



A quick and qualitative assessment of gross motor development in preschool children

Tanja H. Kakebeeke¹ · Aziz Chaouch² · Elisa Knaier¹ · Jon Cafilisch¹ · Valentin Rousson² · Remo H. Largo¹ · Oskar G. Jenni^{1,3}

Received: 29 September 2018 / Revised: 17 January 2019 / Accepted: 17 January 2019 / Published online: 7 February 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

There is a need for a quick, qualitative, reliable, and easy tool to assess gross motor development for practitioners. The aim of this cross-sectional study is to present the Zurich Neuromotor Assessment-Q (ZNA-Q), which assesses static and dynamic balance in children between 3 and 6 years of age in less than 5 min. A total of 216 children (103 boys; 113 girls; median age 4 years, 4 months; interquartile range 1 year, 3 months) were enrolled from day-care centers, kindergartens, and schools, and were tested with 5 different gross motor tasks: standing on one leg, tandem stance, hopping on one leg, walking on a straight line, and jumping sideways. All ordinal measures (consisting of qualitative measures and scales) featured a marked developmental trend and substantial inter-individual variability. Test-retest reliability was assessed on 37 children. It varied from .17 for tandem stance to .43 for jumping sideways for the individual tasks, and it was .41 and .67 for the static and dynamic balance components, respectively. For the whole ZNA-Q, test-retest reliability was .7.

Conclusion: Ordinal scales enable practitioners to gather data on children's gross motor development in a fast and uncomplicated way. It offers the practitioner with an instrument for the exploration of the current developmental motor status of the child.

What is Known:

- Measurement of gross motor skills in the transitional period between motor mile stones and quantitative assessments is difficult.
- Assessment of gross motor skills is relatively easy.

What is New:

- Supplementary and quick gross motor test battery for children for practitioners.
- Normative values of five gross motor skills measured with ordinal scales.

Keywords Zurich Neuromotor assessment · Quick and qualitative version · Gross motor skills

Abbreviations

CAMs Contralateral associated movements
DB Dynamic balance

FM Fine motor
PM Pure motor
PMDA Poor man's data augmentation

Communicated by Mario Bianchetti

✉ Tanja H. Kakebeeke
tanja.kakebeeke@kispi.uzh.ch

Aziz Chaouch
aziz.chaouch@chuv.ch

Elisa Knaier
elisa.knaier@kispi.uzh.ch

Jon Cafilisch
jon.cafilisch@kispi.uzh.ch

Valentin Rousson
valentin.rousson@chuv.ch

Remo H. Largo
brlargo@bluewin.ch

Oskar G. Jenni
oskar.jenni@kispi.uzh.ch

- ¹ Child Development Center, University Children's Hospital Zürich, Steinwiesstrasse 75, CH-8032 Zürich, Switzerland
- ² Division of Biostatistics, Institute of Social and Preventive Medicine, University Hospital, Lausanne, Switzerland
- ³ Children's Research Center, University Children's Hospital Zürich, Zürich, Switzerland

SDS	Standard deviation score
SB	Static balance
ZNA	Zurich Neuromotor Assessment
ZNA-2	Zurich Neuromotor Assessment second edition

Introduction

In 2018, an updated version of the ZNA, the ZNA-2 [14], was published with new normative data for children's motor proficiency from 3 to 18 years. The ZNA-2 presents the developmental course and inter-individual variation of timed performance and quality of movements in a set of motor tasks of variable complexity for children between 3 and 18 years [14, 16, 17].

The greatest challenge for the ZNA-2 was to integrate data from children who were not able to perform the tasks because they were too young, performed too poorly, or both. Such missing data were integrated into the estimation of the outcome model with the poor man's data augmentation (PMDA) algorithm [26]. This method enabled the incorporation of those children who could not perform a task due to their young age.

However, sometimes, it occurs that an easier or similar version of the same task can be performed. For instance, a 3-year-old may not be able to stand on either leg for more than 2 s but can stand still for more than 2 s with one foot ahead of the other on a line. This task, called tandem stance [11], can provide information about ability in static balance when standing on one leg is not possible. Likewise, information about dynamic balance can be obtained by asking a child to walk forward on a straight line when the child is not able to perform a standing long jump. For this reason, some easier items for gross motor skills may help to evaluate these skills more precisely in the younger age group and thus gain an overview of the motor developmental status of the child.

