding in a variety of directions. Consequently, such a child may have an advantage given the opportunity to display their acquired skills in basketball or games that require different kinds of passing, catching, and dribbling skills with changeable movement patterns.

Stodden and colleges (2008) proposed the importance of FMS to explain possessing skillful FMS in the motor development of children and to commit to physical activities in individual's lifetime. Especially, children in the adolescent period may have enough opportunities to develop their FMS to enhance the relationship between the proficiency of their

motor skills and physical activity such as various games, sports, and physical activities (Goodway et al, 2019; Haywood & Getchell, 2019). There exists many studies that have not only examined motor skill proficiency and development, and its relationship with practice, encouragement, feedback and instruction (Spodek & Saracho, 2014), but also significant importance for healthy physical and social development and performance in activities of daily living (Deflandre et al., 2001; Holfelder & Schott, 2014; Lopes et al., 2011; Robinson et al., 2015).

Sequential analysis of fundamental motor skills. Researchers in the field of motor development presented two developmental sequences of FMS, which are the total body sequences and the component sequences. The total body sequences were described by the movement of the entire body. Meanwhile, the component approach identified and examined the movements of parts of the body, such as arms and legs. Both sequential approaches illustrate the general patterns of FMS shown by children. According to this sequential approach, children shift their FMS consecutively to competent patterns with no skipping of stages. However, some studies about motor development of children (Garcia, 1994; Clark & Humphrey, 2002) found that the FMS developmental steps of the participants varied from the suggestions of the theory of developmental sequences. Goodway and colleagues (2019) identified three weaknesses of the developmental sequence theory: (1) the linear fashion did not account for developmental regressions in performance; (2) the developmental sequence theory did not account for children who skip the common developmental sequences; and (3) this theory did not explain the means in which children move from one motor performance pattern to another. Because of these weakness of the developmental sequence theory, motor development researchers reconceptualized stages of FMS using the dynamic systems theory in describing the sequence of movement patterns.

Assessment of fundamental movement skills

Purpose of motor skill assessment. There are five specific purposes for assessment of movement skill: (1) categorization or identification, (2) planning, treatment or instruction, (3) evaluating change over time, (4) providing feedback, and (5) prediction (Burton & Miller, 1998). The first purpose of the assessment of movement skill is to categorize or identify the levels of motor skill of a person (Burton & Rodgerson, 2001). The services related to clinic or educational benefits require testing the level of movement skill ability for the eligibility of a student at the various service placements. The service placements are general or adapted physical education (GPE/APE), and group or individual programs for students with special needs (Herr & Bateman, 2013; IDEA, 2004). In addition, a test of movement skills can be used to guide the level of service for students with special needs based on movement performance outcomes.

Second, the purpose of the motor skill assessment is to plan the program with instructional strategies. Program design consists of making skill progressions for students using the assessment outcomes (Kelly, 2011; Kelly & Melograno, 2014). Especially, annual goals and short-term objectives in APE must be constructed with appropriate instructional methods based on the movement skill foundation tests. The third purpose of movement assessment is to evaluate students' performance over time (Block, 2016). Movement assessments could be utilized to monitor the effect of an instructional program for individuals through testing on separate occasions. Using the change of movement outcomes, educators can reflect on specific information to plan the next steps of the program (Kelly, 2011).

Another purpose of movement assessment is to give feedback to individuals who are students, parents, or service providers (Jette, 1995). Feedback based on the movement performance outcomes can enhance the performers' level of movement competence or motivate

them to have positive attitude towards engaging in the program. According to Campbell (1993) and Piper (1994), the final purpose of movement skill assessment is to predict a student's accomplishments at the end of the instructional period, or to predict instructionally applicable resources for a student. Movement skill assessment could be used to predict students' capabilities or practical instructional methods for students. Movement skill assessments support students' various development areas in educational settings (Burton & Miller, 1998).

Key assessment instruments of movement skills. Burton and Miller (1998), and Jirovec, Musalek, and Mess (2019), identified various motor ability assessment instruments commonly used in the U.S.A. Those instruments clearly produce composite or summary scores to presents the levels of movement ability of children, such as the Bruininks-Oseretsky Test of Motor Proficiency (Bruininks, 1978; Bruininks & Bruininks, 2005), the Movement Assessment Battery for Children Test (Henderson & Sugden, 1992; 2007), and the Test of Gross Motor Development (Ulrich, 1985; 2000; 2019). A brief description of each assessment instrument follows.

Bruininks-Oseretsky Test of Motor Proficiency. The Bruininks-Oseretsky Test of Motor Proficiency (Bruininks, 1978; BOT) and its second edition (Bruininks & Bruininks, 2005; BOT-2) were developed to assess various aspects of fine and gross motor development in children and youth 4 to 21 years of age. The BOT-2 is different from its original version including the elimination of a speed subtest, division of visual-motor control subtest into two subtests, change of subtest title, and revision of items in the overall test assessment. The new version has improvements in motor impairment diagnosis, screening, assisting research objectives, determining placement, and creating and appraising motor training for individuals in specialized education services (Bruininks & Bruininks, 2005; Cools, De Martelaer, Samaey, & Andries, 2009). The BOT-2 consists of eight subtests: (a) fine motor precision, (b) fine motor integration,

(c) manual dexterity, (d) upper-limb coordination, (e) bilateral coordination, (f) balance, (g) running speed and (h) agility, and strength. The four areas of motor functions are fine manual control, manual coordination, body coordination, and strength and agility. This assessment has two different forms that depend on the number of items: the complete form (CF) and the short form (SF). The CF uses a total of 53 items to test the motor ability of a participant that takes about 45 to 60 minutes per person. The SF of the BOT-2 has a total of 14 items that takes about 15 to 20 minutes per person. Since the SF takes significantly less time to complete the test, many studies prefer the use of the short from of the BOT-2 (Cairney et al., 2008; Carmosino, Grzeszczak, & McMurray , 2014; Fransen et al., 2014; Lucas et al., 2013; Spironello et al., 2010; Wuang & Su, 2009).

Several studies have been conducted to examine the validity (Deitz, Kartin, & Kopp, 2007; Fransen et al., 2014; Wuang & Su, 2009) and reliability (Carmosino et al., 2014; Lucas et al., 2013; Wuang & Su, 2009) of the BOT-2. Deitz, et al. (2007) investigated the applicability and the suitability of the BOT-2 for both diagnostic and evaluative purposes to offer appropriate educational programs to students in school settings. The researchers presented that the confirmatory factor analysis of the BOT-2 showed a strong support for the four subtests to evaluate the level of motor performance of individuals. According to Fransen et al. (2014), the BOT-2 showed reasonable evidence of convergent and discriminant validity to assess motor competence of children. Additionally, they suggested to use at least one additional motor competence assessment instrument, such as KörperkoordinationsTest für Kinder (KTK; Kiphard & Schilling, 2007), with the BOT-2 to have more accurate results of motor competence measurement in children. Wuang and Su (2009) analyzed the measurement properties of the BOT-2 using Rasch analysis to evaluate its validity among individuals with intellectual

disabilities (ID). In this study, the authors stated that the revised BOT-2 showed overall good fit to the validation analysis.

Regarding the investigation of reliability of the BOT-2, Brahler et al. (2012) studied the associations among the four subtests, which were fine motor integration, fine motor precision, balance, and strength, of the total of eight subtests. The authors found a wide range of correlations in individual subtest items in the complete form with the relevant overall subtest scores ranging from $r = 0.07$ to 0.86. The authors concluded that the investigated items in the four subtests obtained low correlations with the relevant overall subtest scores. According to Carmosino, Grzeszczak, and McMurray (2014), the relations between the overall subtest score in the CF and the four subtests were investigated. They found that the weakest correlation was generally identified between the bilateral coordination overall subtest scores and the items. Then, Carmosino et al. (2014) concluded that with the exception of bilateral coordination, items from the other three subtests, manual dexterity, running speed and agility, and upper limb coordination, were significantly correlated with the overall subtest scores.

Wuang and Su (2009) investigated the internal consistency, test-retest reliability, and the responsiveness of the BOT-2 among children with ID aged 4 to 12 years. The researchers presented that the BOT-2 had reliability of internal consistency ($a = 0.92$), test-retest reliability (*ICC* = 0.99), and responsiveness (*effect size* = 0.67). These results inferred that the BOT-2 had excellent test-retest reliability and good internal consistency, as well as its responsiveness was acceptable to assess motor skills of children with ID. However, although several studies regarding the validity and reliability of the BOT-2 presented substantial evidence to support the use of evaluation instruments of the level of motor skill among children, there were some problems to apply it to screening and placement decisions (Burton & Miller, 1998). This was

because not only did the test items of the BOT-2 focus on motor abilities rather than movement skills, but also some subtests of the test yielded low reliabilities with large confidence intervals. Such limitations would implicate difficulties in accurately evaluating individual motor performance levels compared to the standard scores of the BOT-2, and in effectively planning intervention programs, especially for children with ID (Burton & Miller, 1998).

Movement Assessment Battery for Children. Movement Assessment Battery for Children (MABC-1; Henderson & Sugden, 1992) and the second edition (MABC-2; Henderson, Sugden, & Barnett, 2007) was developed to identify and describe impairments in motor performance among children aged 3 to 17 years old. The MABC has been commonly used in occupational therapy, physiotherapy, and educational services (Barnett & Henderson, 1998; Ellinoudis et al., 2011; Engel-Yeger et al., 2010). The second edition (MABC-2) has numerous different aspects from the first version (MABC-1): the age range extension from 4 – 12 years old to 3 – 16 years old, reduction of age bands from four to three, revision of items, and addition of new items. Moreover, the MABC-2 has provided the score interpretation method, representative standardization sample, and rearrangement of subtests (Henderson, Sugden, & Barnett, 2007).

The MABC-2 consist of three major performance components in each of the three age bands (3–6 years; 7–10 years; and 11–16 years); manual dexterity (three items), aiming and catching (two items) and static and dynamic balance (three items). Each age band is composed of eight different items, respectively. Those items measure different aspects of motor ability accordingly to each of the age ranges.

For confirming the psychometric properties of the MABC-2, many studies have been conducted to examine the validity and reliability of testing motor competence for different age groups of this assessment instrument (Ellinoudis et al., 2011; Schoemaker et al., 2012; Smits-

Engelsman et al. 2008; Smits-Engelsman, Niemeijer, & van Waelvelde, 2011; Valentini, Ramalho, & Oliveria, 2014; Wuang, Su, & Su, 2012). Ellinoudis et al. (2011) and Smits-Engelsman et al. (2011) evaluated the reliability of age band 1 (3–6 years). According to Ellinoudis and colleges (2011), the MABC-2 could be deemed a reliable and valid assessment instrument for the evaluation of motor performances in children aged from 3 to 5 years. Smits-Engelsman et al. (2011) presented that the revised test of the MABC-2 could be utilized to assess motor performance in 3 years old children without disabilities. In the study of test-retest reliability investigation, Smits-Engelsman et al. (2008) and Smits-Engelsman et al. (2011) depicted good (0.80) and excellent intraclass correlation coefficient values (0.94), respectively. These results demonstrated that the MABC-2 was a reliable instrument to measure motor performance in typically developing children aged 3 years. Another reliability study of the MABC-2 (Wuang, Su, & Su, 2012) was to analyze its internal consistency and test-retest reliability for children with developmental coordination disorder (DCD). The researchers recruited 144 children with DCD in Taiwan to evaluate the reliability of the MABC-2. The results of this study generated excellent outcomes of internal consistency ($\alpha = 0.90$) and test-retest reliability (ICC = 0.94). They concluded the MABC-2 was a reliable instrument to measure motor competence among children with DCD.

For testing the validity of the MABC-2, Schoemaker and colleges (2012) investigated construct validity of the assessment tool for children with and without disabilities across age and gender. The author described a significant Cronbach's alpha value ($\alpha = 0.94$), which implies that all items of the MABC-2 measure the same construct. Valentini, Ramalho, and Oliveria (2014) examined the validity of the MABC-2 for Brazilian children with DCD between 3 and 13 years of age. This research found statistically significant construct validity ($\alpha = 0.78$), discriminate

validity (F (2,814) = 722.07, $p < .001$, $\eta^2 = 0.63$), predictive validity ($ICC = 0.88$, $p < .007$), and concurrent validity ($r = 0.30$, $p < .02$) of the MABC-2. Based on these results, the authors concluded that the MABC-2 was a valid instrument to measure motor impairments for designing proper programs for children with DCD. However, several researchers identified problems in utilizing the MABC-2 (Burton & Miller, 1998; Veldhuizen, Rivard, & Cairney, 2017).

According to Veldhuizen and colleges (2017), when all children within an age band are assessed using the same norms, younger children within the group can be evaluated lower levels of motor performance ability than older ones even if they have competent motor abilities. For instance, when children aged 7 years old are compared with a single set of norms, they may be evaluated as possessing poorer performance levels than 8-year-old peers. This problem was coined the 'relative age effect' in sports and athletic settings (Musch & grondin, 2001). The effects have been shown to be a disadvantage to relatively younger children compared with their older peers. This phenomenon can also be supported by the claim that relatively younger children are less likely to reach elite levels in sport and athletics (Musch & grondin, 2001). Even this problem may influence children that are more likely to be diagnosed with motor impairments or attention deficit hyperactivity disorder (ADHD; Morrow et al., 2012). Burton and Miller (1998) argued that using the MABC-2 should be considered when evaluating the effects of intervention on motor abilities or investigating the change of motor abilities for long periods. This is because a child may proceed into the next age band using a different set of performance tasks.

Test of Gross Motor Development. The Test of Gross Motor Development (TGMD; Ulrich, 1985; 2000; 2019) was developed to evaluate the levels of FMS in elementary school students. The TGMD has been commonly used to assess the level of FMS in children with

(Breslin & Rudisill, 2013; Brian et al., 2018; Burns, Brusseau, & Hannon, 2017; Klavina et al., 2017; Nonis & Tan, 2014; Westendorp et al., 2011) and without disabilities (Bastik et al., 2012; Bisi et al., 2017; Eather et al., 2018; Hastie et al., 2018). This assessment can assist educators to design individualized development programs of FMS due to the ease of administration and scoring, norm-referenced scores for diagnostic evaluation, and the criterion-referenced and process-oriented skills (Maeng et al., 2017). Common uses for the TGMD include the measurement of FMS performance in experimental research (Johnstone, Hughes, Janssen, & Reilly, 2017; Logan, Robinson, Wilson, & Lucas, 2012), and in educational settings to assess the current level of motor skill development for children with and without disabilities. This is an indicator to monitor the levels and progress of FMS in children with developmental disabilities, who have physical, cognitive, and social delays, compared to typically developing children (Kim, Park, & Kang, 2012; Kirk & Rhodes, 2011; MacDonald et al., 2017)

The TGMD-3 has been developed with current values and changed items within both the locomotor and object control subtests (Webster & Ulrich, 2017). There was a subtest name change where object control skills were transformed into ball skills. The leap skill in the TGMD-2 was substituted with skip in the locomotor subtest, and the underhand roll in the TGMD-2 was changed to underhand throw in the ball skill subtest. Additionally, the one-hand forehand strike was newly produced to assess one skill in the ball skill subtest in the TGMD-3. Consequently, the TGMD-3 has a total of thirteen FMS skills with six locomotor skills and seven ball skills. The calculated score by summing the total locomotor subtest score and the total ball skills subtest score is the total gross motor test score.

Reliability of the TGMD-3. Estevan et al. (2017), Kim et al. (2012), and Maeng et al. (2017) presented reliability of the instrument for evaluation is an important and requisite

psychometric property of any research instrument. Reliability is associated with an estimate of potential error in a score. Instruments with lower reliability are assumed to have more measurement error in a score. Thus, a high-quality measure should have acceptable reliability across time (i.e., test-retest), across individuals conducting the measurement (i.e., inter-rater), and across repeated scoring attempts (i.e., intra-rater). It must demonstrate an acceptable level of reliability for an assessment to be valid (Burton & Miller, 1998). Once an instrument is developed, there are two types of reliability that are commonly examined: intra-rater and inter-rater reliability. Intra-rater reliability is the degree of consistency of scores on an assessment across at least two occasions by a single rater which reflects the stability of the test score by a rater on different occasions (Shrout, 1998). According to Rousson and colleges (2002), intra-rater reliability relies on the abilities of trained raters and on good calibration of the task or item being assessed. Inter-rater reliability is the consistency of scores obtained from two or more raters independently scoring the same subjects. Barnett and colleges (2014) presented this was a significant aspect of evaluating FMS competence in children.

In terms of the reliability of the TGMD-2, multiple studies have examined not only typically developing children (Barnett et al., 2014; Farrokhi et al., 2014; Kim, Kim, Valentini, & Clark, 2015; Simons et al., 2008; Valentini, 2012) but also children with disabilities (Harvey et al., 2007; Houwen et al., 2010; Kim et al., 2012). Recently, several studies using the TGMD-3 have been conducted to examine the instrument reliability among children with and without disabilities (Allen et al., 2017; Brian et al., 2018; Maeng et al., 2017; Rintala, Sääkslahti, & Iivonen, 2017; Simons & Eyitayo, 2016; Valentini, Zanella, & Webster, 2017; Webster & Ulrich, 2017). Webster and Ulrich (2017) and Allen et al. (2017) assessed test-retest reliability on a subsample of children based on intraclass correlation coefficients (ICCs) have found strong

levels of agreement for subscale scores (*ICC*: 0.81 – 0.97). Studies examining inter-rater reliability of the TGMD-3 presented varied levels of ICCs between poor and excellent (0.51–0.99). Related to poor or fair agreement for specific skills, which measured ICCs less than 0.60, it can be accepted to adequate inter-rater reliability (Maeng et al., 2017; Rintala et al., 2017; Valentini et al., 2017). Intra-rater reliability has also been shown to be good to excellent. According to Maeng and colleges (2017), evaluating certain groups of children who may be more or less consistent in their performance such as children with disabilities, is also important to evaluate rater reliability separately for appropriate sub-groups. Three studies examining the reliability of the TGMD-3 in children with disabilities such as visual impairments (Brian et al., 2018), intellectual disabilities (Simons & Eyitayo, 2016), and autism spectrum disorders (Allen et al., 2017) suggested the reliability properties of this assessment must be re-examined for children with disabilities.

Validity of the TGMD-3. In an educational setting, validation of measurement instrument must be examined to assure the property of the test before collecting data extensively for confirming the standardized data set because validity is a crucial element in the developmental process of a measurement instrument (AERA, APA, & NCME, 2014; Urbina, 2014)). The validity of the TGMD was investigated to examine its psychometric properties in every edition (TGMD-1; 1985, TGMD-2; 2000, TGMD-3; 2019). Though the TGMD-3 has been officially released with the normative data set in 2019, the developer furnished the record form of the TGMD-3 in 2014. Therefore, many researchers interested in studying the latest version of the TGMD conducted validation studies and composed publications related to the TGMD-3 (Brian et al., 2018; Estevan et al., 2017; Magistro et al., 2018; Mohammadi et al., 2019; Valentini et al., 2017; Wagner et al., 2017; Webster & Ulrich 2017). These studies analyzed the construct

validity of the TGMD-3 using exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). Estevan and colleges (2017) studied the psychometric properties by examining the construct validity of the TGMD-3 for Spanish children. The result of the study showed acceptable fit indexes in both subtests, which were locomotor and ball skills. According to Webster and Ulrich (2017), the TGMD-3 has fair construct validity, which instilled confidence to use the measurement instrument for children.

Understanding rater training

The evaluation of human performance in settings, such as work, school, or any environment, has long been an interest of psychological researchers (Arvey & Murphy, 1998). In general, human performance in organizations is evaluated using subjective performance ratings provided by educators or specialists such as psychiatrist and psychologist. The rating accuracy of human performance is important to the success of a performance rating system, and some researchers have suggested that rating accuracy is the primary goal of performance evaluation (Werner & Bolino, 1997). Rating accuracy is typically examined by comparing rating values across dimensions and items scored by expert raters. An evaluation method of rating accuracy was reviewed to compare rating scores of novice raters by expert raters across task elements on an assessment of performance (Sulsky & Balzer, 1988). Two general strategies have been advanced as ways of improving rating accuracy: rating scale development and rater training (Woehr & Huffcutt, 1994). Regarding rating scale development, the general finding from this literature was that the type of rating scale used made little difference in terms of improving ratings (Gomez-Mejia, 1988; Landy & Farr, 1980). According to Woehr and Huffcutt (1994),

several studies compared the results of modified rating scales and found no significant improvements in the accuracy of performance ratings when using such rating scales.

Another strategy to have accurate rating from raters is rater training (Bittner, 1948). There are four types of rater training: (a) rater error training, (b) performance dimension training, (c) behavioral observation training and (d) frame-of-reference training. Each of these strategies have been considered to potentially enhance the accuracy of raters' scoring performance (Smith, 1986; Spool, 1978; Woehr & Huffcutt, 1994). According to Smith (1986) and Spool (1978), these strategies of rater training provide not only a comprehensive framework of assessment but also content of training for raters to improve rating accuracy. A brief discussion of each approach follows.

Rater error training. According to Smith (1986) and Spool (1978), rater training provides not only an elaborate framework of materials of assessment but also content of training for raters. Furthermore, McIntyre and colleagues (1984) asserted that rater training has shown two major benefits:(a) to enhance raters' knowledge and skills for carrying out evaluations, and (b) to motivate raters to use the knowledge and skills learned in the training program. Some studies have found that raters' perceptions of fairness, accuracy, and credibility on rating process of performance were improved by rater training (Bannister, 1986: Fulk, Brief, & Barr, 1985).

The main objective of rater error training is to increase rating accuracy by becoming acquainted with common classification errors and biases such as similarity, contrast, primacy, first impression, leniency, and halo effect (Aguinis, 2017; Bernardin & Buckley, 1981; Woehr & Huffcutt, 1994). Traditionally, raters are trained in the definitions of the involuntary biases which might affect the rating accuracy. The rater error training consists of graphic illustrations with numerical examples how biases may interfere with rating performance. Moreover, this training

includes how to avoid bias in rating performance (Latham et al., 1975). Several studies presented that rater error training helped to reduce the influence of potential biases on rating (Bernardin, 1978; Bernardin & Walter, 1977; Latham et al., 1975). However, researchers have pointed out rater error training may present a narrow focus of rater bias variables affecting rating accuracy and not consider alternate measure that may affect a rater's ratings (Bernardin & Pence, 1980; Arvey & Murphy, 1998; Hedge & Kavanagh, 1988). Rater error training should be established with appropriate and prevalent causes of errors to raise effectiveness of the training.

Performance dimension training. The performance dimension training approach was suggested to resolve accuracy arguments of rater error training (Smith, 1986). The purpose of performance dimension training was to improve rating accuracy on the meaning of performance according to components and dimensions of rating assessment or scale (Smith, 1986; Woehr & Huffcutt, 1994). Sulsky and Balzer (1988) reported that performance dimension training increased degree of rating agreement among raters. The researchers concluded this rater training was a useful procedure to score assessments more precisely and accurately. In addition, Pulakos (1984) investigated the rater error training and performance dimension training that showed the trained group made more precise scores than the untrained group.

Behavioral observation training. According to Noonan and Sulsky (2001), behavior observation training strategy provided close observation practice for raters to enhance their own observation process of behaviors. This strategy was to improve a rater's memory and recognition of specific behavioral events using multiple observation methods, such as notes and diaries (Sulsky & Day, 1992; Thornton & Zorich, 1980). Especially, Sulsky and Day (1992) examined measures of behavioral recognition, in which raters were asked to indicate the list of behaviors that came up in a given situation. In a different way for improving a memory of specific behavior

in observation, Thornton and Zorich (1980) used a questionnaire method to raters, which were true or false responses, multiple choice questions, or hypothetical case studies to examine a sample of behaviors. Although Woehr and Huffcutt (1994) presented the effectiveness of behavioral observation training (d = .77), Hedge and Kavanagh (1988) were uncertain about the effectiveness of this training.

Frame-of-reference training. This last approach was proposed as a method to train raters' about common conceptualization and multidimensional aspects of performance in a given measurement (Bernardin & Buckley, 1981). This rater training strategy was to help familiarize raters with identifying performance with the correct performance dimensions. Furthermore, it was for diverse raters to evaluate ratees' different performance under natural circumstances through sharing a common framework and conceptualization of the performance (McIntyre et al., 1984; Roch et al., 2012). Thus, this strategy is used as the training method most effective in systematically collecting information to evaluate performance (Roch & O'Sullivan, 2003). Frame-of-reference training has been used to meet more accurate rating using training content with the presentation of sample performance based on the correct performance dimension (D. E. Smith, 1986). Primarily, frame-of-reference training was developed to supplement the inaccurate results of rater error training (Rosales Sánchez et al., 2019).

Since the frame-of-reference strategy was proposed, it has been widely used in various studies (Aguinis et al., 2009; Bernardin & Pence, 1980; Cardy & Keefe, 1994; Chirico et al., 2004; Gorman & Rentsch, 2009; Keown-Gerrard & Sulsky, 2001; Lievens, 2001; Loignon et al., 2017). These studies showed the positive effect of the frame-of-reference training on rating accuracy when scoring individual performance. Especially, Lievens and Sanchez (2007) found that trained raters with the frame-of-reference training could have significantly higher values in

validity, interrater reliability, and rating accuracy compared with the untrained raters. Woehr and Huffcutt (1994) conducted a meta-analysis of frame-of-reference training. The researchers reported a large average effect size (d = .83) of frame-of-reference training studies that compared trained and untrained groups. Those studies support that frame-of-reference training positively impacts accurate rating of performance by influencing raters' memories and perceptions to score ratee's performance. The frame of reference training will be the strategy used in this study to develop a rater training program to score the TGMD-3.

Online Training

Online training is a form of instruction that is completed on the internet. It involves a variety of multimedia components including video, audio, graphics, and web-links as educational technology (Mangal & Mangal, 2009). Online training technology has been introduced to be an effective and efficient method to distribute knowledge and enhance individuals' intellectual context (Chafouleas et al., 2015). In the context of development for instruction, studies examining outcomes associated with face-to-face instruction and online instruction found that online instruction showed comparably positive results (Moore et al., 2011; Russell et al., 2009). Especially, online training for raters can support improvement of rating accuracy through mitigating the impact of personal characteristic and bias (Aguinis et al., 2009). According to Chafouleas and colleges (2015), a study in rating individuals' behavior using online training module showed positive impact in improving rating accuracy. This training method may help enhancing raters' confidence in scoring performance.

Online training opportunities have been presented as a practical and efficient strategy in educational environment for both educators and learners (Brown & Green, 2003; Olsen,

Donaldson, & Hudson, 2010). Online training has various advantages which are the convenience with training accessibility, the ability to serve more learners, relatively low cost, and less environmental barriers compared to face-to-face training. Brown and Green (2003) stated that online training programs may be a feasible and effective method to enhance rating accuracy of behavioral performance even though this method had a disadvantage in which it might offer limited opportunities to interact with educators.

Recently, online programs for education as well as training have been developed to provide effective learning and teaching methods using a variety of resources to various types of learners such as students, practitioners, and educators. Bartolotta and colleges (2017) stated that usability of an online training program should consider how easily learners can find and access program materials, such as assessment for testing, readings, modules, and other resources through videos, sounds, or other online media. In addition, an online program must be readable, have technical usability, and access to provide training methods effectively (Warner & Hewett, 2017).

Researchers conducting studies examining the development of online programs and its effectiveness have performed the analysis of online program validation without any standardized evaluation scale or systematic procedures based on online teaching and educational learning theories. The lack of information provided by researchers regarding the process of developing their online training program and its validation is a limitation when trying to understand and identify what comprises a good online video training program. Under these circumstances, using what was identified above regarding the ease of access to online educational content, this study will clearly articulate operational descriptors of an effective online rater training program and will attempt to validate these criteria using a content validity evaluation survey.

Delphi Method

The Delphi method is described as a structured communication process for collecting knowledge and producing consensus from a group of experts through multiple rounds of questionnaires with controlled opinion feedback (Ziglio, 1996). This experimental method has become a widely used and recognized technique to help predictions and decision-making in social science such as business, education, medicine, public health, and public policy (Clayton, 1997; Landeta, 2006). The Delphi method has been the best-known predicting technique to present fields of development and application based on the opinion of experts through objective criticism and numerous evaluations. It provides valuable solutions with statistical results, flexible methodology, and simple execution to inherent problems in the traditional group opinion, the influence of undesirable psychological effects among participants, selective feedback of the relevant information, and more extensive consideration of ideas (Landeta, 2006).

