ORIGINAL ARTICLE

WILEY

The validity of the 4-Skills Scan A double-validation study

W. G. van Kernebeek[1] [iD] | M. L. A. de Kroon[2] | G. J. P. Savelsbergh[1,3] | H. M. Toussaint[1]

[1]Faculty of Sports and
Nutrition, Amsterdam University of Applied
Sciences, Amsterdam, The Netherlands

[2]Department of Health Sciences, University
Medical Center Groningen, Groningen, The
Netherlands

[3]Faculty of Behavioural and Movement
Sciences, Department of Human Movement
Sciences, Vrije Universiteit, Amsterdam,
The Netherlands

Correspondence
Willem van Kernebeek, Faculty of Sports
and Nutrition, Amsterdam University
of Applied Sciences, Amsterdam, The
Netherlands.
Email: w.g.van.kernebeek@hva.nl

Funding information
This study is funded by RAAK-PRO/SIA
ref. 2014−01−114PRO

Adequate gross motor skills are an essential aspect of a child's healthy development. Where physical education (PE) is part of the primary school curriculum, a strong curriculum-based emphasis on evaluation and support of motor skill development in PE is apparent. Monitoring motor development is then a task for the PE teacher. To fulfill this task, teachers need adequate tools. The 4-Skills Scan is a quick and easily manageable gross motor skill instrument; however, its validity has never been assessed. Therefore, the purpose of this study was to assess the construct and concurrent validity of both 4-Skills Scans (version 2007 and version 2015). A total of 212 primary school children (6-12 years old) was requested to participate in both versions of the 4-Skills Scan. For assessing construct validity, children covered an obstacle course with video recordings for observation by an expert panel. For concurrent validity, a comparison was made with the M-ABC-2, by calculating Pearson correlations. Multivariable linear regression analyses were performed to determine the contribution of each subscale to the construct of gross motor skills, according to the M-ABC-2 and the expert panel. Correlations between the 4-Skills Scans and expert valuations were moderate, with coefficients of .47 (version 2007) and .46 (version 2015). Correlations between the 4-Skills Scans and the M-ABC-2 (gross) were moderate (.56) for version 2007 and high (.64) for version 2015. It is concluded that both versions of the 4-Skills Scans are satisfactory valid instruments for assessing gross motor skills during PE lessons.

KEYWORDS
motor skill assessment, physical education, primary school children, psychometrics, validity

1 | INTRODUCTION

Motor skill development is an essential aspect in a child's healthy development and forms the basis for sports participation and an active and healthy lifestyle.[1-4] Therefore, acquisition of fundamental movement skills (FMS) is an important part of the primary school curriculum.[2,5] As learning is a cumulative process,[6] a lack of physical play and exercise, and thus motor practice, may lead to an increased skill gap with peers. While some decades ago, children would acquire adequate skills by playing outside, nowadays, school seems increasingly important when it comes to the development of gross motor skills. Physical education (PE) consequently plays a major role in motor skill development,[7,8] and more and more evidence is found that PE has a positive influence on FMS development.[9]

In the Netherlands, PE is part of the primary school curriculum. Consequently, a strong curriculum-based emphasis on evaluation and support of motor skill development in PE is apparent. The rationale behind monitoring the development of gross motor skills is threefold. First, monitoring gross motor skills enables the PE teacher to adapt the lessons in terms of diversity and difficulty, so that it matches everyone's "zone of proximal development."[10] Hence, in a good learning

environment, it is essential that each child can engage in the presented activity while being challenged to extend his or her skill ability. This is especially important in mixed-ability classes. Second, it allows for the possibility of evaluating the effectiveness of a given series of lessons, which creates a beneficial feedback loop for the professional PE teacher that is usually missing in the PE context. Thirdly, it enables the recommended early detection of derailed or delayed motor development,[11,12] allowing for further assessment (and treatment) after referral to a preventive child healthcare physician (PCHC physician), general practitioner, or pediatric physiotherapist.

