

Effect of timing of school enrollment on physical fitness in third graders

Thea Fühner (✉ fuehner@uni-potsdam.de)

University of Potsdam

Urs Granacher

University of Potsdam

Kathleen Golle

University of Potsdam

Reinhold Kliegl

University of Potsdam

Research Article

Keywords:

Posted Date: December 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1124241/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Timing of school enrollment may vary considerably due to children who were enrolled early / delayed or skipped / repeated a school year. Accordingly, the age range within school grades includes older-than-keyage (OTK) and younger-than-keyage (YTK) children. Studies examining academic performance of OTK and YTK children versus keyage children reported large differences in academic performance. However, there are no studies available that examined physical fitness of OTK and YTK versus keyage children. The aim of this study was to compare physical fitness of OTK (N = 25,086) and YTK (N = 1,885) children versus keyage children (N = 108,295) in a sample of German third graders. Physical fitness tests comprised cardiorespiratory endurance, coordination, speed, lower, and upper limbs muscle power. Predictions of physical fitness performance for YTK and OTK children were estimated using data from keyage children by taking age, sex, school, and cohort into account. The difference between observed and predicted z-scores yielded a *delta* z-score that was used as a dependent variable in the linear mixed models. Findings indicate that OTK children showed impaired performance compared to keyage children, especially in coordination, and that YTK children outperformed keyage children, especially in coordination. Teachers should be aware that despite their age advantage, OTK children often show impaired fitness compared to keyage children.

Introduction

The importance of physical fitness for children's health is undisputed¹. According to Caspersen et al.², physical fitness can be categorized as health- (e.g., cardiorespiratory endurance, muscular endurance, muscular strength, body composition, and flexibility) or skill-related fitness (e.g., agility, balance, coordination, speed, [muscle] power, and reaction time). There is evidence from original research³, systematic reviews, and meta-analyses^{4,5} that cardiorespiratory endurance and muscular strength are positively associated with markers of physical health (e.g., body mass index, waist circumference, skinfold thickness, cardiovascular disease risk score) in youth. Accordingly, it is important to regularly monitor and evaluate children's physical fitness to timely identify potential adverse health events. Recent studies on global secular declines in different measures of physical fitness (e.g., cardiorespiratory endurance) additionally illustrate the importance of fitness testing^{6,7}.

Physical fitness tests represent an easy-to-administer, reliable, and valid means to assess and evaluate children's physical fitness in large scale studies conducted in sport clubs or schools⁸. Several studies from around the globe⁸⁻¹⁷ showed developmental increases in physical fitness from childhood to adolescence^{8,10-12,15,17}. Irrespective of age, boys outperform girls in most components of physical fitness⁹⁻¹⁷, except for balance^{10,17} and flexibility^{9-15,17}.

The available studies on physical fitness development have been conducted in youth aged 5-18 years. In these studies, children and adolescents were matched into whole-year age groups. This age group classification can lead to an age-difference between children of almost one year¹⁸. Assuming that the

cut-off day is January 1st, children who are born on January 1st or slightly later are older than their peers who are born later during the year. These differences in the birthdate may have an impact on anthropometrics such as body height and mass as well as physical fitness (e.g., muscular strength, power, cardiorespiratory endurance, or speed)¹⁹. The consequences of this within-year age effect have particularly been examined in youth elite sports and is a well-known phenomenon called *Relative Age Effect* (RAE). The RAE describes the uneven distribution of young athletes with regards to their birthdates who were selected for regional or national squads. In other words, those athletes born near the cut-off date are overrepresented in team squads compared with children born later to the cut-off date²⁰. However, sporting talent should be equally distributed across the year and independent of birthdates. Thus, within a certain age-group the relatively older ones (born near to the cut-off date) outperform their relative younger peers (born later to the cut-off date) and are therefore more likely to be selected and become an elite athlete¹⁹.

Similar to the age-grouping system in elite sport, there is an age-grouping system in public schools as well. In general, timing of school enrollment is specific to the country under investigation. For instance, in the Federal State of Brandenburg, Germany the official school enrollment date is September 30th. Accordingly, children are enrolled to school if they are aged between 6.00 and 6.99 years on September 30th (i.e., keyage children). However, due to individual developmental processes, parents have the option to enroll a gifted child earlier or a slow developer later²¹. Further, gifted children can skip a school year or children with learning deficits may have to repeat a school year. According to official statistics of the Federal State of Brandenburg (Germany), 1% of primary school children start school earlier, i.e. are younger than 6.00 years²², 15% start school late, i.e., are older than 6.99 years²², and 1% must repeat a school year²³. All these aspects imply that the age range within school grades can vary considerably including older-than-keyage (OTK) and younger-than-keyage (YTK) children. A previous study has shown that in keyage children aged 8.00 to 8.99 years, physical fitness increased linearly with chronological age and that similar to the RAE even within their ninth year of life, older children (8.00 to 8.49 years) outperform younger (8.50 to 8.99 years) ones²⁴.

However, to the authors' knowledge there are no studies available that examined physical fitness of OTK and YTK children versus keyage children. Of note, studies examining academic performance of OTK and YTK children versus keyage children reported large differences in academic performance²⁵⁻²⁷. For instance, in a study including 1,144 German third graders, Urschitz et al.²⁷ reported that especially OTK children aged > 9.00 years compared with keyage children showed impaired academic performance (i.e., grades in mathematics, science, reading, spelling, and handwriting). In a study including 3,684 Australian high school students aged 14 years, Martin²⁶ reported that YTK children scored significantly better in academic performance (i.e., performance in literacy and numeracy) than keyage children. However, this has not yet been examined for physical fitness.

Therefore, the aim of this cross-sectional study was to compare physical fitness of OTK and YTK children versus keyage children in a sample of German third graders taking age, sex, school, and cohort into

account. With reference to the relevant school-based studies on differences in academic performance of OTK and YTK children versus keyage children²⁵⁻²⁷, we hypothesized that OTK children show impaired and YTK children better physical fitness compared with keyage children. These findings could be of interest for physical education teachers who aim to perform an individualized assessment of their student's physical fitness according to age and sex.

Methods

Experimental approach

This cross-sectional study is part of the ongoing EMOTIKON research project (www.uni-potsdam.de/en/emotikon). Physical fitness tests were conducted every year between September and November starting in 2011 and lasting until today. Physical fitness tests were also administered in 2009 and 2010, but later in the school year that is between March and April. Due to the seasonal variation in physical fitness these data were not included.

Population

Starting in 2009, all third graders living in the Federal State of Brandenburg, Germany are tested annually for their physical fitness. This cross-sectional study was mandated and approved by the Ministry of Education, Youth and Sport of the Federal State of Brandenburg, Germany. The Brandenburg School Law requires that parents are comprehensively informed prior to the start of the study. Consent is not needed given that the tests are obligatory for both, children and schools²¹.

