RoSE
Research on Steiner Education

Volume 9 Number 2
pp. 44-61

Hosted at www.rosejourn.com

Psychometric Evaluation of the Preschool Health Examination at German Steiner Schools. Results of IPSUM, a Multicentre Cross-Sectional Validation Study

Christiane Diefenbach[1]*, Martina F. Schmidt[1,2]*, Jochem König[1], Rainer Patzlaff[2], Michael S. Urschitz[1]

[1] Division of Paediatric Epidemiology, Institute of Medical Biostatistics, Epidemiology, and Informatics, University Medical Centre of the Johannes Gutenberg University; Mainz, Germany

[2] Institute for Pedagogy, Sensory, and Media Ecology (IPSUM); Stuttgart, Germany

** Shared first authorship*

ABSTRACT. *Background*: School readiness, which is assessed as part of the obligatory preschool health examination (PHE), may be an important predictor of educational and health outcomes during primary school. As part of the IPSUM research project "Age at school entry and health outcomes", we developed a standardized PHE suitable for German Steiner schools (Waldorf schools).
Objective: The aim of the present study was to validate the items and scales of the PHE in order to determine how well they measure school readiness.
Methods: The PHE instrument comprised 35 items assessing cognitive, sensory, and motor skills as well as anthropometric measures and was validated in 93 German Steiner schools using a cross-sectional study design. 4,789 preschoolers (mean age: 6.1 years; 46.3 % girls) who were examined by the PHE school teams were included.
Results: Analysis of the factor structure identified 6 factors covering anthropometric measures, second dentition, gross and fine motor skills, auditory memory, and cognitive/sensory skills. With a few exceptions, results showed satisfactory internal consistency (.62-.86), discriminatory power (.30-.78), and inter-rater reliability (.66-1.0). Linear and logistic regression analyses demonstrated that with the exception of fine motor skills, all factors were significantly related to age at the PHE and school readiness as determined by the school team.
Conclusion: The proposed PHE instrument is generally reliable and valid. We suggest some adaptations to further improve the quality of the PHE at German Steiner schools.

Keywords: school readiness, assessment at school entry, developmental status, age, validity

ZUSAMMENFASSUNG. *Hintergrund*: Die Schulreife, die im Rahmen der obligatorischen Schuleingangsuntersuchung beurteilt wird, könnte ein wichtiger Prädiktor für den Bildungserfolg und die gesundheitliche Entwicklung in der Grundschule sein. Im Rahmen des IPSUM-Forschungsprojekts „Einschulungsalter und Gesundheitsentwicklung" entwickelten wir eine standardisierte Schuleingangsuntersuchung für deutsche Waldorfschulen.
Ziel: Mit der vorliegenden Studie sollen die Items und Skalen der Schuleingangsuntersuchung validiert werden, um zu ermitteln, wie gut sie die Schulreife erfassen.
Methode: Die Schuleingangsuntersuchung umfasst 35 Items zur Beurteilung kognitiver, sensorischer und motorischer Fähigkeiten sowie anthropometrischer Größen und wurde anhand der Querschnittsdaten von 93

deutschen Waldorfschulen validiert. In die Analysen wurden 4789 Vorschüler einbezogen (mittleres Alter 6,1 Jahre, 46,3 % Mädchen), die von den Aufnahmegremien der Waldorfschulen untersucht wurden.
Ergebnisse: Die Analyse der Faktorenstruktur ergab 6 Faktoren, die anthropometrische Größen, Zahnwechsel, Grob- und Feinmotorik, auditives Gedächtnis und kognitive/sensorische Fähigkeiten beinhalten. Bis auf wenige Ausnahmen zeigten die Ergebnisse eine zufriedenstellende interne Konsistenz (.62-.86), Trennschärfe (.30-.78) und Interrater-Reliabilität (.66-1.0). Lineare und logistische Regressionsanalysen zeigten, dass mit Ausnahme der Feinmotorik alle Faktoren signifikant mit dem Untersuchungsalter und der vom Aufnahmegremium ermittelten Schulreife zusammenhängen.
Schlussfolgerung: Das vorgestellte Instrument ist weitgehend reliabel und valide. Wir schlagen einige Anpassungen vor, um die Qualität der Schuleingangsuntersuchung an deutschen Waldorfschulen noch weiter zu verbessern.

