Reliability and validity of children's basic motor skills testing: a systematic literature review

Fiabilidad y validez de las pruebas de habilidades motoras básicas en niños: una revisión sistemática de la literatura

Authors

Manshuralhudlori[1*]
Agus Kristiyanto[2]
Rony Syaifullah[3]
Febriani Fajar Ekawati[4]
Fadilah Umar[5]
Slamet Riyadi[6]

[1,2,3,4,5,6] Sebelas Maret University, Indonesia

Corresponding author:
Manshuralhudlori
manshuralhudlori87@staff.uns.ac.id

How to cite in APA

Manshuralhudlori, kristiyanto, A., RonySyaifullah, Ekawati, F., FadilahUmar, & SlametRiyadi. (2025). Reliability and validity of children's basic motor skills testing: a systematic literature review. *Retos*, *65*, 1156–1168. https://doi.org/10.47197/retos.v65.112787

Abstract

Background: Movement is considered natural part of life, because it affects cognitive, physical and social development. Physically active lifestyle supported by basic motor skills. There are several tests that can assess performance of basic motor skills.
Purposes: The research aims critically assess, compare, and summarize the quality of measurement properties of each child's fundamental movement skill test. Method: The MEDLINE, Pubmed, SciELO, Taylor & Francis, Science Direct, Springer, SAGE, ResearchGate, and EBSCOhost databases were searched for the period 2014-2024 using specific search terms for those relevant to children's fundamental movement skill testing. The COSMIN Risk of Bias (RoB) Checklist, COSMIN good measurement properties criteria and the GRADE approach were used to assess the quality of several children's fundamental movement skill tests.
Result: The search results resulted in 253 articles identified, 9 articles included. The 9 articles described construct, validity and/or reliability. The assessment of 9 tests has evidence of risk of bias extremely serious (n=3), very serious (n=1), serious (n=1), and no (n=4); evidence of grade high (n=4), moderate (n=1), low (n=1), and very low (n=3), and evidence of measurement property sufficient (n=1), insufficient (n=2), and indeterminate (n=6). Conclusion: MOBAK-3 is recommended as test of children's basic motor skills that has best evidence quality, because there is no bias risk in research methodology, high grade category, and sufficient quality of results.

Keywords

Children aged 4-13 years, fundamental movement skills, reliability, validity.

Resumen

Antecedentes: El movimiento se considera parte natural de la vida, porque afecta el desarrollo cognitivo, físico y social. Estilo de vida físicamente activo respaldado por habilidades motoras básicas. Existen varias pruebas que pueden evaluar el desempeño de las habilidades motoras básicas.
Propósitos: La investigación tiene como objetivo evaluar, comparar y resumir críticamente la calidad de las propiedades de medición de las pruebas de habilidades de movimiento fundamentales de cada niño. Método: Se realizaron búsquedas en las bases de datos MEDLINE, Pubmed, SciELO, Taylor & Francis, Science Direct, Springer, SAGE, ResearchGate y EBSCOhost durante el período 2014-2024 utilizando términos de búsqueda específicos para aquellos relevantes para las pruebas de habilidad de movimiento fundamental de los niños. La lista de verificación de riesgo de sesgo (RoB) de COSMIN, los criterios de buenas propiedades de medición de COSMIN y el enfoque GRADE se utilizaron para evaluar la calidad de varias pruebas de habilidad de movimiento fundamental de los niños.
Resultado: Los resultados de la búsqueda dieron como resultado 253 artículos identificados, 9 artículos incluidos. Los 9 artículos describieron el constructo, la validez y/o la confiabilidad. La evaluación de 9 pruebas tiene evidencia de riesgo de sesgo extremadamente grave (n=3), muy grave (n=1), grave (n=1) y nulo (n=4); evidencia de grado alto (n=4), moderado (n=1), bajo (n=1) y muy bajo (n=3), y evidencia de propiedad de medición suficiente (n=1), insuficiente (n=2) e indeterminada (n=6).
Conclusión: Conclusión: MOBAK-3 recomienda como prueba de habilidades motoras básicas de los niños que tiene la mejor calidad de evidencia.

Palabras clave

Fiabilidad, habilidades motoras fundamentales, niños de 4 a 13 años, validez.

Introduction

During childhood, motor skills need to be learned because they are said to be an aspect of movement from an early age as well as being involved in various sports (Kezić et al., 2020). Age influences the development of a person's basic movement skills. Running is the first highest movement development achieved at the age of 4 for boys, while for girls it is achieved at the age of 5 years. Boys can do the throw when they are 5 years old, while girls can do it at the age of 8.5 years. At the next age, i.e. 6.5 years, boys are able to skip while girls at the age of 6 are already able to skip. When girls are 6.5 years old, they can reach catching, while boys can only reach catching at the age of 7 years. There is quite a lot that boys can achieve at the age of 7, apart from catching they can also achieve kicking and striking. Entering 7.5 years they are reach hopping. Girls reach hopping at an earlier age, i.e. 7 years, but they are slower to reach kicking and striking, i.e. at 8 years and 8.5 years. Furthermore, at the age of 9.5 years, boys can achieve jumping, while girls can achieve this at the age of 10 years. This is the reason why age influences the development of a person's movement skills. Every time you get older, there are new movement skills that can be achieved (Jurimae, 2001). On the other hand, at 4 years the start of early childhood, then 7 years is the end of the early childhood stage, 9 years is the end of the middle childhood stage. There are profound behavioral changes in the first and second year transition at the end of childhood. Next, we are in adolescence which is marked by puberty as a significant life milestone. This is due to the release of hormones, either directly or indirectly, causing changes in behavior. This period occurs earlier in girls, i.e. at the age of 11 years, while boys enter it at the age of 13 years. Starting from the age of 4 years when boys are able to run, until the age of 13 years the start of puberty for boys. To identify this, a test or testing of the child's basic movement skills is needed (Payne & Isaacs, 2011).

Different tests to assess children's basic motor skill performance are available. Some tests may be specific to certain targets so that what is tested is more specific to that target group. This is because, it can be seen that every age is able to achieve certain movement skills. Apart from being specific to certain targets, some may be criteria or norm references in assessing each movement skill being tested. If a norm-referenced test is used, each child's results are compared with the norms of the normative group. Meanwhile, if you use a test with criteria, each child who is tested will have their movements assessed using specified criteria. In tests with the concept of achieving criteria, there are movement skill items that need to be carried out. Commonly used in assessing children's basic movement skills is the concept of how a skill is performed, or assessing the correct form of movement, or assessing the correct movement technique (Kezić et al., 2020). Children are often assessed using norm-referenced field tests and meet several criteria such as easy and quick administration, appropriate psychometric properties, and simplicity in calculating the final score (Ayan et al., 2019).