In kinesiology, researchers have applied the Delphi method as a quantitative approach of a structured communication process to make agreements in the evaluation of questionnaire content among informed individuals on topics related to curriculum development, educational effectiveness, and its features (A. S. Brian et al., 2016; Bulger & Housner, 2007; Columna et al., 2014; Dyer et al., 2011; Ross et al., 2014; Taliaferro & Bulger, 2020). Brian and colleges (2016) studied the content and face validity for the modified Test of Perceived Motor Competence of children with visual impairments (TPMC-VI) using a Delphi technique. A panel of experts was involved to evaluate the IPMC-VI and to investigate findings on participant interviews for content and face validity. This study used a Delphi investigation across four phases: (Phase 1) establishing a preliminary understanding of PMC including experts in VI (n = 8); (Phase 2) evaluating modified TPMC-VI with feedback with those experts (n = 8); (Phase 3) completing

the suggested changes and developing the TPMC-VI with comments including the same experts (n = 8); and (Phase 4) completing the TPMC-VI along with the additional inquiry from the child and teacher. The authors presented several strengths when using the Delphi method including development of corroboration with experts, in-depth participant interview and instrument testing. Limitations also were identified specifically that not all experts were equally as dedicated to the study process.

Modified Delphi investigation has also been applied in studies examining essential information in exercise science and motor development related to physical education teacher education curriculum (Bulger & Housner, 2007; Ross et al., 2014). Both studies utilized the use of a two-round modified Delphi protocol involving a group of 20 experts, respectively (10 exercise science specialists, 7 physical education teacher educators, and 3 physical education teachers; Bulger & Housner, 2007; 5 physical education teacher educators, 5 motor development specialists, 5 motor learning specialists, and 5 physical education teachers; Ross et al., 2014). Those Delphi panel members were asked to rate the questionnaire items using a 5-point Likert scale to evaluate the theoretical importance and pedagogical relevance of subjects as well as the effectiveness of instructional method for delivering the contents to preservice teachers in PETE program. The first round of the Delphi investigation contacted panel members through mailing, telephone calls, and a follow-up email about an overview of the study process to get their approval of involvement in the study. During the second round of the Delphi investigation, all the panel members shared responses from the first round then reevaluated the questionnaire regarding theoretical importance and pedagogical relevance. In addition, they were asked to rate the survey of effective instructional method for the delivery of knowledge to learners in PETE program. These studies provided a conceptual framework of content (i.e., exercise science, motor

development) in accordance with discussion among experts on the modified Delphi method that can promote skillful teaching of exercise science and motor development curriculum in PETE programs.

Taliaferro and Bulger (2020) performed a Delphi study of effective practicum experiences in adapted physical education programs. This study was to determine a consensus among experts about adapted physical education practicum experiences for preservice physical education teachers. A 3-round Delphi procedure involving a group of 24 content experts was used to establish an online questionnaire. Throughout a 3-round of the Delphi procedure, 47 items out of initial 70 items in Round 1 were retained on four major themes: program context, teaching and learning activities, outcomes/soft skills, and evaluation of instructor performance. The authors concluded that the finalized 47 questionnaire items for APE practicum experiences will help guide the generation of a checklist or instrument for evaluating APE practicum experiences.

Columna and colleges (2014) developed an instrument to assess parental perceptions toward adapted physical education teachers who work with students with autism. This study implemented the Delphi method in the second step among four phases of development to verify the revised Parent Perceptions toward Adapted Physical Education Teachers (PPTAPET) survey (Columna et al., 2008; Glazer, 2009). In total, 8 experts comprised a panel to analyze the revised survey items using content validity coefficient (V; Aiken, 1985) in the Delphi method. The authors presented that the revised survey had high validity to assess parental perceptions of their child's APE teacher.

Dyer et al. (2011) conducted investigation study that was to define the role of lower-limb running prostheses and stakeholders' perceptions of fairness in relation to their use in

competitive disability sport. A three-round of Delphi method was used to ask opinions from experts over three rounds to reach high levels of consensus in the development of sporting equipment technology and its fairness for athletes in disability sport. A total 22 experts were involved as the panel in the research project. During the first round of the Delphi technique, the experts were asked three open-ended questions focusing on the purpose of this research: (a) the role of lower-limb prostheses in competitive sports, (b) the fairness or unfairness of using a sport protheses, and (c) experts' opinions toward the limitations on the use or none-use of this technology. The author established closed-ended questions using a 4-point Likert scale to assess the respondent's attitudes in the second round. Lastly, analyzed and refined questions were generated in the second round. Then, those questions were integrated into the third round of the Delphi method to calculate the level of consensus. The researchers presented that the Delphi technique was a useful tool not only to help refinement of the questions regarding the prostheses technology but also to review the rules and regulations in competitive disability sport.

Collectively, these studies provide information and support for the overall procedures for using a modified Delphi for the present study including how many rounds of review may be needed, what procedure is conducted, how many experts are involved with the panels, and how to finalize the validation of the program with the results obtained.

Understanding developmental disability

Developmental disability. Developmental disability (DD) presents at birth and affects the trajectory of an individual's physical, intellectual, or emotional development. This characteristic of DD affects multiple body parts or systems, especially in "language, mobility, learning, and independent living". DD affects all areas of a child's development (Center for Disease Control

and Prevention [CDC], 2013). Intellectual disability begins any time before a child turns 18 years old and is characterized by problems with intellectual functioning or intelligence, which include the ability to learn, reason, problem solve, and other skills, and adaptive behavior, which includes daily social and life skills (America Psychiatric Association, 2016). ASD consists of a range of conditions classified as neurodevelopmental disorders. Individuals diagnosed with ASD present two types of symptoms: problems in social communication and social interaction, and restricted, repetitive patterns of behavior, interests, or activities. According to the American Psychiatric Association's Diagnostic and Statistical Manual fifth edition (DSM-5; Association, 2013), symptoms are typically recognized between one and two years of age. Approximately 1 in 68 American children are diagnosed to have ASD (Centers for Disease Control and Prevention [CDC], 2014). Long term issues may include difficulties in creating and maintaining relationships, retaining a job, and performing daily tasks (Comer, 2016).

Fundamental movement skills in children with developmental disabilities. Several studies have shown that children with DD represent one such group who have FMS delay (Simons, Daly, Theodorou, Caron, & Andoniadou, 2008; Westendorp, Houwen, Hartman, & Visscher, 2011). Simons and colleges (2008) presented the correlation between age and the development of movement skills that FMS in Flemish children with DD were significantly lower than typically developing children. According to Westendorp et al. (2011), children with DD displayed significantly lower scores in all locomotor and object control skill items on the TGMD-2 compared to children without DD.

To examine the levels of FMS and its changes in children with ASD, diverse studies have been conducted (Edwards et al., 2017; Liu, Breslin, & ElGarhy, 2017; Lloyd, MacDonald, & Lord, 2013; Pan, Tsai, & Chu, 2009). Pan et al. (2009) investigated the difference of FMS level

between children with ASD and typically developing children age ranged from 6 to 10 years. This study presented that children with ASD exhibited significantly lower scores in total, locomotor, and object control subtests on the TGMD-2 than the comparison group. In Edwards and colleges (2017) study, the researchers examined the FMS of children with ASD, especially their object control skills. This study found the overall levels of FMS on the TGMD-3 in children with ASD was generally lower compared to those without ASD. According to Liu et al. (2017), the study examining FMS of children with ASD reported their performance levels of FMS were significantly low or below average on assessments (e.g., BOT-2, MABC-2, TGMD-2). Consistent with other claims, many children with ASD have delays or impairments in the development and performance of gross motor skills (Downey & Rapport, 2012; Lloyd et al., 2013; Vernazza-Martin et al., 2005). According to Liu et al. (2014) and Berkeley et al. (2001), above 70% of children with ASD scored in the poor or very poor range of the gross motor skills on the TGMD-2. This research noted that the developmental trajectory of the FMS in children with ASD was dramatically decreased compared to the other groups in their study (i.e., children with language delays and typically developing children).

There have been numerous studies to examine the development of FMS in individuals with ID (Capio et al., 2013; Rintala & Loovis, 2013; Simons et al., 2008; Westendorp et al., 2011). According to Westendorp et al. (2011), children with ID have been found to perform FMS inferiorly than children without disabilities because of their low cognitive proficiencies that make it harder to conduct pertinent FMS performances. Simons and colleges (2008) compared the standard scores of the TGMD-2 between children with and without ID. The results of this study showed children with ID achieved significantly lower scores of locomotor, object control, and total than those without ID.

Another study to investigate the level of FMS in children with ID was conducted by Rintala and Loovis (2013). This study examined differences in the development of FMS between children with ID and typically developing children. The researchers claimed the group with ID performed significantly lower level of FMS on the TGMD-2 compared to the typically developing group. Capio et al. (2013) presented that not only did children with ID have inferior FMS proficiencies caused by their limited cognitive processing abilities, but also diminished errors in the educational settings facilitate to develop their FMS than the settings which allow errors. Therefore, the development of FMS of children with DD bares deficits and delays in appropriate FMS progression based on previous studies.

3 METHODOLOGY

Instruments

Test of Gross Motor Development-Third Edition. The TGMD-3 (Ulrich, 2019) measures gross FMS performance of children with and without disabilities aged three to ten years. The TGMD-3 is divided into two subtests: locomotor and ball skills. The locomotor subtest is comprised of six skills: run, gallop, hop, skip, horizontal jump, and slide. The ball skills subtest consists of seven skills: two-hand strike of a stationary ball, one-hand forehand strike of self-bounced ball, one-hand stationary dribble, two-hand catch, kick a stationary ball, overhand throw, and underhand throw.

Each skill has 3 to 5 performance criteria used in scoring a child's performance. Criteria are largely process-oriented, with a few product-oriented criteria that specify the number of repetitions (e.g., slide) or certain distance (e.g., underhand throw) needed to complete the skill. If the child performs the criterion correctly, the rater scores a '1' for that trial. If a child performs the criterion incorrectly, the rater scores a '0' on the performance criterion. All performance criteria are scored for each skill over two consecutive trials. The six locomotor skill scores and seven ball skill scores are summed to get the locomotor and ball skill subtest scores, respectively. The total gross motor score is calculated by summing the locomotor subtest and ball skills subtest scores.

Study 1: Methodology for Developing the Online Frame-of-Reference Training Program

Target Audience. The target population of this online training module was novice raters who are in kinesiology or teacher education programs preparing to teach physical education to children with and without disabilities such as intellectual or developmental disabilities (i.e.,

autism spectrum disorders). The TGMD, developed to assess the level of fundamental motor skills (FMS) among children with disabilities, is widely used to create the Individual Education Plan (IEP) for students with disabilities in school settings. Correctly scoring FMS on the TGMD-3 among students with disabilities is necessary to design program content and instructional methods appropriately based on the student's current performance level and needs.

Identifying Frame of Reference Training Criteria. The present study identified frame-of-reference training criteria for the development of the online modules. The criteria on the training program applying frame-of-reference strategies were as follow: (a) identifies correct performance dimension under natural circumstances, (b) provides information to score the variables in question, (c) describes correct performance on variables in question, (d) systematically compares information regarding actual versus desired performance, (e) provides information according to characteristics of performer, (f) provides correct scoring feedback on actual performance, and (g) offers practical questions with answers. These criteria were used to establish effective training strategies for novice raters to score FMS of children with DD considering their behavioral and psychomotor characteristics. The rater training program derived from these frame-of-reference training strategies were anticipated to reduce errors in scoring FMS on the TGMD-3 among children with DD.

Identifying Online Training Criteria. Online frame-of-reference training modules should be developed to confer the benefits of a training program which can maximize the impact of learning on the accurate scoring of the instrument. This study included the following components to guide the development of the online training modules: (a) training program was easy to access, (b) utilized visual resources (angle of camera, angle of performance front and side

view), (c) voice narration is clear and understandable, (d) voice narration is at an appropriate pace, (e) tests are available online, and (f) written material in video is easy to read.

Outlining Learning Outcomes for Training Modules. This study created learning outcomes for training module 1 and 2, respectively. For module 1 in which the general information of the TGMD-3 was introduced, four learning outcomes were selected: (a) the rater understands the TGMD-3 and its components, (b) the raters can identify the skills in the locomotor subtest, (c) the raters can identify the skills in the ball skills subtest, and (d) the raters understand how to score each criterion within a skill.

The second module included information about developmental disability (DD) and behavioral and psychomotor characteristics of children with DD as well as the correct performance cues for 2 skills on the TGMD-3 one locomotor skill (i.e., run) and one ball skill (i.e., two-hand strike). The learning outcomes for module 2 were the following: (a) the rater can explain DD, (b) the rater can list behavior and movement characteristics of children with DD, (c) the rater can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3, (d) the rater can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3, and (e) the rater can score FMS among children with DD according to each criterion on the TGMD-3.

Developing Tests to Check for Novice Rater Understanding. The tests for checking understanding of each module among novice raters were required to encourage their engagement in the training as well as maximize effectiveness of the rater training. Thus, the tests for modules 1 and 2 were made up of the most important concepts from each module aligned with the learning outcomes of the respective model (see Appendix F & H). All tests consisted of

questions to evaluate learners' understanding of the training module of the TGMD-3. Each test had 10 questions consisting of true or false and multiple-choice.

Developing Evaluation Forms for Expert Validation. An evaluation survey was developed for expert raters to evaluate the effectiveness of the online frame-of-reference training modules. The evaluation examined the degree of alignment between the components of a frame of reference training and online training and the content in each of the modules. Second experts evaluated the effectiveness of training content and the tests. All ratings used a 5-point Likert type scale (see Appendix C).

Skill Selection on the TGMD-3 for the Training Module. Based on previous studies reporting the low reliability coefficients of skills on the TGMD-2 and 3 (Barnett et al., 2014; Kim et al., 2014; Kim et al., 2012; Maeng et al., 2017), the present study included one locomotor skill (i.e., run) and one ball skill (i.e., two-hand strike) of the TGMD-3 for the rater training program. A rationale for the selection of these four skills was provided in a previous section of this paper.

Content Development

Development of Module #1 Content. Using a multimedia online platform, a brief description of the TGMD-3 using presentation format slides and narration was provided to help raters understand the general information of this movement instrument (approx. 3 mins). The learning outcomes for module #1 were (a) understand the TGMD-3 and its components, (b) identify the skills in the locomotor subtest, (c) identify the skills in the ball skill subtest, and (d) understand how to score the TGMD-3. Using the administration of the TGMD-3 on YouTube (Webster, 2014), this module demonstrated thirteen fundamental movement skills on the TGMD-

3 separated into locomotor and ball skills to help identifying the characteristics of each skill one by one (approx. 4 mins). The developer of the TGMD-3 demonstrated 6 locomotor and 7 ball skills. Lastly, this section consisted of a presentation of the TGMD-3 record form and scoring methods (approx. 8 mins). Presentation slides explained how to score FMS on the TGMD-3 in module #1 with an illustrative example using the two-hand catch skill. It took approximately 15 minutes to complete Module #1. An evaluation test with 10 questions was required to ensure novice rater's understanding of content in module #1 (see Appendix F).

Development of Module #2 Content. Module #2 had a similar structure as Module #1. The learning outcomes for module 2 were as follows: (a) explain about developmental disability (DD), (b) understand behavior and movement characteristics of children with DD, (c) recognize the performance criteria of the run skill in the locomotor subtest, (d) recognize the performance criteria of the two-hand strike in the ball skill subtest, and (e) get information to correctly score FMS according to each criterion on the TGMD-3 among children with DD. The second module included information about developmental disability (DD; 4 mins) based on the American Psychiatric Association (2016) and behavioral and psychomotor characteristics of children with DD (approx. 5 mins; see Appendix B). Presentation slides with example videos were used to introduce the definition of DD and its general information. Also, this module uses videos of typical behavioral and psychomotor characteristics (e.g., challenging behavior, low motor function) among children with DD to provide not only an understanding of DD but also their unexpected motor skill behaviors when scoring FMS on the TGMD-3.

The correct performance and cues for 2 skills on the TGMD-3 one locomotor skill (i.e., run; approx. 4 mins) and one ball skill (i.e., two-hand strike; approx. 6 mins) were provided to score actual performance among children with DD. Based on the frame-of-reference training

approach for novice raters, this module explained how to correctly score each performance criterion of the run and two-hand strike skills on the TGMD-3 among children with DD. For instance, a child with DD may come out of the path when s/he is required to perform the run skill. Another example could be a child with DD who might push a ball using a bat instead of striking the ball with two-hands when performing the two-hand strike. In addition, training content in this module described how to score specific components of each performance criterion either 1 or 0 on a trial. Moreover, this module demonstrated comments to address complicated or controversial issues according to the behavioral characteristics of children with DD. Also, potential questions with directions in terms of scoring dimension were provided on the rater training program to help correctly understand the specific components of each performance criterion on the TGMD-3. It took about 19 minutes to complete Module #2. An evaluation test with 10 questions was required to ensure novice rater's understanding of content in module 2 after watching the module (see Appendix G).

Study 2: Methodology of validation for an online frame-of-reference training program

modified Delphi method procedure

The present study examined content validity of the rater training program from the expert panel members using a modified Delphi method. The modified Delphi method was primarily developed by Ziglio (1996) to effectively generalize research protocols. Two rounds of administrative procedure were applied for the modified Delphi method.

Selection of the Expert Panel. Prospective expert panel members were contacted to ask for their involvement in the study through e-mail or telephone and provided with a description of the study, its procedures, and their contribution to the study. Experts were given an honorarium for their participation in the validation phase of this study. A total of 12 experts were recruited

from across the USA and Europe. A total of eight professionals all from the USA agreed to participate in the validation of the rater training modules and tests. The demographic information of the expert panel members in this study is presented in Table 1. A total of six experts (75%) were university professors, three with expertise in motor development and three specialized in teaching adapted physical education. An additional two experts were K-12 APE teachers with at least 5-year experience teaching FMS to children with DD. Four of the experts were men.

Table 1. Demographic Information of Expert Panels in the Modified Delphi Method

Expert No.	Area of Expertized	Gender	Affiliation
1	Motor Development	M	Professor*
2		M	Assistant Professor*
3		F	Assistant Professor
4	Adapted Physical Education	F	Associate Professor*
5		M	Associate Professor
6		F	Research Fellow*
7	Adapted Physical Education Teacher	M	APE teacher*
8		F	APE teacher

*: completed both rounds of evaluation

Round 1. An email was sent to all consenting expert panel members containing a description of the study, the modified Delphi method process, the module evaluation questionnaire, and a timeline for completion. Delphi panel members were asked to rate the modules using a 5-point Likert type scales (see Appendix C). Each expert was sent reminder emails a total of 3 to 5 times to encourage completion of the evaluation forms. The degree between 1 and 5 on the module validation form indicated very poor to very strong alignment or content representation, respectively. The expert panel provided not only independent ratings of subject-matter content of the TGMD-3 but also how well the training protocol provided information to score individuals' performance on the TGMD-3 among children with DD with respective feedback to enhance the rater training program (Lawshe, 1975; see Appendix C). The

intraclass correlation coefficient (Koch, 1982; ICC) statistic was used to analyze the agreement between expert panel members on each module and test. All ICC analyses utilized two-way random effects analysis of variance models (ICC 2,1), and coefficients were calculated for single evaluations of consistency with absolute agreement among panel experts (Eliasziw et al., 1994; Maeng et al., 2017; McGraw & Wong, 1996; Shrout & Fleiss, 1979). This type of ICC was selected to account for systematic and random variance between and within the experts. The above procedures were repeated until ICC coefficient reached at least 80% or higher on each module. The ICC analysis was calculated using SPSS version 28.0 for Windows (IBM Corp., Armonk, NY, USA).

Round 2. A response list of revised comments and revisions to both the content in the modules and tests from round 1 was provided to the expert panel members to clearly identify what parts the author modified and why, and how it was adapted. The alignment with learning outcomes of Modules 1 and 2 were evaluated a second time for content validation (see Appendix F and H). Moreover, every module and question required feedback if the rating scored by experts was 3 or below reflecting poor or moderate alignment with learning outcomes. Five out of 8 expert panel members who participated in round 1 evaluated the modules and tests in round 2. These 5 experts included two in motor development, two in adapted physical education, and 1 adapted physical education teacher (see Table 1). The remaining three experts did not respond to repeated reminders to complete round 2 evaluations. The author revised the modules and test questions to address the feedback provided after round 2. The intraclass coefficient statistic was used to analyze the agreement between expert panel members for calculating their agreement on each module and test. The above procedures were repeated until ICC coefficient reached at least 80% or higher on each module.

Study 3: Methodology of evaluating rating accuracy

The videos of one male child with DD performing the run and two hand strike skills was shared with three expert raters for scoring. Eligibility criteria for these expert raters were: (a) a graduate degree in motor development or adapted physical education/activity, (b) experience administering and scoring the TGMD-3, and (c) a minimum of 5-years of experience teaching FMS curricular content to children with DD in physical education, adapted physical education, or physical activity programs. Expert raters independently scored the video.

Table 2. Foundation Scores of Two Skill Performance on the TGMD-3 by Experts

Skill	Performance Criteria	Trial 1	Trial 2	Score
1. Run	1. Arms move in opposition to legs with elbows bent	1	1	2
	2. Brief period where both feet are off the surface	1	1	2
	3. Narrow foot placement landing on heel or toes (not flat-footed)	1	1	2
	4. Non-support leg bent about 90 degree so foot is close to buttocks	1	1	2
			Skill Score	8
2. Two-hand strike of a stationary ball	1. Child's preferred hand grips bat above non-preferred hand	1	1	2
	2. Child's non-preferred hip/shoulder faces straight ahead	1	1	2
	3. Hip and shoulder rotate and derotate during swing	1	1	2
	4. Steps with non-preferred foot	1	1	2
	5. Hits ball sending it straight ahead	1	0	1
			Skill Score	9
			Total score	17

A total of 4 individual or group meetings were held with experts to establish 100% agreement on the correct scoring of the run and two-hand strike skills on the TGMD-3 for a child

with DD. The scores by the expert raters on the TGMD-3 served as the foundation on which to compare the accuracy of scoring between novice versus expert raters (see Table 2).

Participants

Novice raters. The novice raters were undergraduate students from universities in the U.S. and recruited according to the following inclusion criteria: (a) majoring in health and physical education or kinesiology, (b) a maximum of one course in motor development, motor behavior, motor control or movement assessment, and (c) no experience using the TGMD to score fundamental movement skills (FMS) among children with and without disabilities. A priori sample size was calculated using G*Power (version 3.1.9.7; Franz Faul, Kiel, Germany) with a medium effect size (0.5) to support detected significance in the results. The sample size of this study required a minimum sample size of 35 participants with a power of 95% and an alpha of 0.05. The primary researcher contacted instructors and professors who teach college students in kinesiology programs (i.e., exercise science, general or adapted physical education, sports coaching, pre occupational/physical therapy) in the U.S. for participant recruitment. Instructors were given a recruitment script to share with students in their classes. Interested students completed the online training. Instructors were sent reminder emails to forward to their students to encourage participation in this research project. A total of three email reminders were sent after which point students were deemed to be not interested and were no longer recruited.

A total of 84 novice raters agreed to participate in this study. However, forty-three participants did not complete all three rounds of training and scoring. The number of completed participants in each of the three rounds were as follows: a) participation consent: 84; b) 1st round: 58; c) 2nd round: 50; and d) 3rd round: 41. Table 3 presents the descriptive result of the

participants who completed the novice rater training program after each round. A total 58 participants completed the first round. After round 2, eight participants did not complete the scoring of the two skills. The data presented in the current study reflected those 41 participants who completed all three rounds. There was no remarkable difference in participant demographics throughout each round. Among the participants, the majority were male, between the ages of 22-25 years, black or African American, and Juniors majoring in health and physical education. A total of 14 participants completed a motor development module related to fundamental motor skills.

Table 3. Demographic Information of Participants

Category	Component	Number of Participants (#)		
		Round 1	Round 2	Round 3
Gender	Male	35	29	24
	Female	23	21	17
Age	18-21	17	13	8
	22-25	30	27	24
	26-29	5	4	2
	30-33	3	3	2
	34 or above	3	3	3
Ethnicity	White Caucasian	18	14	8
4	Black or African American	29	27	25
	Asian	6	5	5
	American Indian or Alaska Native	1	1	1
	Other	4	3	2
Major	Health and Physical Education	22	20	20
	Exercise science	23	18	16
	Sports coaching	6	5	2
	Pre OT/PT	5	5	2
	Other	2	2	1
Year	Freshman	10	7	5
	Sophomore	9	8	7
	Junior	21	19	14
	Senior	12	11	10
	Graduate student	6	5	5
Total		58	50	41

Data Collection Procedures

The video of one 9-year-old male subject with DD was used to assess the accuracy of scoring between novice and expert raters. The male subject with DD had above average performance on the run and average performance for the two-hand strike skills. According to Ulrich (2019), the '*Average*' performance level was defined as the descriptive term for a performance between the 21st to 73rd percentile ranks. A total of 7-points out of 8 points for the run skill and 9-points out of 10 points on the two-hand strike skill equated to (see Appendix G) an '*Average*' performance level for the data subject. Content and videos for the modules were housed on the University Qualtrics system. Novice raters were given an individual access code to the Qualtrics content. Novice raters completed a total of three rounds of scoring (see Figure 3). Each training is described below.

Round 1 – Module 1 (intro TGMD-3) with demographic questionnaires and 1st scoring.

Novice raters completed module 1 (see Appendix A) and demographic questionnaires (see Appendix E). Module 1 introduced the TGMD-3 and described the performance components of each skill item and how to complete the TGMD-3 examiner record form using the YouTube resource of the TGMD-3 (Webster, 2014). This module included the standardized performance of 13 skills based on the performance criteria of the TGMD-3. Those 13 skills consist of 6 locomotor skill and 7 ball skills. It took about 10 minutes to watch the module. For checking understanding of module 1, raters completed a 10-question test (see Appendix F). Raters passed module 1 when they scored above 80 percent on the test. Novice raters had unlimited number of trials to pass the test for module 1. A total of fifty-eight participants took the test of module 1 a total 72 for an average of 1.24 times per person. Immediately following completion of module 1, the researcher provided raters with the video of the data subject and ask

raters to independently score the performance videos within five days (i.e., 1st scoring; see Appendix G).

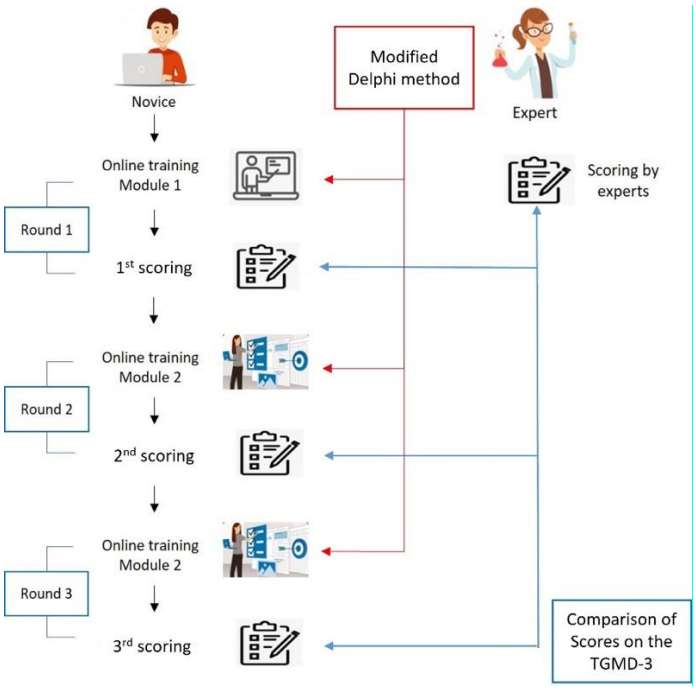

Figure 3. Study procedures

Round 2 – Module 2 (rater training) and 2nd scoring (intervention).