Recent findings from the main study included a sample of 108,295 children from 515 different schools²⁴. The study sample comprised children who were enrolled within the legal key date of the Federal State of Brandenburg, Germany. In other words, children aged at least 6.00 and 6.99 years on September 30th are normally enrolled to school. In grade three, these keyage children are aged 8.00 to 8.99 years.

In this study, we focused on third graders who are younger or older than 8.00 to 8.99 years. Younger children are enrolled in third grade for reasons such as early enrollment into school or if they skip a grade (i.e., younger-than-keyage [YTK] children). In contrast, older children are enrolled in third grade for reasons such as delayed enrollment into school or if they repeat a grade (i.e., older-than-keyage [OTK] children). Due to limited sample sizes we were only able to include YTK children aged 7.50 to 7.99 years and OTK children aged 9.00 to 9.99 years. Children outside of these age ranges were excluded from study participation. We excluded children with signs of emotional (e.g., autism) and/or physical disorders (e.g., disabilities like infantile cerebral palsy). Furthermore, we applied a ± 3 SD criterion to detect extreme outliers. Data were taken from the annual statewide physical fitness testing of third graders between the years 2011 and 2019. Accordingly, 1,885 children constituted the group representing YTK children and 25,086 children the group representing OTK children.

Physical fitness tests

Physical fitness was assessed using the specific EMOTIKON test battery (www.uni-potsdam.de/en/emotikon/projekt/methodik - for further information on the test protocols). These tests evaluated cardiorespiratory endurance (i.e., 6-minute-run test), coordination (i.e., star-run test), speed (i.e., 20-m linear sprint test), lower (powerLOW [i.e., standing long jump test]), and upper limbs muscle power (i.e., powerUP [ball-push test]). The EMOTIKON test battery officially includes six tests. In 2016, the assessment of flexibility (i.e., stand-and-reach test) was stopped and the assessment of balance (i.e., single-leg balance test with eyes closed) was included. Due to the much smaller number of scores and their confound with cohort, these two tests were not included in the analyses.

Physical fitness tests were realized by qualified physical education teachers and conducted during the regular physical education classes. All teachers received standardized test instructions for the assessment (www.uni-potsdam.de/en/emotikon/projekt/methodik - for further information on the test protocols). Furthermore, all teachers participated in advanced training programs about standardized physical fitness assessment. Tests were always conducted in the morning between 8 and 12 am. Prior to testing, all third-graders performed a standardized warm-up program consisting of different running exercises (e.g., side-steps) and small games (e.g., playing tag).

Cardiorespiratory endurance

Cardiorespiratory endurance was assessed using the 6-minute-run test. Participating children had to run the furthest distance during the six minutes test time around a volleyball field (54 m) at a self-paced velocity. The test instructor provided split times every minute. After the six minutes, maximal distance covered in meters to the nearest nine meters was recorded and used as dependent variable. High test-retest reliability was reported for the 6-minute-run test with an intraclass correlation coefficient (ICC) of 0.92 in children aged 7 to 11 years²⁸.

Coordination

Coordination under time pressure was evaluated using the star-run test. During the star-run test, the participating children had to complete a parkour with different running techniques (i.e., running forwards, running backwards, side-steps) as quickly as possible (see Figure 1). The star shaped parkour (9 m x 9 m) consisted of four spikes. Each spike and the center of the star were marked with a pylon. The participants started in the middle of the star. First, they had to run forward to the first pylon and backward to the middle. Next, they had to do side-steps to the second pylon on the right side and side-steps back to the middle. Then, they had to run backward to the third pylon and forward to the middle. Finally, they had to do side-steps to the fourth pylon on the left side and side-steps back to the middle. The participants had to touch each pylon within the parkour with the hand. The whole covered distance was 50.912 m. Time for test completion in seconds to the nearest 1/10 second was taken using a stopwatch and used as dependent variable in the analysis. The participants had two test trials of which the best test trial in terms of time until test completion was kept for analysis. The star run test was reliable (test-retest) for children aged 8 to 10 years with an ICC of 0.68²⁹.

Speed

Speed was assessed using the 20-m linear sprint test. The participating children started from a standing position with one foot right behind the starting line. After an acoustic signal, they had to sprint as fast as possible over a distance of 20 m. Time for test completion in seconds to the nearest 1/10 second was taken using a stopwatch and used as dependent variable in the analysis. The participants had two test trials of which the best trial was taken for further analysis in terms of the time until test completion. Test-retest reliability has been reported to be high for children aged 7 to 11 years with an ICC of 0.90²⁸.

Lower limbs muscle power (PowerLOW)

PowerLOW was assessed through the standing long jump test. The participating children had to jump as far as possible from a frontal position. Arm swing prior to and during the jump was allowed. Jump distance in centimeters between the starting line and heel of the posterior foot was recorded to the nearest one centimeter using a measuring tape. The participants had two test trials of which the best trial in terms of the longest jump distance was taken for further analysis. The standing long jump test showed high test-retest reliability for children aged 6 to 12 years with an ICC of 0.94³¹.

Upper limbs muscle power (PowerUP)

Power up was evaluated with the ball-push test. From a standing position, the participating children had to push a 1 kg medicine ball that was held tight right in front of the chest. The participants had to push the ball at maximal effort with both hands. The pushing distance in meters was recorded to the nearest ten centimeters using a measuring tape. The participants had two test trials of which the best test trial in terms of the longest pushing distance was taken for further analysis. The ball-push test was reliable (test-retest) for children aged 8 to 10 years with an ICC of 0.81²⁹.

Statistics

Pre- and post-processing of data were carried out in the R environment of statistical computing³² using the *tidyverse* package³³. For measures of cardiorespiratory endurance (i.e., 6 min run test), powerLOW (i.e., standing long jump test) and powerUP (i.e., ball push test), higher scores indicate better physical fitness. For measures of coordination (i.e., star run test) and speed (i.e., 20-m linear sprint test), a Box-Cox distributional analyses indicated that a reciprocal transformation brought scores in line with the assumption of a normal distribution³⁴. Therefore, we converted scores from seconds to meters/seconds (i.e., pace scores; star run test = 50.912 [m] / time [s]; 20-m linear sprint test = 20 [m] / time [s]). These transformations had the advantage that a large value was indicative of good physical fitness for all five tests. Finally, we converted scores within tests (aggregated over boys and girls) to z-scores.

To compare physical fitness development of YTK and OTK children with that of keyage children, we generated predictions for the new data based on the Linear Mixed Model (LMM) parameters recently reported by Fühner et al.²⁴. Specifically, we predicted the performance of the YTK and OTK children, taking age, sex, school, and cohort into account. The prediction was carried out with the *MixedModels*

package³⁵ in the *Julia* programming language³⁶. The model parameters of the reference LMM comprised fixed effects for age, tests, sex, and their interactions, variance components (VCs) and correlation parameters (CPs) for GM and four test contrasts for the random factor child, VCs and CPs for GM, four test contrasts, sex, and age for the random factor school, and VCs for test and, age for the random factor cohort. Further details about model specification are provided in the reference paper²⁴ and in script: FKGG21.jl in the repository.