Schlüsselwörter: Schulreife, Einschulungsuntersuchung, Entwicklungsstand, Alter, Validität

Introduction

School entry is a major change in a child's life, combined with many new demands the child has to cope with (Griebel & Niesel, 2002). As there is great heterogeneity in school-relevant skills at the time of school entry, the preschool health examination (PHE) in Germany serves to assess school readiness and thus the developmental status of all children of compulsory school age in order to ensure that they will be able to meet the school requirements (Oldenhage, Daseking, & Petermann, 2009). Children with developmental deficits relevant for school readiness are identified so that targeted interventions can be initiated or school entry deferred. In this way, the emergence of school difficulties are prevented, disadvantages for young children are reduced, and differences in development between the children starting school are diminished (Horstschräer & Muehler, 2014).

The criteria for school readiness collected during the PHE are characteristics and skills that predict school performance in the future. These usually include cognitive skills, motor skills, and social-emotional skills (Kammermeyer, 2000; Oldenhage et al., 2009). With regard to cognitive skills, domain-specific precursors of writing, reading, and mathematics have turned out to be important: These include phonological awareness (e.g., Bus & van Ijzendoorn, 1999; Schneider & Näslund, 1999), which is the awareness of the most basic speech units of a language (i.e., phonemes) as well as larger units such as rhymes and syllables (Castles & Coltheart, 2004), auditory memory, and grammatical skills (e.g., Daseking & Petermann, 2008) as well as quantity-number competencies like counting, quantity discrimination, and linking number-words with quantity (Duncan et al., 2007; Krajewski & Schneider, 2006, 2009). Moreover, selective and visual attention are a prerequisite for acquiring academic skills (e.g., Kinsey, Rose, Hansen, Richardson, & Stein, 2004). Further predictors of academic achievement are fine motor skills such as visual-spatial integration (e.g., Cameron et al., 2012) and visual-motor coordination (Oberer, Gashaj, & Roebers, 2018), while gross motor skills are especially influential on psychosocial development (because children with motor skill problems are more likely to be excluded or teased) (Bejerot, Plenty, Humble, & Humble, 2013). Social-emotional skills that are related to school performance are independence, responsibility, engaging in conversation and cooperation, as well as persistence in tasks (McClelland, Acock, & Morrison, 2006; Petermann, Petermann, & Krummrich, 2008), but also more domain-general characteristics such as working memory (e.g., Berg, 2008) and self-regulation abilities. In particular, executive functions are important indicators of school readiness (Blair & Raver, 2015; Oberer et al., 2018).

Due to the time constraints of the PHE, only a selection of school-relevant skills can be assessed. There are differences in the tested skills and the underlying concept of school readiness between the PHEs of German public schools and Steiner schools. According to the prevailing view of school readiness on which the PHEs of public schools are based, school readiness is not a characteristic of a child, but is developed jointly by child and family, nursery school, and primary school (Kammermeyer, 2000). Hence, the PHE does not aim at selecting school-ready children, but rather at determining the need for compensatory special support regarding school-relevant skills and at avoiding deferred entry to school (Oldenhage et al., 2009). In contrast, Steiner education holds the view of school readiness that children who are not mature in relevant developmental skills should wait until they are ready for school, resulting in a higher proportion of deferred entries to school compared to public schools (Statistisches Bundesamt, 2013). Steiner education argues that the same structural forces first drive growth and physical development in preschoolers and subsequently work on cognitive development. If these forces are used for cognitive development too early because of a young age at school entry, they are no longer available for physical developmental processes, possibly leading to an impairment of the child's health and performance. Therefore, more emphasis is placed on maturity and physical development in the context of school readiness, so that during the PHEs in Steiner schools, particular attention is paid to characteristics such as the change of physical proportions (form changes) and the onset of second dentition (Patzlaff, Boeddecker, & Schmidt, 2006).

The IPSUM research project investigated whether the age at school entry and the child's school readiness or developmental status as assessed by PHE is relevant to the long-term development of its health and school performance (Patzlaff et al., 2006). Within this project, a standardized PHE suitable for Steiner schools was developed for participating schools over several years. The aim of the present work was the psychometric evaluation of the final PHE version. This cross-sectional study examined the factor structure of the instrument, its reliability, specifically its internal consistency and inter-rater reliability, as well as discriminatory power of the items. However, our main focus was on the validity of the instrument. It was determined how well the criteria used for school readiness actually measure school readiness as understood by Steiner education: Acceptable criteria for measuring school readiness are those that i) depend on age, meaning skills that will improve as children grow older, and ii) are related to school readiness as determined by experienced school teams that carry out the PHE and decide whether a child is ready for school entry.