Previously, a review of 7 motor skill tests explored the potential use of these tests to assess movement performance (Cools et al., 2009). However, the assessment of methodological quality and test results has not been reported in previous reviews. In systematic reviews that refer to COSMIN regarding measurements, high-quality systematic reviews need to collect and evaluate information on both methodology and the relevance of the target population to the nature of the measurement (Mokkink et al., 2018a). Therefore, it is necessary to assess the methodological quality of a test (i.e. how the test will be assessed) and the quality of the test results (i.e. how the assessment results can prove that the test can be used), carried out separately. It is important to assess methodological quality first because it relates to the research results to be presented and minimizes the risk of bias. From the quality of the methodology, it can be seen what kind of process the researcher has carried out to compile the test items and what the researcher will do to assess the test items that have been prepared.

Therefore, the current systematic literature review aims to critically assess, compare, and summarize the methodological quality and measurement properties of each child's basic motor skills test so that it can recommend the best quality test that is able to assess children's basic motor skills.

Method

Design

This review was conducted using the systematic literature review method.

Eligibility criteria

Articles reviewed must adequately meet the inclusion criteria. Exclusion criteria excluded articles that were not included in the review. Eligibility criteria explain these two criteria. Inclusion criteria using the PICO framework (P-Participant; I-Intervention; C-Comparison; O-Outcome) must first be met. The PICO framework helps group search terms into thematic groups (Sayer, 2008).

1. Participants (P). Participants (P). The study was conducted on children aged 4 to 13 years as participants.
2. Intervention (I). Articles that contained basic motor skill testing and were applied to children. Only articles that covered all basic movement skills or one of the three basic categories of movement skills (locomotor/ballistic/manipulative) (Haywood & Getchell, 2014).
3. Comparison (C). Studied validity aspects and/or reliability aspects.
4. Outcome (O). The primary outcomes were validity aspects and/or reliability aspects. Secondary outcomes were assessments of the validity and reliability of each skill.

Second, an article was considered eligible for inclusion in the review with regard to the characteristics of the report such as language and reporting status (Page et al., 2021). Only English language articles and published status reports were included.

Exclusion criteria included: (1) Articles referring to the same report (duplicates; (2) Articles that did not discuss testing of children's basic motor skills; (3) Articles that did not explain the design of the test construct; (4) Articles whose primary purpose was not to evaluate the validity aspect and/or reliability aspect.

Literature search

The databases MEDLINE, Pubmed, SciELO, Taylor & Francis, Science Direct, Springer, SAGE, ResearchGate, and EBSCOhost were searched for the period 2014-2024 (the last 10 years) using specific search terms for those relevant to children's basic motor skills testing. The search for the last 10 years was intended to find newly developed tests. To assess children's basic movement skills, new observational assessment tools continue to be developed (Eddy et al., 2020). The keywords used in this search were "basic motor skill test", "basic motor skill testing", "basic movement skill test", "basic movement skill testing".

Study selection and data extraction

Five researchers (M-A, A-K, R-S, F-F-E, R-I-D) independently reviewed the title and abstract according to established criteria. If there are differences in perception regarding articles submitted to the full text selection stage, discussions are held until agreement is reached. After agreeing to select articles based on the title and abstract, the full text of the articles was then selected. Selection according to predetermined criteria. Discussions will be held again if there is a lack of perception regarding articles that have been removed or entered.

Evaluation of article quality

This section explains how the quality evaluation of articles was carried out by the review team using COSMIN guidelines.

The first evaluation carried out was an evaluation of methodological quality. The COSMIN Risk of Bias Checklist was used by the review team for this first evaluation. There are 4 scales used to assess methodological quality. The lowest value is inadequate, and the highest is very good. The two scales above inadequate are doubtful and adequate. The assessment concept used to evaluate methodological quality is the principle of "lowest or worst value" (Mokkink et al., 2018a).

The second evaluation carried out was an evaluation of the quality of the research results. COSMIN (measurement property) criteria for good measurement properties were used by the review team for this second evaluation. Each result submitted is compared with the standard criteria in COSMIN, then assessed using a scale. There are 3 scales used to assess the quality of results. If the article reports and the results do not comply with the standard COSMIN criteria, get a scale of (-) or insufficient. If the article reports and the results are in accordance with the COSMIN standard criteria, it gets a scale of (+) or

sufficient. The third scale is (?) or indeterminate if the article does not report the COSMIN criteria standards (Prinsen et al., 2018).

The third evaluation carried out was to summarize the quality evidence and assess it using the GRADE approach. There are 4 grades explained by COSMIN. The highest grade level is high, and the lowest grade is very low. The two grade levels above very low are low grade and moderate grade. The GRADE approach is carried out by lowering the grade by one level, two levels, or three levels if there is a risk of bias (Mokkink et al., 2018b).

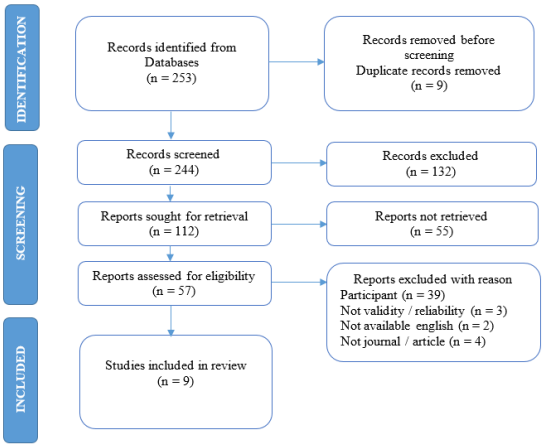
After evaluating the quality of the article, then provides recommendations for good testing. The recommendation criteria were tests/instruments that had sufficient (+) measurement properties and at least had low grade evidence quality. Tests/instruments with insufficient measurement properties and high grade evidence quality were not recommended. In addition to these two criteria, potential tests/instruments were recommended but required further evidence (Mokkink et al., 2018b).

Results

Include studies

The electronic search initially identified 253 articles for review. The 9 articles were excluded due to duplicates, 132 articles were excluded due to irrelevant titles, 55 were excluded due to the absence of the origin of the construct being measured, 39 articles were excluded because the study sample used was not appropriate and age data were not reported, 3 articles were excluded because they did not test aspects of validity or reliability, 2 articles were excluded because the full-paper was not in English, and 4 articles were excluded because they were not published in a journal, leaving 9 articles. Figure 1 shows the review process that resulted in the selection of 9 studies.