In 2012, we presented a way of testing seven basic gross motor skills in preschool children with the purpose of bridging the gap between the measurement of motor milestones for toddlers and quantifying motor behavior in a competitive way [12]. Motor performance was quantified on an ordinal scale with 5 categories. Improvements over age were presented for walking on a beam, running, taking stairs, standing on one leg, hopping on one leg, rising, and jumping down. This investigation showed that developmental trends differed considerably over the seven tasks. For instance, hopping on one leg exhibited a highly significant trend over age, whereas for running, changes over age were non-significant. Therefore, downscaling gross motor tasks precisely required that some new items were added while those without a significant developmental trend were omitted. The revised version of this easier test for gross motor skills using ordinal scales was labeled ZNA-Q.

The ZNA-Q might provide a substantial advantage for practitioners (clinicians, physiotherapists, and sports teachers) who seek a quick overview of the developmental status of a child. The data from this 5 min test allow a rough estimate of the child's gross motor development. For instance, a 3-year-old child who is able to stand still one foot ahead of the other for more than 10 s and is able to hop on one leg more than 5 times does not immediately need a complete neuromotor assessment, which depending on the tool takes about 30 to 45 min [6, 9, 10, 14].

In this study, a limited set of ordinal measures for gross motor development in children below 6 years of age is explored. In using this tool in the future, we hope to detect quickly and easily children with large gross motor delays in the transitional period from the assessment of motor milestones to quantitative measurement of gross motor skills.

Materials and method

Participants

A total of 216 children (103 boys; 113 girls) between the ages of 3 and 6 years (median age 4 years, 4 months; interquartile range 1 year, 3 months) participated in this cross-sectional study. Additionally, 37 children participated in the study that investigated the test-retest reliability of the tool. All children were enrolled from day-care centers and kindergartens in the greater Zurich area. Children with evident medical or behavioral conditions (e.g., physically/mentally challenged) were excluded from the analysis. All the children lived in the Zurich area and together constitute a representative sample of the general local population. Special attention was given to sampling from districts with low, medium, and high socioeconomic backgrounds. The study was approved by the institutional review board of the Canton of Zurich (KEK-ZH-Nr. StV-40/07) and performed according to the Declaration of Helsinki. All families received a study description and provided written informed consent.

Measurements

The original ZNA is a standardized procedure that was specifically designed to describe neuromotor development in typically developing children from 5 to 18 years of age. It focuses on variability and age changes in motor proficiency [14, 16–18] by measuring the speed of motor tasks and the quality of movements (i.e., the intensity of contralateral associated movements, CAMs). Motor proficiency is measured on five components: fine motor adaptive tasks (FM-motor tasks with visuospatial perception), pure motor tasks (PM-motor tasks with as little visuo-perceptual challenges as possible), dynamic balance (DB), static balance (SB), and movement quality

(CAMs). The assessment of motor proficiency was later extended to children between 3 and 5 years of age using a customized version of the test [13].

In the updated version of the ZNA, the ZNA-2 [14], we use essentially the same items and components as in the original [18]. To take account of improvements with age in performance, the number of repetitions was increased for easier comparison of younger with older children. However, the task was sometimes too difficult for the younger age group. The FM and PM tasks could not be made easier, and we had sufficient data with a total of 9 tasks (3 × FM and 6 × PM).

However, for the gross motor tasks, especially SB, we had only one task. For this reason, tandem stance was added to the component of SB with the aim of providing a better estimation of the static balance competencies of a child. For DB, 3 ordinal scales with 5 levels ranging from 0 (best possible performance) to 4 (worst possible performance) were added to provide us with more information on poorly performing children. For consistency, performance on static balance was also expressed on an ordinal scale with 5 levels. This enabled us to describe the gross motor development on 5 different tasks and grouped these 5 items under the name ZNA-Q. All ZNA-Q items, the descriptions of the ordinal scales, and pictures of children doing the tasks are presented in Table 1 and Fig. 1. The ZNA-Q was performed during the same session as the ZNA-2.

Static balance

Standing on one leg (D/ND)

The child is asked to stand for as long as possible on one leg on a board marked with stripes. The stopwatch is started as soon as one foot is lifted and stopped when the child loses balance or shifts the standing foot beyond one stripe. The child performs the task first with the dominant (D) foot, thereafter the non-dominant (ND) foot. This routine is performed twice if the child does not reach the maximal time of 30 s during the first round. If the child performs the task twice, the longer time is used for subsequent calculations. The recorded time is then converted to an ordinal score (see Fig. 1).