Methods

The IPSUM research project

The IPSUM project comprised the development and validation of a PHE suitable for German Steiner schools and the initiation of a nation-wide population-based multicentre prospective cohort study with an open cohort design in cooperation with the German Association of Steiner Schools (Patzlaff et al., 2006). In 2003, all German Steiner schools (with the exception of special education schools) were asked to implement a newly developed, standardized PHE and to document the test results. Following pre-tests since 2004 and a large pilot study in 2007 (65 schools), the definitive cohort study was initiated in 2008 (88 schools). The study protocol was reviewed and approved by the ethics committee of the Federal Physician Chamber in Frankfurt/Main (Hesse; Germany). Written informed consent was obtained from parents/legal guardians prior to study enrolment.

Participants

All preschoolers registered for the PHE in 2007 or 2008 at a participating school underwent the PHE, and those whose parents provided informed consent were included in the study. Ultimately, 6171 children from 93 German Steiner schools were enrolled. For the present analysis, only children who were born within the school entry cut-off dates for German Steiner schools were selected (for 2007, children born between 30 June 2000 and 1 July 2001, and for 2008 between 30 June 2001 and 1 July 2002). This restriction was chosen because the PHE was designed for children in this age range. Furthermore, excluding particularly old children with a deferred entry to school or children who were too young for school entry could prevent bias in the results. Thus, data from 1382 children were excluded from the analysis, leaving a final study sample of 4789 participants (46.3 % girls, 53.7 % boys, aged 5.4 - 7.1 years). The distribution of school readiness categories, gender, and age are given in Table 1.

Inter-rater reliability of the PHE items was investigated in 2008 in one of the participating schools in a subsample of 105 preschoolers. Of these, 50 were girls (47.6 %) and 55 were boys (52.4 %), aged 5.6 to 7.0 years (mean age = 6.2 years).

Table 1
Sample characteristics: Distribution of school readiness categories, gender, and age (N=4789)

School readiness judgement	Gender	*N*	Mean age at PHE (Standard deviation)
School ready	Girls	1928	6.15 (0.29)
	Boys	2022	6.19 (0.28)
	Total	3950	6.17 (0.28)
Questionable school readiness	Girls	145	5.87 (0.24)
	Boys	308	5.96 (0.27)
	Total	453	5.93 (0.26)
Not school ready	Girls	83	5.80 (0.20)
	Boys	163	5.83 (0.20)
	Total	246	5.82 (0.20)
Missing data	Girls	59	6.05 (0.27)
	Boys	81	6.13 (0.29)
	Total	140	6.10 (0.28)

Abbreviations: PHE = preschool health examination.

Development of the PHE

The primary aim of the PHE was the accurate estimation of a child's school readiness. In this context, items of the PHE should meet two major requirements: i) items should reflect developmental aspects (thus, item responses should vary with biological age) and ii) items should be able to predict judgements of an expert committee regarding school readiness of an individual child. Therefore, items of already established and validated instruments such as the MoMo (Motorik-Modul; Woll, Kurth, Opper, Worth, & Bos, 2011), the Mottier-Test of the ZLT (Zürcher Lesetest; Linder & Grissemann, 2000), and the FEW-2 (Frostigs Entwicklungstest der visuellen Wahrnehmung – 2; Büttner, Dacheneder, Schneider, & Weyer, 2008) were reviewed with regard to their suspected ability to meet these requirements.

From 2004 to 2006, preliminary PHE versions were elaborated, evaluated, and revised in an iterative circle. Finally, 35 items covering three major domains were identified and selected for the PHE: The first 10 items are indicators of physical development, for example the onset of second dentition. The subsequent 12 items refer to motor skills – of these, 8 items capture gross motor skills (performed with legs) and 4 items fine motor skills (performed with hands). The last 13 items refer to sensory and cognitive skills – among them are items capturing auditory memory (copying rhythm patterns, repeating sequences of syllables, also referred to as pseudoword repetition), instant recognition of quantities and counting, working memory (remembering quantities), visual perception (grasping a form), and visual-spatial integration (copying a figure). A more detailed description of each item and the rating scales is given in the Appendix. Parallel to item selection and the development of response categories, a PHE documentation sheet containing the examination items and standard operating procedures detailing the appraisal of the child's developmental status were prepared.