Figure 1. PRISMA Flow for a systematic review that includes a database search.



The included articles explored the construct being measured including the purpose of the test, study sample, origin of the construct, age of the test target, and skill design (item number). The nine basic motor skills tests include Battery Motor Tests (Popeska et al., 2015), Motorische Basiskompetenzen (MOBAK-1) instrument first grader (Hermann, Gerlach, & Seelig, 2015), Canadian Agility and Movement Skill Assessment (CAMSA) (Longmuir et al., 2017), Motorische Basiskompetenzen (MOBAK-3) instrument third grader (Hermann & Seelig, 2017), Bilateral Manipulative Skills Tests (Lovric et al., 2019), Performance and Fitness (PERF-FIT) test battery (Smits-Engelsman et al., 2020), Four Station Fundamental Motor Test (4-SFMT) (Leutar et al., 2023), Motor Skills Tests (Changsheng, 2024), and Selbstwahrnehmung motorischer Basiskompetenzen (SEMOK-1-2) first and second graders instrument (Bretz et al., 2024) (Table 1).

Table 1. Construct 9 tests of children's basic motor skills

No	Assessment Tool	Purposes	Research Samples	Origin of construct	Target age	Design of children's basic motor skills (Item number)
1	Battery Motor Tests (Popeska et al., 2015)	Assessing the motor skills and achievements of 7 year old children in grade 2 of elementary school	123 boys aged 7 years	The choice of tests was made based on the results of previous studies and the recommendations of previous researchers.	7 years old	8 items test Rolling, running, tapping, jumping, throwing medicine ball, throwing tennis ball, walking, and leading
2	MOBAK-1 (Herrmann, Gerlach, & Seelig, 2015)	Assessing the basic motor competencies of first graders	317 children Average age 7 years (144 boys and 173 girls)	Test items were compiled based on expert discussions (n=7) and in accordance with elementary school physical education standards.	6-7 years old	8 items test Locomotion: Balancing, Rolls, Jumping, Side stepping Object control: Dribbling, Bouncing, Catching, Throwing
3	CAMSA (Longmuir et al., 2017)	Evaluate children's motor skills and their ability to combine simple movement skills and more complex movement skills in response to environmental changes.	995 children (469 boys and 526 girls)	Test items were prepared based on Delphi expert review (n=19)	8-12 years old	7 movement skill tasks Jumping, sliding, catching, throwing, skipping, hopping, kicking
4	MOBAK-3 (Herrmann & Seelig, 2017)	Assessing the basic motor competencies of third graders	323 children Average age 9.2±0.39 years (158 boys and 165 girls)	Test items were prepared based on discussions with experts (n=6) and in accordance with the elementary school PE curriculum.	8-10 years old	8 items test Object control: Throwing, Catching, Bouncing, Dribbling Locomotion: Moving, Skipping, Balancing, Rolling
5	Bilateral Tests Manipulative Skills (Lovric et al., 2019)	Simultaneous assessment of manipulative skills in 7-year-old children	78 children Average age 7.34±0.53 years (43 boys and 35 girls)	Items adopted from fundamental movement skill manipulative literature	7 years old	3 items test Throwing, bouncing, dribbling
6	PERF-FIT test battery (Smits-Engelsman et al., 2020)	Assessing physical fitness related to motor skills in children in low-resource settings.	80 children aged 7-12 years	Items were compiled based on a literature search of the skill domain related to children's physical fitness. Then identified and reviewed by a panel of experts.	5-12 years old	9 items test Agility dan Power: Long jump, side jump Motor skill performance: Locomotor terdiri dari Jump, hop (right dan left), static balance (right dan left), dynamic balance Object control terdiri dari Bounce dan catch, throw dan catch, overhand throw
7	4-SFMT (Leutar et al., 2023)	Assessing motor skills of preschool and early school age children	30 children aged 5-6 years	Items are compiled based on expert panel discussions, designed according to minimum preschool physical education curriculum standards.	5-6 years old	4 items test Rolling (ROLL), pulling (PULL), pushing ball (BALL), climbing (CLIMB)
8	Motor Skills Tests (Changsheng, 2024)	Assessing preschool children's motor skills	391 children aged 5-6 years	Items are compiled based on expert discussions (n=17) and in accordance with child learning and development guidelines, preschool healthy physical fitness standards, preschool education outline guidelines. Journal and book literature are also used.	5-6 years old	9 items test 15.2m running Double legged jumps Jump on one leg Getting started with tennis throwing One-handed slapping ball Standing one leg eyes open Standing one leg eyes closed Walk the balance beam Walk backward
9	SEMOK-1-2 (Bretz et al., 2024)	Assessing PMC (Perceived Motor Competencies) of first and second grade children	404 children average age 7.8±0.69 years	Items refer to children's perceptions regarding the ability to perform basic motor needs.	6-9 years old	8 PMC test items Throwing, catching, bouncing, dribbling, balancing, rolling, jumping, running

Validity and reliability aspect

Each test is assessed for validity and reliability, or only for validity. As in the Battery Motor Tests, CAMSA, Bilateral Tests Manipulative Skills, and 4-SFMT instruments, their validity and reliability are assessed. While MOBAK-1, MOBAK-3, PERF-FIT test battery, Motor Skill Tests, and SEMOK-1-2 are only assessed for validity. The validity aspects measured are structural validity, hypothesis testing for construct validity, cross-structural validity, content validity, criterion validity. Structural validity is used in seven tests, i.e. the battery motor test (Popeska et al., 2015), MOBAK-1 (Hermann, Gerlach, & Seelig, 2015) MOBAK-3 (Herman & Seelig, 2017), bilateral tests of manipulative skills (Lovric et al., 2019), PERF-FIT (Smits-Engelsman et al., 2020), 4-SFMT (Leutar et al., 2023), and SEMOK-1-2 (Bretz et al., 2024). Several researchers report structural validity using confirmatory factor analysis. RMSEA and CFI are calculated. The reported RMSEA is between 0.023 to 0.037 while the reported CFI is between 0.940

to 0.980 (Hermann, Gerlach, & Seelig, 2015; Herman & Seelig, 2017; Bretz et al., 2024). Structural validity is also reported using exploratory factor analysis. One test, i.e. the battery motor test, only reports KMO (Popeska et al., 2015). Factor loading is clearly reported by two tests, i.e. PERF-FIT (Smits-Engelsman et al., 2020) and 4-SMFT (Leutar et al., 2023). While the bilateral manipulative test reports the factor structure matrix coefficient, unfortunately the factor loading is not reported (Lovric et al., 2019).