Tandem stance

Tandem stance is a new item for the component SB. Tandem stance is a clinical measure of static balance considered to assess postural steadiness by a time measurement [11]. The child stands for as long as possible with one foot in front of the other, heel-to-toe. The stopwatch is started as soon as tandem stance is adopted and stopped when the child loses equilibrium or shifts at least one foot out of the correct position. As for standing on one leg, the recorded time is converted to an ordinal score (see Fig. 1).

Dynamic balance

Hopping on one leg (D/ND)

For hopping on one leg, the examiner stands in front of the child and demonstrates how to hop on one leg. The child has two chances to hop on each leg separately. A hop is considered successful when take-off and landing is achieved on the same foot.

Walking forward on a straight line

The examiner demonstrates to the child how to walk on a straight line of 4 m consisting of an elastic band stretched along the floor and the child is asked to do so likewise afterwards. The child has two chances to complete the task.

Jumping sideways

The examiner demonstrates to the child how to lift off the ground with two feet parallel to the elastic band, make a touch down and jump back. An important feature is lift off and touch down with two feet simultaneously. The child has two chances to complete the task.

Procedure

Subjects were tested in their own day-care center or kindergarten and recorded individually on digital video. The examiners were five experienced ZNA testers who had all been

Table 1 Items measured on ordinal scales of the Zurich Neuromotor Assessment-Q. D, dominant leg; ND, non-dominant leg

Components	Tasks	3–6 years
Static balance	Standing on one leg (D/ND)	Time in scales (0 to 4)
	Tandem stance (eyes open)	Time in scales (0 to 4)
Dynamic balance	Hopping on one leg (D/ND)	Count in scales (0 to 4)
	Walking on a straight line forward	Scale (0 to 4)
	Jumping sideways	Scale (0 to 4)