Validation procedures

In 2007 and 2008, the final version of the PHE was used by the PHE school teams of the respective Steiner schools. Teams usually consist of a school doctor, a teacher, and/or a therapeutically working staff member. Depending on the school, the PHE was executed by the doctor and/or the teacher. Altogether the examination usually lasted 20 - 25 minutes. As a result of the examination, the PHE school team rated the preschooler as being school ready, not school ready, or having an intermediate or questionable school readiness status. This final judgement was based on their expert opinion and was reached after a short discussion. Criteria for judging school readiness were not provided, so the judgement of the committee was not influenced by the study team.

In the PHEs in which inter-rater reliability was investigated, all preschoolers were assessed independently by the same two raters. Both raters were present during the PHEs without mutual interaction, observed the respective participant fulfilling the tasks, and noted their assessments down in two separate documentation sheets without discussing their assessments.

Data analysis

Eleven items were inverted so that higher scores always mirror higher levels of features or skills. Missing values of ratio-scaled items were imputed using single imputation by regression with the independent variables gender and age at PHE. Age at PHE was computed for each child as the difference between the date of PHE and the date of birth.

As the items of the PHE address three domains, a principal component analysis with varimax rotation was carried out to examine the factor structure of the items. Discriminatory power (corrected item-total correlation coefficient r_{itc}) of the items was calculated to assess how well the single items reflect the subscales resulting from the principal component analysis and how well the items are able to distinguish between children with different levels of features or skills. Values lower than .30 indicate poor discriminatory power (Fisseni, 1997).

To evaluate the reliability of the PHE, internal consistency and inter-rater reliability were investigated. Internal consistency was determined by calculating Cronbach's α for the subscales, which should not be lower than .70 for a satisfying result (Schmitt, 1996). Since items have different result scales, all item values were z-transformed in order to obtain a standardized result scale (with mean = 0 and standard deviation = 1). These standardized values were used for the principal component analysis as well as for calculating discriminatory power and internal consistency.

Regarding inter-rater reliability, different measures of agreement in the absolute values raters assigned to the same participant were computed: For dichotomous items, Cohen's kappa was determined, and for ordinal- and ratio-scaled items as well as for the school readiness judgement, intra-class correlation coefficients (two-way random, absolute agreement) were computed. Poor inter-rater reliability is indicated by Cohen's kappa coefficients lower than .40 (Bortz & Döring, 2006) and intra-class correlation coefficients lower than .50 (Wirtz, 2004). To detect significant differences between the ratings, McNemar's tests were used for dichotomous items, sign tests for ordinal-scaled items, and t-tests for ratio-scaled items. Inter-rater reliability was assessed for all but four items: Anthropometric measures were removed from the analysis, because they were not measured independently by the two raters. Instead, a third person from the PHE school team took the measurements and reported the results, which were written down into the documentation sheets by the raters.

In order to investigate how a child's subscale score individually influences the school readiness judgement of the PHE school team, subscale scores were formed from item scores based on the result of the principal component analysis: The standardized values of those items loading highest on a common factor were averaged (if < 50 % of the item values were missing) and the resulting subscale scores were again z-transformed. Initially an ordinal regression analysis was performed with the above mentioned three categories of the school readiness judgement as dependent variable and the standardized scores of the six subscales as independent variables. The ordinal regression model turned out to be invalid because the test of parallel lines was significant, which indicated that the regression coefficients change across the three categories of the school readiness variable. Hence, we conducted two multiple binary logistic regression analyses instead: the first one with the categories school ready vs. not school ready or questionable school readiness as the dependent variable and the second one with the categories school ready or questionable school readiness vs.

not school ready as the dependent variable. Effect estimates (regression coefficients *B*) and standard errors were adjusted for gender and age on the 1st of July[1].

Age dependence of the subscale scores was investigated using linear regression analyses with age at PHE as an independent variable and the standardized scores of each subscale as the respective dependent variable. Results were adjusted for gender. All statistical analyses were performed using IBM SPSS Statistics version 23.