Hypothesis testing for construct validity is used in CAMSA and motor skills tests, both of which hypothesize age. Only CAMSA hypothesizes gender (Longmuir et al., 2017; Changsheng, 2024). Criterion validity is used in 4-SMFT and SEMOK-1-2. Both calculate the correlation with the gold standard test. TGMD-2 is used as the gold standard test in 4-SMFT, while MOBAK is used as the gold standard test in SEMOK-1-2 (Leutar et al., 2023; Bretz et al., 2024). Only MOBAK-1 uses cross-cultural validity to demonstrate measurement invariance (Hermann, Gerlach, & Seelig, 2015). Content validity is also only used in PERF-FIT by calculating CVI (Smits-Engelsman et al., 2020). The reliability aspects measured are intra-rater/inter-rater/test-retest and internal consistency reliability. Internal consistency is used in battery motor tests (Popeska et al., 2015) and bilateral manipulative tests (Lovric et al., 2019). Both calculate Cronbach alpha on each test component (Popeska et al., 2015; Lovric et al., 2019). Raters are used in CAMSA (Longmuir et al., 2017), 4-SFMT (Leutar et al., 2023), and bilateral manipulative tests (Lovric et al., 2019). Three raters were used in 4-SFMT (Leutar et al., 2023), seven raters were used in CAMSA (Longmuir et al., 2017), while the bilateral manipulative test did not explain how many raters were used (Lovric et al., 2019). All three calculate the ICC for each test component. Only 4-SMFT and CAMSA use inter-rater and intra-rater, while the bilateral manipulative test only uses inter-rater (Leutar et al., 2023; Longmuir et al., 2017; Lovric et al., 2019). The test-retest method is used in CAMSA and 4-SMFT. 4-SMFT uses a two-week time interval while CAMSA uses a time interval of 2-4 days and 8-14 days. Both also calculate ICC for each test component (Leutar et al., 2023; Longmuir et al., 2017) (table 2).

Table 2. Validity and reliability aspects

No	Assessment Tool	Validity	Result	Reliability	Result
1	Battery Motor Tests (Popeska et al., 2015)	Structural validity	KMO 0.78 for two balls slalom rolling KMO 0.78 for cries–cross running 4 x 5 m KMO 0.50 for both feet–tapping on wall KMO 0.50 for standing broad jump KMO 0.50 for throwing medicine ball 1 kg from standing position KMO 0.50 for walking on upturned Swedish bench KMO 0.68 for throwing tennis ball in vertical goal with arm KMO 0.70 for leading with long stick	Internal consistency	Cronbach's α = 0.98 for rolling Cronbach's α = 0.99 for running Cronbach's α = 0.93 for tapping Cronbach's α = 0.95 for jump Cronbach's α = 0.91 for throwing medicine ball Cronbach's α = 0.93 for walking Cronbach's α = 0.76 for throwing tennis ball Cronbach's α = 0.80 for leading
2	MOBAK-1 (Hermann, Gerlach, & Seelig, 2015)	Structural validity	Exploratory factor analysis (EFA) Model 1b (2 factor) F1 Balancing = 0.704 Rolls = 0.552 Jumping = 0.787 Sidestepping = 0.641 F2 Bouncing 0.706 Dribbling 0.408 Throwing 0.521 Catching = 0.643 Confirmatory factor analysis (CFA) Model 2a CFI = 0.96, RMSEA = 0.036. Model 2b CFI = 0.98, RMSEA = 0.024.		
		Cross-cultural	Invariant factor loading shows no significant difference $\chi 2 = 6.95$, p = 0.33		
3	CAMSA (Longmuir et al., 2017)	Construct validity	Older age ($p < 0.001$, $\eta 2 = 0.15$) and male gender ($p < 0.001$, $\eta 2 = 0.02$) were associated with higher total scores.	Inter-rater	ICC = 0.99 for completion time ICC = 0.69 for skill scores
				Intra-rater	ICC = 0.52 for skill scores ICC = 0.99 for completion time
				Test-retest	Test-retest results at completion time showed ICC = 0.84 in the 2-4 day interval, and ICC = 0.82 in the 8-14 day interval. Test-retest results on skill scores showed ICC = 0.46 in the 2-4 day interval, and ICC = 0.74 in the 8-14 day interval.
4	MOBAK-3 (Herrmann & Seelig, 2017)	Structural validity	Confirmatory factor analysis (CFA) Model 1a CFI = 0.97, RMSEA = 0.037, TLI = 0.96		

			Model 1b CFI = 0.97, RMSEA = 0.027, TLI = 0.97		
5	Bilateral Tests Manipulative Skills (Lovric et al., 2019)	Structural validity	Factor structure matrix coefficients Standing ball throwing = -0.97 - -0.99 Bouncing ball standing = -0.98 - -0.99 Dribbling the ball with foot = -0.96 - -0.99	Inter-rater	Measurement left Standing ball throwing ICC = 0.90-0.93 Bouncing ball standing ICC = 0.89-0.92 Dribbling ball with foot ICC = 0.84-0.92 Measurement right Standing ball throwing ICC = 0.88-0.91 Bouncing ball standing ICC = 0.93 Dribbling ball with foot ICC = 0.89-0.90
				Internal consistency	Measurement left Standing ball throwing Cronbach's α = 0.96 Bouncing ball standing Cronbach's α = 0.94 Dribbling ball with foot Cronbach's α = 0.94 Measurement right Standing ball throwing Cronbach's α = 0.98 Bouncing ball standing Cronbach's α = 0.98 Dribbling ball with foot Cronbach's α = 0.95
6	PERF-FIT test battery (Smits-Engelsman et al., 2020)	Content validity	Content validity index (CVI) Item throw dan catch CVI = 0.86 Other nine items CVI = 1.00 Total CVI = 0.99		
		Structural validity	Exploratory factor analysis (EFA) PCA with Varimax rotation Locomotor/balance: Jump = 0.746 Hop right = 0.770 Hop left = 0.721 Static balance right = 0.741 Static balance left = 0.528 Dynamic balance = 0.631 Object control/ball skills: Bounce dan catch = 0.797 Throw dan catch = 0.711 Overhand throw = 0.762 Agility/power: Long jump = 0.434 Side jump = 0.851 Running = -0.493 Stepping = -0.796		
7	4-SFMT (Leutar et al., 2023)	Structural validity	Bartlett's significant exploratory factorial analysis (EFA) Sphericity test X2 = 11.471, p = 0.075 Chi-square goodness-of-fit test determined that the one-factor model fit the data well, X2 (2) = 0.18, p = 0.916 Factor loading (faktor 1) ROLL = 0.80 PULL = 0.46 BALL = 0.81 CLIMB = 0.51	Inter-rater	ROLL ICC = 0.927 PULL ICC = 0.953 BALL ICC = 0.823 CLIMB ICC = 0.778 Total ICC = 0.914
		Criterion validity	Significant positive correlation between 4-SFMT total score and TGMD-2 score (r = 0.824, R2 = 0.679, t = 7.696 p < 0.001) Negative and significant correlation between 4-SFMT performance time and TGMD-2 scores (r = -0.652, R2 = 0.425, t = -4.574, p < 0.001)	Intra-rater	ROLL ICC = 0.890-0.932 PULL ICC = 0.751-0.911 BALL ICC = 0.714-0.742 CLIMB ICC = 0.703-0.846 Total ICC = 0.961-0.967
				Test-retest	ROLL ICC = 0.877 PULL ICC = 0.889 BALL ICC = 0.711 CLIMB ICC = 0.850 Total ICC = 0.973