Although there is a solid rationale for monitoring the motor development of each child in primary school, daily practice shows that monitoring is often lacking. Rink[13] indicated that the main problem of motor skill assessment during PE lessons is the impracticality or time-consuming character of most instruments. This makes them unsuitable for educational purposes. In general, instruments with known reliability and validity are to be preferred. PE teachers, however, only have a few instruments to choose from, and feasible tests with known reliability and validity are not available. In the Netherlands, this resulted in a situation where many PE teachers use the 4-Skills Scan of Van Gelder,[14,15] which has unknown validity. In contrast, healthcare professionals often use the M-ABC-2,[11,16] which is a diagnostic test. However, due to the necessary test material and the time it takes to conduct it, the M-ABC-2 is an unpractical test for the PE setting.

The 4-Skills Scan is an easily conducted gross motor skill test and was specifically developed for the PE setting. The current version of the 4-Skills Scan was designed carefully, through many iterations and after fifteen years of gathering feedback from hundreds of PE teachers.[15] An important and appealing aspect of the test is that the difficulty levels correspond to calendar age. Hence, the test outcome is expressed in motor age in comparison with the calendar age. This test has received little scientific attention yet; however, in a previous study,[17] we showed that the 4-Skills Scan is a reliable test for assessing gross motor skills during PE lessons. In this follow-up study, our aim was to assess the construct (gold standard) and concurrent validity of the 4-Skills Scan (version 2007) and its updated version (4-Skills Scan 2015).

2 | METHOD

2.1 | Design and study population

Twenty-one primary schools that are part of the Measuring Motor Skills of children in Primary Education (MAMBO) cohort were approached for participation in this validation study. Flyers were spread amongst schools, and information meetings about the purpose and method of the study were held. This resulted in the inclusion of 9 participating schools from different geographical and socioeconomic areas of Amsterdam. After a weighted randomization for motor skill level, 550 children and their parents received an information letter including an invitation to participate in the study. Children gave assent to participation, and a signed informed consent was obtained from their parents. This resulted in 212 included children (50% boys, 50% girls, mean age 9.16 years; SD 1.84, age range 6-12 years). The children were transported by bus to the test location of the Amsterdam University of Applied Sciences (HvA). The study protocol was approved by the Scientific and Ethical Review Board (VCWE) of the Faculty of Behavioural & Movement Sciences, VU University Amsterdam (VCWE-2015-171).

2.2 | The 4-Skills Scan

Van Gelder's 4-Skills Scan[14,15] was used for assessing the children's gross motor skills. This quantitative motor performance test finds its origin in Ayres' motor development theory,[18,19] which links the development of sensory integration with age, to gross motor performance. The development of gross motor skill is reflected in the age-related difficulty levels in the 4-Skills Scan. The 4-Skills Scan covers the main FMS[1] with four subscales: 1. "jumping force," which can be considered a locomotion task; 2. "bouncing ball," which is considered a manipulative or object control task; 3. "standing still," which is considered as a stability task; and 4. "Jumping coordination," which can be considered a dynamic stability task (twisting), involving rhythmic jumping with coordinated arm and leg movements. Subscales for each test item are composed of a series of similar tasks that increase in difficulty with age.[20] For the 4-Skills Scan version of 2007,[14] nine age-related difficulty levels were defined; for the 2015 version, there were eleven difficulty levels.[15] Each subscale score reflects the expected performance level for a certain age. The test outcome is the combination of the subscales in years, calculated as follows:

$$
\text{Motor age} = \frac{\left(\begin{array}{l} \text{level''balance''} + \text{level''jumping force''} + \\ \text{level''jumping coordination''} + \text{level''bouncing ball''} \end{array} \right)}{4}
$$

Motor age is then compared with calendar age in order to determine "motor lead," the motor skill performance relative to calendar age:

$$
\text{Motor lead(years)} = \text{Motor Age(years)} - \text{calendar age(years)}
$$

In a previous study, ICC's of .93 and .97 were found for, respectively, test-retest and inter-rater reliability.[17] Current

research revealed a linear relationship between motor age and calendar age for the age range of 5-10 years, after which a small divergent trend is noticeable.[21]

2.3 | Construct validity

There is no gold standard for motor skills, so in order to get as close to the nonexisting gold standard as possible, Surowiecki's[22] concept "wisdom of the crowds" was introduced as a method to find a good representation of the true motor skill level of children.