Results

Evaluation of the factor structure

Based on the number of factors with eigenvalues greater than 1, ten factors were extracted by principal component analysis. Because there were some factors on which only two items had their highest loadings, we decided for a six-factor solution that was suggested by the scree test. The six extracted factors explained 46% of the total variance and corresponded with the domains of the PHE, but further divided domains into two factors each. The highest loading of each item on the extracted factors is shown in Table 2. The items of the physical form domain loaded highest on two factors addressing anthropometric measures (Factor 2) and second dentition (Factor 5). The items of the motor skills domain loaded highest on two factors covering gross and fine motor skills (Factor 3 and Factor 1, respectively) and the items of the sensory and cognitive skills domain on two factors referring to auditory memory (Factor 4) and cognitive/sensory skills (Factor 6). This structure of three domains which can be further subdivided into six factors or subscales are also roughly observable in product-moment correlations between the items. As illustrated in Figure 1, associations between items forming a subscale are stronger than between items of different subscales.

Reliability and item analysis

Inter-rater reliability. Results of the inter-rater reliability analysis are reported in Table 3. Regarding the dichotomous items, Cohen's kappa coefficients showed good to very good agreement between the raters for five of six items ($\varkappa$ > .6 or > .75), and this was confirmed by non-significant McNemar's tests (all *p*s > .2). Solely for the tooth gaps item, agreement was rather low; the McNemar's test also yielded significant differences between the raters ($p < .01$).

The intra-class correlation coefficients (two-way random, absolute agreement) revealed very good inter-rater reliability for the ordinal- and ratio-scaled items (all ICC(2,1) > .9).[2] Accordingly, sign tests showed no significant differences between the raters for the ordinal-scaled items (all *p*s > .3) and t-tests for the ratio-scaled items (backward tight-rope walk, frequency of stepping off: $t(102) = 0.58$, $p = .57$; jumping sideways, number of jumps in 10 seconds: $t(102) = 1.42$, $p = .16$).

Internal consistency. Cronbach's α was calculated for the six subscales (see Table 4). The internal consistency coefficients of four of the subscales (gross motor skills, fine motor skills, anthropometric measures, and auditory memory) were acceptable to good. The second dentition subscale initially showed a questionable internal consistency of $\alpha = .66$, but the reliability analysis revealed that removing loose teeth and tooth gaps items would increase internal consistency of this subscale. Since the tooth gaps item also demonstrated low inter-rater reliability, and the loose teeth and tooth gaps items additionally have poor discriminatory power (as will be shown in the following subsection), we decided to exclude these items from the subscale. With the remaining three items, the second dentition subscale reached acceptable internal consistency. The low internal consistency coefficient of the cognitive/sensory skills subscale cannot be improved by eliminating an item. Thus, all but one of the subscales can be considered to be sufficiently reliable.

1. Because PHEs took place in different temporal distances to the school entry cutoff date, the PHE school team took into account not the age at PHE, but how old children will be at the cutoff date.
2. For items that involve hopping on the right and on the left leg, results refer to corrected data after discovering a documentation mistake.

Table 2

Factor structure of the preschool health examination items (N = 2860)

		Factors					
		1	2	3	4	5	6
	Initial eigenvalue	5.28	3.43	2.26	1.87	1.76	1.52
	Explained variance (rotated)	10.3 %	8.1 %	8.0 %	6.7 %	6.7 %	6.2 %
Items of domain „physical form“	Form changes		.37				
	Second dentition					.86	
	Loose teeth					.45	
	Tooth gaps					.47	
	New incisors					.69	
	New molars					.69	
	Body length		.88				
	Span		.85				
	Head circumference		.58				
	Body weight		.84				
Items of domain „motor skills”	Backward tight-rope walk 1	.34					
	Backward tight-rope walk 2	.43					
	Hopping on right leg 1	.78					
	Hopping on left leg 1	.78					
	Hopping on right leg 2	.81					
	Hopping on left leg 2	.80					
	Jumping sideways 1	.45					
	Jumping sideways 2	.54					
	Finger-thumb opposition 1			.77			
	Finger-thumb opposition 2			.78			
	Rapid hand-turning 1			.80			
	Rapid hand-turning 2			.78			
Items of domain „sensory and cognitive skills“	Copying a rhythm pattern 1				.62		
	Copying a rhythm pattern 2				.71		
	Copying a rhythm pattern 3				.69		
	Repeating syllables 1				.52		
	Repeating syllables 2				.55		
	Repeating syllables 3				.51		
	Recognition of quantities						.61