8	Motor Skills Tests (Changsheng, 2024)	Construct validity	Using the percentile method, samples of different ages did not show significance (p>0.05)
9	SEMOK-1-2 (Bretz et al., 2024)	Structural validity	Model 1a CFI = 0.940, RMSEA = 0.032 Model 1b CFI = 0.969, RMSEA = 0.023
		Criterion validity	There is a correlation between MOBAK and SEMOK-1-2 in terms of object movement (r = 0.88, p < 0.001) and self-movement (r = 0.85, p < 0.001)

Study quality (instrument)

Using the COSMIN checklist, the lowest score is referred. Based on the Risk of Bias related to the research methodology, five tests are classified as having very good methodological quality in explaining the origin of the measured design construct (Hermann, Gerlach, & Seelig, 2015; Longmuir et al., 2017; Hermann & Seelig, 2017; Smits-Engelsman et al., 2020; Bretz et al., 2024) and four others have doubtful methodological quality in explaining the origin of the measured design construct (Popeska et al., 2015; Lovric et al., 2019; Leutar et al., 2023; Changsheng, 2024). Based on the COSMIN checklist for general design requirements, the four instruments actually have constructs that are clearly explained, the target population is also clearly explained, and the sample used also represents the target population. Unfortunately, there is one item that makes the instrument given a doubtful value on the design construct, i.e. the bilateral manipulative skills test does not explain the origin of the construct clearly (Lovric et al., 2019), battery motor tests and motor skill tests related to the context of use are not explained clearly (Popeska et al., 2015; Changsheng, 2014), and 4-SMFT was only carried out on a sample of 30 (men, n=18; women, n=12) (Leutar et al., 2023). From the validity aspect, four tests are classified as having very good methodological quality (Hermann, Gerlach, & Seelig, 2015; Longmuir et al., 2017; Hermann & Seelig, 2017; Bretz et al., 2024), three tests are classified as having inadequate methodological quality (Popeska et al., 2015; Lovric et al., 2019; Changsheng, 2024), and two tests are classified as having adequate methodological quality (Smits-Engelsman et al., 2020; Leutar et al., 2023). The reason for the inadequate methodological quality is because in assessing structural validity, there are methodological weaknesses such as inappropriate rotation methods and no exploratory and confirmatory analysis carried out on battery motor tests and bilateral tests of manipulative skills. Meanwhile, motor skills tests use hypothesis testing for construct validity, i.e. comparisons between subgroups, do not explain the important characteristics of the subgroups and do not explain the data analysis carried out. In fact, in assessing content validity, the quality of the methodology in PERF-FIT was given a very good score because professional experts from several scientific disciplines were included and the number of professional experts used also met the criteria (≥ 7 experts). Likewise with 4-SMFT, in assessing criterion validity, the quality of the methodology is given a very good value because it calculates correlation. Unfortunately, the methodological quality of both instruments was rated adequate because the sample size used in exploratory factor analysis was <100 (80 samples used in PERF-FIT and 30 samples used in 4-SMFT).

From the reliability aspect, only four tests report the reliability of the measured design construct. As a result, there are three tests that have very good methodological quality (Longmuir et al., 2017; Lovric et al., 2019; Leutar et al., 2023), while one test has doubtful methodological quality (Popeska et al., 2015). Doubtful marks are given due to minor methodological flaws. Next, an evaluation is carried out on the research results to determine the criteria for good measurement properties. Six tests do not meet the COSMIN requirements because there is an indeterminate value from one of the measurement areas because they do not report confirmatory factor analysis results (Popeska et al., 2015; Lovric et al., 2019; Smits-Engelsman et al., 2020; Leutar et al., 2023). Even though 4-SMFT reports very good methodological quality, unfortunately there are correlation criteria that are not met, i.e. a negative and significant correlation between 4-SFMT performance time and TGMD-2 scores (correlation value r = -0.652 < 0.70) (Leutar et al., 2023). An indeterminate value is also given because McFadden's R2 on MOBAK-1 is not calculated.

Although the methodological quality is very good and CFA is reported with RMSEA<0.06 and CFI>0.95 (Hermann, Gerlach, & Seelig, 2015). Because there is no hypothesis specified in the motor skills tests, it is given an indeterminate value (Changsheng, 2024). Two tests also do not meet the COSMIN requirements because there is an insufficient value from one of the measurement areas. Even though the results of construct validity are reported well and meet the criteria, unfortunately there is an ICC value <0.70

(ICC = 0.52 for intra-rater skill score; ICC = 0.69 for inter-rater skill score; ICC = 0.46 for test-retest skill score interval 2 -4 days) on CAMSA so it is given an insufficient value (Longmuir et al., 2017). SEMOK-1-2 was also given an insufficient value because the CFI value was <0.95 in model 1a confirmatory factor analysis (CFI = 0.940) (Bretz et al., 2024). Only one test, namely MOBAK-3, met the COSMIN requirements in structural validity results, where MOBAK-3 reported confirmatory factor analysis (CFA) results with CFI values>0.95 and RMSEA<0.06 in models 1a and 1b (table 3).