To test this innovative method in the PE setting, a pilot study was performed where the outcome of the test was compared with the valuation of a panel of experts who were presented with video footage of children covering an obstacle course. The pilot study focussed on the practicalities of the obstacle course's layout, its difficulty, and the instructions the children were given. The pilot study revealed that the valuation of experts using video footage is appropriate.[23] The experience also led to a series of changes to the set up of the experiment: 1. build the obstacle course in one central location in order to be able to enrich its landscape of motor affordances; 2. tailor the difficulty of the obstacle course to smaller age groups; 3. implement a time-trial section in the obstacle course; 4. present clear and standardized (video) instruction to the participating children; and 5. provide clear instructions and information to the expert panel regarding the instruction to the children.

For this study, a larger expert panel was formed, based on the three conditions defined by Surowiecki: a diverse, independent group with a certain degree of geographical decentralization. The expert panel consisted of 36 PCHC physicians, 28 PE teachers, and 30 pediatric physiotherapists from different geographical areas in the Netherlands. They were asked to rate children's gross motor skills based on 3-minute video clips of individual children covering an obstacle course. Valuation of skill level by video clips has been successfully used in studies before.[24,25]

Three versions of an obstacle course were developed with varying degrees of difficulty (see Appendix 1). The obstacle course consisted of several fundamental movement tasks (locomotion, balance, and object manipulation[1]) and was based on seven of twelve defined learning objectives for teaching PE in primary education in the Netherlands as well, namely balancing, climbing, swinging, forward rolling, jumping, running, and aiming.[26] Besides overall gross motor skills, experts were asked to score on the following subdomains of gross motor skills: balance skills, locomotion skills, ball skills, and jumping skills. The scoring was done according to a 100-points slider scale from 0 to 10.

2.4 | Concurrent validity

The Movement Assessment Battery for Children—2nd edition (M-ABC-2) is the revised version of the original M-ABC.[27] It is one of the most frequently used motor skills tests by healthcare professionals worldwide and often applied in order to detect delayed or derailed motor development and for diagnosing Development Coordination Disorder.[11,16,27,28] As there is no true gold standard and given the well-known status, researchers often see the M-ABC-2 as an alternative to a gold standard. This is why several studies have applied the M-ABC-2 to examine the validity of new assessment tools.[29,30] However, the M-ABC-2 was designed for application in a clinical setting, whereas the 4-Skills Scan was designed for the PE setting as a screening tool for all children at primary school that participate in PE lessons. Here, an important purpose of motor testing is detecting children with motor delays. Therefore, a comparison with the M-ABC-2 for concurrent validity seems legitimate. For our study, the Dutch version of the M-ABC-2 was used.[31] The combined standard scores of the categories "aiming & catching" and "balance" of the M-ABC-2 were used as a representation of the gross motor skills. The total score of the M-ABC-2 was analyzed as well.

2.5 | Procedure

All measurements took place in a large sports hall (24 m × 22 m) that was split into three parts (see Appendix 1). Throughout half a day, 17 trained test conductors would welcome about 27 children from a single school. This was followed by a short explanation of the purpose of this special "sports day." Children were then allowed to experience the obstacle course one time before being split into three age groups. Subsequently, one age group began performing the two versions of the 4-Skills Scan, while the second age group started out with the M-ABC-2, and the third age group covered the age corresponding version of the obstacle course. After finishing, children proceeded to the next study element in another part of the sports hall, until every child had performed both versions of the 4-Skills Scan, the M-ABC-2, and the obstacle course.

2.6 | Data collection and analyses

Data collection for the 4-Skills Scans and the M-ABC-2 was performed digitally with dedicated FileMaker Pro[32] apps on iPads Air. To collect expert valuations, video clips were uploaded in Qualtrics software[33] and experts were invited to participate in a video questionnaire by email, accompanied with a personal login and password. Each expert viewed a total of 31 videos. Each video showed the performance of one child on the obstacle course. The

videos were presented to the experts in random order, but the order was clustered by age-band and predicted motor skill level. For each video, the expert filled in a questionnaire and rated the motor skill level of the child performing the obstacle course.