Through physical fitness testing in EMOTIKON, we obtained the actual physical fitness status of YTK children aged 7.50 to 7.99 years and OTK children aged 9.00 to 9.99 years, that is the children's physical fitness z-scores, computed with means and standard deviations from the reference population. The difference between observed and predicted z-scores yielded a *delta* z-score that was used as dependent variable in the following LMMs. If fitness performance of YTK and OTK children can be extrapolated from the performance of the reference population *delta* z-scores should not differ significantly from zero.

We analyzed the data with separate LMMs for OTK and YTK children. The starting LMM was similar to the one for the reference population. Specifically, there were four sequential-difference fixed-effect contrasts for the five tests: (H1) coordination vs. cardiorespiratory endurance, (H2) speed vs. coordination, (H3) powerLOW vs. speed, and (H4) powerUP vs. powerLOW. Also included were the effect of age (centered at 8.50 years) as a second-order polynomial trend as well as the quadratic trend of age, the effect of sex (boys – girls), and all interactions between contrasts, age, and sex. We used a two-sided z-value > 2.0 as significance criterion for the interpretation of fixed effects.

LMM for older-than-keyage (OTK) children

The LMM for OTK children was estimated with the *MixedModels* package³⁵ in the *Julia* programming language³⁶. The initial LMM included child (N = 25,086), school (N = 512), and cohort (N = 9) as three random factors; the total number of observations (i.e., max = 5 per child) was 121,745. With three random factors, there was a need for selecting a random-effect structure that included theoretically relevant and reliable variance components (VCs) and correlation parameters (CPs) but was also still supported by the data (i.e., was not overparameterized). Tests varied within children, schools, and cohorts; age and sex varied between children, but within schools and within cohorts. Therefore, in principle, VCs and CPs of linear and quadratic effects of age and sex could be estimated for schools and cohorts, but not for children.

Parsimonious model selection occurred in two major steps without knowledge or consideration of fixed-effect estimates³⁷; details are provided in script: FKGG21_late.jl in the repository. The random-effect structure of the parsimonious LMM of *delta* z-scores was expected to be simpler than the one for the LMM of the reference population²⁴ because the much smaller number of children and because many of the effects, VCs and CPs were already part of the predicted z-scores. We started with a model including Grand Mean (varying intercepts) for all three random factors and, given the relatively large numbers of 25,086 children and 512 schools and the small number of nine cohorts, we also included test-related VCs

and CPs for child and school and age-related and sex-related VCs and CPs for school, but not for the cohort. This LMM was well supported by the data. Increasing the complexity of the random-effect structure by adding test-related VCs for cohort did not improve the goodness of fit. As in Fühner et al.²⁴, we also estimated the final model with an alternative parameterization to obtain correlations between test scores rather than correlations between test contrasts.

In a final step, we fitted two control LMMs to test the significance of fixed effects of sex and sex x age interactions separately for each fitness component (i.e., we specified sex x age as nested within the five levels of the factor test).

LMM for younger-than-keyage (YTK) children

The LMM for YTK children was estimated with the *lme4* package³⁸ in the R environment of statistical computing³². Results were replicated in the *Julia* programming language³⁶. The LMM included child (N = 1,885), school (N = 407), and cohort (N = 9) as three random factors; the total number of observations (i.e., max = 5 per child) was 9,214. In the model selection process, we followed the model of OTK described above.

Parsimonious model selection occurred without knowledge or consideration of fixed-effect estimates³⁷; details are provided in script: `fggk21_early.rmd` in the repository. First, we applied the LMM of OTK to the data of YTK. This model was not supported by the data (i.e., overparameterized) because of the relatively small sample size of YTK (n = 1,885) compared to OTK (n = 25,086). Moreover, effects associated with the random-effect structure of the reference model were, we extensively reduced the complexity of the LMM by deleting the effects of sex, linear and quadratic age, the CPs for school and cohort in the random effects. This LMM was well supported by the data.

In a final step, we fitted control LMMs to test the significance of fixed effects of sex, linear and quadratic effects of age as well as the significance of CPs for school and the significance for cohort as a random factor.

Results

Figure 2. Observed z-scores for physical fitness development for boys (closed circles) and girls (open circles) aged 7.50 to 10.0 years. The lines represent the predicted z-scores for physical fitness development for boys (grey line) and girls (dashed grey line). Data were z-transformed. Endurance = cardiorespiratory endurance (i.e., 6-minute-run test), Coordination = star-run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball-push test). Note that *delta* z-scores for younger-than-keyage boys and girls were aggregated over 7.50 to 7.99 years and that *delta* z-scores for older-than-keyage boys and girls were aggregated over 9.50 to 9.99 years. Points are binned observed child means. Note that for coordination and speed, scores were converted from seconds to meters/seconds (i.e., pace scores; star-run test =

50.912 [m] / time [s]; 20-m linear sprint test = 20 [m] / time [s]). These transformations had the advantage that a large value is indicative of better physical fitness.

Figure 3. *Delta* z-score between observed and predicted physical fitness development for boys (closed circles) and girls (open circles) aged 7.50 to 10.0 years. Data were z-transformed. Endurance = cardiorespiratory endurance (i.e., 6-minute-run test), Coordination = star-run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball-push test). Note that *delta* z-scores for younger-than-keyage boys and girls were aggregated over 7.50 to 7.99 years and that *delta* z-scores for older-than-keyage boys and girls were aggregated over 9.50 to 9.99 years. Points are binned *delta* child means. Note that for coordination and speed, scores were converted from seconds to meters/seconds (i.e., pace scores; star-run test = 50.912 [m] / time [s]; 20-m linear sprint test = 20 [m] / time [s]). These transformations had the advantage that a large value is indicative of better physical fitness.

Physical fitness of older-than-keyage children (OTK)

Table 1 displays statistics for fixed effects of age (linear and quadratic) and sex as well as their interactions with the four test contrasts for LMM of OTK children.