Table 3. Methodological Quality and Criteria for Good Measurement Properties

No	Instrument	Design	Validity	Methodological Quality	Criteria for Good Measurement Properties	Reliability	Methodological Quality	Criteria for Good Measurement Properties
1	Battery Motor Tests (Popeska et al., 2015)	Doubtful	Structural validity	Inadequate	? (Did not report results for confirmatory analysis factor)	Internal consistency	Doubtful	+ (cronbach's alpha value ≥ 0.70)
2	MOBAK-1 (Herrmann, Gerlach, & Seelig, 2015)	Very good	Structural validity	Very good	+ (Reported result of confirmatory analysis factor, with CFI>0.95 and RMSEA<0.06 at model 2a and model 2b			
			Cross-cultural	Very good	? (Reported no difference but McFadden's R2 not calculated)			
3	CAMSA (Longmuir et al., 2017)	Very good	Construct validity	Very good	+ (The results are in accordance with the hypothesis)	Reliability (inter-rater, intra-rater, test-retest	Very good	- There is an ICC value <0.70)
4	MOBAK-3 (Herrmann & Seelig, 2017)	Very good	Structural validity	Very good	+ Reporting the results of confirmatory factor analysis with CFI>0.95 and RMSEA<0.06 in models 1a and 1b.			
5	Bilateral Tests Manipulative Skills (Lovric et al., 2019)	Doubtful	Structural validity	Inadequate	? (Did not report the results of confirmatory factor analysis CFI and RMSEA)	Reliability (intra-rater)	Very good	+ (ICC value ≥0.70)
						Internal consistency	Very good	+ (Cronbach's alpha value ≥ 0.70)
6	PERF-FIT test battery (Smits-Engelsman et al., 2020)	Very good	Content validity	Very good				
			Structural validity	Adequate	? (Did not report the results of confirmatory factor analysis CFI and RMSEA)			
7	4-SFMT (Leutar et al., 2023)	Doubtful	Structural validity	Adequate	? (Did not report the results of confirmatory factor analysis CFI and RMSEA)	Reliability (inter-rater, intra-rater, test-retest	Very good	+ (ICC value ≥0.70)
			Criterion validity	Very good	- (Correlation criteria ≥0.70 is not met)			
8	Motor Skills Tests (Changsheng, 2024)	Doubtful	Construct validity	Inadequate	? (No hypothesis specified)			
9	SEMOK-1-2 (Bretz et al., 2024)	Very good	Structural validity	Very good	- (Reporting the results of confirmatory factor analysis, but model 1a does not meet the CFI>0.95 criteria)			
			Criterion validity	Very good	+ (Correlation with gold standard≥0.70)			

"+" = sufficient, "–" = insufficient, "?" = indeterminate

Discussion

The importance of basic motor skills is well known for children's development. Health is positively related to basic movement skills, participation in physical activities carried out by children is supported by children's well-developed basic movement skills. Because of the importance of the development of children's basic movement skills, many practitioners develop test items or tests to assess children's basic movement skills. The purpose of this systematic literature review is to critically assess, compare,

and summarize the quality of measurement properties of each test of children's basic motor skills. In general, nine tests published in the last 10 years (2014-2024) were revealed, to assess the quality of their measurement properties so that they can be the best choice when used to assess children's basic motor skills (Bremer & Cairney, 2018).

Design construct

Each test has a construct origin of how a series of task items are arranged. Some researchers use expert panel discussions and literature support such as elementary/preschool physical education curriculum standards, books, or previous research studies in compiling task items. Regarding the use of expert panels, test/task items need to be reviewed by experts. Reviews by other qualified individuals (experts) need to be used to study the test and provide constructive criticism. What may be clear or obvious may not be to others (Miller, 2010). Using experts aims to solicit input to prevent problems after administering the test. Several tests that use expert panels have been shown to have very good methodological quality because they are able to clearly explain the origin of the construct in the design aspects described in the COSMIN criteria (Herrmann, Gerlach, & Seelig, 2015; Longmuir et al., 2017; Herrmann & Seelig, 2017; Smits-Engelsman et al., 2020). However, there are several tests that, although using expert panels, have doubtful methodological quality in terms of design (Leutar et al., 2023; Changsheng, 2024). This is because the COSMIN criteria, namely a clear description given about the context of use, are not met. In line with this, test/item instructions must be clear and precise so as to prevent confusion for users (Miller, 2010).

Methodological quality and results reported

Good testing criteria include validity and reliability (Miller, 2010). Validity is the most important criterion to consider when evaluating a test, because it refers to the extent to which the test can measure what is measured. While reliability refers to the consistency of the test in measuring what is measured.

There are 44% of studies assessed both (validity and reliability) while 56% of studies only assessed validity. Commonly measured aspects of validity and reliability are structural validity (78%), hypotheses testing for construct validity (22%), criterion validity (22%), cross-structural validity (11%), content validity (11%), intra-rater/inter-rater/test-retest reliability (33%), and internal consistency reliability (22%). In the COSMIN guidelines, the validity and reliability aspects are also assessed to determine the quality of the measurement. In the validity aspect, each test assesses one or more of the validities in the COSMIN guidelines. Of the nine tests, only the PERF-FIT test battery evaluates content validity, and shows a very good methodology related to content validity (Smits-Engelsman et al., 2020). Only MOBAK-1 also evaluates cross-structural validity, and shows a very good methodology related to cross-structural validity. However, it has indeterminate results because McFadden's R2 was not reported (Hermann, Gerlach, & Seelig, 2015). According to the COSMIN checklist, using only McFadden's R2, can achieve sufficient results. Two tests evaluate criterion validity by determining correlation with other basic motor skill test standards, both showing very good methodology (Leutar et al., 2023; Bretz et al., 2024). Like 4-SMFT tested for correlation with TGMD-2, but the results were insufficient because a negative correlation was reported with a value of r = -0.652 (Leutar et al., 2023) while SEMOK-1-2 tested for correlation with MOBAK-1, the results were sufficient because it reported a correlation with a value of r > 0.70 (Bretz et al., 2024). Still related to the validity aspect, two tests evaluated construct validity with hypothesis testing. One tests have very good methodology and report sufficient results (Longmuir et al., 2017), the rest are inadequate with indeterminate results (Changsheng, 2024). Two tests evaluate construct validity with hypothesis testing on age and gender. Age is used because motor skills can develop at every age. Motor coordination performance in the older age group (10-12 years) was much better than the younger age group (7-9 years) ($p = 0.001$) even though both were obese (Mardiansyah et al., 2024). Meanwhile, gender is used because there are differences in the level of movement proficiency. One study reported significant differences between gender categories and motor stages.