As the 4-Skills Scans, the M-ABC-2, and expert valuations all have continuous scores, correlation is the preferred statistical method.[34] Assumptions for normality were met, and Pearson correlations were calculated between the 4-Skills Scans and the expert valuations for construct validity and between the 4-Skills Scans and the M-ABC-2 for concurrent validity. Hence, a double validation was carried out.

In order to define the contribution of each subscale to the construct of gross motor skills, a multivariable linear regression analysis was performed. In this way, strong or possible redundant subscales could be identified.

2.6.1 | Missing data

Multiple imputation (MI) was applied by means of linear regression with the Markov chain Monte Carlo algorithm. Missing values were imputed at subscale level as recommended by Eekhout et al[35] for the two versions of the 4-Skills Scan as well as the M-ABC-2. The maximal percentage of missing values for any of the imputed variables was 8.9%. The dataset was imputed 10 times with 10 iterations each. Values were only imputed if no more than one item per measurement instrument was missing. Otherwise, the child was excluded from analyses. Expert valuation for motor skills was excluded as a predictor for missing values to avoid any imputed dependency. For all analyses, SPSS version 21 statistical analysis package for Macintosh was used.

3 | RESULTS

The gross motor skill ability data measured for the 212 children with the 4-Skills Scans, expert panel, and the M-ABC-2 are presented in Table 1. It should be noted that the metrics to express the motor ability differ for the different tests.

In studying the construct validity, moderate positive overall pooled correlations were found between both versions of the 4-Skills Scan and the experts' valuation (.47 for version 2007 and .46 for version 2015).

With respect to the assessment of the concurrent validity, moderate-to-strong pooled correlation (.56 for version 2007 and .64 for version 2015) was found between the 4-Skills Scan and the gross motor section of the M-ABC-2. With correlations of .58 (version 2007) and .62 (version 2015), the total M-ABC-2 gave similar results (see Table 2).

Table 3 presents the results of linear regression analyses with expert valuations and M-ABC-2 (gross) as the dependent variables. It shows that, with regard to the entire group, all β-coefficients and thus all subscales of the 2015 version significantly contribute to the explained variance of the M-ABC-2 (gross) as the dependent variable, whereas for the 2007 version, only bouncing (ball) and jumping force significantly contribute to predicting the M-ABC-2 (gross) outcome. Also, for the expert valuations as the dependent variable, bouncing (ball) and jumping force are found to be the strongest predictors.

4 | DISCUSSION

Validity is an important psychometric property of a gross motor test. The absence of a gold standard presented a challenge to determine that property for the two versions of the 4-Skills Scan. This study opted for a dual approach. In line with many previous validation studies, concurrent validity was assessed by comparing test results with the M-ABC-2. The valuations of a panel of experts were used to assess construct validity.

4.1 | Concurrent validity

For version 2007, correlations of .58 were found with the M-ABC-2 and .56 with the gross motor skill section of the M-ABC-2. The 2015 version showed a correlation of .62

TABLE 1 Pooled results for the 4-Skills Scans, expert valuation, and M-ABC-2 (gross)

Study sample	N	4-Skills Scan (2015) Motor age (y)		4-Skills Scan (2007) Motor age (y)		Experts valuation Score (1-10)		M-ABC-2 Std score (1-19)		M-ABC-2 (gross) Std score (1-19)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total group	212	8.92	1.99	8.51	2.02	7.03	1.02	7.54	3.22	8.53	2.38
Junior grades (age 5-9)	103	7.51	1.56	7.08	1.23	6.82	0.97	8.12	3.20	8.85	2.41
Senior grades (age 9-12)	109	10.27	1.30	9.88	1.64	7.24	1.03	7.00	3.15	8.23	2.32

with M-ABC-2 total score and .64 with the gross motor skill section of the M-ABC-2. This compares favorably to other validation studies of motor skill assessment instruments. For example, in a review study performed by Cools et al,[36] a correlation of −.53 was reported between the M-ABC and the BOTMP. An even weaker relationship (.30) was found between TGMD-2 and M-ABC-2 for 3- to 13-year-old children.[37] Fransen et al[38] reported correlations of .44 regarding outcomes of the KörperkoordinationsTest für Kinder (KTK) with the Bruininks-Oseretsky Test of Motor Proficiency 2 (BOT-2) and .64 with the gross motor skill section of the BOT-2.