Table 1

Fixed-effect estimates of linear mixed model for older-than-keyage (OTK) children

Source of variance	Fixed-effect estimates	Standard error	z-values
Main effects			
Grand mean (intercept)	0.404	0.067	6.00*
H1: coordination vs. endurance	0.275	0.103	2.67*
H2: speed vs. coordination	-0.001	0.095	-0.10
H3: powerLOW vs. speed	0.088	0.087	1.02
H4: powerUP vs. powerLOW	0.009	0.105	0.08
Age (linear)	-1.117	0.146	-7.65*
Age (quadratic)	0.398	0.075	5.32*
Sex	-0.284	0.131	-2.17*
Age (linear) x Sex	0.570	0.286	1.99
Age (quadratic) x Sex	-0.265	0.147	-1.80
Age (linear) x Fitness component			
H1: coordination vs. endurance	-0.765	0.225	-3.40*
H2: speed vs. coordination	0.082	0.207	0.40
H3: powerLOW vs. speed	-0.242	0.189	-1.28
H4: powerUP vs. powerLOW	-0.011	0.230	-0.05
Age (quadratic) x Fitness component			
H1: coordination vs. endurance	0.328	0.115	2.84*
H2: speed vs. coordination	0.014	0.106	0.13
H3: powerLOW vs. speed	0.085	0.097	0.87
H4: powerUP vs. powerLOW	0.036	0.118	0.31
Sex x Fitness component			
H1: coordination vs. endurance	0.142	0.205	0.69
H2: speed vs. coordination	-0.257	0.189	-1.36
H3: powerLOW vs. speed	0.121	0.173	0.70
H4: powerUP vs. powerLOW	-0.082	0.201	-0.39

Source of variance	Fixed-effect estimates	Standard error	z-values
Age (linear) x Sex x Fitness component			
H1: coordination vs. endurance	-0.443	0.449	-0.99
H2: speed vs. coordination	0.568	0.413	1.37
H3: powerLOW vs. speed	-0.370	0.379	-0.98
H4: powerUP vs. powerLOW	0.196	0.459	0.43
Age (quadratic) x Sex x Fitness component			
H1: coordination vs. endurance	0.210	0.231	0.91
H2: speed vs. coordination	-0.287	0.212	-1.35
H3: powerLOW vs. speed	0.197	0.195	1.01
H4: powerUP vs. powerLOW	-0.054	0.236	-0.23

H1 to H4 = hypothesis 1 to 4, endurance = cardiorespiratory endurance (i.e., 6 min run test), coordination = star run test, speed = 20-m linear sprint test, powerLOW = lower limbs muscle power (i.e., standing long jump test), powerUP = upper limbs muscle power (i.e., ball push test), * = z-value > 2.0, linear mixed model random factors: cohorts (9), schools (512), children (25,086), observations = 121,745 (missing = 2.9%). For estimates of variance components and correlation parameters see Table 2.

The overall negative linear trend for age ($z = -7.65$) and positive quadratic trend of age ($z = 5.32$) were significant. The positive quadratic trend of age indicates a plateau for the oldest children (see Figure 3). Furthermore, the main effect of contrast H1 was significant ($z = 2.67$) indicating that the main effect was larger for coordination than for cardiorespiratory endurance. The LMM tested the interactions of linear and quadratic age with the four test contrasts, that is whether slopes in neighboring panels in Figure 3 (averaged across sex) were parallel. The slope can be equated with the developmental rate. Indeed, one of four interaction was significant (see second and third block of Table 1): the linear age developmental rate was larger for cardiorespiratory endurance than coordination (H1; $z = -3.40$) and the quadratic age developmental rate was larger for coordination than cardiorespiratory endurance (H1; $z = -2.84$).

The overall sex effect was significant with $z = -2.17$. However, there were no significant interactions between sex and the tests contrasts (see fourth block of Table 1). Furthermore, neither the linear / quadratic age x sex interactions nor any of the interactions of these terms with the four test contrasts were significant (all $|z| \leq 1.99$).

Table 2

Variance components and correlation parameters for test scores for older-than-keyage (OTK) children

Fitness component	End	Coord	Speed	PowerLOW	PowerUP	Age	Sex	
Child - OTK								
Endurance	0.71	1.00						
Coordination	0.64	0.56	1.00					
Speed	0.68	0.64	0.69	1.00				
PowerLOW	0.73	0.62	0.70	0.83	1.00			
PowerUP	0.67	0.25	0.49	0.44	0.52	1.00		
School								
Endurance	0.01	1.00						
Coordination	0.02	0.44	1.00					
Speed	0.01	0.58	0.67	1.00				
PowerLOW	0.01	0.59	0.69	0.68	1.00			
PowerUP	0.00	-0.34	0.08	0.07	0.51	1.00		
Age	0.01	-0.30	-0.47	-0.52	-0.54	-0.24	1.00	
Sex	0.00	-0.26	-0.40	-0.28	-0.24	0.10	-0.56	1.00
Cohort (Intercept)	0.00							
End = cardiorespiratory endurance (i.e., 6 min run test), Coord = star run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball push test), VC = variance component, CP = correlation parameter, OTK = older-than-keyage children. Theoretically relevant correlations discussed in the text are set in bold . VC for Residual = 0.32.								

Table 2 lists estimates of VCs and CPs between tests scores for OTK children and for school. The test-related VCs were large for children (0.64 to 0.73) and small for schools (0.00 to 0.02). VCs for the age-related gains (slopes 0.01) and the sex effect (0.00) were also small for schools. In the first block of Table 2, CPs between tests scores for children are shown. In the second block, we list corresponding results for schools.

All child-related CPs between test scores were positive. Thus, the five tests represent a latent construct “physical fitness” for differences between children. However, the tests did not correlate equally highly with each other. Most notably, the CPs of cardiorespiratory endurance, coordination, speed, and powerLOW correlated very highly between 0.56 and 0.83, but their correlations with powerUP were distinctly smaller

(CPs: 0.25 to 0.52). For the school-related CPs one of them was negative (Endurance with PowerUP $r = -0.34$).

Physical fitness of younger than keyage (YTK) children

Table 3 displays estimates and test statistics for fixed effects of the four test contrasts. Figure 3 displays the *delta* z-scores between observed and predicted physical fitness development for YTK boys and girls aggregated over 7.50 to 7.99 years.

Table 3
Fixed-effect estimates of linear mixed model for younger-than-keyage (YTK) children

Source of variance	Fixed-effect estimates	Standard error	z-values
Main effects			
Grand mean (intercept)	0.071	0.015	4.90*
H1: coordination vs. endurance	0.077	0.025	3.09*
H2: speed vs. coordination	-0.057	0.023	-2.46*
H3: powerLOW vs. speed	0.015	0.021	0.72
H4: powerUP vs. powerLOW	0.086	0.026	3.33*

H1 to H4 = hypothesis 1 to 4, endurance = cardiorespiratory endurance (i.e., 6 min run test), coordination = star run test, speed = 20-m linear sprint test, powerLOW = lower limbs muscle power (i.e., standing long jump test), powerUP = upper limbs muscle power (i.e., ball push test), * = z-value > 2.0, linear mixed model random factors: schools (407), children (1,885), observations = 9,214 (missing = 2.2%). For estimates of variance components and correlation parameters see Table 4.