Within 5 days of finishing round 1, the novice raters completed module 2 (see Appendix B). This module described information on the behavioral and FMS characteristics of children with DD (e.g., challenging behavior, low motor function; see Appendix B), as well as sample performance videos of different children with DD and how to score their performance of the 2 skills on the TGMD-3 (i.e., run, two-hand strike). It took about 20 minutes to complete module 2. For checking understanding of module 2, raters completed a 10-question test (see Appendix I). Raters passed module 2 when they scored above 80 percent on the test. Fifty participants took the test of module 2 a total 63 times to achieve a passing score for an average of 1.26 times per

person. Upon successful completion of module 2, the researcher resent a video link of the data subjects and ask novice raters to complete scoring the same performance videos within five days (i.e., 2nd scoring; see Appendix I).

Round 3 – Module 2 (same rater training) and 3rd scoring (intervention).

Within 5 days of finishing round 2, novice raters completed the same process as round 2 then conducted scoring again within five days of completing module 2 a second time (i.e., 3rd scoring; see Appendix I). In the result data of round 3, forty-one participants took the test of module 2 a total 46 times for an average of 1.12 times/person. Table 4 presented how many days of the interval between rounds. Also, this table showed the length of time participants took to complete the three rounds of training and scoring.

Table 4. Participants' Completion Period of Three Rounds

Period (days) & Number of Participants (N)					
Round 1 to 2	N	Round 2 to 3	N	All three rounds	N
1 – 2	8	1 – 2	6	3 – 5	4
3 – 4	10	3 – 4	9	6 – 10	11
5 – 6	9	5 – 6	12	11 – 15	17
7 – 8	5	7 – 8	4	16 – 20	5
9 – 10	4	9 – 10	3	21 – 25	2
Above 10	5	Above 10	7	26 – 30	2
Total	41	Total	41	Total	41

Data Analysis

The descriptive statistics of the TGMD-3 scores was calculated for total gross motor scores (i.e., selected two skills), one locomotor (i.e., run) and one ball skills (i.e., two-hand strike) for each of the one data set at 1st, 2nd, and 3rd scoring, respectively. A repeated measure analysis of variance (RM-ANOVA) was used to analyze the changes in scoring of the novice raters among three different occasions (i.e., 1st, 2nd, and 3rd scoring) compared to expert raters.

The RM-ANOVA was used to compare means across different occasion variables that were based on repeated scoring by the raters. Additionally, *t*-test was conducted to investigate the score difference between expert and novices after each round. Finally, effect sizes were calculated for each of the run, two-hand strike, and total score. Partial eta squared (η_p^2) was interpreted using the following recommendations: .01 = small, .25 = medium, and .4 = large effect (Cohen's f ; Cohen, 1992). All analyses were conducted using SPSS version 28.0 for Windows (IBM Corp., Armonk, NY, USA).

4 RESULTS

Delphi Study Validation of the TGMD-3 Online Training Modules

Round 1

Module 1 Evaluation of Frame of Reference Training Criteria.

Descriptive statistics of experts' mean evaluation score was 4.50 and a standard deviation (SD) of 0.55 with individual scores on the 7 frame-of-reference training criteria ranging from 4.17 to 4.83 (see Table 3). Experts' evaluation agreement was analyzed using the intraclass correlation coefficient (ICC; Koch, 1982). The ICC value of Module 1 was excellent (ICC = 0.94, 95% CI: 0.83 - 0.99). A detailed description of the feedback from module 1 with the researcher's response to the reviewers' comments was provided in Appendix K. The following is a general overview of these recommendations and a summary of how they were addressed in the revision of module 1.

(a) The length of the module, alignment and redundancy of content (reviewer #2, #3 and #6; see Appendix K #1_4 and #1_5). Module 1 was revised to increase fluidity and alignment of content and help novice raters stay focused. Also, there was some redundant content related to explaining the TGMD-3 and scoring methods. Content was combined to reduce playing time.

(b) Clarify and use terminology consistently (reviewer #3, and #6; see Appendix K # 1_5). Experts commented on the use of terminology such as motor and movement, process and product, and defining locomotor. The primary researcher investigated the use of those terminologies in motor development textbooks and research articles to support the terminology selected for use in both modules 1 and 2. According to the literature (Newell, 2020), the term 'motor' was selected for use throughout the modules when used to reference skill performance,

especially “fundamental motor skill”. Other terminologies related to process-oriented, product-oriented, locomotor, and object manipulation skills (i.e., ball skill) were defined in the video.

Module 1 Evaluation of Learning Outcomes.

Table 5 presents evaluation mean and standard deviation scores of each learning outcome and ICC value among the expert panel. All evaluation scores of the learning outcome of Module 1 were above 4-points on a 5-point scale which means strong alignment between Module 1 and the learning outcomes. The agreement value of the experts in this evaluation category was excellent (ICC = .97, 95% CI: .93 - .99). The learning outcome #1 “The rater understands the TGMD-3 and its components” had the highest evaluation score (Mean ± SD = 4.67 ± 0.52). Learning outcome #4 “The raters understand how to score the TGMD-3” received the lowest evaluation score (Mean ± SD = 4.00 ± 0.63) among the learning outcomes. Reviewer #2 and #6 suggested providing the actual criteria of each skill on the TGMD-3 (see Appendix K #2_2, and #2_5). All reviewers had not yet seen module 2 and therefore misunderstood the purpose of module 1. The purpose of Module 1 was to introduce the TGMD-3 and what skills are on it rather than explain each performance criterion for the novice raters. As such, no changes were made to module 1 about how to score the FMS on the TGMD-3.

Module 1 Evaluation of the Test for Understanding.

The results of the test for understanding for module 1 is reported in Table 6. The ICC value of the test of Module 1 was excellent (ICC = 0.97, 95% CI: 0.93 - 0.99). Only question #7 showed relatively lower mean scores on meaningfulness in content knowledge (Mean ± SD = 3.63 ± 1.30) and alignment with the learning outcome (Mean ± SD = 3.75 ± 1.49) compared to other test questions. All other questions were scored above 4 points on a 5-point scale indicating

clarity and appropriateness of the question, meaningfulness in content knowledge, and alignment with the learning outcome.

The test of Module 1 was revised according to comments from the experts in Appendix N. Experts #3 and #7 pointed out that terminology regarding skill performance should align with the module and test questions (see Appendix N #1_1 and #1_5). When the terminology 'motor' was substituted to 'movement' in the module to explain fundamental skill performance, the question-and-answer choices on the test had to match the same terminology. In addition, the types of questions were modified to ensure clear wording and meaning (reviewer #2, #3, #4, and #6; Appendix N #1_2, #1_5, #4_1, #6_5, #7_5, #8_5, and #10_5). For instance, some questions that included written scenarios to describe a child's fundamental motor skill performance were replaced with performance videos.

Module 2 Evaluation of Frame-of Reference Training Criteria.

Descriptive statistics of experts' evaluation scores resulted in a mean score of 4.47 with a standard deviation of 0.74 (see Table 3). The ICC value of Module 2 was excellent (ICC = 0.92, 95% CI: 0.84 - 0.98). A critical comment provided by reviewers #6 and #8 pertained to the perceived controversial nature of terms used to describe characteristics of developmental disability (DD) that could be perceived to exaggerate or cause misunderstanding about DD (See Appendix L). For example, the initial video portrayed the self-injurious behaviors of a child with DD with sample behavior management strategies used by a teacher. Experts #6 and #8 were concerned that the scene may mislead the novice rater about behavioral characteristics of children with DD and behavior management strategies. These video clips were removed from this module. Regarding the two fundamental motor skills (FMS), experts #2, #3, #6, and #8 suggested showing examples of each skill to provide a visual aid to correctly score the respective

performance criterion (see Appendix L #4_1, #4_4, #4_5, #5_2, and #5_4). The incorrect or unexpected skill performance on the TGMD-3 among children with DD is hard to correctly score. Providing performance examples to score '1' or '0', respectively, could help the learners understand performance criteria and characteristics of performance among children with DD. Performance videos of children with DD and sample scoring strategies were added to facilitate understanding of correct scoring.

Module 2 Evaluation of Learning Outcomes

Table 5 presents the evaluation mean and standard deviation scores of each learning outcome, and the agreement value (ICC) among the experts. All evaluation scores of the learning outcome of Module 2 were above 4 points which means strong alignment between Module 2 and the learning outcomes. The agreement value of the experts in this evaluation category was excellent (ICC = .92, 95% CI: .82 - .98). Learning outcome #3 was "The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3" and #4 "The raters can recognize the performance criteria of the two-hand strike skill in the ball skill subtest on the TGMD-3" had the highest evaluation score (Mean ± SD = 4.50 ± 0.55) among the learning outcomes. Learning outcome #1 "The raters can explain developmental disability" and #5 "The raters get information to score fundamental motor skills (FMS) among children with DD according to each criterion on the TGMD-3" had the lowest evaluation score (Mean ± SD = 4.17 ± 0.75). Regarding learning outcome #2 "The rater can list behavior and movement characteristics of children with DD", expert #2 pointed out there was much more information on autism spectrum disorders (ASD) than intellectual disability (ID) (see Appendix L #6_2). The primary researcher responded to the experts' comments explaining why Module 2 had more content about ASD than ID. This was because children with ASD may show low performance

levels and challenging behaviors on the TGMD assessment. Also, they may have multiple characteristics associated with both ASD and cognitive limitations of ID.

Module 2 Evaluation of Test for Understanding.

Table 7 shows the evaluation result including the mean and standard deviation of each question from the expert panel in the study as well as the overall ICC of the test of Module 2. Question #1 was rated low for clarity, appropriateness, and meaningfulness. Question #3 was rated lower for clarity, meaningfulness, and alignment with learning outcomes. All other questions were rated 4.0 or above out of 5-point. Thus, the author revised several questions in the test of Module 2 according to the comments of the expert panels. For example, expert reviewers #3, #4, #6, and #7 suggested replacing written scenarios with video examples of each skill component demonstrated by a child with DD (see Appendix O #7_5 and #9_5). The written scenarios of a child with DD's performance were changed to performance videos on the test. There was a misunderstanding of expert reviewer #3 about the concepts of Module 2. Expert reviewer #3 suggested that a separate module be developed to address not just examples of motor delays in children with DD but how to address these performance behaviors and testing issues. The primary researcher responded to this expert that the purpose of Module 2 was not to train novice raters how to correct behaviors that affect FMS performance but rather to focus on correctly scoring these differences in performance of children with DD.

Evaluation of Online Training Criteria.

The training modules in this study applied online instruction methods including a variety of multimedia components using video, audio, graphics, and web links as educational technology (Mangal & Mangal, 2009). Those components were evaluated to investigate their effectiveness to deliver information according to the five evaluation questions (see Table 8). All five

evaluation scores from the experts were above 4 points supporting the alignment between the instructional methods used in the modules with recommended online training criteria. The highest evaluation score among the five evaluation questions was #1 “Training program is easy to access” (Mean ± SD = 4.63 ± 0.52; see Table 7). Question #3 “Voice narration is clear and understandable” had the lowest evaluation score (Mean ± SD = 4.25 ± 0.89). The agreement value of the experts in the online training evaluation category was excellent (ICC = .92, 95% CI: .82 - .98).

The experts evaluated the appropriateness (i.e., correct or incorrect performance) of the video examples (see Appendix N #7_2). However, some videos in the module were hard to recognize the desired example of skill performance due to the quality or size of the video. Module revisions focused on providing better quality and larger size video scenes to better see the motor performance of the child with DD. According to the comment from the experts #3, #4, and #8 (see Tables 6 and 8) the lowest evaluation score related to the quality of the voice throughout Module 1 and 2. Some of the audio qualities were poor or inconsistent (see Appendix #7_3 and #7_4). For example, the audio sound of the word “derotate” was not clear to provide precise information of the performance criterion of the two-hand strike skill. The revised module adapted a better pronunciation of this word using the video editing program. Another point in the evaluation of online training was the written materials in the module. The experts commented that some slides were difficult to focus on the subject due to a large amount of text on the slide. Both Modules 1 and 2 were revised reducing the amount of text while ensuring understanding and attention to the training video.

Table 5. Statistic Results from Round 1 of Frame-of-Reference Training Module Content

Module	Category	Question	Mean ± SD	ICCs (95% CI)
Module 1	Identify correct performance dimension under natural circumstances	Introduce to the TGMD-3 and the scope	4.50 ± 0.55	.94 (.83 - .99)
		Introduce 6 locomotor skills	4.50 ± 0.55	
		Introduce 7 ball skills	4.50 ± 0.55	
	Provides information to score the variables in question	Introduce scoring FMS on the TGMD-3	4.50 ± 0.55	
Module 2	Describes correct performance on variables in question	Discuss scoring of the run skill	4.33 ± 0.82	.92 (.84 - .98)
		Show good and bad performance of the run skill	4.17 ± 0.98	
	Systematically compare information regarding actual versus desired performance	Provide scoring feedback on the run skill	4.50 ± 0.55	
	Provides information according to characteristics of performer	Discuss scoring of the two-hand strike skill	4.83 ± 0.41	
		Shows good and bad performance of the two-hand strike skill	4.50 ± 0.84	
	Provides correct scoring feedback on actual performance	Provided scoring feedback on the two-hand strike skill	4.50 ± 0.84	

Table 6. Result of Alignment of Content with Learning Outcome from Round 1

Category	Question	Mean ± SD	ICCs (95% CI)
Learning Outcome of Module 1	#1: The rater understands the TGMD-3 and its components	4.67 ± 0.52	.97 (.93 - .99)
	#2: The raters can identify the skills in the locomotor subtest	4.50 ± 0.55	
	#3: The raters can identify the skills in the ball skills subtest	4.50 ± 0.55	
	#4: The raters understand how to score on the TGMD-3	4.00 ± 0.63	
Learning Outcome of Module 2	#1: The rater can explain about developmental disability (DD)	4.17 ± 0.75	.92 (.82 - .98)
	#2: The rater can list behavior and movement characteristics of children with DD	4.33 ± 0.52	
	#3: The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3	4.50 ± 0.55	
	#4: The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3	4.50 ± 0.55	
	#5: The raters get information to score FMS among children with DD according to each criterion on the TGMD-3	4.17 ± 0.75	

Table 7. Round 1 Question Conformance & Alignment with Learning outcome of the Test of Module 1

Question #	Category				ICCs (95% CI)
	Clarity and understanding (Mean ± SD)	Appropriateness of Answer response options (Mean ± SD)	Meaningfulness of question to assess content knowledge (Mean ± SD)	Degree to which question aligns with learning outcome (Mean ± SD)	
1	4.38 ± 1.06	4.00 ± 1.07	4.38 ± 0.52	4.63 ± 0.52	.97 (.93 - .99)
2	4.75 ± 0.46	4.50 ± 1.07	4.63 ± 0.52	4.75 ± 0.46	
3	4.50 ± 1.07	4.75 ± 0.71	4.25 ± 1.04	4.63 ± 0.52	
4	4.50 ± 1.07	4.38 ± 1.06	4.13 ± 0.99	4.13 ± 1.13	
5	4.75 ± 0.46	4.75 ± 0.46	4.38 ± 0.74	4.63 ± 0.52	
6	4.50 ± 1.07	4.50 ± 1.07	4.25 ±1.04	4.25 ± 1.17	
7	4.00 ± 1.31	4.13 ± 1.36	3.63 ± 1.30	3.75 ±1.49	
8	4.38 ± 1.41	4.75 ± 0.46	4.88 ± 0.35	4.88 ± 0.35	
9	4.75 ± 0.46	4.75 ± 0.46	4.88 ± 0.35	4.88 ± 0.35	
10	4.38 ± 1.06	4.88 ± 0.35	4.88 ± 0.35	4.88 ± 0.35	

Table 8. Round 1 Question Conformance & Alignment with Learning outcome of the Test of Module 2

Question #	Category				ICCs (95% CI)
	Clarity and understanding (Mean ± SD)	Appropriateness of Answer response options (Mean ± SD)	Meaningfulness of question to assess content knowledge (Mean ± SD)	Degree to which question aligns with learning outcome (Mean ± SD)	
1	3.57 ± 1.40	3.57 ± 1.27	3.86 ± 1.35	4.00 ± 1.41	.93 (.83 - .99)
2	4.00 ± 0.82	4.00 ± 0.82	4.14 ± 0.69	4.14 ± 0.69	
3	3.43 ± 1.27	4.00 ± 1.16	3.71 ± 1.38	3.71 ± 1.38	
4	4.57 ± 0.54	4.57 ± 0.54	4.29 ± 1.11	4.43 ± 1.13	
5	4.57 ± 0.54	4.43 ± 0.79	4.71 ± 0.49	4.71 ± 0.49	
6	4.86 ± 0.38	4.71 ± 0.49	4.86 ± 0.38	4.14 ± 0.69	
7	4.14 ± 0.69	4.86 ± 0.38	4.86 ± 0.38	4.86 ± 0.38	
8	4.57 ± 0.79	4.71 ± 0.76	4.71 ± 0.76	4.71 ± 0.76	
9	4.29 ± 0.49	4.71 ± 0.49	4.71 ± 0.49	4.71 ± 0.49	
10	4.43 ± 0.79	4.86 ± 0.38	4.86 ± 0.38	4.14 ± 0.69	

Table 9. Round 1 Result of Online Training Module Content

Category	Question	Mean ± SD	ICCs (95% CI)
Online training	Training program is easy to access	4.63 ± 0.52	.85 (.60 - .97)
	Utilizes visual resources (Angle of camera, angle of performance front and side view)	4.50 ± 0.54	
	Voice narration is clear and understandable	4.25 ± 0.89	
	Voice narration is at an appropriate pace	4.63 ± 0.74	
	Written material in video is easy to read	4.63 ± 0.74	

Round 2

***Module 1 Evaluation of Frame of Reference Training Criteria*.**

The mean score in round two was similar to that of round 1 (Mean ± SD = 4.33 ± 0.55) (see Table 10). Experts' evaluation score agreement using intraclass correlation coefficient was good (ICC = 0.81, 95% CI: 0.67 - 0.97; see Table 10). The ICC value while lower in round 2 satisfied the apriori criteria. The smaller number of experts (5 instead of 8 from round 1) completed the evaluation of module 1 in round 2 likely impacting the results of the ICC in round 2. There was a suggestion from experts #2, #6, and #8 that the audio quality (i.e., artificial computer voice; see Appendix O #1_5) was still not clear despite the changes made after round 1. The primary researcher changed the audio from a digital voice to a human voice narrating the content in both modules. Another comment of experts #2 and #6 was the redundancy of training contents in Module 1. Duplicated contents were revised to reduce the running time of this module.

***Module 1 Evaluation of Learning Outcomes*.**

Table 11 presents evaluation mean and standard deviation scores of each learning outcome, and ICC value among the expert panels. All evaluation scores of the learning outcome of Module 1 were above 4-points which means strong alignment with the learning outcomes of Module 1. The evaluation scores of learning outcome #1 "The rater understands the TGMD-3 and its components" (*Mean* ± *SD* = 4.33 ± 0.58) and learning outcome #4 "The raters understand how to score the TGMD-3" (*Mean* ± *SD* = 4.33 ± 1.16) had the same mean scores as each other with different standard deviations. Meanwhile, the other two learning outcomes (i.e., #2 and #3) showed the same mean and standard deviation scores. The agreement value of the experts in this evaluation category was moderate (ICC = 0.77, 95% CI: 0.53 - 0.82). The ICC values did not

meet the initial criteria set for .08 or above. Therefore, the primary researcher revised contents in module 1 to effectively deliver core information about the TGMD-3, such as the skill items, components, and the scoring method. The important factors of revision were clear auditory and visual resources as well as simple explanation with examples to maintain learners' focus on the training.

Module 1 Evaluation of the Test for Understanding.

Regarding the test of Module 1, the experts' evaluation results including the mean and standard deviation of each question as overall ICC are presented in Table 12. The ICC value of the test of Module 1 was excellent (ICC = 0.96, 95% CI: 0.88 - 0.99). Question #7 showed the lowest mean scores of all evaluations on clarity and understanding (*Mean* ± *SD* = 3.35 ± 1.26), appropriateness of answer response options (*Mean* ± *SD* = 3.25 ± 1.26), meaningfulness in content knowledge (*Mean* ± *SD* = 3.75 ± 0.96) and alignment with the learning outcome (*Mean* ± *SD* = 3.75 ± 0.96). All other questions were scored above 4 points on the four criteria on the evaluation indicating strong clarity and appropriateness of the question, appropriateness of answer response options, meaningfulness in content knowledge, and alignment with the learning outcome. The experts #2, #5, #6, and #8 suggested that each question should be clear and understandable from the Module 1 to learn about the TGMD-3 (see Appendix O #3_4, #4_5, #6_5, and #7_5). The finalized test of Module 1 was constructed according to the experts' comments in Round 2. The questions and answer choices on the test, for example were revised to consider novice raters' background and motivation for continued participation in this project. For example, some questions were revised to reduce the number of response choices from 4 to 3. Also, the answer selections were revised to reduce the level of difficulty.

Module 2 Evaluation of Frame of Reference Training Criteria.

Descriptive statistics of experts' evaluation scores ranged from 4.00 and 4.33 with standard deviations ranging from 0.0-1.0 (see Table 10). The ICC value of Module 2 was excellent (ICC = 0.90, 95% CI: 0.76 - 0.99). Experts #2 and #8 commented that overall slides with performance example videos were useful to correctly score the performance criteria of the skills on the TGMD-3 (see Appendix P #3_3, #4_4, and #5_4). However, experts #6 and #8 suggested that additional locomotor and ball skills on the TGMD-3 with more performance examples in children with DD would be helpful to improve novice raters' scoring accuracy in Module 2. After round 2, the primary researcher added sample correct and incorrect skill performance videos on each performance criterion of the run and the two-hand strike skills on the TGMD-3 among children with DD.

Module 2 Evaluation of Learning Outcomes.

Table 11 presents the evaluation means and standard deviation scores of each learning outcome, and the agreement value (ICC) among the experts. Five learning outcomes in Module 2 to provide behavioral and movement characteristics of developmental disabilities (DD) and correct scoring method of each performance criterion of the run and the two-hand strike skills on the TGMD-3 among children with DD showed excellent ICC values (ICC = .91, 95% CI: .58 - .99) among the experts in Round 2. The evaluation mean scores were all above 3 out of 5 with different standard deviations (SD: 0.58 – 1.16). The third learning outcome was "The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3" and #4 "The raters can recognize the performance criteria of the two-hand strike skill in the ball skill subtest on the TGMD-3" had the highest evaluation scores (*Mean* ± *SD* = 4.33 ± 0.58) among the learning outcomes of Module 2. Learning outcome #2 was "The rater can list behavior and

movement characteristics of children with DD" (*Mean* ± *SD* = 3.67 ± 0.58) and #5 "The raters get information to score fundamental motor skills (FMS) among children with DD according to each criterion on the TGMD-3" (*Mean* ± *SD* = 3.67 ± 1.16) showed the lowest evaluation scores. The expert #8 suggested that each skill performance video should be used to describe more clearly how to score a certain performance criterion (see Appendix P #6_5). The primary researcher revised each performance video to point out the skill performance on the specific scene according to each performance criterion in module 2.

Module 2 Evaluation of the Test for Understanding

The evaluation of the alignment of test questions with the learning outcomes for module 2 can be found in Table 12 and Appendix S (ICC = .96, 95% CI: .88 - .99). Experts #2 and #8 suggested to revise some challenging performance videos on the test questions. For example, a child with a developmental disability changed his hand grip right before the two-hand strike skill on the video. This performance video might be difficult to score correctly because it required highly sensitive observation of novice raters on the multiple-choice question with six response options. Test questions were revised to reduce not only the number of skill performance criteria being assessed in the test questions, but also the reduced the number response options.

Evaluation of Online Training Criteria

A variety of multimedia components using video, audio, graphics, and web links as educational technology were evaluated in Round 2. Those components are important factors in effectively delivering information in online education (Mangal & Mangal, 2009). All five evaluation scores on the survey were above 4 out of total of 5 points for each question. The mean scores of the four evaluation questions (i.e., #1, 3, 4, & 5) were 4.33 with the exception of #2 "Utilizes visual resources" (*Mean* ± *SD* = 4.00 ± 0.58; see Table 14; see Appendix #7_2). The

agreement value of the experts in the online training evaluation category was good (ICC = .85, 95% CI: .70 - .98). Expert #2, #6, and #8 pointed out the audio issue that was difficult to hear and understand throughout the module (see Appendix Q #7_3). There was difficulty in using computer voice audio in both round 1 and 2 due to the limitations with the video editing program. The primary researcher decided to replace the artificial computer voice with the human voice throughout both module 1 and 2. Regarding visual resources, overall statements, pictures, and videos were revised to provide clear information to improve understanding.

Table 10. Statistic Results from Round 2 Frame-of-reference Training Module Content

Module	Category	Question	Mean ± SD	ICCs (95% CI)
Module 1	Identify correct performance dimension under natural circumstances	Introduce to the TGMD-3 and the scope	4.33 ± 0.47	.81 (.67 - .97)
		Introduce 6 locomotor skills	4.33 ± 0.47	
		Introduce 7 ball skills	4.33 ± 0.47	
	Provides information to score the variables in question	Introduce scoring FMS on the TGMD-3	4.33 ± 0.94	
Module 2	Provides information about developmental disabilities	Introduce developmental disability (DD)	4.33 ± 0.58	.90 (.76 - .99)
		Introduce motor performance of children with DD	4.00 ± 0.00	
		Use video to effectively illustrate motor delay in children with DD	4.00 ± 0.82	
	Provides information to score the run skill of a child with DD	Discuss scoring of the run skill	4.33 ± 0.58	
		Shows good and bad performance of the run skill	4.33 ± 0.58	
		Provided scoring feedback on the run skill	4.33 ± 0.58	
	Provides information to score the two-hand strike skill of a child with DD	Discuss scoring of the run skill	4.33 ± 0.58	
		Shows good and bad performance of the run skill	4.33 ± 0.58	
		Provided scoring feedback on the run skill	4.33 ± 0.58	

Table 11. Result of Alignment of Content with Learning Outcome from Round 2

Category	Question	Mean ± SD	ICCs (95% CI)
Learning Outcome of Module 1	#1: The rater understands the TGMD-3 and its components	4.33 ± 0.58	.77 (.53 - .82)
	#2: The raters can identify the skills in the locomotor subtest	4.00 ± 0.00	
	#3: The raters can identify the skills in the ball skills subtest	4.00 ± 0.00	
	#4: The raters understand how to score on the TGMD-3	4.33 ± 1.16	
Learning Outcome of Module 2	#1: The rater can explain about developmental disability (DD)	4.00 ± 1.00	.91 (.58 - .99)
	#2: The rater can list behavior and movement characteristics of children with DD	3.67 ± 0.58	
	#3: The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3	4.33 ± 0.58	
	#4: The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3	4.33 ± 0.58	
	#5: The raters get information to score FMS among children with DD according to each criterion on the TGMD-3	3.67 ± 1.16	

Table 12. Round 2 Question Conformance & Alignment with Learning outcome of Test of Module 1

Question #	Category				ICCs (95% CI)
	Clarity and understanding (Mean ± SD)	Appropriateness of Answer response options (Mean ± SD)	Meaningfulness of question to assess content knowledge (Mean ± SD)	Degree to which question aligns with learning outcome (Mean ± SD)	
1	4.50 ± 0.58	4.00 ± 1.41	4.50 ± 0.58	4.50 ± 0.58	.96 (.88 - .99)
2	4.50 ± 0.58	4.50 ± 0.58	4.25 ± 0.50	4.50 ± 0.58	
3	4.25 ± 0.50	4.50 ± 0.58	4.25 ± 0.50	4.00 ± 0.82	
4	4.50 ± 0.58	4.50 ± 0.58	4.50 ± 0.58	4.50 ± 0.58	
5	4.25 ± 0.50	3.75 ± 1.26	4.25 ± 0.50	4.00 ± 0.82	
6	4.50 ± 0.58	4.50 ± 0.58	4.50 ± 0.58	4.50 ± 0.58	
7	3.35 ± 1.26	3.25 ± 1.26	3.75 ± 0.96	3.75 ±0.96	
8	4.75 ± 0.50	4.25 ± 1.50	4.75 ± 0.50	4.75 ± 0.50	
9	4.75 ± 0.50	4.75 ± 0.50	4.75 ± 0.50	4.75 ± 0.50	
10	4.50 ± 1.00	4.25 ± 1.50	4.50 ± 1.00	4.50 ± 1.00	

Table 13. Round 2 Question Conformance & Alignment with Learning outcome of Test of Module 2

Question #	Category				ICCs (95% CI)
	Clarity and understanding (Mean ± SD)	Appropriateness of Answer response options (Mean ± SD)	Meaningfulness of question to assess content knowledge (Mean ± SD)	Degree to which question aligns with learning outcome (Mean ± SD)	
1	4.40 ± 0.55	3.80 ± 1.30	4.20 ± 0.84	4.60 ± 0.55	.97 (.93 - .99)
2	4.40 ± 0.55	4.20 ± 1.30	4.40 ± 0.55	4.60 ± 0.55	
3	4.00 ± 0.71	4.20 ± 0.84	4.00 ± 0.71	4.40 ± 0.55	
4	4.60 ± 0.55	4.20 ± 1.30	4.20 ± 1.30	4.60 ± 0.55	
5	4.80 ± 0.45	4.40 ± 1.34	4.40 ± 0.89	4.60 ± 0.55	
6	4.40 ± 0.55	4.40 ± 0.55	4.20 ± 0.84	4.40 ± 0.55	
7	4.60 ± 0.55	4.40 ± 1.34	4.40 ± 0.55	4.60 ± 0.55	
8	4.60 ± 0.55	4.40 ± 1.34	4.60 ± 0.55	4.60 ± 0.55	
9	4.80 ± 0.45	4.40 ± 1.34	4.60 ± 0.55	4.80 ± 0.45	
10	4.40 ± 0.89	4.40 ± 1.34	4.20 ± 0.84	4.80 ± 0.45	

Table 14. Round 2 Result of Online Training Module Content

Category	Question	Mean ± SD	ICCs (95% CI)
Online training	Training program is easy to access	4.33 ± 0.58	.85 (.70 - .98)
	Utilizes visual resources (Angle of camera, angle of performance front and side view)	4.00 ± 0.58	
	Voice narration is clear and understandable	4.33 ± 0.58	
	Voice narration is at an appropriate pace	4.33 ± 0.58	
	Written material in video is easy to read	4.33 ± 1.16	

Effect of a TGMD-3 Online Training Program on Rater Accuracy

The following sections summarize the results from the data collected on the scoring of the run and two-hand strike by the 41 novice raters.