Motor slowness is indicated by a low level of movement proficiency. Men show more effective motor criteria (Kucharski et al., 2021). One test using differences in age and gender has a very good methodology and reports sufficient results (Longmuir et al. 2017), the rest using age has an inadequate methodology with indeterminate results (Changsheng, 2024). Most evaluations using structural validity were

carried out on seven tests. Unfortunately, only those who use confirmatory factor analysis (CFA) in accordance with the COSMIN checklist can achieve very good quality, i.e. in three tests they have a very good methodology (Herrmann, Gerlach, & Seelig, 2015; Herrmann & Seelig, 2017; Bretz et al., 2024).

CFA is used because it shows that the construct design of the instrument has a clear and strong theoretical or empirical basis. Meanwhile, EFA is performed when there is no or only little prior knowledge about the data structure and needs to explore its dimensions and patterns. Both can be done separately or choose one or combine both sequentially. As MOBAK-1 uses sequential EFA and CFA, initially using EFA to determine factors and test items, two factors are generated with each factor consisting of four test items. Then CFA was carried out to confirm the validity and reliability of the model (Herrmann, Gerlach, & Seelig, 2015). Two tests only use EFA so they have adequate methodology (Smits-Engelsman et al., 2020; Leutar et al., 2023). The other two tests also only use EFA, but only report KMO (Popeska et al., 2015), and do not explain how many factors are produced (Popeska et al., 2015; Lovric et al., 2019).

Even though three tests used CFA according to COSMIN criteria, unfortunately only two tests reported sufficient results by reporting CFA results, i.e. CFI>0.95 and RMSEA<0.06 (Herrmann, Gerlach, & Seelig, 2015; Herrmann & Seelig, 2017). Regarding the reliability aspect, four tests evaluate reliability, only those that report Cronbach alpha values > 0.70 for internal consistency or ICC > 0.70 for reliability (inter-rater, intra-rater, test-retest) comply with the COSMIN checklist. Bilateral tests of manipulative skills evaluate intra-rater and internal consistency, have very good methodology for both types of reliability and report sufficient results with Cronbach's alpha values > 0.70 for internal consistency and ICC values > 0.70 for intra-rater (Lovric et al., 2019). Internal consistency is also reported to be sufficient with Cronbach's alpha values > 0.70 in the Battery Motor Tests (Popeska et al., 2015). Only CAMSA reported ICC values < 0.70 for intra-rater, inter-rater, and test-retest (Longmuir et al., 2017). The rest, 4-SFMT reported sufficient results with ICC values > 0.70 for inter-rater, intra-rater, and test-retest (Leutar et al., 2023).

Test recommendations children's basic motor skills

Based on the risk of bias, this systematic literature review assesses the grade of each test into the categories of high, moderate, low, and very low. Using the lowest score referred to, three tests have an extremely serious risk of bias, so the grade is lowered to very low. The extremely serious score was given because the quality of the research methodology was inadequate in terms of structural validity (Popeska et al., 2015; Lovric et al., 2019) and hypothesis testing for construct validity (Changsheng, 2024). There is one test, i.e. the 4-SFMT, has a very serious risk of bias, so the grade is lowered to low, because the quality of the research methodology is doubtful in the design (Leutar et al., 2023). There is one test, i.e. the PERF-FIT test battery, has a serious risk of bias, so the grade is lowered to moderate, because the quality of the research methodology is adequate in terms of structural validity (Smits-Engelsman et al., 2020). There are four tests that do not have a risk of bias, so the grade of the four tests remains in the high category, all of them have very good research methodology quality (Hermann, Gerlach, & Seelig, 2015; Longmuir et al., 2017; Hermann & Seelig, 2017); Bretz et al., 2024). After the quality of the evidence has been produced, it has been seen from the research methodology using the grade approach and evidence has been produced on the quality of the measurement property, seen from the reported research results. Then determine which testing recommendations have the best quality of evidence based on methodology and reported results. Finally, the evidence shows that MOBAK-3 is recommended as a test of children's basic motor skills because it has the best quality of evidence, with no risk of bias in the research methodology, included in the category with a high grade, and has sufficient quality results by reporting the results of confirmatory factor analysis according to COSMIN criteria.

Several tests such as Battery Motor Tests, MOBAK-1, Bilateral Tests Manipulative Skills, PERF-FIT test battery, 4-SFMT, and Motor Skills Tests have the potential to be recommended as tests of children's basic motor skills, but require further research to obtain adequate evidence in terms of methodology and reported results. While CAMSA and SEMOK-1-2 are not recommended, although in terms of research methodology there is no risk of bias, but the reported results do not comply with the COSMIN criteria, i.e. the ICC value <0.70 reported in CAMSA (Longmuir et al., 2017), CFI<0.95 was reported in SEMOK-1-2 (Bretz et al., 2014) (table 4).

Table 4. Test recommendations based on GRADE quality of evidence and measurement property

No	Instrument Name	Risk of Bias	Quality of envidence	Measurement Property	Recommendation

1	Battery Motor Tests (Popeska et al., 2015)	Extremely Serious	Very low	Indeterminate	Recommended, but needs further research
2	MOBAK-1 (Herrmann, Gerlach, & Seelig, 2015)	No	High	Indeterminate	Recommended, but needs further research
3	CAMSA (Longmuir et al., 2017)	No	High	Insufficient	Not recommended
4	MOBAK-3 (Herrmann & Seelig, 2017)	No	High	Sufficient	Recommended
5	Bilateral Tests Manipulative Skills (Lovric et al., 2019)	Extremely Serious	Very low	Indeterminate	Recommended, but needs further research
6	PERF-FIT test battery (Smits-Engelsman et al., 2020)	Serious	Moderate	Indeterminate	Recommended, but needs further research
7	4-SFMT (Leutar et al., 2023)	Very Serious	Low	Indeterminate	Recommended, but needs further research
8	Motor Skills Tests (Changsheng, 2024)	Extremely Serious	Very low	Indeterminate	Recommended, but needs further research
9	SEMOK-1-2 (Bretz et al., 2024)	No	High	Insufficient	Not recommended

Conclusions

There are several new observational tests that have been and continue to be developed to measure children's basic motor skills. A total of 9 articles producing 9 tests were considered. Overall, MOBAK-3 is recommended as a test of children's basic motor skills that has the best quality of evidence, with no risk of bias in the research methodology, included in the category with a high grade, and has sufficient quality results by reporting the results of confirmatory factor analysis according to the COSMIN criteria.