The grand mean was significant ($z = 4.90$). Furthermore, three of the four main effects of contrasts were significant: the main effect was larger for coordination than cardiorespiratory endurance (H1; $z = 3.09$), larger for coordination than speed (H2; $z = -2.46$) and larger for powerUP than powerLOW (H4; $z = 3.33$), which can also be seen in Figure 3.

Table 4

Variance components and correlation parameters for test scores for younger-than-keyage (YTK) children

	VC	CP				
Fitness component		End	Coord	Speed	PowerLOW	PowerUP
Child - YTK						
Endurance	0.46	1.00				
Coordination	0.52	0.57	1.00			
Speed	0.52	0.59	0.68	1.00		
PowerLOW	0.54	0.56	0.66	0.76	1.00	
PowerUP	0.49	0.25	0.44	0.43	0.51	1.00
School						
Endurance	0.01					
Coordination	0.01					
Speed	0.00					
PowerLOW	0.00					
PowerUP	0.02					
End = cardiorespiratory endurance (i.e., 6 min run test), Coord = star run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball push test), VC = variance component, CP = correlation parameter, YTK = younger-than-keyage (YTK) children. Theoretically relevant correlations discussed in the text are set in bold . VC for Residual = 0.27.						

Table 4 lists estimates of VCs between tests scores for children and for school as well as CPs between tests scores for YTK children. The test-related VCs were large for children (0.46 to 0.54) and small for schools (0.00 to 0.02). In the first block of Table 4, CPs between tests scores for children are shown. All child-related CPs between test scores were positive. Thus, the five tests represent a latent construct “physical fitness” for differences between children. However, the tests did not correlate equally high with each other. Most notably, the CPs of cardiorespiratory endurance, coordination, speed, and powerLOW correlated very highly between 0.56 and 0.76, but their correlations with powerUP were distinctly smaller (CPs: 0.25 to 0.51).

Discussion

The aim of this cross-sectional study was to examine physical fitness of YTK and OTK children versus keyage children in a sample of German third graders. Our findings indicate that (i) OTK children show

impaired performance compared to keyage children, especially in coordination, (ii) OTK girls outperform OTK boys, (iii) YTK children show better results than keyage children, especially in coordination, and (iv) four of the five tests represent a common construct (i.e., correlate strongly and positively with each other), except for the ball push test that tests upper limbs muscle power.

The child-related CPs between test scores were the same for YTK and OTK children. Consequently, the five tests represent a latent construct “physical fitness” for differences between children. However, the correlations with powerUP were distinctly smaller. Of note, the ball-push test is the only non-weight-bearing test within the test battery. In other words, the participants do not have to accelerate their own body mass during the test. It seems that foremost OTK children are able to compensate their delayed fitness or cognitive development through other factors which positively affect the performance in a non-weight-bearing test such as the ball-push test. It is possible that the onset of maturation associated with increased body mass, body height and muscle mass³⁹ may further the performance in the ball push test especially for OTK children. Ceschia et al.⁴⁰ examined associations between performance in physical fitness and the body mass index in 2,411 healthy Caucasian children aged 7 to 11 years. The analysis revealed that performance in weight-bearing tests (e.g., 500 m run test, 20-m linear sprint test, standing long jump test) was negatively affected by high body mass. In contrast, performance in non-weight-bearing tests (e.g., medicine ball forward throw test, handgrip test) was positively associated with high body mass, irrespective of sex. The results were confirmed by other studies^{41,42}. Therefore, it can be concluded that performance for upper limbs muscle power was positively influenced by high body mass. Unfortunately, anthropometric factors such as body mass were not assessed in this study which is why we were unable to confirm this hypothesis based on our own data.

Several studies confirmed a linear increase in physical fitness performance with age^{9-11, 15}. For instance, in a study with 424,328 Greek children and adolescents aged 6 to 18 years, Tambalis et al.¹⁵ reported a linear increase in physical fitness performance with age for cardiorespiratory endurance (i.e., 20-m shuttle run test), lower limbs muscle power (i.e., standing long jump test), flexibility (i.e., sit-and-reach test), muscular strength (i.e., sit-ups test), and agility (i.e., 10 × 5 m shuttle run test). The development of physical fitness of keyage children (see predicted gray lines in Figure 2) is in accordance with the above reported results. For keyage children, physical fitness performance increased linearly with age. However, the development of physical fitness for OTK children is different. Impaired performance was found in OTK children aged 9.00 to 9.99 years compared with keyage children (i.e., 8.00 to 8.99 years) for all components of physical fitness, especially for coordination. This could be due to the fact that third graders aged 9.00 to 9.99 years (i.e., OTK children) are not representative for the “average” third grader which is why we observed a deviation from the typically reported fitness development with age in this cohort^{9-11, 15}. We do not know the exact circumstances which lead to the delayed enrollment into first grade or to the repetition of a school year. According to our results, we can only speculate that maybe a delay in cognitive development might be the reason why children are late enrolled into first grade or must repeat a school year. Of note, the coordination test requires high levels of motor coordination. OTK children achieved the lowest scores compared with the other two groups in the coordination tests. This is

in line with a study by Urschitz et al.²⁷. In a cohort of 1,144 German third graders, these authors observed that impaired academic performance significantly increased with age for all school subjects (i.e., mathematics, science, reading, spelling, and handwriting) and that especially children who had repeated a school year were more prone to impaired academic performance. These results were confirmed by other studies^{25,26}. Interestingly, in our study OTK girls showed better performance compared to OTK boys which is in accordance with Urschitz et al.²⁷. These authors reported that except for mathematics, boys showed a larger prevalence for poor academic performance compared with girls²⁷. As girls mature approximately two years earlier than boys, the better performance of girls compared to boys might be influenced by biological maturation. Girls enter the adolescent growth spurt at approximately ten years of age and peak height velocity at 12 years, whereas boys enter the growth spurt on average at age 12 and peak height velocity at 14³⁹.

In contrast, YTK children outperformed keyage children especially in tests requiring motor coordination. Again, we do not know the exact circumstances which resulted in early enrollment into first grade or reasons for skipping a school year. According to our results, we speculate that accelerated cognitive development could be a reason why early enrolled children skip a school year. This is supported by the fact that in this study, YTK children showed the best performance in the coordination test which has an inherent large cognitive demand. Moreover, findings from Martin²⁶ point in a similar direction by showing that in a cohort of 3,684 Australian high school students, YTK children outperformed keyage children in academic performance.