Run

The descriptive results of the run skill scores between the novice rater and experts are reported in Table 15. There was a significant difference in the reduction of errors between novice and expert raters for the run skill ($F(1, 39)= 56.431$, $p < .001$). The scoring difference of the run skill between the expert and the novice raters was positively reduced from the 1st scoring to 3rd scoring (Hypothesis #1, 2 and 3; see Figure 4). On the 1st scoring of the run skill, the mean score of novice raters was 5.55, and the expert score was 8.00. The mean score difference between those raters was about -2.45 with 1.29 SD. After the module 2 training, the mean score of novice rater was 6.97 on the 2nd scoring. The mean difference with the expert was -1.03 with 1.05 SD. On the 3rd scoring of this skill, novice raters' mean score was 7.32. The mean difference was -0.68 with 0.96 SD compared to the expert. The mean score differences of the run skill between novice and expert raters decreased as the rater training progressed. Also, the standard deviations decreased from the 1st to 3rd scoring of the run skill. The effect size of a TGMD-3 online rater training on the run skill was strong (η_p^2 = .65).

Table 15. Descriptive result of the run skill

Rater	Time	Mean ± SD	Mean Difference ± SD
Expert	Scoring	8.00	
Novice	1st scoring	5.55 ± 1.29	-2.45 ± 1.29
	2nd scoring	6.97 ± 1.05	-1.03 ± 1.05
	3rd scoring	7.32 ± 0.96	-0.68 ± 0.96

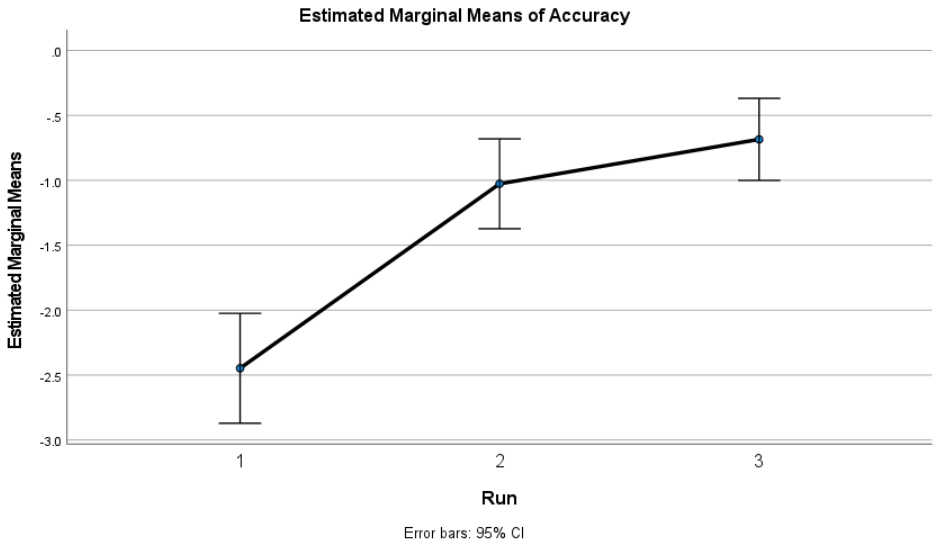

Figure 4. Scoring Difference of the run skill between the expert and novice raters among 3 rounds

Pairwise comparison table of multiple paired t-test between each pair of time points (see Table 16) confirms that errors for scoring the run skill significantly decreased and hence accuracy of scoring the run skill significantly improved by 1.42 points between 1st and 2nd scoring ($p < .001$; Hypothesis #2, #3, and #4). The difference between expert and novice raters significantly reduced by 0.34 points between 2nd and 3rd scoring ($p < .05$).

Table 16. Pairwise comparison result of the run skill

Comparison		Mean change	SE	95% CI		*P*
				Lower	Upper	
1st scoring	2nd scoring	1.42	0.19	-1.91	-0.93	< .001**
2nd scoring	3rd scoring	0.34	0.11	-0.63	-0.05	.02*
1st scoring	3rd scoring	1.76	0.21	1.25	2.28	< .001**

*$p < .05$; **$p < .001$*

The *t*-test value of all three rounds was significant ($p < .001$). Each *t*-test value showed a difference between expert and novice raters in the run skill 11.10, 5.90, and 4.31 for the first, second and third rounds respectively. While the scoring differences between those two groups

were reduced, novice raters' scores still differed from the expert score at the end of the 3rd scoring.

Two-hand strike.

The descriptive results of the two-hand strike skill scores between the novice rater and experts are represented in Table 17. There was a significant decrease in scoring errors between novice and expert raters supporting the improvement in scoring the two-hand strike skill ($F(1, 39)= 35.549$, $p < .001$). The scoring difference of the two-hand strike skill between the expert and the novice raters was positively reduced from the 1st scoring to 3rd scoring (Hypothesis #1, 2 and 3; see Figure 5).

Table 17. Descriptive result of the two-hand strike skill

Rater	Time	Mean ± SD	Mean Difference ± SD
Expert	Scoring	9.00	
Novice	1st scoring	6.37 ± 1.36	-2.63 ± 1.36
	2nd scoring	7.68 ± 1.17	-1.32 ± 1.17
	3rd scoring	8.16 ± 0.89	-0.84 ± 0.89

Figure 5 shows the change of scoring mean difference of the two-hand strike skill on the TGMD-3 between novice and expert raters. The mean score of novice raters was 6.37 on the 1st scoring of the two-hand strike skill, and the expert score was 9.00. The mean score difference between those raters was about -2.63 with 1.36 SD. The mean score of novice rater was 7.68 after module 2 training for the 2nd scoring. The mean difference with the expert was -1.32 with 1.17 SD. The novice raters' mean score of this skill was 8.16 on the 3rd scoring. The mean difference was -0.84 with 0.89SD compared to the expert. As the rater training progressed, the mean score differences of the two-hand strike skill between the two raters decreased. The standard deviations also decreased from the 1st to 3rd scoring of the two-hand strike skill. The

effect size of a TGMD-3 online rater training to score the two-hand strike skill was strong (η_p^2 = .76).

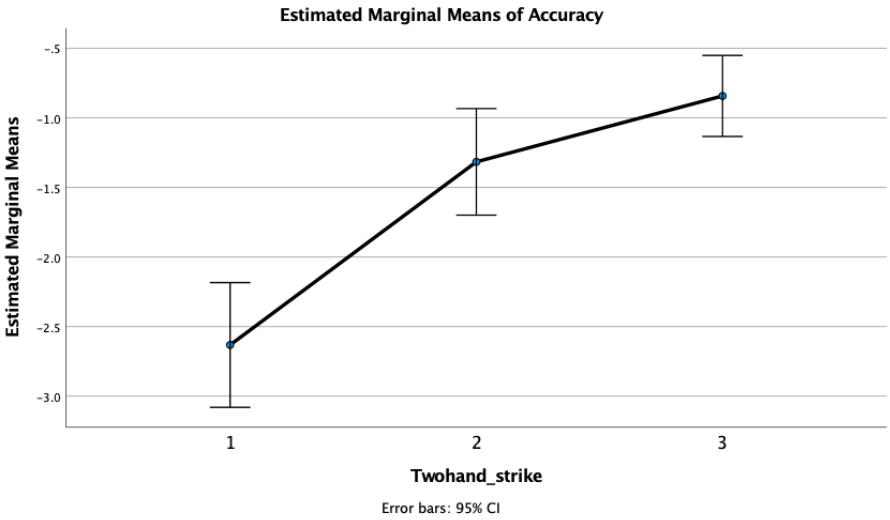

Figure 5. Scoring Difference of the two-hand strike skill between the expert and novice raters among 3 rounds

Pairwise comparison of multiple paired t-test between each pair of time points illustrates significant improvement by 1.32 points between 1st and 2nd scoring ($p < .001$; see Table 18). The difference was significantly reduced by 0.47 points between 2nd and 3rd scoring ($p < .05$; Hypothesis #1, #2, and #3).

Table 18. Pairwise comparison result of the two-hand strike skill

Comparison		Mean change	SE	95% CI		*P*
				Lower	Upper	
1st scoring	2nd scoring	1.32	0.25	0.69	1.94	< .001**
2nd scoring	3rd scoring	0.47	0.15	0.09	0.86	.01*
1st scoring	3rd scoring	1.79	0.24	1.18	2.40	< .001**

*$p < .05$; **$p < .001$*

Total skill score

Table 19 shows the descriptive results of the two skills combined between the novice rater and experts. There was a significant difference in the reduction of scoring errors and thus a significant improvement in scoring accuracy on the total skill scores ($F(1, 39)= 64.323$, $p < .001$). The scoring difference of the total skill score between the expert and the novice raters was positively reduced from the 1st scoring to 3rd scoring (Hypothesis #1; see Figure 6).

On the 1st scoring of the total score, the mean score of novice raters was 11.92, and the expert score was 17.00. The mean score difference between those raters was about -5.08 with 2.20 SD. After the module 2 training, the mean score of novice rater was 14.66 on the 2nd scoring. The mean difference with the expert was -2.34 with 1.92 SD. On the 3rd scoring of this skill, novice raters' mean score was 15.45. The mean difference was -1.55 with 1.57SD compared to the expert. The mean score differences of the two skills between novice and expert raters decreased through the rater training. Standard deviations also showed decreased scores from the 1st to 3rd scoring of the total skills in a child with DD. The scoring difference of the total skill score between the expert and the novice raters positively improved from the 1ststoring to 3rd scoring (see Figure 6). The effect size of a TGMD-3 online rater training on total skill score was strong (η_p^2 = .77).

Table 19. Descriptive result of the total skill score

Rater	Time	Mean ± SD	Mean Difference ± SD
Expert	Scoring	17.00	
Novice	1st scoring	11.92 ± 2.20	-5.08 ± 2.20
	2nd scoring	14.66 ± 1.92	-2.34 ± 1.92
	3rd scoring	15.45 ± 1.57	-1.55 ± 1.57

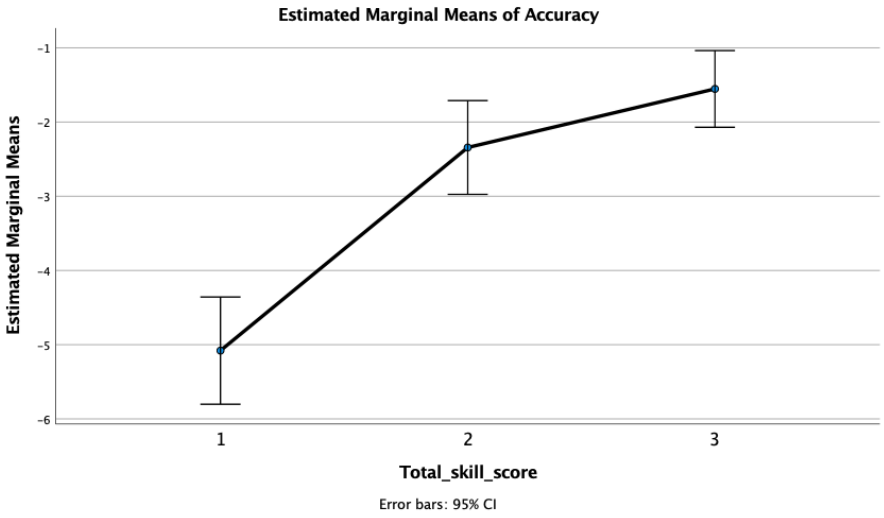

Figure 6. Scoring Difference of the total skill score between the expert and novice raters among 3 rounds

Table 20 presents multiple paired t-test between each pair of time points. The errors in scoring significantly decreased such that scoring accuracy of the total skill score significantly improved by 2.74 points between 1st and 2nd scoring ($p < .001$; Hypothesis #1). The difference was significantly reduced by 0.79 points between 2nd and 3rd scoring ($p < .001$; Hypothesis #2). The pairwise comparison results between 1st and 3rd scoring showed significant improvement on scoring accuracy of the total skill score through the TGMD-3 online rater training ($p < .001$; Hypothesis #3).

Table 20. Pairwise comparison result of the total skill score

Comparison		Mean change	SE	95% CI		*P*
				Lower	Upper	
1st scoring	2nd scoring	2.74	0.38	1.80	3.68	< .001**
2nd scoring	3rd scoring	0.79	0.15	0.31	1.27	< .001**
1st scoring	3rd scoring	1.79	0.24	1.18	2.40	< .001**

***p < .001*

5 DISCUSSION

TGMD-3 Online Training Program on Rater Accuracy

The purpose of this study was to investigate the impact of a TGMD-3 online rater training program for novice raters on the scoring accuracy of FMS of children with DD on the TGMD-3. Previous research by Palmer and Brian (2017) reported significant differences in scores on both locomotor and object control skill (i.e., the ball skill in the TGMD-3) subtests on the TGMD-2 between novice and expert raters. In the study by Palmer and Brian (2017), on the locomotor skill subtest, all skills were significantly different, with the exception of the gallop skill (p = .09). Similarly, Kim and colleagues (2012) investigated rater effects in scoring FMS items on the TGMD-2 among children with intellectual disabilities. They found that the run skill had relatively large error variance by rater effects (17.89%) compared to other locomotor skills on the instrument. Palmer and Brian (2017) and Kim et al (2012) recommended the development of training protocols to correctly score FMS on the TGMD assessment instrument. In the present study, though there were scoring differences on the TGMD-3 between expert and novice raters, the difference in scores between expert and novice raters decreased suggesting that the overall scoring accuracy of the run, two-hand strike, and total skill scores were significantly improved following two rounds of the training program. These findings support the recommendation that a training program to score the TGMD-3 can reduce scoring errors of novice raters and improve scoring accuracy of FMS of children with DD. The TGMD-3 training program in the present study could be used as a basic resource and as a sample training program for the development of additional training programs for other locomotor skills (i.e., gallop, hop, skip, horizontal jump,

and slide skills) to possibly reduce scoring differences between expert and novice raters and improve the scoring accuracy of novices on these latter skills.

The effectiveness of the online training modules may be attributed in part to the apriori criteria set for acceptable interrater reliability in the validation of the training module content and assessments. Previous Delphi studies defined various degrees of agreement over multiple rounds of validation ranging from 0.66 (Dyer and colleagues, 2011), 0.70 (Hasson et al., 2000) to above 0.80 (Finger et al., 2006). According to references of ICC assessment, Cicchett (1994) presented an interpretation guideline that ICC inter-rater agreement was excellent with scores between 0.75 and 1.00. A different guideline from Koo and Li (2016) defined agreement values as good ($0.75 < ICC \leq 0.90$) and excellent ($ICC > 0.90$). The present study defined the eligible agreement value (ICC) equal to or above 0.80. All evaluation categories were above .90 in round 1. In round 2, however, one evaluation category did not meet the standard value of ICC (i.e., 0.80; learning outcome of module 1 (ICC: 0.77). This lower score may be attributed to a smaller number of experts participating in round 2 of the Delphi study. Overall, the expert panels' agreement in the present study across the two rounds of validation were considered good or excellent based on these ICC interpretation guidelines (Cicchett, 1994; Koo & Li, 2016). The ICC values provide confirmation of the degree to which the content in the training modules addressed the FOR criteria, the criteria for online learning, and the learning outcomes that guided the development of each module and test for understanding were appropriate to reduce the scoring errors of novice raters.

A second element contributing to the success of the rater training modules may be the application of the 'frame-of-reference' (FOR) elements in the modules. Different strategies of rater training have been used to expand knowledge, evaluation skills, and ultimately enhance

rating accuracy of novice raters (McIntyre et al., 1984). Specifically, the FOR approach emphasizes the importance of rater's awareness of the multidimensional performance criteria along with the systematic comparison between actual and desired movement outcomes (Bernardin & Buckley, 1981; Roch & O'Sullivan, 2003). The FOR training' elements used in the present study appear to have been effective for improving scoring accuracy of the run and two-hand strike skills on the TGMD-3 among children with DD. These findings are consistent with those of Bernardin et al. (1981), Lievens et al. (2007), and Rosales Sánchez et al.,(2019) who used FOR training and reported a significant decrease in rating errors and increased accuracy of scores between novice and expert raters. The findings in the present study provide additional support that FOR training can contribute to ensuring raters understand the multidimensional performance criteria of FMS in comparison with the common errors often seen among children with DD.

Liu (2014) and Lozovoy et al (2019) presented that online education has several disadvantages for learners, such as lack of motivation for academic engagement and a home or familiar environment that presents increased distractions from attending to the educational content to be learned. The online training in the present study was designed to address several of the known limitations of online learning. First, the online module used strategies to address different learning styles including videos, audio narration, transcribed narration. Second, the training videos used a variety of multimedia components including videos from multiple angles, and graphics to cue and focus the learners' attention on specific elements of motor skill performance. All these techniques were intended to reduce distractions and focus the learner's attention on the content being presented. Third, a test with a pass rate of 80% or higher was used to check for understanding at the end of each module providing an additional external motivator

to attend to the content in the modules. Lastly, the content was aligned with apriority learning outcomes that provided clear directions for the development of content in each module. The strategies used in the online modules in the present study appear to have contributed to a successful rater training program for the TGMD-3. The present study showed a reduction in score differences between novice and expert raters and hence a positive effect on the improvement of scoring accuracy of FMS among children with DD through a TGMD-3 online FOR training program (Hypothesis #1, #2, #3). Similar findings were reported by Chafouleas et al. (2015) who used a web-based FOR training module and found a significant impact on the rating accuracy of behavioral performance from a varied group of raters.

Conclusion

The use of the online training program in the present study can be used by novice physical education teachers, preservice teachers, or practitioners to learn and improve their scoring competency in skill analysis (i.e., specifically the run and two hand strike) of children with DD. Reduced errors leading to improved accuracy in skill evaluation may contribute to ensuring adequate placement decisions for children with special needs in physical education (Akuffo and Hodge, 2008; Columna et al., 2010) and to the development of PE/APE curriculum to teach the performance criteria of the different FMS in the TGMD-3 (Lytle et al., 2010).

Limitations

There are several limitations that impacted the development and validation of the online training modules as well as the rater training portion of this study. With regards to the Delphi study, of the 8 experts who completed round one in the process, despite providing compensation

to the experts for completing the Delphi evaluation, only 5 completed round 2. The smaller number of experts may have lowered rather than increased the ICC values. In so doing, the goal of achieving and ICC of .80 or higher was achieved in round 1 with 8 experts but was not achieved in round 2 with only 5 experts though the ICC results were excellent above .75 (Cicchetti, 1994). These findings highlight the importance of recruiting, retaining, and if necessary, recruiting new members to ensure a broad and diverse pool of experts when conducting multiple rounds of a Delphi study. Second, only one subject with DD was scored to investigate the effect of the TGMD-3 rater training intervention on scoring accuracy. Due to varying behavior and movement characteristics of children with DD, the results may differ for children with DD who have different behavior and movement characteristics than that presented in the current sample video. Third, the length of the rater training programs (15-min for Module 1 and 20-min of Module 2) limited the amount of information provided to score sample videos of children with more diverse challenging behaviors and levels of motor skill performance. Third, the online training process in the present study could not be controlled without any influential factors according to individual environment settings as the discussed disadvantage of online education.

Future Research Directions

The following recommendations are provided to guide future research on the development of online FMS training programs.

1. A TGMD-3 online rater training program for novice raters could be conducted using multiple children with DD in sample training videos so that novice raters see and practice scoring children with a range of movement behavior patterns to facilitate increased accuracy in scoring FMS among children with DD.

2. This was a pilot study to investigate the effect of a TGMD-3 online training intervention. Therefore, the comparison between an experiment and a control group could be implemented to sufficiently verify the impact of the training intervention.
3. Regarding the novice participants, this study recruited only college students who had not taken a motor development class or those with no experience scoring the TGMD-3. However, kinesiology program curriculums differ from university to university and the personal experiences of college students with persons with a disability (e.g., dealing with individuals with disabilities, siblings with disabilities) were not considered in this study. Future research should consider applying more specific eligibility (for recruitment) and may consider looking at demographics and previous experience of novice raters with individuals with a disability as a mediator on scoring accuracy.
4. The current study used a test for understanding separate from the online training modules. Online training modules that embed checks for understanding within the online module are recommended to engage learners with the content and to reduce the use of multiple platforms for online training.

References

Aguinis, H. (2017). *Performance Management*. CreateSpace Independent Publishing Platform.

Aguinis, H., Mazurkiewicz, M. D., & Heggestad, E. D. (2009). Using Web-Based Frame-of-Reference Training to Decrease Biases in Personality-Based Job Analysis: An Experimental Field Study. *Personnel Psychology*, *62*(2), 405–438. https://doi.org/10.1111/j.1744-6570.2009.01144.x

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Allen, K. A., Bredero, B., Van Damme, T., Ulrich, D. A., & Simons, J. (2017). Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. *Journal of Autism and Developmental Disorders*, *47*(3), 813–833. https://doi.org/10.1007/s10803-016-3005-0

America Psychiatric Association. (2016). *DSM-5*. https://www.psychiatry.org/psychiatrists/practice/dsm

Annotated Bibliography | Pearson Assessment. (n.d.). Retrieved March 21, 2020, from https://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildPerceptionandVisuomotorAbilities/MABC-2/ForThisProduct/AnnotatedBibliography.aspx

Arvey, R. D., & Murphy, K. R. (1998). Performance Evaluation in Work Settings. *Annual Review of Psychology*, *49*(1), 141–168. https://doi.org/10.1146/annurev.psych.49.1.141

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Barnett, L., Hinkley, T., Okely, A. D., & Salmon, J. (2013). Child, family and environmental correlates of children's motor skill proficiency. *Journal of Science and Medicine in Sport*, *16*(4), 332–336. https://doi.org/10.1016/j.jsams.2012.08.011

Barnett, L. M., Minto, C., Lander, N., & Hardy, L. L. (2014). Interrater reliability assessment using the Test of Gross Motor Development-2. *Journal of Science and Medicine in Sport*, *17*(6), 667–670. https://doi.org/10.1016/j.jsams.2013.09.013

Bartolotta, J., Bourelle, T., & Newmark, J. (2017). Revising the Online Classroom: Usability Testing for Training Online Technical Communication Instructors. *Technical Communication Quarterly*, *26*(3), 287–299. https://doi.org/10.1080/10572252.2017.1339495

Bastik, C., Kalkavan, A., Yamaner, F., Sahin, S., & Gullu, A. (2012). Investigation of Basic Motor Skills According to TGMD-2 Test on Male Athletes of 10 Ages Group Who Participated to Competitions in Different Sports Branches. *Procedia - Social and Behavioral Sciences*, *46*, 4741–4745. https://doi.org/10.1016/j.sbspro.2012.06.328

Berkeley, S. L., Zittel, L. L., Pitney, L. V., & Nichols, S. E. (2001). Locomotor and Object Control Skills of Children Diagnosed with Autism. *Adapted Physical Activity Quarterly*, *18*(4), 405–416. https://doi.org/10.1123/apaq.18.4.405

Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology*, *63*(3), 301–308. https://doi.org/10.1037/0021-9010.63.3.301

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *The Academy of Management Review*, *6*(2), 205–212. https://doi.org/10.2307/257876

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65*(1), 60–66. https://doi.org/10.1037/0021-9010.65.1.60

Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology, 62*(1), 64–69. https://doi.org/10.1037/0021-9010.62.1.64

Bisi, M. C., Pacini Panebianco, G., Polman, R., & Stagni, R. (2017). Objective assessment of movement competence in children using wearable sensors: An instrumented version of the TGMD-2 locomotor subtest. *Gait & Posture, 56*, 42–48. https://doi.org/10.1016/j.gaitpost.2017.04.025

Bittner, R. H. (1948). Developing an industrial merit rating procedure. *Personnel Psychology, 4*, 403–432. https://doi.org/10.1111/j.1744-6570.1948.tb01319.x

Block, M. E. (2016). *Teacher's Guide to Adapted Physical Education: Including Students with Disabilities in Sports and Recreation*. Brookes Publishing.

Breslin, C. M., & Rudisill, M. E. (2013). Relationships Among Assessment Time, Time on Task, and Motor Skill Performance in Children With Autism Spectrum Disorder. *Adapted Physical Activity Quarterly, 30*(4), 338–350. https://doi.org/10.1123/apaq.30.4.338

Brian, A. S., Haegele, J. A., Lieberman, L., & Bostick, L. (2016). The content and face validity for the Test of Perceived Motor Competence for children with visual impairments: A Delphi investigation. *British Journal of Visual Impairment, 34*(3), 238–247. https://doi.org/10.1177/0264619616658072

Brian, A., Taunton, S., Lieberman, L. J., Haibach-Beach, P., Foley, J., & Santarossa, S. (2018). Psychometric Properties of the Test of Gross Motor Development-3 for Children With

Visual Impairments. *Adapted Physical Activity Quarterly*, *35*(2), 145–158. https://doi.org/10.1123/apaq.2017-0061

Brian et al. - 2018—Psychometric Properties of the Test of Gross Motor.pdf. (n.d.). Retrieved March 21, 2020, from https://www.researchgate.net/profile/Ali_Brian/publication/323679993_Psychometric_Properties_of_the_Test_of_Gross_Motor_Development-3_for_Children_With_Visual_Impairments/links/5d3f3f6c299bf1995b55fdf4/Psychometric-Properties-of-the-Test-of-Gross-Motor-Development-3-for-Children-With-Visual-Impairments.pdf

BRONFENBRENNER, U. (1979). *The Ecology of Human Development*. Harvard University Press.

Bronfenbrenner, U. (1992). Ecological systems theory. In *Six theories of child development: Revised formulations and current issues* (pp. 187–249). Jessica Kingsley Publishers.

Bronfenbrenner, U. (2005). *Making Human Beings Human: Bioecological Perspectives on Human Development*. SAGE.