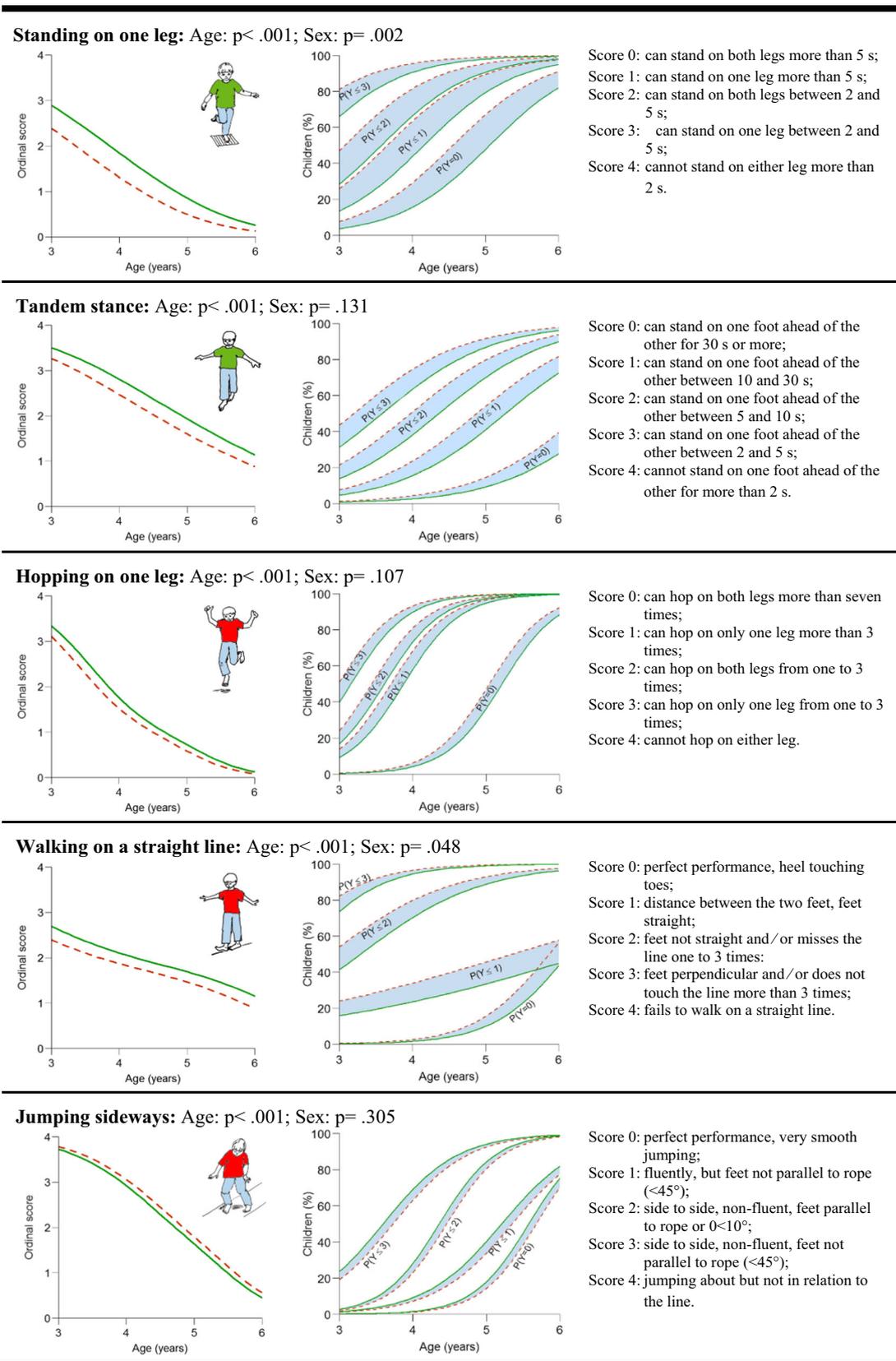


Fig. 1 Developmental trend of the five gross motor skills tasks of the ZNA-Q. Top row: expected ordinal score. Bottom row: cumulative probabilities to obtain a score below or equal to k , with $k \in \{0, 1, 2, 3\}$

trained and supervised by THK and JC. Tests always took place individually in a separate room. All tasks were performed in the same order by all children. The qualitative gross motor skills (SB and DB) were always tested in the same order in relation to the quantitative items. The examiner explained verbally and demonstrated how to perform the tasks. If the child did not understand a task and did something different, a second demonstration was provided. Total test time for the 5 items was less than 5 min.

Statistical analyses

Ordered data from the five motor tasks was modeled as a function of age and sex using a multinomial logit regression model (see [Appendix](#) for more details). Simpler models (e.g., models with partial or full proportional odds) were also investigated and the quality of their fit was compared using likelihood ratio tests. As in Kakebeeke et al. [12], the developmental trend was visualized by plotting the expected ordinal score as a function of age and sex and raw data were then converted into standard deviation scores (SDS) (see [Appendix](#) for more details). An SDS is a standardized measure of motor performance adjusted for age and sex that is approximately normally distributed with a zero mean and a unit variance in the normative sample. Positive values are associated with above-average performance and negative values with below-average performance. Component SDSs were calculated by summing SDSs from individual tasks and standardizing this sum to have zero mean and unit variance in the normative sample.

The test-retest reliability of individual task and component SDSs was estimated using Spearman's rank correlation coefficient. This coefficient provides a robust alternative to the intra-class correlation coefficient (ICC) [23] when the systematic error (i.e., bias due to a learning or fatigue effect) between two SDSs measured on the same child is not penalized. A 95% confidence interval (CI) for this rank correlation was calculated using Fisher's transformation [8].

Tasks for the ZNA-Q and the ZNA-2 were performed during the same session. This enabled us to investigate the association between SDSs from the two tests at the component level. However, we note that ordinal scores for the task standing on one leg in the ZNA-Q were not obtained from a separate assessment. Rather, they were obtained by applying a categorization to the time measurements observed in the same task in the ZNA-2. It follows that a naïve correlation analysis between SDSs from the two tests on the static balance component would produce biased results with enlarged correlation coefficients because the test-retest variability is not accounted for. In order to properly incorporate the test-retest variability into our correlation analysis, results from the ZNA-Q obtained at the first visit were correlated with those results from the ZNA-2 obtained at the second visit. Similarly, results from the ZNA-Q obtained at the second visit were correlated with

those results of the ZNA-2 obtained at the first visit. This was performed both for the SB and DB components. Each child thus contributed two pairs of measurements in the correlation analysis and the rank correlation between SDSs of the two tests was calculated after pooling the two pairs of measurements for each child. A 95% CI for this rank correlation was calculated using a non-parametric bootstrap approach by resampling individuals with replacement in the original dataset. This ensured a proper handling of the dependence in the data since each individual had two pairs of measurements.

Results

Modeling

Motor performance for standing on one leg, tandem stance, and hopping on one leg was adequately described by the full proportional odds model ($p \geq .547$; likelihood ratio test compared to a multinomial logit model). However, a partial proportional odds model was used for the tasks walking on a straight line and jumping sideways ($p \leq .015$; likelihood ratio test compared to a full proportional odds model). No evidence of a significant lack of fit was observed for any of the five tasks ($p \geq .548$).