Remembering quantities	.56
Continuing counting	.54
Number memory	.57
Grasping a form	.30
Copying a figure (cross)	.50
Copying a figure (fish)	.46

***Note*. The highest loading of each item on the extracted factors is presented.**

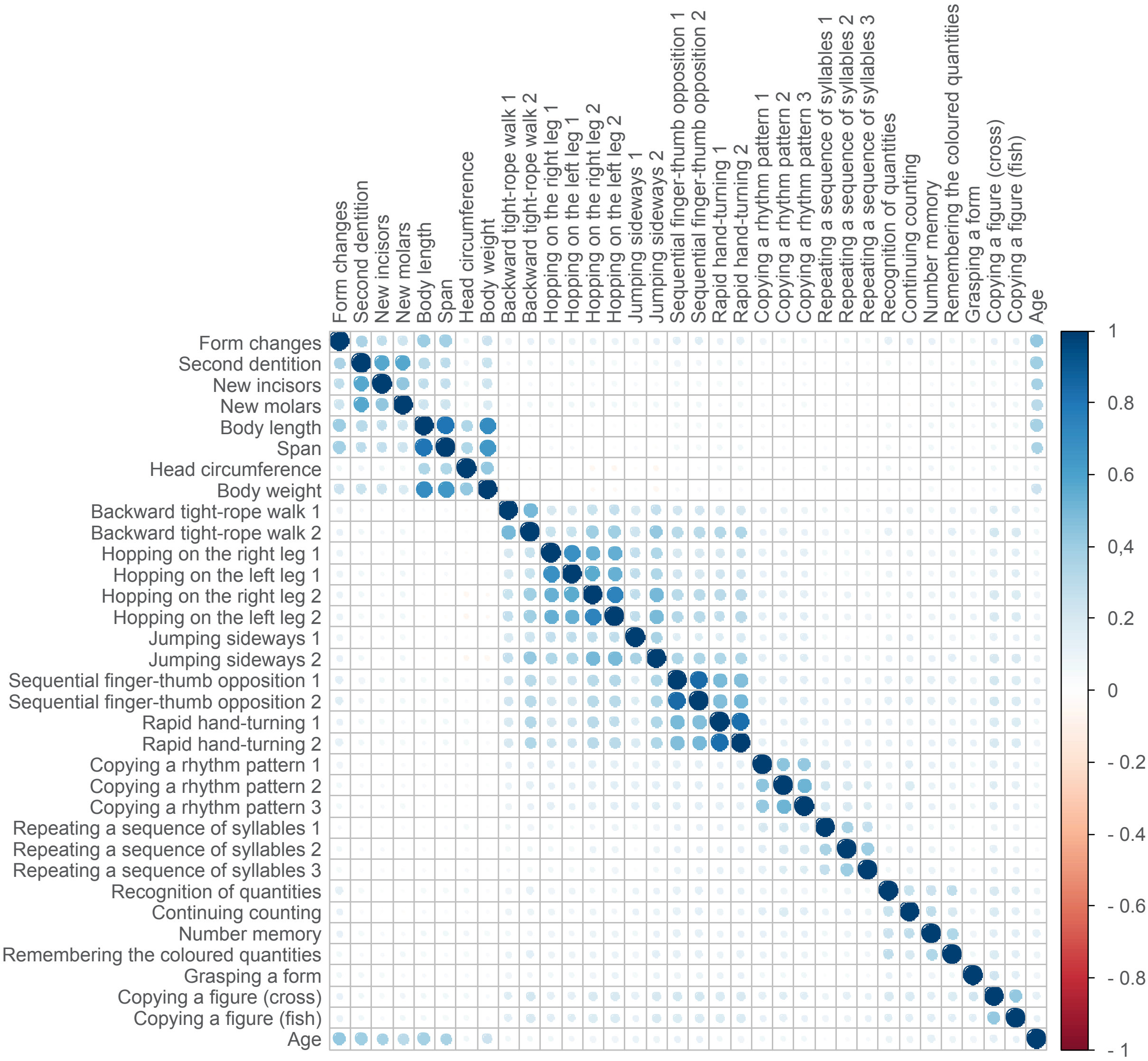

Figure 1. Correlation matrix for the relationship between age at preschool health examination and item scores as well as for the relationship between the item scores (*N* = 2860, list-wise deletion), presented without later excluded loose teeth and tooth gaps items.