4.2 | Construct validity

The expert panel approach was used as an alternative method to validate the 4-Skills Scan. However, somewhat less strong correlations were observed between the expert valuations and both versions of the 4-Skills Scans. Moderate correlations were found for version 2007 (.47) and version 2015 (.46) of the 4-Skills Scan. Nevertheless, the results suggest that the 4-Skills Scan is apt to fulfill a role in early detection of derailed or delayed motor development in children.

Our findings were not quite in line with previous studies involving video observation.[24,25] It could be that the more holistic quantification of the children's gross motor skills by the experts caused lower correlation than the correlation between two more objective quantitative tests (ie, M-ABC-2). The expert panel in the current study was made up of experts from three different fields, each with their own particular educational background and practical experience. Hence, the diversity in professional background of the panel may have contributed to a more diverse assessment of similar motor skill ability. This may adversely affect our findings. For instance, PE teachers and PCHC physicians may have a good sense of motor skill levels of the general population. In contrast, the reference for a typically skilled

child may be biased for pediatric physiotherapists, as the majority of the children they see may not have a typical development. It could also be that the experts' reference perspective of motor performance is susceptible to shifting, after viewing a number of children with above or below average gross motor skills. However, we believe that their reference point is relatively stable due to their thorough education and years of experience.

The regression analyses showed that the subscales standing still and jumping coordination are less predictive for the M-ABC-2 (gross) and expert valuations. However, there seems to be little room for an ultrashort version of the test that is limited to the two subscales "jumping force" and "bouncing ball." Additional analyses showed that with the M-ABC-2 (gross) as the dependent variable, explained variance significantly drops from .43 to .41 when leaving out the subscale jumping coordination, and to .39 when leaving out the subscale standing still. With respect to the panel of experts, this would mean the explained variance to stay .30.

In general, part of the lower-than-expected relationships and agreement between motor skill instruments may be explained by the absence of a gold standard. Many motor skill instruments are believed to measure the same construct. However, as the study by Fransen et al[38] pointed out, the construct of motor skills instrument might not be well-aligned in most cases. For instance, the KTK was found suitable for monitoring gross motor skills in typically developing children, albeit that dynamic balance skills seem over-represented.[36,39] In addition, it is difficult to fully remove measuring physical fitness from the equation and it may differ to what extent physical fitness is included by motor skill instruments. Moreover, some instruments are particularly sensitive with regard to detecting children that have motor performance difficulties.[36,39,40] For monitoring purposes, however, a sufficient responsiveness for a wide range of skill levels of the children is desired.

TABLE 2 Pearson correlations on pooled results between the 4-Skills Scans (motor lead), expert valuation, and M-ABC-2 (gross)

Study sample	4-Skills Scan (2015)	Expert valuation	M-ABC-2	M-ABC-2 (gross)
4-Skills Scan (2007)				
Total group	0.75[*]	0.47[*]	0.58[*]	0.56[*]
Junior grades (age 5-9)	0.75[*]	0.51[*]	0.56[*]	0.53[*]
Senior grades (age 9-12)	0.75[*]	0.55[*]	0.58[*]	0.59[*]
4-Skills Scan (2015)				
Total group	-	0.46[*]	0.62[*]	0.64[*]
Junior grades (age 5-9)		0.53[*]	0.65[*]	0.67[*]
Senior grades (age 9-12)		0.50[*]	0.58[*]	0.60[*]

N range from 194 to 211.

*P < .001.