Developmental course

Panels in the top row of Fig. 1 illustrate the evolution of the expected ordinal score calculated on the five tasks as a function of age and sex. Lower scores indicate better performance. All tasks featured a marked developmental trend ($p < .001$). With the exception of jumping sideways, girls performed better than boys in all gross motor tasks, but the gender difference was statistically significant only for standing on one leg ($p = .002$) and walking on a straight line ($p = .048$). Panels in the bottom row of Fig. 1 illustrate how the corresponding expected cumulative probability of obtaining a score below or equal to some value k , with $k \in \{0, 1, 2, 3\}$, varies as a function of age and sex.

Reliability

Test-retest reliabilities of all tasks are provided in Table 2 (see [Appendix](#) for more details) and were estimated from the data on 37 children between 3 and 6 years of age. Children were tested twice with just 1 week in between the two tests. The same experimenter did the two assessments and the same room was used in the child care centers and kindergartens. At the individual task level, the test-retest reliability was relatively low in this cohort, with rank correlation coefficients ranging from .17 for tandem stance to .43 for jumping sideways. However, at the component level, the rank correlation coefficients reached .41 for SB and .67 for DB, with the total

Table 2 Test-retest reliabilities (rank correlation coefficients) for single tasks and components of the ZNA-Q and rank correlations between components of the ZNA-Q and ZNA-2, with 95% confidence intervals (brackets)

	Test-retest reliability (<i>N</i> = 37)	Rank correlations with ZNA-2
Static balance		
Standing on one leg (D/ND)	.39 [.08; .64]	
Tandem stance	.17 [−.17; .47]	
Dynamic balance		
Hopping on one leg (D/ND)	.41 [.10; .65]	
Walking forward on a straight line	.39 [−.08; .64]	
Jumping sideways	.43 [.12; .66]	
Components		
Static balance	.41 [.09; .65]	.48 [.03; .68]
Dynamic balance	.67 [.44; .82]	.50 [.20; .75]
Total composite score	.70 [.48; .84]	.61 [.24; .75]

score of the whole test reaching .70 (95% CI [.48; .84]). Intra- and inter-tester reliability values on the items are available online in an earlier work of the same group (Table SIII; as referenced in [13]).

Correlation with SB and DB from the ZNA-2

The rank correlation between the composite scores for SB of the ZNA-Q and the ZNA-2 was estimated at .48 while the rank correlation for DB between ZNA-Q and ZNA-2 was .5. The rank correlation for the SB and DB scores together between the two tests was .61. Due to the limited sample size, the uncertainty associated with these estimates was large, with the lower bound of the 95% CI being .03, .20, and .24 for SB, DB, and the total score, respectively.

Discussion

Developmental trends for 5 gross motor tasks of typically developing children are reported in this study. These 5 tasks form the core of the ZNA-Q, a tool which enables a rough estimate of gross motor development in children between 3 and 6 years in a time of only 5 min. No similarly quick tool is currently used in pediatric practice, and we hope that this simplified test allows over- and underachievers to be identified more easily.

It is known that children undergo huge developmental leaps in their motor development between being a toddler and becoming a child. Below 2 years, motor development is described predominantly with motor milestones [4, 5, 21], which can only inform us about the first occurrence of certain

tasks. Two commonly used assessment tools provide meaningful measurements from age three [10, 13]. However, these measurements take at least half an hour to complete and are therefore not practical for screening purposes. Other instruments cannot be used as they do not include children of age three [6, 9, 15]. An investigation into the construct of motor competence in early childhood used Rasch modeling as a test for motor proficiency in 4- to 6-year-old children [27]. This modeling can also be used to assess 3-year-olds [25]. However, even the short form consists of 12 items comprising fine motor tasks and tasks with balls and therefore taking too long. As we were aiming to make a quick developmental test on gross motor development for 3–6-year-olds, we could not use any of the existing tests.