Discriminatory power. For all items, discriminatory power was determined. Table 4 shows the range of coefficients of the items separately for each subscale. The second dentition subscale and the cognitive/sensory skills subscale contain single items with low discriminatory power, the items of the other subscales have acceptable to good discriminatory power (r_{itc} > .3 or > .5). Regarding the second dentition subscale, only items registering loose teeth and tooth gaps cannot sufficiently distinguish between children with a different second dentition status (r_{itc} = .24 and .27, respectively) and finally were removed from further analysis. All items within the cognitive/sensory skills subscale except for the grasping a form item (r_{itc} = .20) have satisfying discriminatory power (r_{itc} > .3). Thus, apart from very few items, discriminatory power is sufficiently high.

Table 3
Inter-rater reliability of the preschool health examination items (N = 105)

Item	Cohen's kappa / Intraclass correlation [a]	p^b
Dichotomous items:		
Second dentition	1.0	1.0
Loose teeth	.66	1.0
Tooth gaps	.40	.008
New incisors	.98	1.0
New molars	.94	.25
Number memory	1.0	1.0
Ordinal-scaled items:		
Form changes	.92	1.0
Backward tight-rope walk 2	1.0	1.0
Hopping on right leg 1	.99	1.0
Hopping on left leg 1	.98	1.0
Hopping on right leg 2[c]	.98	.63
Hopping on left leg 2[c]	.97	.63
Jumping sideways 2	.99	1.0
Finger-thumb opposition 1	.97	1.0
Finger-thumb opposition 2	1.0	1.0
Rapid hand-turning 1	.98	1.0
Rapid hand-turning 2	.97	.38
Copying a rhythm pattern 1	.97	.50
Copying a rhythm pattern 2	1.0	1.0
Copying a rhythm pattern 3	1.0	1.0
Repeating syllables 1	1.0	1.0
Repeating syllables 2	.99	1.0
Repeating syllables 3	.99	1.0
Recognition of quantities	1.0	1.0
Remembering quantities	1.0	1.0
Continuing counting	.96	1.0
Grasping a form	1.0	1.0
Copying a figure (cross)	1.0	1.0
Copying a figure (fish)	1.0	1.0

School readiness judgement	.96	.50
Ratio-scaled items:		
Backward tight-rope walk I	.997	.57
Jumping sideways I	.999	.16

[a] Cohen's kappa for dichotomous items, intraclass correlation (two-way random, absolute agreement) for ordinal- and ratio-scaled items.
[b] McNemar's test for dichotomous items, sign test for ordinal-scaled items, t test for ratio-scaled items.
[c] Corrected data after discovering a documentation mistake.

Table 4
Internal consistency and range of discriminatory power of the six subscales

Subscales	*N*	Number of items	Cronbach's α	Range of discriminatory power
Gross motor skills	3968	8	.82	.37 - .70
Anthropometric measures	4485	5	.79	.30 - .78
Fine motor skills	4529	4	.86	.70
Auditory memory	4442	6	.69	.35 - .49
Second dentition[a]	4545	3	.75	.55 - .66
Cognitive/sensory skills	3928	7	.62	.20 - .39

[a] Coefficients after elimination of two items: Loose teeth and Tooth gaps.

Criterion validity

Relationship between subscale scores and school readiness judgement. Coefficients from the multiple binary logistic regressions of subscale scores on school readiness judgement are presented in Table 5. The first regression analysis with the categories school ready vs. not school ready or questionable school readiness as a dependent variable (see the left side of Table 5) yielded highly significant effects on the school readiness judgement for all six subscales (all $ps < .01$). The scores on the cognitive/sensory skills subscale and the anthropometric measures subscale have the strongest influence on whether a child is rated as school ready. Smaller regression coefficients, that is weaker effects on the school readiness judgement, than for these two subscales are obtained for the gross motor skills subscale and the second dentition subscale, and the weakest effects are found for the auditory memory subscale and the fine motor skills subscale.