Our study is not without limitations. First, anthropometric factors such as body mass, body height, and sitting height were not assessed in this study so that associations between anthropometric factors, biological maturation, and physical fitness could not be calculated. These factors would have provided additional insight as there is strong evidence that children's physical fitness is associated with anthropometric characteristics⁴⁰⁻⁴² and biological maturation³⁹. One explanation of the deviation of YKT and OKT children might be a difference between chronological and biological age. It appears plausible to argue that YKT children may be more mature and that OKT children are biologically somewhat younger than indicated by their chronological age. Thus, in a hypothetical plot of performance over biological age, the linear trend may well hold for all children. Second, we predicted the performance of the YTK and OTK children based on a linear extrapolation recently reported by Fühner et al.²⁴. However, we do not know if this linear extrapolation exactly fits to the data of keyage children aged 7.50 to 7.99 / 9.00 to 9.99 years as we do not have such cross-sectional data.

To sum up, this study is the first study that examined differences in physical fitness development of YTK and OTK children compared to keyage children. Our study findings complement results reported in the literature on the development of academic performance in youth²⁵⁻²⁷. Politicians and decision makers, schools, (physical education) teachers, and parents should be aware that despite their age advantage, OTK children often show impaired fitness compared to keyage children. Therefore, it is important to

specifically target this “minority” within the population of third graders to specifically offer health and fitness programs to compensate the gap in fitness experienced during prior years.

The EMOTIKON test battery is easy-to-administer, cost effective, requires only minimal school specific equipment (e.g., stopwatch, measuring tape, medicine ball, pylons), and the tests represent a common construct of physical fitness. Therefore, physical education teachers, coaches, or researchers can use the EMOTIKON test battery to evaluate children’s physical fitness. However, they should be careful with the interpretation of the ball push test since this test does not measure physical fitness to the same degree than the other tests.

Declarations

Availability of material, data, and code

The datasets generated and analyzed during the current study as well as Julia and R scripts are available in the Open Science Framework (OSF) repository: https://osf.io/dmu68/?view_only=240bdab8f1be4d8384acf9356ee50f8b

Funding

The study was commissioned and supported by the Ministry of Education, Youth, and Sport of the Federal State Brandenburg. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Reinhold Kliegl was supported by the Center for Interdisciplinary Research, Bielefeld (ZiF)/Cooperation Group "Statistical models for psychological and linguistic data".

Author contributions

TF, RK, and UG: made substantial contributions to conception and design; KG and TF: contributed to data collection; TF and RK: carried out data analysis; TF, RK, and UG: interpreted the data; TF: wrote the first draft of the manuscript and all authors were involved in revising it critically for important intellectual content; all authors provide final approval of the version to be published and agreed to be accountable for all aspects of the work and agreed with the order of presentation of the authors.

Competing Interests Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ortega, F. B., Ruiz, J. R., Castillo, M. J. & Sjörström, M. Physical fitness in childhood and adolescence: A powerful marker of health. *Int. J. Obes*, **32**, 1–11 (2008).

2. Caspersen, C., Powell, K. & Christenson, G. Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research CARL. *Notes Queries*, **100**, No. (2), 125–131 (1985).
3. Hussey, J., Bell, C., Bennett, K., O'Dwyer, J. & Gormley, J. Relationship between the intensity of physical activity, inactivity, cardiorespiratory fitness and body composition in 7-10-year-old Dublin children. *Br. J. Sports Med*, **41**, 311–316 (2007).
4. Mintjens, S. *et al.* Cardiorespiratory Fitness in Childhood and Adolescence Affects Future Cardiovascular Risk Factors: A Systematic Review of Longitudinal Studies. *Sport. Med*, **48**, 2577–2605 (2018).
5. García-Hermoso, A., Ramírez-Campillo, R. & Izquierdo, M. Is Muscular Fitness Associated with Future Health Benefits in Children and Adolescents? A Systematic Review and Meta-Analysis of Longitudinal Studies. *Sport. Med*, **49**, 1079–1094 (2019).
6. Fühner, T., Kliegl, R., Arntz, F., Kriemler, S. & Granacher, U. An Update on Secular Trends in Physical Fitness of Children and Adolescents from 1972 to 2015: A Systematic Review. *Sport. Med*, **51**, 303–320 (2021).
7. Tomkinson, G. R., Lang, J. J. & Tremblay, M. S. Temporal trends in the cardiorespiratory fitness of children and adolescents representing 19 high-income and upper middle-income countries between 1981 and 2014. *Br. J. Sports Med*, **53**, 478–486 (2019).
8. Niessner, C. *et al.* Representative Percentile Curves of Physical Fitness From Early Childhood to Early Adulthood: The MoMo Study. *Front. Public Heal*, **8**, 1–9 (2020).
9. Tomkinson, G. R. *et al.* European normative values for physical fitness in children and adolescents aged 9-17 years: Results from 2 779 165 Eurofit performances representing 30 countries. *Br. J. Sports Med*, **52**, 1445–1456 (2018).
10. De Miguel-Etayo, P. *et al.* Physical fitness reference standards in European children: The IDEFICS study. *Int. J. Obes*, **38**, S57–S66 (2014).
11. Santos, R. *et al.* Physical fitness percentiles for Portuguese children and adolescents aged 10-18 years. *J. Sports Sci*, **32**, 1510–1518 (2014).
12. Ortega, F. B. *et al.* Physical fitness levels among European adolescents: The HELENA study. *Br. J. Sports Med*, **45**, 20–29 (2011).
13. Catley, M. J. & Tomkinson, G. R. Normative health-related fitness values for children: Analysis of 85347 test results on 9-17-year-old Australians since 1985. *Br. J. Sports Med*, **47**, 98–108 (2013).
14. Roriz de Oliveira, M. S., Seabra, A., Freitas, D., Eisenmann, J. C. & Maia, J. Physical fitness percentile charts for children aged 6-10 from Portugal. *J. Sports Med. Phys. Fitness*, **54**, 780–792 (2014).
15. Tambalis, K. D. *et al.* Physical fitness normative values for 6–18-year-old Greek boys and girls, using the empirical distribution and the lambda, mu, and sigma statistical method. *Eur. J. Sport Sci*, **16**, 736–746 (2016).
16. Castro-Piñero, J. *et al.* Percentile values for running sprint field tests in children ages 6-17 years: Influence of weight status. *Res. Q. Exerc. Sport*, **81**, 143–151 (2010).