TABLE 3 Results of a multivariable linear regression with expert valuation or M-ABC-2 (gross) as dependent variable

	Bouncing (ball)			Standing still			Jumping coordination			Jumping force		
	β	CI-95%	P-value	β	CI-95%	P-value	β	CI-95%	P-value	β	CI-95%	P-value
Expert valuation												
4-Skills Scan (2007)												
Total group	0.04	−0.04 to 0.11	.32	0.01	−0.04 to 0.06	.58	0.01	−0.05 to 0.07	.76	**0.31**	0.24 to 0.39	.00
Junior grades (age 5 to 9)	**0.26**	0.05 to 0.46	.02	0.01	−0.08 to 0.10	.87	0.06	−0.01 to 0.14	.09	**0.29**	0.16 to 0.42	.00
Senior grades (age 9 to 12)	**0.08**	0.00 to 0.16	.04	0.02	−0.03 to 0.08	.41	−0.07	−0.17 to 0.02	.13	**0.31**	0.22 to 0.40	.00
4-Skills Scan (2015)												
Total group	**0.22**	0.13 to 0.30	<.001	0.00	−0.06 to 0.07	.89	−0.01	−0.08 to 0.07	.83	**0.20**	0.10 to 0.29	<.001
Junior grades (age 5 to 9)	**0.21**	0.10 to 0.33	<.001	0.01	−0.08 to 0.10	.86	0.06	−0.03 to 0.15	.17	**0.18**	0.06 to 0.31	<.01
Senior grades (age 9 to 12)	**0.16**	0.04 to 0.28	<.01	0.00	−0.08 to 0.08	.97	−0.05	−0.17 to 0.07	.42	**0.32**	0.17 to 0.47	<.001
M-ABC-2 (gross)												
4-Skills Scan (2007)												
Total group	**0.23**	0.05 to 0.41	.02	0.10	−0.02 to 0.22	.09	**0.32**	0.18 to 0.47	.00	**0.36**	0.18 to 0.54	.00
Junior grades (age 5 to 9)	0.29	−0.31 to 0.88	.34	0.18	−0.08 to 0.43	.17	**0.33**	0.13 to 0.52	.00	**0.41**	0.06 to 0.75	.02
Senior grades (age 9 to 12)	**0.22**	0.02 to 0.42	.03	0.08	−0.05 to 0.21	.22	**0.29**	0.06 to 0.53	.01	**0.35**	0.14 to 0.57	.00
4-Skills Scan (2015)												
Total Group	**0.44**	0.26 to 0.62	.00	**0.24**	0.10 to 0.38	.00	**0.19**	0.02 to 0.35	.02	**0.34**	0.13 to 0.55	.00
Junior grades (age 5 to 9)	**0.39**	0.10 to 0.67	<.01	**0.39**	0.16 to 0.62	<.001	**0.25**	0.04 to 0.46	.02	0.22	−0.08 to 0.53	.15
Senior grades (age 9 to 12)	**0.44**	0.19 to 0.70	<.001	0.16	−0.01 to 0.33	.07	0.04	−0.22 to 0.30	.75	**0.43**	0.11 to 0.76	<.01

β-coefficients, CI-95%, and P-values for motor lead per 4-Skill Scan subscale for the total group, junior grades, and senior grades, based on pooled results
Significant (P <.05) β-coefficients in bold; β: slope estimate; CI-95%: 95% confidence interval.

4.3 | Perspectives

The obtained findings indicate that the 4-Skills Scan is a reliable test for PE teachers to use in their lessons. Outcomes of the 4-Skills Scans can be considered a valid reflection of gross motor skills. Thus, using this test, PE teachers collect valuable information regarding gross motor skill levels. This helps setting up adequate PE lessons and follow-up trajectories for each individual child. In this sense, the current study can also be seen as an effort to support PE teachers to close the gap in gross motor skills between competent and less competent primary school children. This study demonstrated the value of using a panel of diverse experts as well. To our best knowledge, no study has managed to compare and report expert valuations with quantitative motor skills instrument before.

Future research could focus on determining cutoff values, discriminative ability, and norm scores for the 4-Skill Scans. In addition, further optimization of flexible assessment regarding testing time and feasibility would be valuable to the field of PE.

5 | CONCLUSION

In conclusion, moderate-to-strong positive correlations were found between the 4-Skills Scans and expert valuation, and the 4-Skills Scan and the M-ABC2. This indicates a sufficient construct validity and a good concurrent validity. Therefore, the 4-Skills Scan can be regarded as a valid instrument for assessing gross motor skills in primary school children. This is especially so when taking into account that during the PE lessons only a limited amount of time can be spent on the assessment itself and that children are being tested in a noisy environment.