Brown, A., & Green, T. (2003). Showing up to Class in Pajamas (or Less!): The Fantasies and Realities of On-Line Professional Development Courses for Teachers. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *76*(3), 148–151. https://doi.org/10.1080/00098650309601992

Bruininks-Oseretsky Test of Motor Proficiency | Second Edition. (n.d.). Retrieved March 21, 2020, from https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-

Assessments/Motor-Sensory/Bruininks-Oseretsky-Test-of-Motor-Proficiency-%7C-Second-Edition/p/100000648.html

Bruininks-Oseretsky test of motor proficiency: Examiner's manual (Book, 1978) [WorldCat.org]. (n.d.). Retrieved March 21, 2020, from https://www.worldcat.org/title/bruininks-oseretsky-test-of-motor-proficiency-examiners-manual/oclc/4491119

Bulger, S. M., & Housner, L. D. (2007). Modified Delphi Investigation of Exercise Science in Physical Education Teacher Education. *Journal of Teaching in Physical Education*, *26*(1), 57–80. https://doi.org/10.1123/jtpe.26.1.57

Burton, A. W., & Miller, D. E. (1998). *Movement Skill Assessment*. Human Kinetics.

Burton, A. W., & Rodgerson, R. W. (2001). New Perspectives on the Assessment of Movement Skills and Motor Abilities. *Adapted Physical Activity Quarterly*, *18*(4), 347.

Cairney, J., Schmidt, L. A., Veldhuizen, S., Kurdyak, P., Hay, J., & Faught, B. E. (2008). Left-Handedness and Developmental Coordination Disorder. *The Canadian Journal of Psychiatry*, *53*(10), 696–699. https://doi.org/10.1177/070674370805301009

Campbell, S. K. (1993). Commentary. *Physical Therapy*, *73*(2), 110–113. https://doi.org/10.1093/ptj/73.2.110

Cantell, M. H., Smyth, M. M., & Ahonen, T. P. (2003). Two distinct pathways for developmental coordination disorder: Persistence and resolution. *Human Movement Science*, *22*(4–5), 413–431. https://doi.org/10.1016/j.humov.2003.09.002

Capio, C. M., Mak, T. C. T., Tse, M. A., & Masters, R. S. W. (2018). Fundamental movement skills and balance of children with Down syndrome. *Journal of Intellectual Disability Research*, *62*(3), 225–236. https://doi.org/10.1111/jir.12458

Cardy, R. L., & Keefe, T. J. (1994). Observational Purpose and Evaluative Articulation in Frame-of-Reference Training: The Effects of Alternative Processing Modes on Rating Accuracy. *Organizational Behavior and Human Decision Processes*, *57*(3), 338–357. https://doi.org/10.1006/obhd.1994.1019

Carmosino, K., Grzeszczak, A., McMurray, K., Olivo, A., Slutz, B., Zoll, B., Donahoe-Fillmore, B., & Brahler, C. J. (2014). *Test Items in the Complete and Short Forms of the BOT-2 that Contribute Substantially to Motor Performance Assessments in Typically Developing Children 6-10 Years of Age*. *7*(2), 14.

Chafouleas, S. M., Riley-Tillman, T. C., Jaffery, R., Miller, F. G., & Harrison, S. E. (2015). Preliminary Investigation of the Impact of a Web-Based Module on Direct Behavior Rating Accuracy. *School Mental Health*, *7*(2), 92–104. https://doi.org/10.1007/s12310-014-9130-z

Chirico, K. E., Buckley, M. R., Wheeler, A. R., Facteau, J. D., Bernardin, H. J., & Beu, D. S. (2004). A Note on the Need for True Scores in Frame-of-Reference (FOR) Training Research. *Journal of Managerial Issues*, *16*(3), 382–395. JSTOR.

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, *6*(4).

Clark, J. E., & Humphrey, J. H. (2002). *Motor Development: Research and Reviews*. National Association for Sport and Physical Education.

Clark, J. E., & Metcalfe, J. S. (n.d.). *THE MOUNTAIN OF MOTOR DEVELOPMENT: A METAPHOR*. 32.

Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology, 17*(4), 373–386. https://doi.org/10.1080/0144341970170401

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, *1*(3), 98–101.

Colombo-Dougovito, A. M. (2017). The role of dynamic systems theory in motor development research: How does theory inform practice and what are the potential implications for autism spectrum disorder? *International Journal on Disability and Human Development, 16*(2). https://doi.org/10.1515/ijdhd-2016-0015

Columna, L., Cook, A., Foley, J. T., & Bailey, J. (2014). Survey development to assess parental satisfaction with adapted physical education teachers' abilities working with children with autism. *Physical Education & Sport Pedagogy*, *19*(5), 481–493. s3h.

Columna, L., Pyfer, J., Senne, T., Velez, L., Bridenthrall, N., & Canabal, M. Y. (2008). Parental Expectations of Adapted Physical Educators: A Hispanic Perspective. *Adapted Physical Activity Quarterly*, *25*(3), 228–246. https://doi.org/10.1123/apaq.25.3.228

Comer, R. J. (2016). *Fundamentals of Abnormal Psychology 8e & Launchpad for Fundamentals of Abnormal Psychology 8e (Six-Months Access)*. WORTH PUBL Incorporated.

Cools, W., Martelaer, K. D., Samaey, C., & Andries, C. (2009). Movement Skill Assessment of Typically Developing Preschool Children: A Review of Seven Movement Skill Assessment Tools. *Journal of Sports Science & Medicine*, *8*(2), 154–168.

Crain, W. (2015). *Theories of Development: Concepts and Applications: Concepts and Applications*. Psychology Press.

Deflandre, A., Lorant, J., Gavarry, O., & Falgairette, G. (2001). Determinants of Physical Activity and Physical and Sports Activities in French School Children. *Perceptual and Motor Skills*, *92*(2), 399–414. https://doi.org/10.2466/pms.2001.92.2.399

Developmental Disabilities | CDC. (n.d.). Retrieved March 22, 2020, from https://www.cdc.gov/ncbddd/developmentaldisabilities/index.html

Downey, R., & Rapport, M. J. K. (2012). Motor Activity in Children With Autism: A Review of Current Literature. *Pediatric Physical Therapy*, *24*(1), 2–20. https://doi.org/10.1097/PEP.0b013e31823db95f

Dyer, B., Noroozi, S., Sewell, P., & Redwood, S. (2011). The Fair Use of Lower-Limb Running Prostheses: A Delphi Study. *Adapted Physical Activity Quarterly*, *28*(1), 16–26. https://doi.org/10.1123/apaq.28.1.16

Eather, N., Bull, A., Young, M. D., Barnes, A. T., Pollock, E. R., & Morgan, P. J. (2018). Fundamental movement skills: Where do girls fall short? A novel investigation of object-control skill execution in primary-school aged girls. *Preventive Medicine Reports*, *11*, 191–195. https://doi.org/10.1016/j.pmedr.2018.06.005

Ecological Psychology. Concepts and Methods for Studying the Environment of Human Behavior. Roger G. Barker. (1969). *The Quarterly Review of Biology*, *44*(3), 332–332. https://doi.org/10.1086/406221

Edwards, J., Jeffrey, S., May, T., Rinehart, N. J., & Barnett, L. M. (2017). Does playing a sports active video game improve object control skills of children with autism spectrum disorder? *Journal of Sport and Health Science*, *6*(1), 17–24. https://doi.org/10.1016/j.jshs.2016.09.004

Einspieler, C., Bos, A. F., Libertus, M. E., & Marschik, P. B. (2016). The General Movement Assessment Helps Us to Identify Preterm Infants at Risk for Cognitive Dysfunction. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00406

Eliasziw, M., Young, S., Woodbury, M. G., & Fryday-Field, K. (1994). Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example. *Physical Therapy*, *74*, 777–788. https://doi.org/10.1093/ptj/74.8.777

Ellinoudis, T., Evaggelinou, C., Kourtessis, T., Konstantinidou, Z., Venetsanou, F., & Kambas, A. (2011). Reliability and validity of age band 1 of the Movement Assessment Battery for Children – Second Edition. *Research in Developmental Disabilities*, *32*(3), 1046–1051. https://doi.org/10.1016/j.ridd.2011.01.035

Engel-Yeger, B., Rosenblum, S., & Josman, N. (2010). Movement Assessment Battery for Children (M-ABC): Establishing construct validity for Israeli children. *Research in Developmental Disabilities*, *31*(1), 87–96. https://doi.org/10.1016/j.ridd.2009.08.001

Estevan, I., Molina-García, J., Queralt, A., Álvarez, O., Castillo, I., & Barnett, L. (2017). Validity and Reliability of the Spanish Version of the Test of Gross Motor Development-3. *Journal of Motor Learning & Development*, *5*(1), 69–81.

Finger, M. E., Cieza, A., Stoll, J., Stucki, G., & Huber, E. O. (2006). Identification of Intervention Categories for Physical Therapy, Based on the International Classification of Functioning, Disability and Health: A Delphi Exercise. *Physical Therapy*, *86*(9), 1203–1220. https://doi.org/10.2522/ptj.20050134

Fleishman, E. A., Quaintance, M. K., & Broedling, L. A. (1984). *Taxonomies of human performance: The description of human tasks* (pp. xvi, 514). Academic Press.

Fleiss, J. L. (2011). *Design and Analysis of Clinical Experiments*. John Wiley & Sons.

Fransen, J., D'Hondt, E., Bourgois, J., Vaeyens, R., Philippaerts, R. M., & Lenoir, M. (2014). Motor competence assessment in children: Convergent and discriminant validity between the BOT-2 Short Form and KTK testing batteries. *Research in Developmental Disabilities*, *35*(6), 1375–1383. https://doi.org/10.1016/j.ridd.2014.03.011

Frontiers | Test of Motor Proficiency Second Edition (BOT-2): Compatibility of the Complete and Short Form and Its Usefulness for Middle-Age School Children | Pediatrics. (n.d.). Retrieved March 20, 2020, from https://www.frontiersin.org/articles/10.3389/fped.2019.00153/full

Gagen, L. M., & Getchell, N. (2006). Using 'Constraints' to Design Developmentally Appropriate Movement Activities for Early Childhood Education. *Early Childhood Education Journal*, *34*(3), 227–232. https://doi.org/10.1007/s10643-006-0135-6

Garcia, C. (1994). Gender Differences in Young Children's Interactions When Learning Fundamental Motor Skills. *Research Quarterly for Exercise and Sport*, *65*(3), 213–225. https://doi.org/10.1080/02701367.1994.10607622

Gesell, A. (1928). *Infancy and human growth* (pp. xvii, 418). MacMillan Co. https://doi.org/10.1037/14664-000

Getting the fundamentals of movement: A meta-analysis of the effectiveness of motor skill interventions in children—Logan—2012—Child: Care, Health and Development—Wiley Online Library. (n.d.). Retrieved March 21, 2020, from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2214.2011.01307.x?casa_token=xbyZlPi0yTcAAAAA:czO7Np7lG4adjj7c5jKAdqLTV9kf2cU7B3JUmAibPvYbIucAuvsmQnwnm8rax_pfjP4srNAcALdIUw

Glazer, D. D. (2009). Development and Preliminary Validation of the Injury-Psychological Readiness to Return to Sport (I-PRRS) Scale. *Journal of Athletic Training*, *44*(2), 185–189. https://doi.org/10.4085/1062-6050-44.2.185

Goldstein, S., & Naglieri, J. A. (2010). *Encyclopedia of Child Behavior and Development*. Springer Science & Business Media.

Gomez-Mejia, L. R. (1988). Evaluating Employee Performance: Does the Appraisal Instrument Make a Difference? *Journal of Organizational Behavior Management*, *9*(2), 155–172. https://doi.org/10.1300/J075v09n02_10

Goodway, J. D., Ozmun, J. C., & Gallahue, D. L. (2019). *Understanding Motor Development: Infants, Children, Adolescents, Adults*. Jones & Bartlett Learning.

Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology*, *94*(5), 1336–1344. https://doi.org/10.1037/a0016476

Gould, T. W., & Enomoto, H. (2009). Neurotrophic Modulation of Motor Neuron Development. *The Neuroscientist*, *15*(1), 105–116. https://doi.org/10.1177/1073858408324787

Gross Motor Performance by Children with Autism Spectrum Disorderand Typically Developing Children on TGMD-2 | Semantic Scholar. (n.d.). Retrieved March 22, 2020, from https://www.semanticscholar.org/paper/Gross-Motor-Performance-by-Children-with-Autism-on-Liu-Hamilton/b5bff117a1a4508f3ee42bfa4486df93673fb7b7

Hardy, L. L., Barnett, L., Espinel, P., & Okely, A. D. (2013). Thirteen-Year Trends in Child and Adolescent Fundamental Movement Skills: 1997–2010. *Medicine & Science in Sports & Exercise*, *45*(10), 1965–1970. https://doi.org/10.1249/MSS.0b013e318295a9fc

Harvey, W. J., Reid, G., Grizenko, N., Mbekou, V., Ter-Stepanian, M., & Joober, R. (2007). Fundamental Movement Skills and Children with Attention-Deficit Hyperactivity Disorder: Peer Comparisons and Stimulant Effects. *Journal of Abnormal Child Psychology*, *35*(5), 871–882. https://doi.org/10.1007/s10802-007-9140-5

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Hastie, P. A., Valentini, N. C., Rudisill, M. E., & chiviacowsky, S. (n.d.). Children's knowledge of skill cues and the enhancements of motor skill performance. *Journal of Physical Education and Sport*, *2018*(03).

Haubenstricker, J., & Seefeldt, V. (Eds.). (1986). *Physical activity & well-being*. American Alliance for Health, Physical Education, Recreation, and Dance.

Haywood, K. M., & Getchell, N. (2019). *Life Span Motor Development*. Human Kinetics.

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, *73*(1), 68–73. https://doi.org/10.1037/0021-9010.73.1.68

Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2007). *Movement Assessment Battery for Children-2. Second edition (MABC-2). Examiner's manual*. https://www.scienceopen.com/document?vid=c2ddd6db-c244-425d-b971-acdbc8bddd33

Hernandez, B. L. M., Gober, D., Boatwright, D., & Strickland, G. (2009). Jump Rope Skills for Fun and Fitness in Grades K-12. *Journal of Physical Education, Recreation & Dance*, *80*(7), 15–41. https://doi.org/10.1080/07303084.2009.10598352

Herr, C. M., & Bateman, B. D. (2013). *Handbook of Learning Disabilities, Second Edition*. Guilford Press.

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. https://doi.org/10.1016/j.psychsport.2014.03.005

Houwen, S., Hartman, E., Jonker, L., & Visscher, C. (2010). Reliability and Validity of the TGMD-2 in Primary-School-Age Children With Visual Impairments. *Adapted Physical Activity Quarterly*, *27*(2), 143–159. https://doi.org/10.1123/apaq.27.2.143

Hutzler, Y. (2007). A Systematic Ecological Model for Adapting Physical Activities: Theoretical Foundations and Practical Examples. *Adapted Physical Activity Quarterly*, *24*(4), 287–304. https://doi.org/10.1123/apaq.24.4.287

Interrater Reliability of the Movement Assessment Battery for Children | Physical Therapy | Oxford Academic. (n.d.). Retrieved March 21, 2020, from https://academic.oup.com/ptj/article/88/2/286/2742387

Ives, J. C. (2013). *Motor Behavior: Connecting Mind and Body for Optimal Performance*. Lippincott Williams & Wilkins.

Jaakkola, T., Hillman, C., Kalaja, S., & Liukkonen, J. (2015). The associations among fundamental movement skills, self-reported physical activity and academic performance during junior high school in Finland. *Journal of Sports Sciences*, *33*(16), 1719–1729. https://doi.org/10.1080/02640414.2015.1004640

Johnstone, A., Hughes, A. R., Janssen, X., & Reilly, J. J. (2017). Pragmatic evaluation of the Go2Play Active Play intervention on physical activity and fundamental movement skills

in children. *Preventive Medicine Reports*, *7*, 58–63. https://doi.org/10.1016/j.pmedr.2017.05.002

Kail, R. V., & Cavanaugh, J. C. (2012). *Human Development: A Life-Span View*. Cengage Learning.

Kalaja, S. (2012). *Fundamental Movement Skills, Physical Activity, and Motivation toward Finnish School Physical Education. A Fundamental Movement Skills Intervention*. 160.

Kelly, L. E. (2011). Designing and Implementing Effective Adapted Physical Education Programs. In *Sagamore Publishing LLC*. Sagamore Publishing LLC.

Kelly, L. E., & Melograno, V. J. (2014). *Developing the Physical Education Curriculum: An Achievement-Based Approach*. Waveland Press.

Kennedy, J., Brown, T., & Stagnitti, K. (n.d.). *Top-down and bottom-up approaches to motor skill assessment of children: Are child-report and parent-report perceptions predictive of children's performance-based assessment results?* 10.

Keown-Gerrard, J. L., & Sulsky, L. M. (2001). The Effects of Task Information Training and Frame-of-Reference Training With Situational Constraints on Rating Accuracy. *Human Performance*, *14*(4), 305–320. https://doi.org/10.1207/S15327043HUP1404_2

Kephart, N. C. (1960). *The slow learner in the classroom* (pp. xi, 292). Charles E. Merrill.

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

Kim, Y., Park, I., & Kang, M. (2012). Examining rater effects of the TGMD-2 on children with intellectual disability using G-theory and Many-facet Rasch model. *Adapted Physical Activity Quarterly: APAQ*, *29*, 346–365. https://doi.org/10.1123/apaq.29.4.346

Kirk, M. A., & Rhodes, R. E. (2011a). Motor Skill Interventions to Improve Fundamental Movement Skills of Preschoolers With Developmental Delay. *Adapted Physical Activity Quarterly, 28*(3), 210–232.

Kirk, M. A., & Rhodes, R. E. (2011b). Motor Skill Interventions to Improve Fundamental Movement Skills of Preschoolers With Developmental Delay. *Adapted Physical Activity Quarterly, 28*(3), 210–232. https://doi.org/10.1123/apaq.28.3.210

Klavina, A., Ostrovska, K., & Campa, M. (2017). Fundamental movement skill and physical fitness measures in children with disabilities. *European Journal of Adapted Physical Activity, 10*(1), 28–37. https://doi.org/10.5507/euj.2017.004

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Landers, R. Q., Carson, R. L., & Blankenship, B. T. (2010). Special Section: The promise and pitfalls of sport specialization in youth sport. *Journal of Physical Education, Recreation & Dance, 81*(8), 14–39.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change, 73*(5), 467–482. https://doi.org/10.1016/j.techfore.2005.09.002

Langley.pdf. (n.d.). Retrieved March 20, 2020, from http://www.vasa.abo.fi/users/jromar/Biamne38/langley.pdf

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology, 60*(5), 550–555. https://doi.org/10.1037/0021-9010.60.5.550

Lawshe, C. H. (1975). A QUANTITATIVE APPROACH TO CONTENT VALIDITY. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leonard, H. C. (2016). The Impact of Poor Motor Skills on Perceptual, Social and Cognitive Development: The Case of Developmental Coordination Disorder. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00311

Lerner, R. M., Liben, L. S., & Mueller, U. (2015). *Handbook of Child Psychology and Developmental Science, Cognitive Processes*. John Wiley & Sons.

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, *86*(2), 255–264. https://doi.org/10.1037/0021-9010.86.2.255

Lievens, F., & Sanchez, J. I. (2007). Can training improve the quality of inferences made by raters in competency modeling? A quasi-experiment. *Journal of Applied Psychology*, *92*(3), 812–819. https://doi.org/10.1037/0021-9010.92.3.812

Linstone, H. A., & Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. Addison-Wesley Publishing Company, Advanced Book Program.

Lloyd, M., MacDonald, M., & Lord, C. (2013). Motor skills of toddlers with autism spectrum disorders. *Autism*, *17*(2), 133–146. https://doi.org/10.1177/1362361311402230

Logan, S. W., Robinson, L. E., Wilson, A. E., & Lucas, W. A. (2012). Getting the fundamentals of movement: A meta-analysis of the effectiveness of motor skill interventions in children. *Child: Care, Health & Development*, *38*(3), 305–315. https://doi.org/10.1111/j.1365-2214.2011.01307.x

Loignon, A. C., Woehr, D. J., Thomas, J. S., Loughry, M. L., Ohland, M. W., & Ferguson, D. M. (2017). Facilitating Peer Evaluation in Team Contexts: The Impact of Frame-of-Reference Rater Training. *Academy of Management Learning & Education*, *16*(4), 562–578. https://doi.org/10.5465/amle.2016.0163

Lopes, V. P., Rodrigues, L. P., Maia, J. a. R., & Malina, R. M. (2011). Motor coordination as predictor of physical activity in childhood. *Scandinavian Journal of Medicine & Science in Sports*, *21*(5), 663–669. https://doi.org/10.1111/j.1600-0838.2009.01027.x

Lubans, D. R., Morgan, P. J., Cliff, D. P., Barnett, L. M., & Okely, A. D. (2010). Fundamental Movement Skills in Children and Adolescents. *Sports Medicine*, *40*(12), 1019–1035. https://doi.org/10.2165/11536850-000000000-00000

Lucas, B. R., Latimer, J., Doney, R., Ferreira, M. L., Adams, R., Hawkes, G., Fitzpatrick, J. P., Hand, M., Oscar, J., Carter, M., & Elliott, E. J. (2013). The Bruininks-Oseretsky Test of Motor Proficiency-Short Form is reliable in children living in remote Australian Aboriginal communities. *BMC Pediatrics*, *13*(1), 135. https://doi.org/10.1186/1471-2431-13-135

MacDonald, M., Lord, C., & Ulrich, D. (2013). The relationship of motor skills and adaptive behavior skills in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *7*(11), 1383–1390. https://doi.org/10.1016/j.rasd.2013.07.020

MacDonald, M., Ross, S., McIntyre, L. L., & Tepfer, A. (2017). Relations of Early Motor Skills on Age and Socialization, Communication, and Daily Living in Young Children With Developmental Disabilities. *Adapted Physical Activity Quarterly*, *34*(2), 179–194. https://doi.org/10.1123/apaq.2015-0091

Maeng, H., Webster, E. K., Pitchford, E. A., & Ulrich, D. A. (2017). Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition Among Experienced TGMD-2 Raters. *Adapted Physical Activity Quarterly*, *34*(4), 442–455. https://doi.org/10.1123/apaq.2016-0026

Magistro, D., Piumatti, G., Carlevaro, F., Sherar, L. B., Esliger, D. W., Bardaglio, G., Magno, F., Zecca, M., & Musella, G. (2018). Measurement Invariance of TGMD-3 in Children With and Without Mental and Behavioral Disorders. *PSYCHOLOGICAL ASSESSMENT*, *30*(11), 1421–1429. https://doi.org/10.1037/pas0000587

Mangal, S. K., & Mangal, U. (2009). *Essentials of Educational Technology*. PHI Learning Pvt. Ltd.

McGraw, K. O., & Wong, S. P. (1996). *Forming Inferences About Some Intraclass Correlation Coefficients*. *30*, 18.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, *69*(1), 147–156. https://doi.org/10.1037/0021-9010.69.1.147

Mohammadi, F., Bahram, A., Khalaji, H., Ulrich, D. A., & Ghadiri, F. (2019). Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. *Journal of Motor Learning & Development*, *7*(1), 106–121.

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, *14*(2), 129–135. https://doi.org/10.1016/j.iheduc.2010.10.001

Morgan, P. J., Barnett, L. M., Cliff, D. P., Okely, A. D., Scott, H. A., Cohen, K. E., & Lubans, D. R. (2013). Fundamental Movement Skill Interventions in Youth: A Systematic Review

and Meta-analysis. *Pediatrics*, *132*(5), e1361–e1383. https://doi.org/10.1542/peds.2013-1167

Morrow, R. L., Garland, E. J., Wright, J. M., Maclure, M., Taylor, S., & Dormuth, C. R. (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *CMAJ : Canadian Medical Association Journal*, *184*(7), 755–762. https://doi.org/10.1503/cmaj.111619

Motor Skill Assessment in Autism Spectrum Disorder: A Case Study—ProQuest. (n.d.). Retrieved March 22, 2020, from https://search.proquest.com/docview/1883071652?pq-origsite=gscholar

Movement Assessment Battery for Children | Second Edition. (n.d.). Retrieved March 21, 2020, from https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Motor-Sensory/Movement-Assessment-Battery-for-Children-%7C-Second-Edition/p/100000433.html

Movement Assessment Battery for Children (Kit, 1992) [WorldCat.org]. (n.d.). Retrieved March 21, 2020, from https://www.worldcat.org/title/movement-assessment-battery-for-children/oclc/221664826

Multivariate Associations Among Health-Related Fitness, Physical Activity, and TGMD-3 Test Items in Disadvantaged Children From Low-Income Families—Ryan Burns, Tim Brusseau, James Hannon, 2017. (n.d.). Retrieved March 21, 2020, from https://journals.sagepub.com/doi/abs/10.1177/0031512516672118?journalCode=pmsb

Musch__grondin_2001.pdf. (n.d.). Retrieved March 21, 2020, from http://www.cms.fss.ulaval.ca/upload/psy_nouveau/fichiers/musch__grondin_2001.pdf

Myford, C. M., & Wolfe, E. W. (n.d.). *Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I*. 38.

National PE Standards-Highly Effective Physical Education. (n.d.). Retrieved April 17, 2020, from https://www.shapeamerica.org/standards/pe/

NEWELL, K. (1986). Constraints on the development of coordination. *Motor Development in Children : Aspects of Coordination and Control*. https://ci.nii.ac.jp/naid/10016683819/

Newell, K. M., & Jordan, K. (2007). Task constraints and movement organization: A common language. In *Ecological task analysis and movement* (pp. 5–23). Human Kinetics.

Noonan, L. E., & Sulsky, L. M. (2001). Impact of Frame-of-Reference and Behavioral Observation Training on Alternative Training Effectiveness Criteria in a Canadian Military Sample. *Human Performance*, *14*(1), 3–26. https://doi.org/10.1207/S15327043HUP1401_02

Numerous Test Items in the Complete and Short Forms of the BOT-2 Do Not Contribute Substantially to Motor Performance Assessments in Typically Developing Children Six to Ten Years Old: Journal of Occupational Therapy, Schools, & Early Intervention: Vol 5, No 1. (n.d.). Retrieved March 21, 2020, from https://www.tandfonline.com/doi/abs/10.1080/19411243.2012.674746

Ohgi, S., Morita, S., Loo, K. K., & Mizuike, C. (2007). A Dynamical Systems Analysis of Spontaneous Movements in Newborn Infants. *Journal of Motor Behavior*, *39*(3), 203–214. https://doi.org/10.3200/JMBR.39.3.203-214

O'keeffe, S. L., Harrison, A. J., & Smyth, P. J. (2007). Transfer or specificity? An applied investigation into the relationship between fundamental overarm throwing and related

sport skills. *Physical Education and Sport Pedagogy*, *12*(2), 89–102. https://doi.org/10.1080/17408980701281995

Olsen, H., Donaldson, A. J., & Hudson, S. D. (2010). Online Professional Development: Choices for Early Childhood Educators. *Dimensions of Early Childhood*, *38*(1), 12–18.

Outcomes Research: Shifting the Dominant Research Paradigm in Physical Therapy | Physical Therapy | Oxford Academic. (n.d.). Retrieved March 20, 2020, from https://academic.oup.com/ptj/article-abstract/75/11/965/2632823

Palmer, K. K., & Brian, A. (2016). Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. *Journal of Motor Learning & Development*, *4*(2), 142–151.

Pan, C.-Y., Tsai, C.-L., & Chu, C.-H. (2009a). Fundamental Movement Skills in Children Diagnosed with Autism Spectrum Disorders and Attention Deficit Hyperactivity Disorder. *Journal of Autism & Developmental Disorders*, *39*(12), 1694–1705. eue.

Pan, C.-Y., Tsai, C.-L., & Chu, C.-H. (2009b). Fundamental Movement Skills in Children Diagnosed with Autism Spectrum Disorders and Attention Deficit Hyperactivity Disorder. *Journal of Autism and Developmental Disorders*, *39*(12), 1694. https://doi.org/10.1007/s10803-009-0813-5

Piaget, J. (n.d.). *THE ORIGINS OF INTELLIGENCE IN CHILDREN*. 442.

Piaget, J. (1954). *The construction of reality in the child* (pp. xiii, 386). Basic Books. https://doi.org/10.1037/11168-000

Piek, J. P., Baynam, G. B., & Barrett, N. C. (2006). The relationship between fine and gross motor ability, self-perceptions and self-worth in children and adolescents. *Human Movement Science*, *25*(1), 65–75. https://doi.org/10.1016/j.humov.2005.10.011

Piper, M. C. (1994). *Motor Assessment of the Developing Infant*. Saunders.