With the ZNA-Q, we focus on the development of gross motor abilities, described by Burton and Miller [7] as general traits or capacities of an individual, as the basis of a variety of motor skills. Of these general traits, we are interested in gross locomotor skills used in a range of movements to transport the body from one location to another including standing [24]. While fine motor performance is also important for the global evaluation of motor abilities in infants [22], the assessment of fine motor performance is less complicated and easily testable with other tools [10, 14]. As we wanted to build a pre-screening tool for the quick assessment of motor performance, we chose to focus on gross motor abilities in the ZNA-Q. Our interest in the general traits of infants' gross motor development led us to apply the holistic model of movement competence proposed by Rudd et al. [22]. According to this model, locomotor, object control, and body coordination develop as intrinsic aspects of movement competence. However, several articles have reported that cultural background [2, 3] and sex [17, 20] have an impact on motor proficiency. In this article, we explore firstly the general traits of the child, independent of environment or sex, and secondly, the locomotor part (going from A to B) of motor development. Testing children below age six who are not yet competitive and are not very competent on tasks in which speed and distance are measured, requires a tool such as the ZNA-Q to bridge this gap between toddler and child. This implies that assessments which apply a product-oriented approach cannot be used; a process-oriented approach is likely to be more informative [19].

Earlier work by the same group on this topic [12] included some items that are no longer present in the ZNA-Q. The items rising, running, and jumping down do not distinguish sufficiently between children due to too little variability after age three, and for this reason, are not suitable for screening at this developmental age. Conversely, some other items were added as easier versions of tasks in the ZNA-2. Yet, because many preschool children, especially the 3-year-olds, are unable to perform some tasks (especially on DB), the use of ordinal scales allows the inability to perform to be treated simply as the lowest category of motor performance without relying on more

complex approaches such as that used in the ZNA-2 with the PMDA algorithm [14, 26].

The test-retest reliability of the total score in the ZNA-Q was estimated at .70, which we consider acceptable given the context, for two reasons. First, as a simplified and shorter version of the ZNA-2, the ZNA-Q contains fewer tasks and thus may be expected to have a lower test-retest reliability than the complete test battery, which was estimated at .8 [14]. However, one should recall that the ZNA-Q only focuses on gross motor skills while the ZNA-2 also includes fine and pure motor tasks and contralateral associated movements. The SB and DB components in the ZNA-2 were the least reliable dimensions, with test-retest reliabilities of .67 and .78 for SB and DB, respectively [14]. The test-retest reliability of the ZNA-Q is thus comparable to that observed for the corresponding dimensions of the ZNA-2. Secondly, and more importantly, the ZNA-Q only considers preschool children from 3 to 6 years old, while the ZNA-2 was developed for children and adolescents from 3 to 18 years of age. Preschool children are naturally not very stable in their behavior [1], and such behavioral instability has a direct impact on the variability of motor performance. This is especially true for those as young as age three. As a consequence, the test-retest reliability in this population of very young children is anticipated to be lower than that measured in older children and adolescents. It should be noted that, if we exclude 3-year-olds ($n = 4$) from our sample, the test-retest rank correlation increases somewhat, from .70 to .74. This effectively supports the fact that motor performance in younger children is inherently more difficult to assess than that of older children. With these considerations in mind, the test-retest reliability of the ZNA-Q appears to be broadly comparable to that of the ZNA-2, which is used in clinical practice. Moreover, we believe it is difficult to reasonably expect a better result when considering preschool children as young as 3 years of age. Nonetheless, additional investigations of the test-retest reliability of the ZNA-Q should be conducted on a larger sample, because our sample ($N = 37$) was limited in size, which resulted in very large confidence intervals for rank correlations.

The correlation of .61 between composite scores of the ZNA-Q and ZNA-2 indicates that the two tests measure some common traits of gross motor skills but do not necessarily capture the exact same information. This may be partly explained by the tasks included in and excluded from the ZNA-Q. However, we note that while the test-retest reliability measures the agreement between results obtained with the same child at two different occasions using the same test, the correlation of .61 between the ZNA-Q and the ZNA-2 also incorporates between-test variation. Despite this additional source of variability, the inter-test reliability thus appears fairly close to the test-retest reliability, which again means that we cannot expect to achieve a much higher inter-test reliability than we observed in this study. Nevertheless, although the ZNA-Q

may capture slightly different information from the SB and DB components of the ZNA-2, the clear developmental trends observed in all tasks incorporated in the ZNA-Q suggest that this test is applicable as a pre-screening tool for gross motor performance in preschool children to assess the current developmental status of the child in clinical practice.

The ZNA-Q was established with a cohort of typically developing children. Because this cohort included few children with gross motor problems or delay, future research should focus on children with motor difficulties and how reliably the ZNA-Q is able to identify them.