17. Dobosz, J., Mayorga-Vega, D. & Viciano, J. Percentile values of physical fitness levels among polish children aged 7 to 19 years – a population-based study. *Cent. Eur. J. Public Health*, **23**, 340–351 (2015).
18. Sandercock, G. R. H. *et al.* Quantification of the relative age effect in three indices of physical performance. *J. Strength Cond. Res*, **27**, 3293–3299 (2013).
19. Cobley, S., Baker, J., Wattie, N. & McKenna, J. Annual age-grouping and athlete development: A meta-analytical review of relative age effects in sport. *Sport. Med*, **39**, 235–256 (2009).
20. Musch, J. & Grondin, S. Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. *Dev. Rev*, **21**, 147–167 (2001).
21. Ministerium für Bildung Jugend und Sport. Gesetz über die Schulen im Land Brandenburg. <https://bravors.brandenburg.de/gesetze/bbgschulg> (2018)
22. Ministerium für Bildung Jugend und Sport. Einschulungen nach Schulform, Art der Einschulung und Schuljahr. https://mbjs.brandenburg.de/media_fast/bb2.a.5978.de/z_einsch2_2020_21.pdf (2021)
23. Ministerium für Bildung Jugend und Sport. Wiederholer (Nichtversetzte oder freiwillige Wiederholer) nach Schulstufe, Jahrgangsstufe, Schulform und Schuljahr. https://mbjs.brandenburg.de/media_fast/bb2.a.5978.de/z_wied_P_2020_21.pdf (2021)
24. Fühner, T., Granacher, U., Golle, K. & Kliegl, R. Age and sex effects in physical fitness components of 108,295 third graders including 515 primary schools and 9 cohorts. *Sci. Rep.* **11**, (2021)
25. Jaekel, J., Strauss, V. Y. C., Johnson, S., Gilmore, C. & Wolke, D. Delayed school entry and academic performance: A natural experiment. *Dev. Med. Child Neurol*, **57**, 652–659 (2015).
26. Martin, A. J. Age Appropriateness and Motivation, Engagement, and Performance in High School: Effects of Age Within Cohort, Grade Retention, and Delayed School Entry. *J. Educ. Psychol*, **101**, 101–114 (2009).
27. Urschitz, M. S. *et al.* Snoring, intermittent hypoxia and academic performance in primary school children. *Am. J. Respir. Crit. Care Med*, **168**, 464–468 (2003).
28. Bös, K. *Deutscher Motorik-Test 6-18*. (Czwalina, 2009)
29. Schulz, S. *The reliability of the Star coordination run and the 1-kg medicine ball push - physical fitness tests used in the EMOTIKON-study* (University of Potsdam, 2013).
30. Golle, K., Muehlbauer, T., Wick, D. & Granacher, U. Physical fitness percentiles of german children aged 9-12 Years: Findings from a longitudinal study. *PLoS One*, **10**, 9–11 (2015).
31. Fernandez-Santos, J. R., Ruiz, J. R., Cohen, D. D., Gonzalez-Montesinos, J. L. & Castro-Pinero, J. Reliability and validity of tests to assess lower-body muscular power in children. *J. Strength Cond. Res*, **29**, 2277–2285 (2015).
32. Team, R. C. R a language and environment for statistical computing. <https://www.r-project.org/> (2020)
33. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw*, **4**, 1686 (2019).
34. Box, G. E. P. & Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B*, **26**, 211–243 (1964).

35. Bates, D. *et al.* JuliaStats/MixedModels.jl:v3.4.0. vol. 1 20071787–20071787
<https://doi.org/10.5281/zenodo.4589101> (2021)
36. Bezanson, J., Edelman, A., Karpinski, S., Shah, V. B. & Julia A fresh approach to numerical computing. *SIAM Rev*, **59**, 65–98 (2017).
37. Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. Parsimonious Mixed Models. *ArXiv:1506.04967*
<http://arxiv.org/abs/1506.04967> (2015)
38. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, (2015)
39. Malina, R. M., Bouchard, C. & Bar-Or, O. *Growth, Maturation, and Physical Activity* (Human Kinetics, 2004).
40. Ceschia, A. *et al.* Deleterious effects of obesity on physical fitness in pre-pubertal children. *Eur. J. Sport Sci*, **16**, 271–278 (2016).
41. Sacchetti, R. *et al.* Physical fitness of primary school children in relation to overweight prevalence and physical activity habits. *J. Sports Sci*, **30**, 633–640 (2012).
42. Deforche, B. *et al.* Physical fitness and physical activity in obese and nonobese Flemish youth. *Obes. Res*, **11**, 434–441 (2003).

Figures

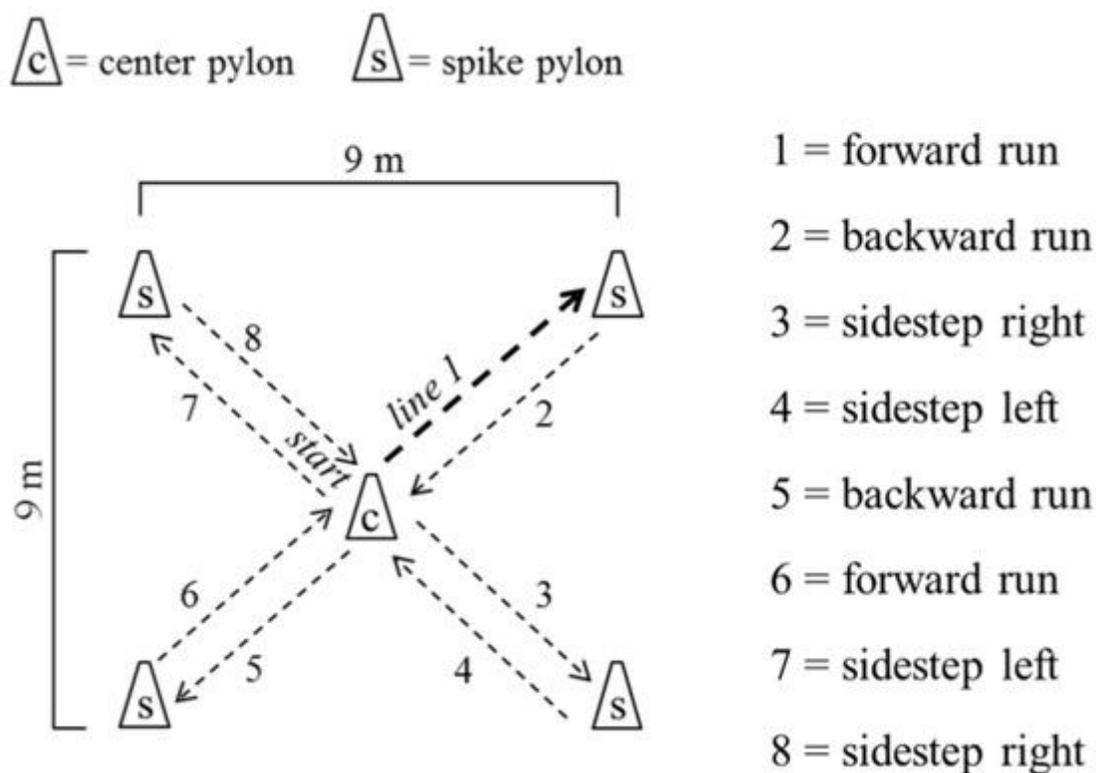


Figure 1

Schematic description of the star-run test (adapted from Golle et al.30).