Provost, B., Lopez, B. R., & Heimerl, S. (2007). A Comparison of Motor Delays in Young Children: Autism Spectrum Disorder, Developmental Delay, and Developmental Concerns. *Journal of Autism and Developmental Disorders*, *37*(2), 321–328. https://doi.org/10.1007/s10803-006-0170-6

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, *69*(4), 581–588. https://doi.org/10.1037/0021-9010.69.4.581

Reduction of errors during practice facilitates fundamental movement skill learning in children with intellectual disabilities—Capio—2013—Journal of Intellectual Disability Research—Wiley Online Library. (n.d.). Retrieved March 19, 2020, from https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2788.2012.01535.x

Reliability and responsiveness of the Movement Assessment Battery for Children–Second Edition Test in children with developmental coordination disorder—WUANG - 2012—Developmental Medicine & Child Neurology—Wiley Online Library. (n.d.). Retrieved March 21, 2020, from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8749.2011.04177.x

Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City | Semantic Scholar. (n.d.). Retrieved March 21, 2020, from https://www.semanticscholar.org/paper/Reliability-and-validity-of-test-of-gross-motor-of-Farrokhi-Zadeh/c84a571482dcc2d778fd39967750de5c16cd5b97

Renshaw, I., Chow, J. Y., Davids, K., & Hammond, J. (2010). A constraints-led perspective to understanding skill acquisition and game play: A basis for integration of motor learning

theory and physical education praxis? *Physical Education & Sport Pedagogy*, *15*(2), 117–137. https://doi.org/10.1080/17408980902791586

Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition...: Georgia State University Library. (n.d.). Retrieved February 11, 2020, from https://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=f7362b39-1298-43a0-82ac-2d410c621567%40sessionmgr4008

Rintala et al. - 2017—Reliability Assessment of Scores From Video-Record.pdf. (n.d.). Retrieved March 22, 2020, from https://jyx.jyu.fi/bitstream/handle/123456789/54823/reliability%20ms%20resubmitted%20august%2026%202016.pdf?sequence=1

Rintala, P., & Loovis, E. M. (2013). Measuring Motor Skills in Finnish Children with Intellectual Disabilities. *Perceptual and Motor Skills*, *116*(1), 294–303. https://doi.org/10.2466/25.10.PMS.116.1.294-303

Rintala, P. O., Sääkslahti, A. K., & Iivonen, S. (2017). Reliability Assessment of Scores From Video-Recorded TGMD-3 Performances. *Journal of Motor Learning and Development*, *5*(1), 59–68. https://doi.org/10.1123/jmld.2016-0007

Roberton, M. A., Thompson, G., & Langendorfer, S. J. (2017). Initial steps in creating a developmentally valid tool for observing/assessing rope jumping. *Physical Education and Sport Pedagogy*, *22*(2), 187–196. https://doi.org/10.1080/17408989.2016.1165193

Robinson, L. E., Rudisill, M. E., & Goodway, J. D. (2009). Instructional Climates in Preschool Children Who Are At-Risk. Part II: Perceived Physical Competence. *Research Quarterly for Exercise & Sport*, *80*(3), 543–551. https://doi.org/10.5641/027013609X13088500159525

Robinson, L. E., Stodden, D. F., Barnett, L. M., Lopes, V. P., Logan, S. W., Rodrigues, L. P., & D'Hondt, E. (2015). Motor Competence and its Effect on Positive Developmental Trajectories of Health. *Sports Medicine*, *45*(9), 1273–1284. https://doi.org/10.1007/s40279-015-0351-6

Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: Recall, time and behavior observation training. *International Journal of Training and Development*, *7*(2), 93–107. https://doi.org/10.1111/1468-2419.00174

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*(2), 370–395. https://doi.org/10.1111/j.2044-8325.2011.02045.x

Rosales Sánchez, C., Díaz-Cabrera, D., & Hernández-Fernaud, E. (2019). Does effectiveness in performance appraisal improve with rater training? *PLOS ONE*, *14*(9), e0222694. https://doi.org/10.1371/journal.pone.0222694

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of Intellectual and Perceptual-Motor Skills. *Annual Review of Psychology*, *52*(1), 453–470. https://doi.org/10.1146/annurev.psych.52.1.453

Ross, S., Metcalf, A., Bulger, S. M., & Housner, L. D. (2014). Modified Delphi Investigation of Motor Development and Learning in Physical Education Teacher Education. *Research Quarterly for Exercise and Sport*, *85*(3), 316–329. https://doi.org/10.1080/02701367.2014.930087

Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test–retest reliability of continuous measurements. *Statistics in Medicine*, *21*(22), 3431–3446. https://doi.org/10.1002/sim.1253

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7

Rowe, G., & Wright, G. (2001). Expert Opinions in Forecasting: The Role of the Delphi Technique. In J. S. Armstrong (Ed.), *Principles of Forecasting* (Vol. 30, pp. 125–144). Springer US. https://doi.org/10.1007/978-0-306-47630-3_7

Russell, M., Carey, R., Kleiman, G., & Venable, J. D. (2019). FACE-TO-FACE AND ONLINE PROFESSIONAL DEVELOPMENT FOR MATHEMATICS TEACHERS: A COMPARATIVE STUDY. *Online Learning*, *13*(2). https://doi.org/10.24059/olj.v13i2.1669

Sackman, H. (1974). *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*. https://www.rand.org/pubs/reports/R1283.html

Santrock, J. (2009). *A Topical Approach to Lifespan Development*. McGraw-Hill Education.

Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-ofreference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 735–746.

Schmidt, R. A., & Lee, T. D. (2005). *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., & Zelaznik, H. N. (2018). *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics.

Schmidt, R. A., & Wrisberg, C. A. (2008a). *Motor Learning and Performance: A Situation-based Learning Approach*. Human Kinetics.

Schmidt, R. A., & Wrisberg, C. A. (2008b). *Motor Learning and Performance: A Situation-based Learning Approach*. Human Kinetics.

Schoemaker, M. M., Niemeijer, A. S., Flapper, B. C. T., & Smits-Engelsman, B. C. M. (2012). Validity and reliability of the Movement Assessment Battery for Children-2 Checklist for children with and without motor impairments. *Developmental Medicine & Child Neurology*, *54*(4), 368–375. https://doi.org/10.1111/j.1469-8749.2012.04226.x

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, *7*(3), 301–317. https://doi.org/10.1177/096228029800700306

Shrout, P. E., & Fleiss, J. L. (1979). *Intraclass Correlations: Uses in Assessing Rater Reliability*. *86*(2), 420.

Simons, J., Daly, D., Theodorou, F., Caron, C., Simons, J., & Andoniadou, E. (2008). Validity and Reliability of the TGMD-2 in 7–10-Year-Old Flemish Children with Intellectual Disability. *Adapted Physical Activity Quarterly*, *25*(1), 71–82. https://doi.org/10.1123/apaq.25.1.71

Simons, J., & Eyitayo, G. B. (2016). Aspects of reliability and validity of the TGMD-3 in 7-10 year old children with intellectual disability in Belgium. *European Psychomotricity Journal*, *8*, 3–16.

Skinner, R. A., & Piek, J. P. (2001). Psychosocial implications of poor motor coordination in children and adolescents. *Human Movement Science*, *20*(1), 73–94. https://doi.org/10.1016/S0167-9457(01)00029-X

Smith, D. E. (1986). Training Programs for Performance Appraisal: A Review. *The Academy of Management Review*, *11*(1), 22–40. JSTOR. https://doi.org/10.2307/258329

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6

Smits-Engelsman, B. C. M., Niemeijer, A. S., & van Waelvelde, H. (2011). Is the Movement Assessment Battery for Children-2nd edition a reliable instrument to measure motor performance in 3 year old children? *Research in Developmental Disabilities*, *32*(4), 1370–1377. https://doi.org/10.1016/j.ridd.2011.01.031

Spironello, C., Hay, J., Missiuna, C., Faught, B. E., & Cairney, J. (2010). Concurrent and construct validation of the short form of the Bruininks-Oseretsky Test of Motor Proficiency and the Movement-ABC when administered under field conditions: Implications for screening. *Child: Care, Health and Development*, *36*(4), 499–507. https://doi.org/10.1111/j.1365-2214.2009.01066.x

Spodek, B., & Saracho, O. N. (2014). *Handbook of Research on the Education of Young Children*. Routledge.

Spool, M. D. (1978). Training Programs for Observers of Behavior: A Review1. *Personnel Psychology*, *31*(4), 853–888. https://doi.org/10.1111/j.1744-6570.1978.tb02128.x

Staples, K. L., & Reid, G. (2010). Fundamental Movement Skills and Autism Spectrum Disorders. *Journal of Autism & Developmental Disorders*, *40*(2), 209–217. https://doi.org/10.1007/s10803-009-0854-9

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A Developmental Perspective on the Role of Motor Skill

Competence in Physical Activity: An Emergent Relationship. *Quest*, *60*(2), 290–306. https://doi.org/10.1080/00336297.2008.10483582

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and Measurement of Performance Rating Accuracy: Some Methodological and Theoretical Concerns. *Journal of Applied Psychology*, *73*(3), 497–506. https://doi.org/10.1037/0021-9010.73.3.497

Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, *77*(4), 501–510. https://doi.org/10.1037/0021-9010.77.4.501

Taliaferro, A. R., & Bulger, S. M. (2020). A Delphi Study of Effective Adapted Physical Education Practicum Experiences. *Adapted Physical Activity Quarterly*, *37*(1), 20–40.

The Gross Motor Skills of Children with Mild Learning Disabilities. (n.d.). Retrieved March 21, 2020, from https://espace.curtin.edu.au/bitstream/handle/20.500.11937/21080/200108_200108.pdf?sequence=2&isAllowed=y

The Standards for Educational and Psychological Testing. (n.d.). Https://Www.Apa.Org. Retrieved March 22, 2020, from https://www.apa.org/science/programs/testing/standards

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Thelen, E., & Ulrich, B. D. (1991). *Hidden Skills: A Dynamic Systems Analysis of Treadmill Stepping During the First Year*. University of Chicago Press.

Thornton, G. C., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, *65*(3), 351–354. https://doi.org/10.1037/0021-9010.65.3.351

Ting Liu, Breslin, C. M., & ElGarhy, S. (2017). Motor Skill Assessment in Autism Spectrum Disorder: A Case Study. *Physical Educator*, *74*(2), 239–254. https://doi.org/10.18666/TPE-2017-V74-I2-7148

Urbina, S. (2014). *Essentials of Psychological Testing*. John Wiley & Sons.

Valentini, N. C., Ramalho, M. H., & Oliveira, M. A. (2014). Movement Assessment Battery for Children-2: Translation, reliability, and validity for Brazilian children. *Research in Developmental Disabilities*, *35*(3), 733–740. https://doi.org/10.1016/j.ridd.2013.10.028

Valentini, N. C., Zanella, L. W., & Webster, E. K. (2017). Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. *Journal of Motor Learning and Development*, *5*(1), 15–28. https://doi.org/10.1123/jmld.2016-0002

Veldhuizen, S., Rivard, L., & Cairney, J. (2017). Relative age effects in the Movement Assessment Battery for Children-2: Age banding and scoring errors. *Child: Care, Health and Development*, *43*(5), 752–757. https://doi.org/10.1111/cch.12459

Vernadakis, N., Papastergiou, M., Zetou, E., & Antoniou, P. (2015). The impact of an exergame-based intervention on children's fundamental motor skills. *Computers & Education*, *83*, 90–102. https://doi.org/10.1016/j.compedu.2015.01.001

Vernazza-Martin, S., Martin, N., Vernazza, A., Lepellec-Muller, A., Rufo, M., Massion, J., & Assaiante, C. (2005). Goal Directed Locomotion and Balance Control in Autistic Children. *Journal of Autism and Developmental Disorders*, *35*(1), 91–102. https://doi.org/10.1007/s10803-004-1037-3

Wagner, M. O., Webster, E. K., & Ulrich, D. A. (2017). Psychometric properties of the Test of Gross Motor Development, third edition (German translation): Results of a pilot study.

Journal of Motor Learning and Development, *5*(1), 29–44. https://doi.org/10.1123/jmld.2016-0006

Warner, R. Z., & Hewett, B. L. (2017). Technical Communication Coaching: A Strategy for Instilling Reader Usability Assurance in Online Course Material Development. *Technical Communication Quarterly*, *26*(3), 300–313. https://doi.org/10.1080/10572252.2017.1339493

Webster and Ulrich—2017—Evaluation of the Psychometric Properties of the T.pdf. (n.d.). Retrieved March 21, 2020, from https://www.researchgate.net/profile/Dale_Ulrich/publication/312232404_Evaluation_of_the_Psychometric_Properties_of_the_Test_of_Gross_Motor_Development_-_3rd_Edition/links/5b645a2aa6fdcc45b30d5d2a/Evaluation-of-the-Psychometric-Properties-of-the-Test-of-Gross-Motor-Development-3rd-Edition.pdf

Welcome to CDC stacks |. (n.d.). Retrieved March 22, 2020, from https://stacks.cdc.gov/view/cdc/22182

Werner, J. M., & Bolino, M. C. (1997). Explaining U.s. Courts of Appeals Decisions Involving Performance Appraisal: Accuracy, Fairness, and Validation. *Personnel Psychology*, *50*(1), 1–24. https://doi.org/10.1111/j.1744-6570.1997.tb00898.x

Westendorp, M., Houwen, S., Hartman, E., & Visscher, C. (2011). Are gross motor skills and sports participation related in children with intellectual disabilities? *Research in Developmental Disabilities*, *32*(3), 1147–1153. https://doi.org/10.1016/j.ridd.2011.01.009

Wiersma, L. D. (2000). Risks and Benefits of Youth Sport Specialization: Perspectives and Recommendations. *Pediatric Exercise Science*, *12*(1), 13–22. https://doi.org/10.1123/pes.12.1.13

Winnick, J., & Porretta, D. L. (2016). *Adapted Physical Education and Sport*. Human Kinetics.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*(3), 189–205. https://doi.org/10.1111/j.2044-8325.1994.tb00562.x

Wolraich, M. (2008). *Developmental-behavioral Pediatrics: Evidence and Practice*. Elsevier Health Sciences.

Wuang, Y.-P., & Su, C.-Y. (2009). Reliability and responsiveness of the Bruininks–Oseretsky Test of Motor Proficiency-Second Edition in children with intellectual disability. *Research in Developmental Disabilities*, *30*(5), 847–855. https://doi.org/10.1016/j.ridd.2008.12.002

APPENDICES

Appendix A

Training Content of Module 1 (1stscoring)

Topic	Content	Time (mins)
Introduction TGMD-3 (Ulrich, 2019)	What is the TGMD-3 and its utilization (with narration)	3
	Components of the TGMD-3 (https://www.youtube.com/watch?v=9WggHyZpXl0)	4
	How to score using the TGMD-3 record form and its multimedia platform (with narration)	8
	Total	15

Appendix B

Training Content of Module 2 (1st and 3rd scoring)

Topic	Content	Time (mins)
Introduction to developmental disabilities (DD)	What is DD (ppt slides with narration)	4
	Behavioral & psychomotor characteristics of children with DD (ppt slides with narration)	5
Correct scoring and practical applications on the TGMD-3 among children with DD	Review of scoring 1 item on locomotor skills (i.e., run)	4
	Review of scoring 1 item on ball skills (i.e., two-hand strike)	6
	Total	19

Appendix C

Part #1: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 1

Thank you for taking the time to review and provide feedback on the video training modules developed to train novice raters to score the TGMD-3 on students with developmental disabilities. These modules were developed using the frame-of-reference training criteria. Frame-of-reference training was developed to familiarize raters with identifying correct performance dimensions and enables raters to systematically compare information regarding actual versus desired performance using training content with the presentation of sample performance based on correct performance dimension.

I am seeking your feedback on the degree to which:

1. The content in the video training modules align with the criteria listed in the left most column for a fame-of-reference training. Please rate the degree of alignment on a scale from 1 = very poor alignment, to 5 = very strong alignment.
2. The effectiveness of the content in the training videos to prepare novice raters to score the TGMD-3 on a child with a developmental disability. Please rate the quality of the content on a scale from 1 = very poor content representation, to 5 = very strong content representation.

Please provide comments or suggestions to improve the videos for any ratings of 3 or below.

Appendix C (cont.…)

Part #1: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 1

	Frame of refence training criteria	Training module content	Frame of reference training criteria alignment	Effectiveness of video training content	Comment
Frame-of-reference training	Identify correct performance dimension under natural circumstances	Module 1 introduction to the TGMD-3 and the scope of application	1 2 3 4 5	1 2 3 4 5	
		Module 1 demonstrates 6 locomotor skills on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
		Module 1 demonstrates 7 ball skills on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
	Provides information to score the variables in question	Module 1 explains how to score FMS on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
	Describes correct performance on variables in question	Module 2 demonstrates accurate performance of run skill on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
		Module 2 demonstrates accurate performance of two-hand strike on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
	Systematically compare information regarding actual versus desired performance	Module 2 shows incorrect performance of each skill among children with DD and explains why it does not meet the performance criteria	1 2 3 4 5	1 2 3 4 5	

Appendix C (Cont.…)

Part #1: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 1

	Component	Training module description	Frame of reference training criteria alignment	Effectiveness of training content	Comment
Frame-of-reference training	Provides information according to characteristics of performer	Module 2 introduces general characteristics of DD	1 2 3 4 5	1 2 3 4 5	
		Module 2 presents sufficient behavioral characteristics of DD	1 2 3 4 5	1 2 3 4 5	
	Provides correct scoring feedback on actual performance	Module 2 explains behaviors you may see that will help you when you score the TGMD-3.	1 2 3 4 5	1 2 3 4 5	

Appendix C (Cont.…)

Part #2: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 1

This evaluation questionnaire is designed to determine the degree of alignment between the video training modules and criteria for effective online training. Please provide feedback on the degree to which the content in the video modules align with the components for online training using a scale of 1 = very poor alignment to 5 = very strong alignment. Please provide feedback to improve the online training videos for scores rated 3 or below.

	Criteria for online training	Degree of alignment between training module and online training criteria	Comment
Online training	Training program is easy to access	1 2 3 4 5	
	Utilizes visual resources (angle of camera, angle of performance front and side view)	1 2 3 4 5	
	Voice narration is clear and understandable	1 2 3 4 5	
	Voice narration is at an appropriate pace	1 2 3 4 5	
	Tests are available on online	1 2 3 4 5	
	Written material in video is easy to read	1 2 3 4 5	

Appendix D

Part #1: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 2

	Frame of refence training criteria	Training module content	Frame of reference training criteria alignment	Effectiveness of video training content	Comment
Frame-of-reference training	Identify correct performance dimension under natural circumstances	Module 1 introduction to the TGMD-3 and the scope of application	1 2 3 4 5	1 2 3 4 5	
		Module 1 demonstrates 6 locomotor skills on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
		Module 1 demonstrates 7 ball skills on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
	Provides information to score the variables in question	Module 1 explains how to score FMS on the TGMD-3	1 2 3 4 5	1 2 3 4 5	
	Provides information about developmental disabilities	Introduce developmental disability (DD)	1 2 3 4 5	1 2 3 4 5	
		Introduce motor performance of children with DD	1 2 3 4 5	1 2 3 4 5	
		Use video to effectively illustrate motor delay in children with DD	1 2 3 4 5	1 2 3 4 5	

Appendix D (cont…)

Part #1: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 2

	Frame of refence training criteria	Training module content	Frame of reference training criteria alignment	Effectiveness of video training content	Comment
Frame-of-reference training	Provides information to score the run skill of a child with DD	Discuss scoring of the run skill	1 2 3 4 5	1 2 3 4 5	
		Shows good and bad performance of the run skill	1 2 3 4 5	1 2 3 4 5	
		Provided scoring feedback on the run skill	1 2 3 4 5	1 2 3 4 5	
	Provides information to score the two-hand strike skill of a child with DD	Discuss scoring of the run skill	1 2 3 4 5	1 2 3 4 5	
		Shows good and bad performance of the run skill	1 2 3 4 5	1 2 3 4 5	
		Provided scoring feedback on the run skill	1 2 3 4 5	1 2 3 4 5	

Appendix D (Cont.…)

Part #2: Expert Evaluation Form for Content Validation of Online Frame-of-Reference Training Module in Round 2

	Criteria for online training	Degree of alignment between training module and online training criteria	Comment
Online training	Training program is easy to access	1 2 3 4 5	
	Utilizes visual resources (angle of camera, angle of performance front and side view)	1 2 3 4 5	
	Voice narration is clear and understandable	1 2 3 4 5	
	Voice narration is at an appropriate pace	1 2 3 4 5	
	Written material in video is easy to read	1 2 3 4 5	

Appendix E

Demographic Questionnaire of Novice Rater

1. Age		2. Gender	☐ male ☐ female ☐ other
3. Race	☐ White or Caucasian ☐ Black or African American ☐ Asian Native ☐ Hawaiian or Pacific Islander ☐ Hispanic or Latino ☐ Multiracial ☐ Other		
4. Affiliation	☐ Health and physical education ☐ Other		
5. Major (if 'Other' above)			
6. Grade	☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior ☐ Graduate program		
7. Have you taken curriculum in college related to fundamental movement skills in children?			Yes / No
8. Have you had experience scoring the Test of Gross Motor Development (TGMD)?			Yes / No

Appendix F

Test of Module 1_1st ver.

Test of Module 1 is to check for understanding of the TGMD-3. Raters will be requested to take the test to pass the module 1 training session. The test consists of 10 questions about the content of the module 1. This test has the following learning outcomes.

1. The rater understands the TGMD-3 and its components
2. The raters can identify the skills in the locomotor subtest
3. The raters can identify the skills in the ball skills subtest
4. The raters understand how to score each criterion within a skill

Please evaluate the test questions for module 1 for their (a) clarity and understanding; (b) appropriateness of answer response options; (c) meaningfulness of question to assess content knowledge, and (d) degree to which the test questions align with the learning outcome. Each category has the criteria of 1 = very poor alignment to 5 = very strong alignment. Please provide feedback for responses rated 3 or below.

Appendix F (cont.…)

Test of module 1_1st ver.

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
1	Q	The Test of Gross Motor Development-3 (TGMD-3) is an assessment to measure _____ of children.	1. The rater understands the TGMD-3 and its components	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) physical fitness; (b) FMS*; (c) body components						
2	Q	The TGMD-3 consists of ____ fundamental movement skills.	1. The rater understands the TGMD-3 and its components	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 2; (b) 7; (c) 12; (d) 13*						

*****Correct answer

Appendix F (cont.…)

Test of module 1_1st ver.

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
3	Q	Locomotor subtest on the TGMD-3 includes the skills except _____.	2. The raters can identify the skills in the locomotor subtest	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) slide; (b) leap*; (c) skip; (d) gallop						
4	Q	Scoring values on the TGMD-3 is ___ or ___.	4. The raters understand how to score each criterion within a skill	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 0, 1*; (b) 1, 2; (c) 0, 2						
5	Q	Ball skill subtest on the TGMD-3 includes the skills except _____.	3. The raters can identify the skills in the ball skills subtest	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) underhand throw; (b) one-hand strike; (c) underhand roll*; (d) two-hand catch						

*****Correct answer

Appendix F (cont.…)

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
6	Q	All skills have the same number of performance criteria on the TGMD-3.	1. The rater understands the TGMD-3 and its components	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) true; (b) false*						
7	Q	The two-hand strike skill on the TGMD-3 is like a _____ skill.	3. The raters can identify the skills in the ball skills subtest	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) basketball; (b) tennis; (c) baseball*; (d) badminton						
8	Q	How many trials are assessed for scoring each skill?	4. The raters understand how to score each criterion within a skill	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 1; (b) 2*; (c) 3; (d) 4						

Test of Module 1_1st ver.

***Correct answer

Appendix F (cont.…)

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
9	Q	How many skills are in the locomotor subtest on the TGMD-3?	2. The raters can identify the skills in the locomotor subtest	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 2; (b) 6*; (c) 7; (d) 13						
10	Q	When a child performed incorrectly, score the value '___'.	4. The raters understand how to score each criterion within a skill	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 0*; (b) 1						

Test of module 1_1st ver.

*Correct answer

Appendix G

Test of Module 1_Final ver.

No	Question / Answer	
1	Q	The Test of Gross Motor Development-3 (TGMD-3) is an assessment to measure _____ of children.
	A	(a) physical fitness; (b) fundamental motor skills*; (c) body components
2	Q	How many skills are on locomotor and ball skills of the TGMD-3 respectively?
	A	(a) 1, 1; (b) 5, 8; (c) 4, 4; (d) 6, 7*
3	Q	What is the skill score of the two-hand catch if a child correctly performed all three performance criteria of the skill in two trials?
	A	(a) 1; (b) 2; (c) 3; (d) 6*
4	Q	Which of the following skills is not a locomotor skill on the TGMD-3?
	A	(a) slide; (b) kick*; (c) skip; (d) gallop
5	Q	The skill score on the TGMD-3 is _____.
	A	(a) an average of performance criteria scores across two trials (b) the sum of the scores form trial one and trial two* (c) the score from the trial with highest score
6	Q	Which of the following ball skills is not on the TGMD-3?
	A	(a) underhand throw; (b) one-hand strike; (c) chest pass*; (d) two-hand catch

*Correct answer

Appendix G (con…)

Test of Module 1_Final ver.

No	Question / Answer	
7	Q	On the first trial, if a student performs a performance criterion correctly for part of the required distance (e.g., run) but then changes to a different skill (e.g., skip) in the middle. How would this trial be scored?
	A	(a) 0; (b) 1*
8	Q	Scoring values for each performance criterion on the TGMD-3 are ___ or ___.
	A	(a) 0, 1*; (b) 1, 2; (c) 0, 2
9	Q	A teacher observes and scores a student perform ___ trials for each skill.
	A	(a) 1; (b) 2*; (c) 3; (d) 4
10	Q	When a child performs a performance criterion correctly, they receive a score of ___.
	A	(a) 0; (b) 1*; (c) 2

*Correct answer

Appendix H

Test of Module 2_1st ver.

Test of module 2 is to check the understanding of developmental disability (DD) how to score the TGMD for children with DD. Raters will be requested to take the test to pass the module 2 training session. The test consists of 10 questions about the content of the module 2. This test has the following learning outcomes.

1. The rater can explain about developmental disability (DD)
2. The rater can list behavior and movement characteristics of children with DD
3. The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3
4. The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3
5. The raters can score FMS among children with DD according to each criterion on the TGMD-3

Please evaluate the test questions for module 2 for their (a) clarity and understanding; (b) appropriateness of answer response options; (c) meaningfulness of question to assess content knowledge, and (d) degree to which question aligns with learning outcome. Each category has the criteria of 1 = very poor alignment to 5 = very strong alignment. Please provide feedback for responses rated 3 or below.

Appendix H (cont.…)

Test of Module 2_1st ver.

No		Question / Answer	Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
1	Q	Scoring FMS on the TGMD-3 among children with DD should consider their characteristics compared to typically developing children.	5. The raters can score FMS among children with DD according to each criterion on the TGMD-3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) true; (b) false*						
2	Q	Developmental disability includes _____.	1. The rater can explain about DD	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) physical disability; (b) psychological disability; (c) both*						

*Correct answer

Appendix H (cont.…)

Test of Module 2_1st ver.

No		Question / Answer	Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
3	Q	Fundamental movement skills of children with DD have to be scored regardless of their performance characteristics.	5. The raters can score FMS among children with DD according to each criterion on the TGMD-3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) true*; (b) false						
4	Q	Raters can score on the TGMD-3 considering performance levels of a child with DD.	5. The raters can score FMS among children with DD according to each criterion on the TGMD-3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) true; (b) false*						
5	Q	What score will be given if a child with DD performed the run skill in which arms move in opposition to legs with one elbow bent?	3 & 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 0*; (b) 1; (c) N/A						

*Correct answer

Appendix H (cont.…)

Test of Module 2_1st ver.