Acknowledgments We gratefully acknowledge the support of the presidents' conference of the public schools in Zurich and the educators and teachers of the child care centers and Kindergartens for their help with recruitment.

Authors' contributions THK: Designed the study, acquired funding, performed data collection, coded and analyzed the corresponding data, and wrote the manuscript

AC: Performed statistical analysis and modeling

EK: Performed data collection and coded the corresponding data

JC: Performed data collection and corrected the manuscript

VR: Assisted in statistical data analysis

RHL: Contributed to the draft and gave advice on the assessment tools

OGJ: Designed the study, acquired funding, corrected final draft of the manuscript

All authors reviewed and edited the manuscript

Funding This study was supported by the Swiss National Science Foundation, grant no. 32003B_153273, the Largo, Maiores and Giedion Risch Foundation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study. Parents provided written informed consent for their participating child and children consented orally.

Statistical appendix

Multinomial regression model

A multinomial regression was used to model observed ordinal scores for each task separately. More specifically and for any particular task, the random variable Y_i describing the ordinal score of child i (whose observed score is y_i) was modeled as a function of linear age and sex, such that

$$\log\left(\frac{p_i(k)}{1-p_i(k)}\right) = \beta_{0k} + \beta_{1k}\text{age}_i + \beta_{2k}\text{sex}_i \quad k \in \{0, 1, 2, 3, 4\}$$

where

$$p_i(k) = P(Y_i \leq k | \text{age}_i, \text{sex}_i) \quad (1)$$

refers to the (cumulative) probability to obtain a score smaller or equal to k given the age and sex of the child, and β_{0k} , β_{1k} , and β_{2k} are regression coefficients that depend on the score level k . Constrained versions of this model that contain fewer parameters were also fitted for comparative purposes. For example, a partial proportional odds model where the effect of sex is assumed constant across levels was fitted by letting $\beta_{2k} = \beta_2$ for all k . A full proportional odds model where the effect of both age and sex is constant across levels was also fitted by letting $\beta_{1k} = \beta_1$ and $\beta_{2k} = \beta_2$ for all k . Goodness-of-fit was assessed using a chi-square test by comparing observed and predicted frequencies in each ordinal category for different age groups (3–4, 4–5, 5–6 years) and sex.

Visualizing the developmental trend

The developmental trend was visualized by plotting the expected ordinal score

$$\hat{S}(\text{age, sex}) = \sum_{k=0}^4 k \cdot \hat{P}(Y = k | \text{age, sex})$$

as a function of age and sex, where $\hat{P}(Y = k | \text{age, sex})$ is the predicted probability (derived from the model) of obtaining a score k given age and sex. Additionally, the estimated cumulative probability

$$\hat{P}(Y \leq k | \text{age, sex}) = \sum_{j=0}^k \hat{P}(Y = j | \text{age, sex})$$

was also plotted as a function of age and sex.

Calculation of standard deviation scores (SDS)

For each particular task, an unbiased estimate of the percentile \hat{p}_i^* associated with the observed score y_i of child i was calculated as

$$\hat{p}_i^* = \begin{cases} 1 - \frac{\hat{p}_i(y_i)}{2} & y_i = 0 \\ 1 - \frac{\hat{p}_i(y_i - 1) + \hat{p}_i(y_i)}{2} & 0 < y_i \leq 4 \end{cases}$$