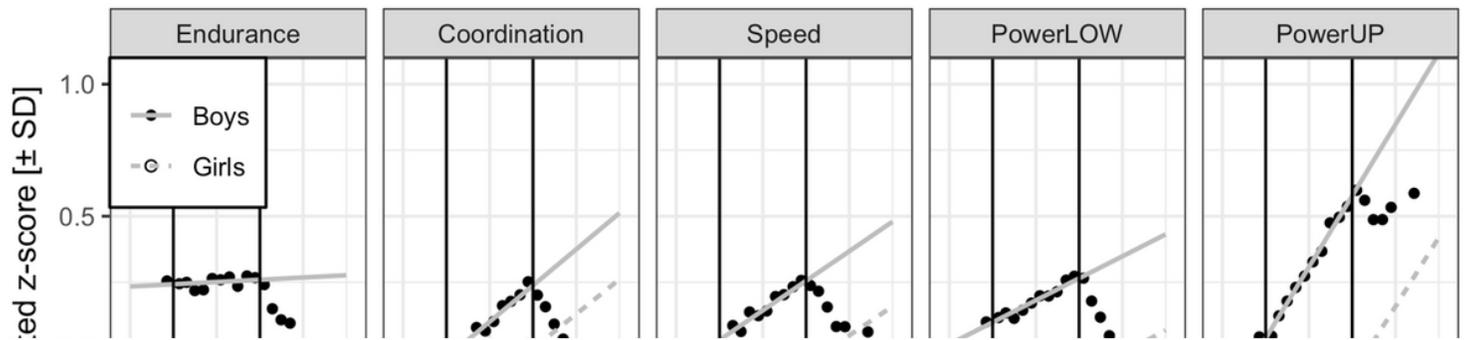


Figure 2

Observed z-scores for physical fitness development for boys (closed circles) and girls (open circles) aged 7.50 to 10.0 years. The lines represent the predicted z-scores for physical fitness development for boys (grey line) and girls (dashed grey line). Data were z-transformed. Endurance = cardiorespiratory endurance (i.e., 6-minute-run test), Coordination = star-run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball-push test). Note that delta z-scores for younger-than-keyage boys and girls were aggregated over 7.50 to 7.99 years and that delta z-scores for older-than-keyage boys and girls were aggregated over 9.50 to 9.99 years. Points are binned observed child means. Note that for coordination and speed, scores were converted from seconds to meters/seconds (i.e., pace scores; star-run test = $50.912 \text{ [m]} / \text{time [s]}$; 20-m linear sprint test = $20 \text{ [m]} / \text{time [s]}$). These transformations had the advantage that a large value is indicative of better physical fitness.

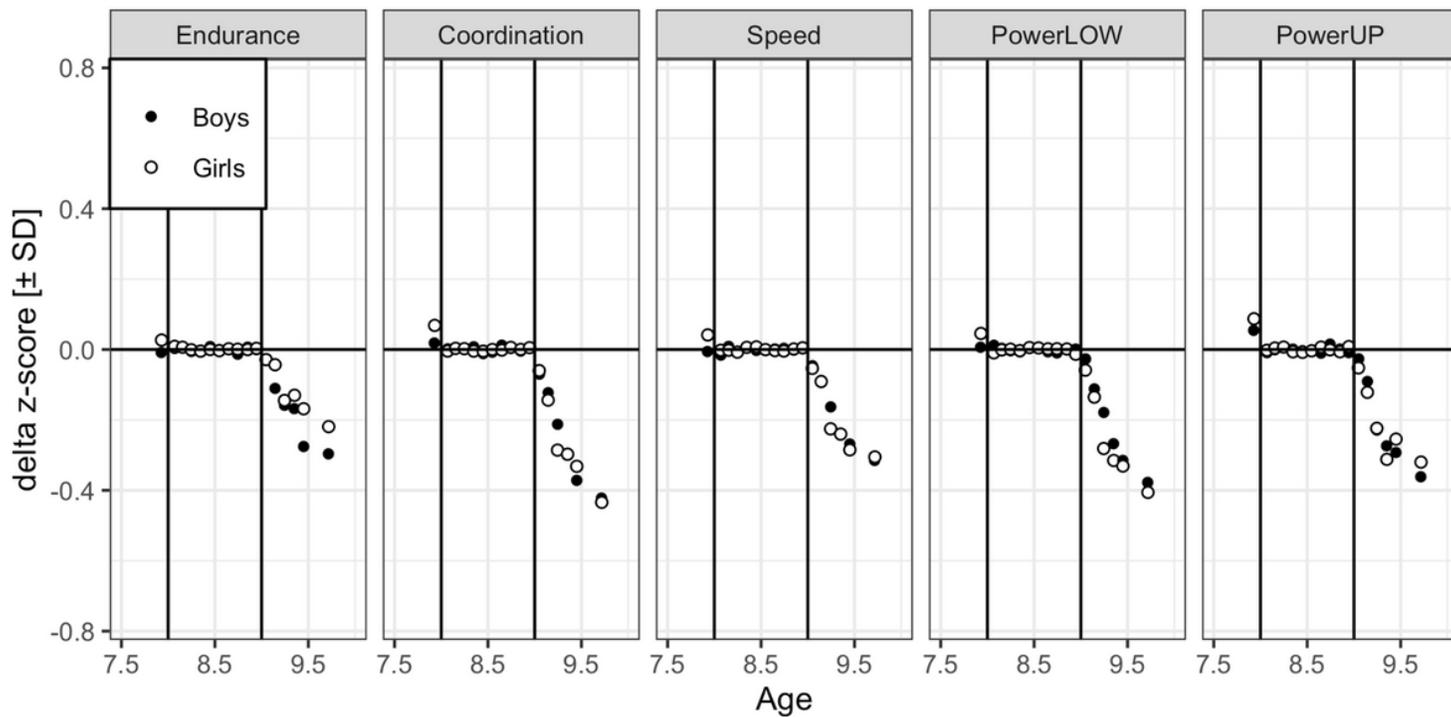


Figure 3

Delta z-score between observed and predicted physical fitness development for boys (closed circles) and girls (open circles) aged 7.50 to 10.0 years. Data were z-transformed. Endurance = cardiorespiratory endurance (i.e., 6-minute-run test), Coordination = star-run test, Speed = 20-m linear sprint test, PowerLOW = lower limbs muscle power (i.e., standing long jump test), PowerUP = upper limbs muscle power (i.e., ball-push test). Note that delta z-scores for younger-than-keyage boys and girls were aggregated over 7.50 to 7.99 years and that delta z-scores for older-than-keyage boys and girls were aggregated over 9.50 to 9.99 years. Points are binned delta child means. Note that for coordination and speed, scores were converted from seconds to meters/seconds (i.e., pace scores; star-run test = $50.912 \text{ [m]} / \text{time [s]}$; 20-m linear sprint test = $20 \text{ [m]} / \text{time [s]}$). These transformations had the advantage that a large value is indicative of better physical fitness.