No		Question / Answer	Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
6	Q	What is correct score on the performance criteria of trial 1 of two-hand strike skill below when a child with DD performed it on the video? (Scenario: The child's preferred hand grips bat above non-preferred hand. And child's non-preferred hip/shoulder faces straight ahead. Hip and shoulder derotate during swing. Both feet are fixed on the ground during swing. Hits ball sending it straight ahead)	2 & 4	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 1-1-0-0-0; (b) 1-1-0-1-1; (c) 1-1-1-0-1; (d) 1-1-0-1-0; (e) 1-1-0-0-1*; (f) 1-1-1-1-0						
	Performance criteria of two hand strike	1	Child's preferred hand grips bat above non-preferred hand					
		2	Child's non-preferred hip/shoulder faces straight ahead					
		3	Hip and shoulder rotate and derotate during swing					
		4	Steps with non-preferred foot					
		5	Hits ball sending it straight ahead					

*****Correct answer

Appendix H (cont.…)

No		Question / Answer	Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
7	Q	(Scenario) A child with DD was way off the running path in the middle of performing the run skill well until half of the requested distance (30/60 feet). What total score will be given on the 4 performance criteria of trial 1?	2 & 3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 4; (b) 2; (c) 1; (d) 0*						
8	Q	Which is the performance criterion of the run skill?	3. The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) Arms flexed and swinging forward (b) Arms flex and swing forward to produce force (c) Arms are flexed and move in opposition to legs to produce force (d) Arms move in opposition to legs with elbows bent*						

Test of Module 2_1st ver.

*Correct answer

Appendix H (cont.…)

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
9	Q	Which is not the performance criteria of two-hand strike?	4. The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) Rotates hip and shoulder to a point where the non-throwing side faces the wall* (b) Child's non-preferred hip/shoulder faces straight ahead (c) Hip and shoulder rotate and derotate during swing (d) Hits ball sending it straight ahead						

Test of Module 2_1st ver.

*Correct answer

Appendix H (cont…)

No	Question / Answer		Alignment with Learning Outcomes	Clarity and understanding	Appropriateness of Answer response options	Meaningfulness of question to assess content knowledge	Degree to which question aligns with learning outcome	Feedback for ratings of 3 or below
10	Q	What is correct score on the performance criteria of trial 1 of two-hand strike skill below when a child with DD performed it on the video? (Scenario: The child's preferred hand grips bat above non-preferred hand. And child's non-preferred hip/shoulder faces straight ahead. Hip and shoulder rotate during swing. Steps with preferred foot. Hits ball sending it straight to the ground)	2 & 4	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
	A	(a) 1-1-1-1-0; (b) 1-1-1-0-1; (c) 1-1-0-1-1; (d) 1-1-0-1-0; (e) 1-1-1-0-0; (f) 1-1-0-0-0*						
	Performance criteria of two hand strike	1	Child's preferred hand grips bat above non-preferred hand					
		2	Child's non-preferred hip/shoulder faces straight ahead					
		3	Hip and shoulder rotate and derotate during swing					
		4	Steps with non-preferred foot					
		5	Hits ball sending it straight ahead					

Test of Module 2_1st ver.

*Correct answer

Appendix I

Test of Module 2_Final ver.

No		Question / Answer		
1	Q	Module 2 presented two representative disabilities among developmental disabilities, such as intellectual disability (ID) and autism spectrum disorder (ASD). Which of the following statements is **correct** regarding ID or ASD?		
	A	(a) Children with ID and ASD may have delayed motor performance (b) ID presents before a child is 18 years old (c) Children with ASD shows limitations in social interactio (d) All of above*		
2	Q	Which of the following statements is a correct example regarding possible movement characteristics among children with DD?		
	A	(a) Poor coordination of arm and leg movements when throwing (b) Poor hand-eye coordination when dribbling (c) Poor motor planning to tun and jump rope (d) All of above*		
3	Q	After watching this video clip, what would be the correct score for the given performance criteria for the two-hand strike skill? Note: You can watch as many times as you want, click the replay arrow on the bottom left corner next to the volume button.		
		Performance criteria of the two-hand strike skill	1	Child's preferred hand grips bat above non-preferred hand
			2	Hip and shoulder rotate and derotate during swing
			3	Hits ball sending it straight ahead
	A	(a)1-1-1*; (b) 0-0-0; (c) 1-0-0		
4	Q	Which statement accurately reflects the arm movement in the run skill?		
	A	(a) Arms move in opposition to legs with elbows bent* (b) Arms flex and swing forward to produce force (c) Arms are extended and move side to side to produce force		

***** correct answer

Appendix I (cont…)

Test of Module 2_Final ver.

No	Question / Answer			
5	Q	After watching this video clip, what would be the correct score for the given performance criteria for the run skill? Note: You can watch as many times as you want, click the replay arrow on the bottom left corner next to the volume button.		
		Performance criteria of run skill	1	Arms move in opposition to legs with elbows bent
			2	Brief period where both feet are off the surface
			3	Narrow foot placement landing on heel or toes (not flat-footed)
	A	(a) 0-1-1*; (b) 0-0-0; (c) 1-1-1		
6	Q	What is ‘Derotate’ in the two-hand strike skill on the TGMD-3?		
	A	(a) Making a backswing (b) Facing hip/shoulder straight ahead (c) Stopping the rotation after the follow through when striking the ball* (d) Stepping with non-preferred foot		
7	Q	After watching this video clip, what would be the correct score for the given performance criteria for the two-hand strike skill? Note: You can watch as many times as you want, click the replay arrow on the bottom left corner next to the volume button.		
		Performance criteria of the two-hand strike skill	1	Child’s preferred hand grips bat above non-preferred hand
			2	Hip and shoulder rotate and derotate during swing
			3	Steps with non-preferred foot
	A	(a)1-1-1; (b) 1-1-0; (c) 0-0-1*		

***** correct answer

Appendix I (cont…)

Test of Module 2_Final ver.

No	Question / Answer			
8	Q	Which statement does not describes the movement when performing the two-hand strike skill?		
	A	(a) Child's non-preferred hip/shoulder faces straight ahead (b) Hip and shoulder rotate and derotate during swing (c) Contacts ball with one hand at about waist level* (d) All of above		
9	Q	After watching this video clip, what would be the correct score for the performance criterion for the run skill? Note: You can watch as many times as you want, click the replay arrow on the bottom left corner next to the volume button.		
		Performance criteria of the run skill	1	Arms move in opposition to legs with elbows bent
	A	(a) 1; (b) 0*; (c) 2		
10	Q	A child with DD engages in repetitive stereotypical movements such as hand flapping while running. How should you score this student?		
	A	(a) Ignore the behavior and assume they can run with arms in opposition when the hand flapping stops and give a score of 1 (b) Score the performance as a 0 since you did not see them perform using the correct arm movements*		

***** correct answer

Appendix J

Scoring Form of the Run and Two-hand Strike Skills on the TGMD-3

Skill	Directions	Performance Criteria	Trial 1	Trial 2	Score
1. Run	Place two cones 50 feet (15.2 meters) apart. Make sure there is at least 8–10 feet (2.4–3.1 meters) of space beyond the cone for a safe stopping distance. Tell the child to run fast from one cone to the other cone when you say, "Go." Repeat a second trial.	1. Arms move in opposition to legs with elbows bent			
		2. Brief period where both feet are off the surface			
		3. Narrow foot placement landing on heel or toes (not flat-footed)			
		4. Non-support leg bent about 90 degrees so foot is close to buttocks			
			Skill Score		
2. Two-hand strike of a stationary ball	Place ball on batting tee at child's waist level. Tell child to hit the ball hard, straight ahead. Point straight ahead. Repeat a second trial.	1. Child's preferred hand grips bat above non-preferred hand			
		2. Child's non-preferred hip/shoulder faces straight ahead			
		3. Hip and shoulder rotate and derotate during swing			
		4. Steps with non-preferred foot			
		5. Hits ball sending it straight ahead			
			Skill Score		

Appendix K. Expert Feedback & Response to Module 1 in Round 1

Questions for Module 1		Reviewer # & Comment	Response
1_1	Introduce to the TGMD-3 and the scope		
1_2	Introduce 6 locomotor skills	#2- Dr. Ulrich's videos provide a nice example of demonstration, but the actual criteria of each skill are overlooked in this training.	The purpose of module 1 is not to demonstrate each criterion of the skills but to introduce what skills are on the TGMD-3.
1_3	Introduce 7 ball skills	#2- Dr. Ulrich's videos provide a nice example of demonstration, but the actual criteria of each skill are overlooked in this training.	It is the same as above.
1_4	Introduce scoring	#2- This section was good, but it was long and more repetitive than needed.	Repetitiveness of the module subjects in the scoring section has been reduced. This helped to the length the module.
1_5	Additional comment	#2- In my opinion, the biggest gap is how to score the 13 skills. The logistics of scoring are well covered, but the biggest hurdle for raters is learning what constitutes each performance criteria.	Module 1 is to introduce what the TGMD-3 and what skills are on it rather than explain each performance criterion for raters.
		#3- Approx 1:35- movement vs motor (back to motor at 5:36). No definition of motor skill provided What does criterion-based scoring system mean? Provide a definition. What is the purpose of the skill demonstrations? Is this to show people how to administer or just to familiarize them with the skills? If it is showing people how to administer, I think you need to also think about verbal cues/prompts. This is a unique opportunity to push for normalized or standardized verbal cues. Small detail but I would make sure the captions are in a sans serif font- cleaner and easier to read.	Movement was used in the terminology 'Fundamental movement skill', meanwhile motor was used int the terminology 'Gross motor skill'. The purpose of this module is not to administrate TGMD-3 but to be familiar with skills on the TGMD-3.
		#3- Approx 6:00- you mention process and product- who is this training aimed for- will they difference between process and product be clear to them?	Added explanation about product and process-oriented criteria on the slide.
		#3- 13:22- says "two-hand strike" when we are looking at the catch. Make sure to change audio and subtitles here.	Revised.
		#4- It was clear and concise.	
		#6- I thought the module was clear - a couple notes: check for spelling and repeated words in the first few slides; consider quickly explaining what locomotor and ball skills are/mean before listing them; For scoring, it may be worth explaining what process and product-oriented means (assuming 'novice' raters may not be clear on what that is) or removing if not relevant for them	Added explanation about product and process-oriented criteria on the slide.

Appendix K (cont…). Expert Feedback & Response to Module 1 in Round 1

Questions for Module 1		Reviewer # & Comment	Response
2_1	LO#1: The rater understands the TGMD-3 and its components		
2_2	LO#2: The raters can identify the skills in the locomotor subtest	#2- The learning can identify the six skills, but is not introduced to the performance criteria	At the beginning, an overview has been added to explain the purpose of module 1 and gave a scoring example of two-hand catch in the performance criteria on the TGMD-3.
		#6- I think clarifying what locomotor means (e.g., moving from A to B) would help raters identify skills if they weren't listed in front of them	Locomotor is identified the definition and skills in the subtest on 2'58" of the module.
2_3	LO#3: The raters can identify the skills in the ball skills subtest	#2- The learning can identify the seven skills, but is not introduced to the performance criteria	Ball skill is identified the definition and skills in the subtest on 3'15" of the module.
		#6- Same comment as above	
2_4	LO#4: The raters understand how to score each criterion within a skill	#2- While the logistics of scoring are addressed, how to actually score each skill is not covered.	The way to correctly score two skills on the TGMD-3 has been presented on the Module 2.
		#3- I like the example of how to score the catch, but you may consider also scoring a continuous LM skill or at least try to discuss how to score those. For example, in the run- the child cannot run flat footed- is this for the whole run or just 50% of the run?	The scoring a continuous locomotor skill like running is introduced in module 2 where the specific criteria are highlighted, and specific scoring procedures addressed.
		#6- Yes and no - From this module, I think raters would understand how to score in general... this part was very clear. But they may not know exactly how to score each criterion (I would not assume a novice rater would know what to look for in an actual assessment video) just based on this. Hope that makes sense.	Module 1 is to introduce what the TGMD-3 and what skills are on it rather than explain each performance criterion for raters.
2_5	Additional comment	#7- I suggest adding a "freeze frame" or still shot from the demonstration video to highlight the specific performance criteria for each skill. Additionally, I would suggest that you produce video demonstrations using a stationary or non-moving (non-panning) camera, in a vantage point identical to the skill illustrations found in the TGMD-3 examiner's manual.	Module 1 is to introduce what the TGMD-3 and what skills are on it rather than explain each performance criterion for raters.
		#2- In addition to this introductory module, novice raters would benefit from a task analysis perspective of each skill with examples and opportunities for practice	The focus of module 2 is on scoring. This suggestion would be great for a module 3 which is to introduce and teach about how to score two of the 13 locomotor skills as a pilot study.

Appendix L. Expert Feedback & Response to Module 2 in Round 1

Questions for Module 2		Reviewer # & Comment	Response
3_1	Introduce developmental disability (DD)	#8- In my opinion, I would not use a video of a student striking another student (Aggression example). I also would not suggest using the "self-injurious" example because it looks like the teacher is dragging the student by the arms after the child flopped.	Of course, every child with DD does not have aggression. However, I would like to explain about behavioral characteristics among children with DD which could help novice raters to understand challenging behaviors of them during assessing the TGMD-3 as well as teaching in the class. Regarding self-injurious behavior has been revised to cut a teacher's controversial performance to the child with DD.
3_2	Introduce motor performance of children with DD	#6- just a note/consideration that some (often including people with DD) do not prefer "typically-developing"; also I see why you chose to describe movements as 'unexpected' I was wondering if there was a different word to get the point across.	The word 'unexpected' has been revised to 'wandering' to describe the challenging behavior of a child with DD on assessment.
3_3	Use video to effectively illustrate motor delay in children with DD	#3- Some videos were small and hard to see (in particular the slide with 4 videos)	Those four videos have been divided into two slides to watch it on the bigger screen than the previous one.
3_4	Additional comment	#2- The start of module 2 sounds as if I just practiced scoring an actual child. Is this part of the training module?	The training module consists of #1 and #2. Module 1 is to give the information what the TGMD-3 is and how to score. Then, Module 2 train how to correctly score two skill performance on the TGMD-3 among children with DD.
		#6- check spelling throughout module 2 video; it may be worth mentioning that some of these behaviors that you have listed to characterize DD/ID/ASD usually have an underlying function - just to avoid negatively or inaccurately representing these behaviors (I know this is not the purpose of your module but still important to consider);	Added to following description "The motor behaviors that are observed may be a response to something in the environment like noise or lights even environmental factors that may influence their motor performance."

Appendix L (cont…). Expert Feedback & Response to Module 2 in Round 1

Questions for Module 2		Reviewer # & Comment	Response
4_1	Discuss scoring of the run skill	#8- In my opinion, I would not use a video of a student striking another student (Aggression example). I also would not suggest using the "self-injurious" example because it looks like the teacher is dragging the student by the arms after the child flopped.	To help understanding about the characteristics of DD.
4_2	Show good and bad performance of the run skill		
4_3	Provided scoring feedback on the run skill (information accurately scoring correct performance)	#2- Good, but it would be better if there was an opportunity to put all four criteria together and score a true trial.	Added a true trial and a false trial each performance criterion.
		#6- Don't know how to rate this because I am not sure what this is referring to in your module - does this refer to the bullets listed under the 0?	Revised to show appropriate score on each slide
4_4	Practical questions and answers about the run skill (Provide practical strategies about common questions and answers)	#2- I don't understand what this criterion is referring to in the module.	
		#6- Don't know how to rate this because I am not sure what this is referring to in your module	Need to discuss (Revised evaluation criteria)
4_5	Additional comment	#2- This section is great, but it is only 1 of 6 locomotor skills. A novice rater will need training in all six.	This is a pilot study to develop rater training module to score FMS accurately on the TGMD-3.
		#3- These break downs were really helpful... even for TD having examples of scores of 1 and 0 for each criterion is good!	This module is to improve scoring accuracy of FMS among children with DD
		#8- I would consider adding a statement that landing on the balls of the feet is also acceptable (as performed during sprints).	Technically Heel and toe is appropriate performance on the TGMD-3 rather than on the balls of the feet.

Appendix L (cont…). Expert Feedback & Response to Module 2 in Round 1

Questions for Module 2		Reviewer # & Comment	Response
5_1	Discuss scoring of the two-hand skill		
5_2	Show good and bad performance of the two-hand skill	#3- The term "derotate" was pronounced oddly.	'dee' rotate was applied to make a better sound.
		#6- maybe it was because there were more pictures than videos shown regarding the two-hand strike versus the run skill but after one watch it felt like there was less focus on 1 and 0 examples	I understand. But, it will take much more time to watch the module if the module shows relevant examples of 0 and 1.
5_3	Provided scoring feedback on the two-hand skill	#2- Good, but it would be better if there was an opportunity to put all four criteria together and score a true trial.	Added a true trial and a false trial each performance criterion.
		#6- Don't know how to rate this because I am not sure what this is referring to in your module - does this refer to the bullets listed under the 0?	Revised to show appropriate score on each slide
5_4	Practical questions and answers about the two-hand skill	#2- I don't understand what this criterion is referring to in the module.	Evaluation criteria have been revised.
5_5	Additional comment	#2- This section is great, but it is only 1 of 7 ball skills. A novice rater will need training in all seven.	This is a pilot study to develop rater training module to score FMS accurately on the TGMD-3.

Appendix L (cont…). Expert Feedback & Response to Module 2 in Round 1

Questions for Learning Outcome of Module		Reviewer # & Comment	Response
6_1	LO#1: The rater can explain about developmental disability (DD)	#2- DD as only ID or ASD is a very conservative definition.	ID or ASD are not a definition of DD. The slide mentioned that there are multiple types of DD. This module only described ID and ASD among DD.
6_2	LO#2: The rater can list behavior and movement characteristics of children with DD	#2- Much more information is provided on ASD than on ID.	This is because ID shows lower performance level than TD whereas ASD may show lower performance level and unexpected behaviors on the TGMD assessment. Moreover, some ASD may have multiple characteristics of ID such as IQ.
6_3	LO#3: The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3		
6_4	LO#4: The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3		
6_5	LO:5: The raters get information to score FMS among children with DD according to each criterion on the TGMD-3	#2- While the information shown helps with identifying "1" and "0" performance on each criterion (of the two skills), this information is not unique to children with DD - but rather is about "poor skill".	One of the purposes of this module is to train scoring performance correctly with no confusing according to the criteria on the TGMD-3 even if children with DD perform poor skill performance or unexpected behaviors as the given examples of performance video.
		#6- I think this has a lot of great information, some additional clarifications (explained throughout quiz and evaluation) may provide additional support here. I wasn't able to see the videos that you would like to send to the raters but are these modules aimed to train people how to rate videotaped performance or live performance? Because that will require some different practice and skills - consider adding that within the module.	Very interesting comment. This module has been designed to train raters how to watch video-taped performance on the TGMD-3 rather than training to score live performance. Next study plan- videotaped vs live performance scoring Children with DD who demonstrate poor skill according to each criterion on the TGMD-3
6_6	Additional comment	#2- A missed aspect of the training is on the administration of the TGMD-3, especially for children with DD. Scoring performances of children of DD is only more challenging when they respond in a way that does not follow the protocol - essentially making it hard to tell if they unable to perform or do not understand. There are plenty of studies available on modified administration techniques – especially in ASD and VI. The training does well on the two skills covered for scoring performance, but I do not know if it addresses the correct challenge related to assessing children with ID.	When children with DD change the protocol and perform a different skill from what they were asked to do, it can be a challenge to assess their ability to perform the criteria of the required skill e.g., run in a circle instead of a straight line, or skip or gallop instead of run, it can be a challenge to assess their ability to perform the criteria of the required skill

Appendix M. Expert Feedback & Response to Online Training in Round 1

Questions for Online-training Module		Reviewer # & Comment	Response
7_1	Training program is easy to access	#2- Google drive was easy to use.	
7_2	Utilizes visual resources (angle of camera, angle of performance front and side view)	#3- A few videos were hard to see. Really appreciated when things were highlighted on videos (e.g., the ball trajectory for the poor strike performance)	
7_3	Voice narration is clear and understandable	#2- Narration is good, but sound quality changes from slide to slide. I appreciate the inclusion of closed captioning text. #6- Some of the audio quality was poor.	#3- "derotate" have been applied to make a better clear sound.
		#6- for the most part yes and I think it was great that CC was added. Some words were mispronounced and said very quickly to the point where I had to watch the CC to make sure. But that may be unavoidable.	That's a current technological limitation.
7_4	Voice narration is at an appropriate pace	#2- The computer assisted voice is good, but a human speaking might be more engaging	That's a current technological limitation. It will be considered to use a human speaking in future research.
		#6- For the most part - as stated above some transitions and words were quick	
7_5	Written material in video is easy to read	#6- Some slides had a lot of text on them and at first it was difficult to know where to focus but I did adjust. Particularly for the slides where criteria and important points for 0 and 1 scores were outlined	Overall, subjects and materials on the slide have been revised.
7_6	Additional comment	#2- There were some minor grammar issues in the voice text that could be addressed.	Overall revised.
		#8- The voice presentation does become a bit boring and struggles to keep the viewers' attention.	Need to discuss to use a different voice.
		#6- this is a cool idea!	

Appendix N. Expert Feedback & Response to the Test of Module 1 in Round 1

Quiz #1: The Test of Gross Motor Development-3 (TGMD-3) is an assessment to measure _____ of children.

Questions for Module 1 Test #1		Reviewer # & Comment	Response
1_1	Clarity and understanding	#3- Several times in the presentation the term motor skills and movement skills were interchanged (1:35 vs 5:36). This discrepancy might make this question unclear.	Overall use of terminology between movement and motor has been fixed to 'motor' according to Newwell (2020).
1_2	Appropriateness of Answer response options	#4- More challenging options/distractions would improve the question.	
		#6- fundamental movement skill performance? - consider revising body components to body composition	Revised to body composition.
1_3	Meaningfulness of question to assess content knowledge		
1_4	Degree to which question aligns with learning outcome		
1_5	Additional comment	#7- I would not abbreviate FMS on test.	Revised to fundamental motor skill.
		#4- Make sure that all of your potential answers match with the grammar of the sentence.	
		#3- Several times in the presentation the term motor skills and movement skills were interchanged (1:35 vs 5:36). This discrepancy might make this question unclear.	The term has been revised to motor skill in the module.
		#8- I think fundamental movement skill is the best choice. this is tough as the idea of motor competence, motor skill, motor skill proficiency, movement vs. motor etc is hotly contested. I would go w/ this for now but it may be something a reviewer questions, just be prepared. there really is no good answer here. movement skill learning, development, and performance?? it all depends on context	

Q2: The TGMD-3 consists of ____ fundamental movement skills.

Questions for Module 1 Test #2		Average / score range	Reviewer # & Comment	Response
2_1	Clarity and understanding	4.75/4-5		

2_2	Appropriateness of Answer response options	4.50/2-5	#4- I would recommend including both 6 and 7 (subscale #s) as potential answers	Both choices have been added on it.
2_3	Meaningfulness of question to assess content knowledge	4.63/4-5		
2_4	Degree to which question aligns with learning outcome	4.75/4-5		
2_5	Additional comment	#4- Will quiz takers be able to move back to previous questions? If so, Q2 provides the answer to Q1.		Q2 question has been changed.
		#1- pretty basic.		Q2 question has been changed to the number of each subtest on the TGMD-3.

Q3: All skills have the same number of performance criteria on the TGMD-3.

Questions for Module 1 Test #3		Average / score range	Reviewer # & Comment	Response
3_1	Clarity and understanding	4.50/2-5	#3- I don't think this was well described on the video.	Q3 question has been revised to the multiple choice for the skill score of the two-hand catch on the given case.
3_2	Appropriateness of Answer response options	4.75/3-5	#2- The T/F question is appropriately written, but poses little challenge.	
3_3	Meaningfulness of question to assess content knowledge	4.25/2-5	#2- The T/F question is appropriately written, but poses little challenge.	
3_4	Degree to which question aligns with learning outcome	4.62/4-5		
3_5	Additional comment			

Q4: Locomotor subtest on the TGMD-3 includes the following skills except _____.

Questions for Module 1 Test #4		Average / score range	Reviewer # & Comment	Response
4_1	Clarity and understanding	4.50/2-5	#6- Consider adding the before locomotor	
4_2	Appropriateness of Answer response options	4.38/2-5	#3- People who used the TGMD-2 would know this, but because the leap is never mentioned in the video, it seems odd to include it here. I think a better test question would include a ball skill instead of the leap.	The purpose of this question is to check novice raters' recognition to locomotor skills on the TGMD-3. Module 1 does not demonstrate leap but describe slide, skip, and gallop.
4_3	Meaningfulness of question to assess content knowledge	4.13/2-5		
4_4	Degree to which question aligns with learning outcome	4.13/2-5	#6- This question alone may not get at the learning outcome completely since it is only asking to identify which one is not in the subtest, not which ones are.	Q4 questions has been changed to check students' learning what skills are in the locomotor subtest on the TGMD-3.
4_5	Additional comment	#4- Tricky because that was on TGMD-2		
		#3- People who used the TGMD-2 would know this, but because the leap is never mentioned in the video, it seems odd to include it here. I think a better test question would include a ball skill instead of the leap.		I guess that any learners can select a correct answer without watching the module if there would be a ball skill instead of the leap.

Q5: How many skills are in the locomotor subtest on the TGMD-3?

Questions for Module 1 Test #5		Average / score range	Reviewer # & Comment	Response
5_1	Clarity and understanding	4.75/4-5		
5_2	Appropriateness of Answer response options	4.75/4-5		
5_3	Meaningfulness of question to assess content knowledge	4.38/3-5	#2- The number of items was already addressed. This adds little to LO2.	Q5 has been revised.
5_4	Degree to which question aligns with learning outcome	4.62/4-5		
5_5	Additional comment			

Q6: Ball skill subtest on the TGMD-3 includes the following skills except _____.

Questions for Module 1 Test #6			Average / score range	Reviewer # & Comment	Response
6_1	Clarity and understanding		4.50/2-5	#6- consider adding the before ball	
6_2	Appropriateness of Answer response options		4.50/2-5		
6_3	Meaningfulness of question to assess content knowledge		4.25/2-5		
6_4	Degree to which question aligns with learning outcome		4.25/2-5	#6- Since you are asking them to identify which skill is not included and not to identify which ones are (all 7).	The purpose of this question is that leaners can figure out what ball skills are on the TGMD-3.
6_5	Additional comment	#2- I assume that the inclusion of TGMD-2 skills as the answers (roll, leap) is intentional.			The purpose of this question is to check novice raters' recognition to ball skills on the TGMD-3 even if they understand the TGMD-2.
		#3- Same comment as on the question with the "leap". I think that while people who used the TGMD-2 would remember the roll, this question would be confusing for someone just learning the TGMD-3 and it is unfair as this content was not covered in the training module.			I expect that learners can realize changes from the TGMD-2 to the TGMD-3 as well as distinguish certain difference between underhand throw and underhand roll performance.
		#8- these are good because the "learner" will then be pushed if they were familiar w/ the TGMD-2			

Q7: The two-hand strike skill on the TGMD-3 is like a _____ skill.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
7_1	Clarity and understanding	4.00/2-5		
7_2	Appropriateness of Answer response options	4.13/2-5		
7_3	Meaningfulness of question to assess content knowledge	3.63/2-5	#6- I don't remember this being in the module - I am sure raters could assess this and think of that answer but in terms of the alignment with the module, I am not sure it aligns.	Q7 has been changed to ask what score is correct on the given case "On the first trial, If a student performs a performance criteria correctly for part of the required distance (e.g. a run) but then changes to a different skill (e.g., skip) in the middle".
7_4	Degree to which question aligns with learning outcome	3.75/2-5	#6- I don't see which learning outcome this might align with given the four you listed.	
7_5	Additional comment		#2- I do not agree with the question as evidence of LO3 because this information was not actually covered in the module. Two-handed strike is presented as a ball skill, but outside of Dr. Ulrich's video, the actual skill is not discussed.	
			#3- I don't understand the relevance of this question. Why does someone need to know this?	
			#7- Question 1. Consider "most resembles".	