References

Ayan, C., Varela, S., Sanchez-Lastra, M. A., & De Quel, Ó. M. (2019). Reliability and validity of the basic motor ability test in preschool children. *Journal of Physical Education and Sport*, *19*, 987-991. http://dx.doi.org/10.7752/jpes.2019.s3142

Bremer, E., & Cairney, J. (2018). Fundamental movement skills and health-related outcomes: A narrative review of longitudinal and intervention studies targeting typically developing children. *American journal of lifestyle medicine*, *12*(2), 148-159. https://doi.org/10.1177/1559827616640196

Bretz, K., Strotmeyer, A., Seelig, H., & Herrmann, C. (2024). Development and validation of a test instrument for the assessment of perceived basic motor competencies in first and second graders: the SEMOK-1-2 instrument. *Frontiers in Psychology*, *15*, 1358170. https://doi.org/10.3389/fpsyg.2024.1358170

Changsheng, L. (2024). Comprehensive Method for Testing and Evaluating the Motor Skills of Children in Guangzhou City. *International Journal of Sociologies and Anthropologies Science Reviews*, *4*(3), 125-140. https://doi.org/10.60027/ijsasr.2024.4100

Cools, W., De Martelaer, K., Samaey, C., & Andries, C. (2009). Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. *Journal of sports science & medicine*, *8*(2), 154. https://pmc.ncbi.nlm.nih.gov/articles/PMC3761481/

Eddy, L. H., Bingham, D. D., Crossley, K. L., Shahid, N. F., Ellingham-Khan, M., Otteslev, A., Figueredo, N. S., Mon-Williams, M., & Hill, L. J. (2020). The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. *PloS one*, *15*(8), e0237919. https://doi.org/10.1371/journal.pone.0237919

Haywood M. Kathleen & Getchell, Nancy. (2014). *Life Span Motor Development*. Human Kinetics Publishers Inc.

Herrmann, C., Gerlach, E., & Seelig, H. (2015). Development and validation of a test instrument for the assessment of basic motor competencies in primary school. *Measurement in Physical Education and Exercise Science*, *19*(2), 80-90. https://doi.org/10.1080/1091367X.2014.998821

Herrmann, C., & Seelig, H. (2017). Structure and profiles of basic motor competencies in the third grade—validation of the test instrument MOBAK-3. *Perceptual and motor skills*, *124*(1), 5-20. https://doi.org/10.1177/0031512516679060

Jurimae, T., & Jurimae, J. (2001). *Growth, Physical Activity, and Motor Development in Prepubertal Children.* New York: CRC Press.

Kezić, A., Šimunović, I., & Kalinski, S. D. (2020). Application of the TGMD-2 test in early school-age children for determining the level of fundamental movement skills in different sports. *Journal of Physical Education and Sport*, *20*(2), 635-639. https://doi.org/10.46733/PESH2090041k

Kucharski, M. M., da Silva Pereira Filho, E., & Molina Saorin, J. (2021). The motor skill of kicking: relations between experience, sex and motor stage. *Retos*, *39*, 210–215. https://doi.org/10.47197/retos.v0i39.77790

Leutar, P. P., Krstulović, S., De Giorgio, A., & Kuvačić, G. (2023). Reliability and Validity of the Four Station Fundamental Motor Test (4-SFMT) for Assessing Motor Skills in Preschool Children—A Pilot Study. *Education Sciences*, *13*(3), 306. https://doi.org/10.3390/educsci13030306

Longmuir, P. E., Boyer, C., Lloyd, M., Borghese, M. M., Knight, E., Saunders, T. J., Boiarskaia, E., Zhu, W., & Tremblay, M. S. (2017). Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age. *Journal of sport and health science*, *6*(2), 231-240. https://doi.org/10.1016/j.jshs.2015.11.004

Lovric, F., Jelaska, I., Clark, C. C., Eyre, E. L., Miletic, D., & Mandic, P. (2019). Bilateral tests for the assessment of manipulative skills in children: development, reliability, and validity. *Human Movement*, *20*(3), 1-9. https://doi.org/10.5114/hm.2019.83991

Mardiansyah, A., Bakhtiar, S., Syafruddin, S., Syahputra, R., Pratama Putri, L., Atradinal, A., Mardela, R., Afrian, H., & Pion, J. (2024). Motor coordination in relation to weight status and age in primary school children in Indonesia. *Retos*, *57*, 72–79. https://doi.org/10.47197/retos.v57.104807

Miller, D. K. (2010). *Measurement by the Physical Educator: Why and How.* New York: Mc-Graw-Hill.

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018a). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*, 1171-1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C. W., & Terwee, C. B. (2018b). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs)-User manual (version 1.0). https://doi.org/10.1007/978-3-319-69909-7_2972-2

Payne, V. G., and Isaacs, L. D. (2012). *Human Motor Development: A Lifespan Approach, Eighth Edition.* New York: McGraw-Hill.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *bmj*, *372*. https://doi.org/10.1136/bmj.n160

Popeska, B., Mitevski, O., Jovanova-Mitkovska, S., & Sivevska, D. (2015). Battery of tests for evaluation of motor abilities at 7 years old children. *Личност, мотивација, спорт*, *20*, 280-293. https://eprints.ugd.edu.mk/14497/

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*, *27*, 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Sayers, A. (2008). Tips and tricks in performing a systematic review. *British Journal of General Practice*, *58*(547), 136-136. https://doi.org/10.3399/bjgp08X277168

Smits-Engelsman, B. C., Bonney, E., Neto, J. L. C., & Jelsma, D. L. (2020). Feasibility and content validity of the PERF-FIT test battery to assess movement skills, agility and power among children in low-resource settings. *BMC Public Health*, *20*, 1-11. https://doi.org/10.1186/s12889-020-09236-w

Authors' and translators' details:

Manshuralhudlori	manshuralhudlori87@staff.uns.ac.id	Autor/a
AgusKristiyanto kristiyanto	agus_k@staff.uns.ac.id	Autor/a
RonySyaifullah	ronysyaifullah@staff.uns.ac.id	Autor/a
FebrianiFajarEkawati Ekawati	febriani@staff.uns.ac.id	Autor/a
FadilahUmar	fadilahumar@staff.uns.ac.id	Autor/a
SlametRiyadi	slametriyadi70@staff.uns.ac.id	Autor/a