ORCID

W. G. van Kernebeek http://orcid.org/0000-0003-1017-2415

REFERENCES

1. Gallahue D, Ozmun J, Goodway J. *Understanding Motor Development: Infants, Children, Adolescents, Adults*, 7th edn. New York, NY: McGraw-Hill Education; 2012.
2. Holfelder B, Schott N. Relationship of fundamental movement skills and physical activity in children and adolescents: a systematic review. *Psychol Sport Exerc*. 2014;15:382-391.
3. Lopes VP, Rodrigues LP, Maia JAR, Malina RM. Motor coordination as predictor of physical activity in childhood. *Scand J Med Sci Sports*. 2011;21:663-669.
4. Stodden DF, Goodway JD, Langendorfer SJD, et al. A developmental perspective on the role of motor skill competence in physical activity: an emergent relationship. *Quest*. 2008;60:290-306.
5. Barnett LM, Van Beurden E, Morgan PJ, Brooks LO, Beard JR. Childhood motor skill proficiency as a predictor of adolescent physical activity. *J Adolesc Health*. 2009;44:252-259.
6. Bellman M, Byrne O, Sege R. Developmental assessment of children. *BMJ*. 2013;346:e8687.
7. Cardon GM, De Bourdeaudhuij IMM. Physical education and physical activity in elementary schools in Flanders. *Phys Educ Sport Pedagogy*. 2002;7:5-18.
8. Shephard RJ, Trudeau F. The legacy of physical education: influences on adult lifestyle. *Pediatr Exerc Sci*. 2000;12:34-50.
9. Morgan PJ, Barnett LM, Cliff DP, et al. Fundamental movement skill interventions in youth: a systematic review and meta-analysis. *Pediatrics*. 2013;132:e1361-e1383.
10. Vygotsky LS. *Mind in Society*. Cambridge, MA: Harvard University Press; 1978:79-91.
11. Blank R, Smits-Engelsman B, Polatajko H, Wilson P. European academy for childhood disability (EACD): recommendations on the definition, diagnosis and intervention of developmental coordination disorder (long version): EACD Recommendations. *Dev Med Child Neurol*. 2012;54:54-93.
12. Polatajko H, Fox M, Missiuna C. An international consensus on children with developmental coordination disorder. *Can J Occup Ther*. 1995;62:3-6.
13. Rink JE. Measuring teacher effectiveness in physical education. *Res Q Exerc Sport*. 2013;84:407-418.
14. Van Gelder W, Stroes H. *Leerlingvolgsysteem Bewegen en Spelen. Over observeren, Registeren En Extra Zorg*, [Pupil tracking system Moving and Playing. About observing, registering, and extra care]. 2nd edn. Amsterdam, The Netherlands: Elsevier; 2010.
15. Stimuliz. Manual 4-Skills Scan, Stimuliz Help Centre 2015. Available at: http://support.stimuliz.nl/handleiding-4-s-en-testen. Accessed: September 2017.
16. Henderson SE, Sugden DA, Barnett AL. *Movement Assessment Battery for Children - second edition [Movement ABC-2]*. London: The Psychological Corporation; 2007.
17. Van Kernebeek WG, De Schipper AW, Savelsbergh GJP, Toussaint HM. Inter-rater and test–retest (between-sessions) reliability of the 4-Skills Scan for Dutch elementary school children. *Meas Phys Educ Exerc Sci*. 2018;22:129-137.
18. Ayres AJ. The development of perceptual-motor abilities: a theoretical basis for treatment of dysfunction. *Am J Occup Ther*. 1963;17:221-225.
19. Roley SS, Mailloux Z, Miller-Kuhaneck H, Glennon T. Understanding Ayres' sensory integration. *OT Pract*. 2007;12:CE1-CE8.
20. Rinne MB, Pasanen ME, Miilunpalo SI, Oja P. Test-retest reproducibility and inter-rater reliability of a motor skill test battery for adults. *Int J Sports Med*. 2001;22:192-200.
21. Van Kernebeek WG, De Kroon MLA, De Schipper AW, Savelsbergh GJP, Toussaint HM. Testing motor skills during

physical education classes in elementary school: The 4-Skills Scan. (submitted).