Q8: Scoring values on the TGMD-3 is ___ or ___.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
8_1	Clarity and understanding	4.38/1-5		
8_2	Appropriateness of Answer response options	4.75/4-5		
8_3	Meaningfulness of question to assess content knowledge	4.88/4-5		
8_4	Degree to which question aligns with learning outcome	4.88/4-5		
8_5	Additional comment		#2- The question could possibly be improved by adding "Scoring values for each performance criteria..."	Q8 has been revised to give the question clearly.
			#3- Scoring values on the TGMD "ARE"... I think this question could be framed better. Potential score for skill criteria are...	
			#6- Change is to are	
			#8- hopefully they get that zero = not correct in my judgement	

Q9: How many trials are assessed for scoring each skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
9_1	Clarity and understanding	4.75/4-5		Q9 questions has been revised to ask clearly to novice raters.
9_2	Appropriateness of Answer response options	4.75/4-5		
9_3	Meaningfulness of question to assess content knowledge	4.88/4-5		
9_4	Degree to which question aligns with learning outcome	4.88/4-5		
9_5	Additional comment	#3- Nowhere in this video were practice trials mentioned. Since it is important to distinguish between a practice and test trial, this might be something to include.		Added '1 Practice trial' in Module 1 (7'43").

Q10: When a child performs incorrectly, score the value '___'.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
10_1	Clarity and understanding	4.38/2-5		
10_2	Appropriateness of Answer response options	4.88/4-5		
10_3	Meaningfulness of question to assess content knowledge	4.88/4-5		
10_4	Degree to which question aligns with learning outcome	4.88/4-5		
10_5	Additional comment	#3- Again, not sure the wording of this question is the most clear. Perhaps- 'When a child performs a SKILL CRITERION correctly, they receive a score of _____". A		Q10 has been revised to 'When a child performs a performance criterion correctly, they receive a score of ___' with adding one more choice '2'.
		#6- I think this is clear but you may consider rephrasing the question to specify/clarify that you would score the "criterion" 0 if they performed that "criterion" incorrectly. E.g. When a child performs a criterion incorrectly, score that criterion with the value '0'		
		#8- ignore my previous point		

Appendix O. Expert Evaluation & Response to the Test of Module 2 in Round 1

Q1: Module 2 presented two representative disabilities among developmental disabilities, such as intellectual disability (ID) and autism spectrum disorder (ASD). Which of the following statements is <u>correct</u> regarding ID or ASD?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
1_1	Clarity and understanding	3.25/1-5	#3- If possible, I would bold or underline "incorrect" in the question.	It has been applied.
			#1- sociality should be changed to a clearer word	It changed to 'sociability'.
1_2	Appropriateness of Answer response options	3.25/1-5	#6- I think this question and responses are clear but suggest revising the response options about ASD to match the way you present characteristics of ID... in that signs are presented or the diagnosis is characterized by xyz, versus that ASD shows or has	
1_3	Meaningfulness of question to assess content knowledge	3.50/1-5		
1_4	Degree to which question aligns with learning outcome	3.63/1-5		
1_5	Additional comment	#7- To my knowledge, ID is a neurodevelopmental disorder		
		#2- In my opinion, this is a poor question. Some definitions of neurodevelopmental disorder includes intellectual disability. More importantly, this question does little to address LO1. A select all correct answers questions might be more appropriate.		The choice has been revised to check the understanding of novice raters about the characteristics of DD.

Q2: Which of the following statements is correct regarding behavior and movement characteristics among children with DD?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
2_1	Clarity and understanding	3.63/1-5	#3- Didn't the video emphasize unexpected performance of gallop and skip not run?	The video shows that inappropriate gallop and skip.
2_2	Appropriateness of Answer response options	3.62/1-5	#6- All these response options seem more indicative of movement characteristics. In general, here and in the module, I am a bit confused by the "unexpected" performance	Q2 choices have been revised.
2_3	Meaningfulness of question to assess content knowledge	3.75/1-5	#6- Same comment as above, it seems to focus more on movement characteristics versus behavior and movement	
2_4	Degree to which question aligns with learning outcome	3.75/1-5	#6- Same comment as previous	
2_5	Additional comment	#2- Again – in my opinion – this is a poor question. The only response that is generalizable and correct is delayed motor skills as we will have plenty of evidence to support. While atypical or unexpected performance in running or throwing are possible, they are not evidence-based attributes of “children with DD”.		
		#6- Just a consideration, why is atypical movement performance specific to overhand throw and not others?		
		#7- Consider rewording (c). Difficulty performing a mature running pattern.		

Q3: What total score will be given on the 4 performance criteria of trial 1?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
3_1	Clarity and understanding	3.38/2-5	#4- A lot of the stem questions are not super clear and I have to re-read them to understand what you want #2- This is a good question, but a video example would be better.	Question#3 has been revised to correctly score the performance on the video.
			#6- I don't understand the connection between the scenario and the 4th performance criteria.	
3_2	Appropriateness of Answer response options	4.00/2-5		
3_3	Meaningfulness of question to assess content knowledge	3.75/2-5	#6- Don't see connection between question/scenario and content knowledge.	
3_4	Degree to which question aligns with learning outcome	3.75/2-5		
3_5	Additional comment	#3- I did not have a good understanding of this question.		
		#6- Responses make sense but this question and scenario was confusing to me - sorry I could not rate. I rated as poor alignment because it was confusing.		
		#7- I think the scenario needs to be more specific than “way off”.		

Q4: Which is the performance criterion of the run skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
4_1	Clarity and understanding	4.50/4-5		
4_2	Appropriateness of Answer response options	4.50/4-5		
4_3	Meaningfulness of question to assess content knowledge	4.29/2-5		
4_4	Degree to which question aligns with learning outcome	4.38/2-5		
4_5	Additional comment	#3- Not sure how this question aligns with DD.		This question is aligned with the locomotor skill on the TGMD-3 rather than DD in module 2.

Q5: What score will be given if a child with DD performed the run skill in which arms move in opposition to legs with one elbow bent?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
5_1	Clarity and understanding	4.38/3-5	#2- Again, good question but would be better with a video example #3- I would bold or underlie "one". I had to read the question twice.	Q5 has been revised to correctly score the performance on the video.
5_2	Appropriateness of Answer response options	4.38/4-5	#6- i think the N/A is fine but not sure it is worth putting in, unless trying to trick the raters.	N/A has been removed.
5_3	Meaningfulness of question to assess content knowledge	4.63/4-5		
5_4	Degree to which question aligns with learning outcome	4.63/4-5		
5_5	Additional comment			

Q6: What is 'Derotate' in the two-hand strike skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
6_1	Clarity and understanding	4.75/4-5		
6_2	Appropriateness of Answer response options	4.62/4-5		
6_3	Meaningfulness of question to assess content knowledge	4.88/4-5		
6_4	Degree to which question aligns with learning outcome	4.75/4-5	#6- can recognize a component of the criteria	
6_5	Additional comment		#2- I like this being an emphasis as it is a new criterion in TGMD-3 and can be easily misunderstood.	
			#3- Thought this was well done in the video. Again... not sure about the alignment with DD vs general TGMD coding per se.	

Q7: What would be the correct score for each criteria for trial 1 for the two-hand strike skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
7_1	Clarity and understanding	3.88/2-5	#2- This is an important question, but was difficult to read - especially the format of the answers	Q7 has been revised to correctly score the performance on the video.
7_2	Appropriateness of Answer response options	4.75/2-5		
7_3	Meaningfulness of question to assess content knowledge	4.75/2-5		
7_4	Degree to which question aligns with learning outcome	4.75/2-5		
7_5	Additional comment		#2- Check the grammar of your question text.	
			#6- This scenario made a lot more sense to me. Regarding criterion 1 clarity/understanding, consider giving a little more context to the scenario though: "The following scenario provides details of xyz"	

Q8: Which is not the performance criteria of two-hand strike skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
8_1	Clarity and understanding	4.50/1-5		
8_2	Appropriateness of Answer response options	4.63/4-5		
8_3	Meaningfulness of question to assess content knowledge	4.63/4-5		
8_4	Degree to which question aligns with learning outcome	4.63/4-5		
8_5	Additional comment	#3- Is this question relevant since you would never score without a score sheet?		

Q9: What would be the correct score for each criteria for trial 1 for the run skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
9_1	Clarity and understanding	4.00/4-5		
9_2	Appropriateness of Answer response options	4.38/4-5		
9_3	Meaningfulness of question to assess content knowledge	4.63/4-5		
9_4	Degree to which question aligns with learning outcome	4.63/4-5		
9_5	Additional comment	#2- This is also a good question, but the description forces the learner to make a lot of assumptions. More description of the movement would be better to test learning understanding of scoring the trial. Moreover, a video to code from would be even better.		Q9 has been revised to correctly score the performance on the video.
		#3- For a newer coder, these questions to score based on descriptions and not videos might be challenging.		
		#6- Similar comment as before - consider giving some direction and context before/after adding the scenario		

Q10: (T/F) Raters should consider the characteristics of children with DD when scoring their motor performance.

Questions for Module 2			Average / score range	Reviewer # & Comment	Response
10_1	Clarity and understanding		4.38/2-5	#6- You may want to clarify scoring versus administering based on the work being done to validate TGMD-3 administration for children with DD/ASD	Q10 has been revised to a multiple-choice question "A child with DD engages in repetitive stereotypical movements such as hand flapping while running. How should you score this student?".
10_2	Appropriateness of Answer response options		4.88/4-5		
10_3	Meaningfulness of question to assess content knowledge		4.75/4-5		
10_4	Degree to which question aligns with learning outcome		4.75/4-5		
10_5	Additional comment	#2- This is an important question - but the setup of your learning module suggests that we might need to score children with DD differently. I think this goes back to the need for addressing administration too.			
		#1- Raters should score what they see, not who is being tested			
		#6- See comment for criterion 1 - this wasn't addressed in the module so may not be relevant here but wondering if it is important for you to acknowledge this more clearly in the module, and distinguish that while some supports (e.g., extra demonstrations, visual supports) may be provided depending on the assessment situation, the scoring still remains the same			

Appendix P. Expert Feedback & Response to Module 1 in Round 2

Questions for Module 1		Reviewer # & Comment	Response
1_1	Introduce to the TGMD-3 and the scope		
1_2	Introduce 6 locomotor skills		
1_3	Introduce 7 ball skills		
1_4	Introduce scoring	#2- Scoring overview continues to focus on the logistics of scoring - something that is clearly explained in the manual. It does not address HOW to properly score skills.	Module 1 described how to properly score skills with performance examples according to each performance criterion of the two-hand catch skill.
1_5	Additional comment	#2- Issues with spelling and grammar on slides. Inconsistent audio between slides. The procedures slide does not mention the use of a non-scored practice trial. Check that red squares that appear match the current dialog. The second pass through the scoring (~12min) could be combined with the first pass through, to reduce redundancy.	Module 1 has been revised to take away spelling and grammar issues. Regarding the comment about audio problem, overall computer voice sounds replaced to a human voice. This module minimized the redundancy issues through editing process.
		#8- This may not require any revisions but just noting that you introduce the TGMD-3 as assessing gross motor skills, including locomotor and object manipulation. But then when you explain the skills in the TGMD-3 you say it includes locomotor and ball skills. Someone brand new to this may think ball skills are something different so it may be worth noting in the beginning (even in writing) that object manipulation skills are sometimes otherwise called object control or ball skills.	
		#6- I suggest pausing the videos in Module 1 at the point of each performance criteria to be rated. I suggest doing this for every ball skill and locomotor skill. I also suggest including a video in module 1 of a child incorrectly performing each performance criteria for each skill. I also suggest using a child to model the skills in Module 1. Additionally, a large amount of the audio is of poor quality and difficult to understand.	Module 1 is to introduce what the TGMD-3 and what skills are on it as well as brief explanation about scoring rather than describing correct and incorrect performances according to criteria for raters. Regarding the comment about audio problem, overall computer voice sounds replaced to a human voice to improve the quality of the audio.

Appendix P (cont…). Expert Feedback to Module 1 in Round 2

Questions for Module 1		Reviewer # & Comment	Response
2_1	LO#1: The rater understands the TGMD-3 and its components		
2_2	LO#2: The raters can identify the skills in the locomotor subtest		
2_3	LO#3: The raters can identify the skills in the ball skills subtest		
2_4	LO#4: The raters understand how to score on the TGMD-3	#2- The logistics of scoring are properly addressed. Raters would not be ready to accurately score items.	Module 1 is to explain briefly about how to score rather than describing how to accurately score items according to each performance criterion.
2_5	Additional comment	#2- Based on the feedback you received in Round 1, it might be useful to present the learning objectives of each module to the learner at the start of each video. Multiple reviewers commented on the issue of teaching the criteria - but your response was that it is beyond the scope. Clear learning objectives for the learner may help resolve that issue.	

Appendix Q. Expert Feedback & Response to Module 2 in Round 2

Questions for Module 2		Reviewer # & Comment	Response
3_1	Introduce developmental disability (DD)		
3_2	Introduce motor performance of children with DD		
3_3	Use video to effectively illustrate motor delay in children with DD	#2- Provides examples of motor delay or behaviors that could affect testing - but not how to address them.	
		#6- I suggest including videos for all skills in the TGMD-3.	
3_4	Additional comment	#2- Issues with spelling and grammar on slides. Inconsistent audio between slides. My suggestion would be to turn this into three modules. Module 1 as currently planned as an introduction to the TGMD-3. Module 2 to practice scoring (all skills would be preferable). Module 3 to address DD and relevant testing issues.	
Questions for Module 2		Reviewer # & Comment	Response
4_1	Discuss scoring of the run skill		
4_2	Show good and bad performance of the run skill		
4_3	Provided scoring feedback on the run skill		
4_4	Additional comment	#2- Audio on Run #3 would be useful to hear foot strike. On all slides, some description to direct the learners attention to a 1 or 0 would be useful. This was done better and more consistently with striking.	
		#6- Provide examples like this for all locomotor skills. Improve audio quality.	

Appendix Q (cont…). Expert Feedback & Response to Module 2 in Round 2

Questions for Module 2		Reviewer # & Comment	Response
5_1	Discuss scoring of the two-hand skill		
5_2	Show good and bad performance of the two-hand skill		
5_3	Provided scoring feedback on the two-hand skill		
5_4	Additional comment	#2- Strike #3 could also show an example of a 0 due to insufficient rotation - not just derotating. Strike #4 does not include an example.	
		#6- I suggest providing examples like this for all of the TGMD-3 ball skills. Improve audio quality in this section as well.	

Appendix Q (cont…). Expert Feedback & Response to Module 2 in Round 2

Questions for Learning Outcome of Module		Reviewer # & Comment
6_1	LO#1: The rater can explain about developmental disability (DD)	
6_2	LO#2: The rater can list behavior and movement characteristics of children with DD	
6_3	LO#3: The raters can recognize the performance criteria of the run skill in the locomotor subtest on the TGMD-3	
6_4	LO#4: The raters can recognize the performance criteria of the two-hand strike in the ball skills subtest on the TGMD-3	
6_5	LO:5: The raters get information to score FMS among children with DD according to each criterion on the TGMD-3	#2- The content on DD is insufficient to train a novice rater. While potential issues or pitfalls are presented, no solutions are provided.
		#8- Not clear on what this learning objective means. I think that this video will give important information to raters to score run and two-hand strike. But it likely will still be difficult for new raters to understand the criteria for any other skills and then score.
6_6	Additional comment	#2- Even for a pilot project - more than two skills would be needed to train raters.
		#8- I like the addition of the slide that says that environmental or other factors may lead to some of the atypical or behavioral characteristics seen from children with DD. Overall I think the video module has improved but I think there is still a clear issue that it will be difficult for raters to code skills that aren't described in this video. The scoring process is very clear. But is there any plan for understanding the performance criteria of all the other skills? Maybe it will be done in a class? Do raters review and try to understand the criteria on their own? Not clear to me based on the video so I am just noting it here.
		#6- When showing the example of aggressive behaviors, consider using another example for the self-injurious part. The restraint used in the video may not be appropriate.

Appendix R. Expert Feedback & Response to Online Training in Round 2

Questions for Online-training Module		Reviewer # & Comment	Response
7_1	Training program is easy to access		
7_2	Utilizes visual resources (angle of camera, angle of performance front and side view)	#2- Replacing Dale's videos with examples that don't require panning or short angles would help Module 1.	
7_3	Voice narration is clear and understandable	#2- Audio level and quality is inconsistent.	
		#6- Many sections were difficult to hear and understand. Other sections the audio was muted.	
		#8- the volume/clarity varies throughout the videos. And there were a couple points in the video where it seemed like the sound cut out.	
7_4	Voice narration is at an appropriate pace		
7_5	Written material in video is easy to read	#2- Easy to read, but in need of editing for grammar and spelling	
		#8- yes, but check spelling throughout (e.g., introduce on first slide of module 1)	
7_6	Additional comment	#8- check spelling throughout.	

Appendix S. Expert Feedback & Response to the Test of Module 1 in Round 2

Questions for Module 1 Test #1		Average / score range	Reviewer # & Comment	Response
1_1	Clarity and understanding	4.50/4-5		
1_2	Appropriateness of Answer response options	4.00/2-5	#1- More challenging distractors. Add a fourth distractor to increase difficulty.	
1_3	Meaningfulness of question to assess content knowledge	4.50/4-5		
1_4	Degree to which question aligns with learning outcome	4.50/4-5		
1_5	Additional comment			

Quiz #1: The Test of Gross Motor Development-3 (TGMD-3) is an assessment to measure _____ of children.

Q2: The TGMD-3 consists of ____ fundamental movement skills.

Questions for Module 1 Test #2		Average / score range	Reviewer # & Comment	Response
2_1	Clarity and understanding	4.50/4-5		
2_2	Appropriateness of Answer response options	4.50/4-5		
2_3	Meaningfulness of question to assess content knowledge	4.25/4-5		
2_4	Degree to which question aligns with learning outcome	4.50/4-5		
2_5	Additional comment			

Q3: All skills have the same number of performance criteria on the TGMD-3.

Questions for Module 1 Test #3		Average / score range	Reviewer # & Comment	Response
3_1	Clarity and understanding	4.25/4-5		
3_2	Appropriateness of Answer response options	4.50/4-5		
3_3	Meaningfulness of question to assess content knowledge	4.25/4-5		
3_4	Degree to which question aligns with learning outcome	4.00/3-5	#5- Seems to be related more to understanding the scoring process than to the TGMD	
3_5	Additional comment			

Q4: Locomotor subtest on the TGMD-3 includes the following skills except _____.

Questions for Module 1 Test #4		Average / score range	Reviewer # & Comment	Response
4_1	Clarity and understanding	4.50/4-5		
4_2	Appropriateness of Answer response options	4.50/4-5		
4_3	Meaningfulness of question to assess content knowledge	4.50/4-5		
4_4	Degree to which question aligns with learning outcome	4.50/4-5		
4_5	Additional comment	#2- I am ok with the leap here. It could confuse people that know the TGMD-2, but would also distinguish learners as leap is not in the list presented.		I agree with your comment. This question has been revised to remove any confusing issue.

Q5: How many skills are in the locomotor subtest on the TGMD-3?

Questions for Module 1 Test #5		Average / score range	Reviewer # & Comment	Response
5_1	Clarity and understanding	4.25/4-5		
5_2	Appropriateness of Answer response options	3.75/2-5	#1- Add fourth distractor to increase difficulty	This question has been revised to meet the learning objective in terms of the locomotor subtest.
5_3	Meaningfulness of question to assess content knowledge	4.25/4-5		
5_4	Degree to which question aligns with learning outcome	4.00/3-5	#5- Seems more relevant to the learning objective about scoring. Not locomotor subtest	This question has been revised to meet the learning objective in terms of the locomotor subtest.
5_5	Additional comment			

Q6: Ball skill subtest on the TGMD-3 includes the following skills except _____.

Questions for Module 1 Test #6		Average / score range	Reviewer # & Comment	Response
6_1	Clarity and understanding	4.50/4-5		
6_2	Appropriateness of Answer response options	4.50/4-5		
6_3	Meaningfulness of question to assess content knowledge	4.50/4-5		
6_4	Degree to which question aligns with learning outcome	4.50/4-5		
6_5	Additional comment	#2- Same point as the question with the leap		This question has been revised to check the understanding of learners easily.

Q7: The two-hand strike skill on the TGMD-3 is like a _____ skill.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
7_1	Clarity and understanding	3.25/2-5	#2- This point was not clearly established in Module 1.	I agree with your comment.
7_2	Appropriateness of Answer response options	3.25/2-5	#1- Add a fourth distractor to increase difficulty.	Q7 has been replaced a new question to select correct scoring on a performance video of a child with DD.
7_3	Meaningfulness of question to assess content knowledge	3.75/3-5		
7_4	Degree to which question aligns with learning outcome	3.75/3-5	#5- More relevant to the scoring learning objective than ball skills.	
7_5	Additional comment	#2- This question is important, but is by far the hardest item due to the lack of focus in Module 1.		
		#6- I'm not sure I understand this question or that I agree with the answer. And I don't remember this being covered in module 1. You may want to clarify what "part" means both here and in the video module.		
		#8- I do not recall if this was addressed in the instructional video. I think most eaters would score this as a zero because the student did not perform the skill for the appropriate distance as required in the TGMD-3		

Q8: Scoring values on the TGMD-3 is ___ or ___.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
8_1	Clarity and understanding	4.75/4-5		
8_2	Appropriateness of Answer response options	4.25/2-5		
8_3	Meaningfulness of question to assess content knowledge	4.75/4-5		
8_4	Degree to which question aligns with learning outcome	4.75/4-5		
8_5	Additional comment			

Q9: How many trials are assessed for scoring each skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
9_1	Clarity and understanding	4.75/4-5		
9_2	Appropriateness of Answer response options	4.75/4-5		
9_3	Meaningfulness of question to assess content knowledge	4.75/4-5		
9_4	Degree to which question aligns with learning outcome	4.75/4-5		
9_5	Additional comment	#3- More emphasis in Module 1 about a practice trial is needed.		

Q10: When a child performs incorrectly, score the value '___'.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
10_1	Clarity and understanding	4.50/3-5		
10_2	Appropriateness of Answer response options	4.25/2-5		
10_3	Meaningfulness of question to assess content knowledge	4.50/3-5		
10_4	Degree to which question aligns with learning outcome	4.50/3-5		
10_5	Additional comment	#2- Repetitive from Q8. Could replace with something more challenging.		Q10 has been replaced to a new question.

Appendix T. Expert Feedback & Response to the Test of Module 2 in Round 2

Q1: Module 2 presented two representative disabilities among developmental disabilities, such as intellectual disability (ID) and autism spectrum disorder (ASD). Which of the following statements is <u>correct</u> regarding ID or ASD?

Questions for Module 1 Test #1		Average / score range	Reviewer # & Comment	Response
1_1	Clarity and understanding	4.40/4-5		
1_2	Appropriateness of Answer response options	3.80/2-5	#6- Separate questions for ID and ASD with meaningful distractors would be more challenging and discriminating of learning	I recommended revising the options to state "children with ID and ASD" as necessary
			#2- I recommended revising the options to state "children with ID and ASD" as necessary	
1_3	Meaningfulness of question to assess content knowledge	4.20/3-5		
1_4	Degree to which question aligns with learning outcome	4.60/4-5		
1_5	Additional comment			

Q2: Which of the following statements is correct regarding behavior and movement characteristics among children with DD?

Questions for Module 1 Test #2		Average / score range	Reviewer # & Comment	Response
2_1	Clarity and understanding	4.40/4-5		
2_2	Appropriateness of Answer response options	4.20/2-5	#8- More challenging option, such as which item is NOT a correct example could be more challenging.	Q2 choices have been revised with response options to help understand of behavior and movement characteristics among children with DD.
2_3	Meaningfulness of question to assess content knowledge	4.40/4-5		
2_4	Degree to which question aligns with learning outcome	4.60/4-5		
2_5	Additional comment	#2- check spelling of response options, and should be a.		

Q3: What total score will be given on the 4 performance criteria of trial 1?

Questions for Module 1 Test #3		Average / score range	Reviewer # & Comment	Response
3_1	Clarity and understanding	4.00/3-5	#2- I found it a little difficult to clearly see his hands/grip	The video on Q3 has been magnified to see a little clearly.
3_2	Appropriateness of Answer response options	4.20/3-5		
3_3	Meaningfulness of question to assess content knowledge	4.00/3-5	#5- why not provide all performance criteria? Seems like that would better assess content knowledge.	c
3_4	Degree to which question aligns with learning outcome	4.40/4-5		
3_5	Additional comment	#3- Why are they only scoring 3 of 5 performance criteria? Identifying the other 2 is the most challenging aspect of that example.		If there are all performance criteria of the skill, this question would be difficult.
		#8- I would personally score this 1-0-1. Especially given the clear rotate video that you provided in the module. I like the inclusion of the video.		This question was intended to watch a performance video carefully.
		#7- I would consider using another example. It was difficult to see the student switch his hand placement immediately before the skill was initiated.		

Q4: Which is the performance criterion of the run skill?

Questions for Module 1 Test #4		Average / score range	Reviewer # & Comment	Response
4_1	Clarity and understanding	4.60/4-5	#2- C and d are basically saying the same thing.	The response option has been revised.
4_2	Appropriateness of Answer response options	4.20/2-5		
4_3	Meaningfulness of question to assess content knowledge	4.20/2-5		
4_4	Degree to which question aligns with learning outcome	4.60/4-5		
4_5	Additional comment			

Q5: What score will be given if a child with DD performed the run skill in which arms move in opposition to legs with one elbow bent?

Questions for Module 1 Test #5		Average / score range	Reviewer # & Comment	Response
5_1	Clarity and understanding	4.80/4-5		
5_2	Appropriateness of Answer response options	4.40/2-5	#3- Why is the 4th criteria omitted?	If there are all performance criteria of the skill, this question would be difficult.
5_3	Meaningfulness of question to assess content knowledge	4.40/3-5	#3- Same feedback as last question. Why not provide all performance criteria here?	
5_4	Degree to which question aligns with learning outcome	4.60/4-5		
5_5	Additional comment	Does not match the skill term between the tile and question		Revised the skill term to the run skill.

Q6: What is 'Derotate' in the two-hand strike skill?

Questions for Module 1 Test #6		Average / score range	Reviewer # & Comment	Response
6_1	Clarity and understanding	4.40/4-5		
6_2	Appropriateness of Answer response options	4.40/4-5		
6_3	Meaningfulness of question to assess content knowledge	4.20/3-5	#2- I would ask a question that includes both elements of rotation and derotation.	This question pointed out only the term in the performance criterion of the two-hand strike skill.
6_4	Degree to which question aligns with learning outcome	4.40/4-5		
6_5	Additional comment			

Q7: What would be the correct score for each criteria for trial 1 for the two-hand strike skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
7_1	Clarity and understanding	4.60/4-5		
7_2	Appropriateness of Answer response options	4.40/2-5	#2- Why are the other two criteria not included?	If there are all performance criteria of the skill, this question would be difficult.
7_3	Meaningfulness of question to assess content knowledge	4.40/4-5		
7_4	Degree to which question aligns with learning outcome	4.60/4-5		
7_5	Additional comment			

Q8: Which is not the performance criteria of two-hand strike skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
8_1	Clarity and understanding	4.60/4-5		
8_2	Appropriateness of Answer response options	4.40/2-5	#1- The "non-throwing side" is a give-away of the correct answer. A different distractor would be better.	Q8 has been revised to remove confusing statements.
8_3	Meaningfulness of question to assess content knowledge	4.60/4-5		
8_4	Degree to which question aligns with learning outcome	4.60/4-5		
8_5	Additional comment			

Q9: What would be the correct score for each criteria for trial 1 for the run skill?

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
9_1	Clarity and understanding	4.80/4-5		
9_2	Appropriateness of Answer response options	4.40/2-5	#1- More options needed. Why only 1 of 4 criteria addressed?	This question was to only point out the first performance criterion of the run skill.
9_3	Meaningfulness of question to assess content knowledge	4.60/4-5		
9_4	Degree to which question aligns with learning outcome	4.80/4-5		
9_5	Additional comment			

Q10: (T/F) Raters should consider the characteristics of children with DD when scoring their motor performance.

Questions for Module 2		Average / score range	Reviewer # & Comment	Response
10_1	Clarity and understanding	4.40/3-5		
10_2	Appropriateness of Answer response options	4.40/2-5	#1- Add additional distractor	
10_3	Meaningfulness of question to assess content knowledge	4.20/3-5	#8- I may have missed this in the module but this should be emphasized in the module.	Module 2 emphasized to score the skill performance regardless of behavioral and movement characteristics of children with DD.
10_4	Degree to which question aligns with learning outcome	4.80/4-5		
10_5	Additional comment	#2- This is a really important question- but received little attention or time within Module 2.		