with $\hat{p}_i(k)$ defined in Eq. (1). This percentile was then converted into a standard deviation score (SDS) $\hat{z}_i = \Phi^{-1}(\hat{p}_i^*)$ with Φ^{-1} denoting the standard normal cumulative distribution function.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Anderson P (2002) Assessment and development of executive function (EF) during childhood. *Child Neuropsychol* 8(2):71–82. <https://doi.org/10.1076/chin.8.2.71.8724>
- Barnett L, Hinkley T, Okely AD, Salmon J (2013) Child, family and environmental correlates of children's motor skill proficiency. *J Sci Med Sport* 16(4):332–336. <https://doi.org/10.1016/j.jsams.2012.08.011>
- Barnett LM, Telford RM, Strugnelli C, Rudd J, Olive LS, Telford RD (2018) Impact of cultural background on fundamental movement skill and its correlates. *J Sports Sci*:1–8. <https://doi.org/10.1080/02640414.2018.1508399>
- Bax M, Gillberg C (2009) Development: normal/delayed/disordered. In: Aicardi J (ed) *Diseases of the nervous system in childhood*, 3rd edn. Mac Keith Press, London, pp 891–901
- Bayley N (2006) *Bayley scales of infant and toddler development*, 3rd edn. Pearson, San Antonio
- Bruininks RH, Bruininks BD (2005) *Bruininks-Oseretsky test of motor proficiency - second edition (BOT-2)*. Pearson, Minneapolis
- Burton AW, Miller DA (1998) *Movement skill assessment*. Human Kinetics, Leeds
- Fisher RA (1936) *Statistical methods for research workers*, 6th edn. Oliver and Boyd, Edinburgh
- Folio M, Fewell R (2000) *Peabody developmental motor scales-2. PRO-ED*, Austin
- Henderson SE, Sugden DA, Barnett AL, Petermann F, Bös K, Kastner J (2007) *Movement assessment battery for children - second edition (Movement ABC-2) - Deutschsprachige adaptation*. Harcourt Assessment, London
- Jonsson E, Seiger A, Hirschfeld H (2005) Postural steadiness and weight distribution during tandem stance in healthy young and elderly adults. *Clin Biomech (Bristol, Avon)* 20(2):202–208. <https://doi.org/10.1016/j.clinbiomech.2004.09.008>
- Kakebeeke TH, Locatelli I, Rousson V, Cafilisch J, Jenni OG (2012) Improvement in gross motor performance between 3 and 5 years of age. *Percept Mot Skills* 114(3):795–806. <https://doi.org/10.2466/1.13.25.PMS.114.3.795-806>
- Kakebeeke TH, Cafilisch J, Chaouch A, Rousson V, Largo RH, Jenni OG (2013) Neuromotor development in children. Part 3: motor performance in 3- to 5-year-olds. *Dev Med Child Neurol* 55(3):248–256. <https://doi.org/10.1111/dmcn.12034>
- Kakebeeke TH, Knaier E, Chaouch A, Cafilisch J, Rousson V, Largo RH, Jenni OG (2018) Neuromotor development in children. Part 4: new norms from 3 to 18 years. *Dev Med Child Neurol*. <https://doi.org/10.1111/dmcn.13793>
- Kiphard EJ, Schilling F (2017) *KTK - Körperkoordinationstest für Kinder*. Hogrefe Verlag, Bern
- Largo RH, Cafilisch JA, Hug F, Muggli K, Molnar AA, Molinari L (2001) Neuromotor development from 5 to 18 years. Part 2: associated movements. *Dev Med Child Neurol* 43(7):444–453
- Largo RH, Cafilisch JA, Hug F, Muggli K, Molnar AA, Molinari L, Sheehy A, Gasser ST (2001) Neuromotor development from 5 to 18 years. Part 1: timed performance. *Dev Med Child Neurol* 43(7):436–443
- Largo RH, Rousson V, Cafilisch JA, Jenni OG (2007) *Zurich neuromotor assessment*. AWE Verlag, Zurich
- Logan SW, Robinson LE, Wilson AE, Lucas WA (2012) Getting the fundamentals of movement: a meta-analysis of the effectiveness of motor skill interventions in children. *Child Care Health Dev* 38(3):305–315. <https://doi.org/10.1111/j.1365-2214.2011.01307.x>
- Morris AM, Williams JM, Atwater AE, Wilmore JH (1982) Age and sex-differences in motor-performance of 3 through 6 year old children. *Res Q Exerc Sport* 53(3):214–221. <https://doi.org/10.1080/02701367.1982.10609342>
- Piper MC, Darragh J (1994) *Motor assessment of the developing infant*. Saunders, Philadelphia
- Rudd J, Butson ML, Barnett L, Farrow D, Berry J, Borkoles E, Polman R (2016) A holistic measurement model of movement competency in children. *J Sports Sci* 34(5):477–485. <https://doi.org/10.1080/02640414.2015.1061202>
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428

24. Ulrich DA (2000) Test of gross motor development, 2nd edn. Pro-ED. Inc., Austin
25. Utesch T, Bardid F, Huyben F, Strauss B, Tietjens M, De Martelaer K, Seghers J, Lenoir M (2016) Using Rasch modeling to investigate the construct of motor competence in early childhood. *Psychol Sport Exerc* 24:179–187. <https://doi.org/10.1016/j.psychsport.2016.03.001>
26. Wei G, Tanner M (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J Am Stat Assoc* 85(411):699–704
27. Zimmer R, Volkamer M (1987) Motoriktest für vier- bis sechsjährige Kinder [Motor Test for four- to six-year-olds]. Beltz Test, Weinheim