22. Surowiecki J. *The Wisdom of Crowds*. New York, NY: Anchor Books; 2004.

23. Brocken JEA. Examining the validity of the 4-Skills Scan: A pilot study. (Master's thesis). University of Amsterdam, The Netherlands; 2016. Retrieved from http://arno.uva.nl/document/644607

24. Chatfield SJ. A test for evaluating proficiency in dance. *J Dance Med Sci*. 2009;13:108-114.

25. Gorter H, Lucas C, Groothuis-Oudshoorn K, Maathuis C, Van Wijlen-Hempel R, Elvers H. The use of video clips in teleconsultation for preschool children with movement disorders. *Pediatr Phys Ther*. 2013;25:323-329.

26. Mooij C, Van Berkel M, Consten A, et al. *Basisdocument Bewegingsonderwijs Voor het Basisonderwijs [Handbook Physical Education in Primary Education]*, 6th edn. Nieuwegein, Arko Sports Media BV: The Netherlands; 2011.

27. Henderson SE, Sugden DA, Barnett AL. *Movement Assessment Battery for Children*. Kent, UK: The Psychological Corporation; 1992.

28. Brown T, Lalor A. The Movement Assessment Battery for Children - second edition (M-ABC-2): a review and critique. *Phys Occup Ther Pediatr* 2009;29:86-103.

29. Chambers ME, Sugden DA. The identification and assessment of young children with movement difficulties. *Int J Early Years Educ*. 2002;10:157-176.

30. Rosenblum S. The development and standardization of the Children Activity Scales (ChAS-P/T) for the early identification of children with Developmental Coordination Disorders. *Child Care Health Dev*. 2006;32:619-632.

31. Smits-Engelsman BC, Henderson SE, Sugden DA, Barnett AL. *Movement Assessment Battery for Children, Second Edition, Manual, Nederlandse Bewerking*. Amsterdam, The Netherlands: Pearson Assessment and Information BV; 2010.

32. Filemaker Pro (version 2014). Santa Clara, Canada: Filemaker Inc, 2015. Retrieved from https://www.filemaker.com

33. Qualtrics. Provo, UT: Qualtrics, 2015. Retrieved from www.qualtrics.com

34. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.

35. Eekhout I, De Vet HCW, Twisk JWR, Brand JPL, De Boer MR, Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol*. 2014;67:335-342.

36. Cools W, De Martelaer K, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. *J Sports Sci Med*. 2009;8:154-168.

37. Valentini NC, Ramalho MH, Oliveira MA. Movement Assessment Battery for Children-2: translation, reliability, and validity for Brazilian children. *Res Dev Disabil*. 2014;35:733-740.

38. Fransen J, D'Hondt E, Bourgois J, Vaeyens R, Philippaerts RM, Lenoir M. Motor competence assessment in children: convergent and discriminant validity between the BOT-2 Short Form and KTK testing batteries. *Res Dev Disabil*. 2014;35:1375-1383.

39. Vandorpe B, Vandendriessche J, Lefevre J, et al. The KörperkoordinationsTest für Kinder: Reference values and suitability for 6-12-year-old children in Flanders: KTK in 6-12-year-old children. *Scand J Med Sci Sports*. 2011;21:378-388.

40. Wuang YP, Su JH, Su CY. Reliability and responsiveness of the Movement Assessment Battery for Children-Second Edition Test in children with developmental coordination disorder. *Dev Med Child Neurol*. 2012;54:160-165.

APPENDIX 1

Floor plan of the experimental setup

Explanation of the obstacle course: (1) running and climbing over the vaulting boxes; (2) jumping of the vaulting box; (3) running around the bench; (4) slaloming around pylons; (5) hopping on floor mat; (6) walking over the wide and small balance bench; (7) swaying underneath the parallel bar; (8) throwing and catching different sized balls; (9) climbing on the vaulting box; (10) jumping of the vaulting box; and (11) performing a forward roll.

M-ABC-2

4-Skills Scan

jumping coordination

standing still

bouncing (ball)

jumping force

Obstacle course

2 jumping

climbing   1 running

Start

3

4 slaloming

CAM 1

7 swinging

6 balance

5 hopping

CAM 2

CAM 3

9 climbing   10 jumping   11

8 ball skills

Finish