The second regression analysis with the categories school ready or questionable school readiness vs. not school ready as dependent variable (see the right side of Table 5) showed slightly different results. Although in both analyses fine motor skills have the smallest effect on the school readiness judgement, the effect missed statistical significance in the second analysis ($p > .05$). For the other five subscales, effects are highly significant ($ps < .01$). To distinguish children who are not school ready from the rest of the children, anthropometric measures appear to be more important than cognitive/sensory skills – that is, the order is reversed compared to the first analysis. The same holds for the gross motor skills subscale and the second dentition subscale: Whereas in the first analysis, the regression coefficient is larger for the gross motor skills subscale, it is now smaller than that for the second dentition subscale. In summary, to judge a child as school ready, cognitive/sensory skills appear to matter the most, while to judge a child as not school ready, anthropometric measures are most relevant.

Table *5*

Relationship between the six subscale scores and the school readiness judgement (dependent variable) as assessed by binary logistic regression analyses (N = 4267)

	School ready[a]			School ready or questionable school readiness[b]		
Independent variable[c]	*B*[d]	Standard error	p	*B*[d]	Standard error	p
Gross motor skills	0.36	0.06	<.001	0.25	0.09	.004
Anthropometric measures	0.52	0.06	<.001	0.58	1.0	<.001
Fine motor skills	0.23	0.06	<.001	0.15	0.08	.076
Auditory memory	0.25	0.05	<.001	0.23	0.08	.004
Second dentition	0.29	0.06	<.001	0.28	1.0	.004
Cognitive/sensory skills	0.72	0.06	<.001	0.56	0.08	<.001

[a] vs. not school ready or questionable school readiness.
[b] vs. not school ready.
[c] Standardized subscale scores.
[d] Non-standardized regression coefficient; results are adjusted for age at cut-off date (1st July) and gender.

Age dependence of the subscale scores. Table 6 shows the results of the linear regression analyses that examined how a child's score on each subscale depends on his/her age at the PHE. Age has a highly significant effect on all six subscales (all *ps* < .01). Regression coefficients reveal that the largest effects are obtained for second dentition and anthropometric measures. Age dependence is lower for cognitive/sensory skills and gross motor skills, and the lowest age dependence is found for fine motor skills and auditory memory. This finding is also reflected in product-moment correlations between item scores and age at PHE, depicted in the first row/column of the correlation matrix in Figure 1: The strongest positive associations with age at PHE are obtained for the first six items that belong to the second dentition subscale and the anthropometric measures subscale. Solely the head circumference and body weight items of the anthropometric measures subscale show no clear association with age.

Table 6

Relationship between age at preschool health examination and the subscale scores as assessed by linear regression analyses

Dependent variable[a]	*N*	*B*[b]	Standard error	*p*
Gross motor skills	4707	0.42	0.05	< .001
Anthropometric measures	4789	1.09	0.05	< .001
Fine motor skills	4696	0.34	0.05	< .001
Auditory memory	4690	0.37	0.05	< .001
Second dentition	4545	1.25	0.05	< .001
Cognitive/sensory skills	4717	0.69	0.05	< .001

[a] Standardized subscale scores.
[b] Non-standardized regression coefficient; results are adjusted for gender.

Discussion

In the present study, we evaluated the psychometric properties of a newly developed, standardized PHE suitable for German Steiner schools and investigated how well the subscales and items of the instrument measure school readiness. Analysis of the factor structure revealed that the 35 items can be grouped into six subscales covering gross motor skills, anthropometric measures, fine motor skills, auditory memory, second dentition, and cognitive/sensory skills. Inter-rater reliability of the items is for the most part very good, with the exception of the tooth gaps item where agreement between the raters is poor. The items show satisfactory discriminatory power apart from items referring to loose teeth, tooth gaps, and grasping a form. Internal consistency of the subscales is sufficiently high; only for the cognitive/sensory skills subscale is it questionable. Following some minor adjustments to the present PHE, this validated examination may be used henceforth by German Steiner schools.

Subscales or items measuring school readiness should be related to the school readiness concept of Steiner schools as well as depend on the age of the preschooler. Our results confirm that all subscales fulfil these requirements except the fine motor skills subscale, for which the relationship with the school readiness judgement was partly not significant. The anthropometric measures subscale belongs to the two subscales with the strongest effects on school readiness judgement and has the second strongest age dependence. The other subscale that belongs to the two subscales with the strongest effect on school readiness judgement is the cognitive/sensory skills subscale. In terms of age dependence, this subscale is also among the better half of the subscales. The second dentition subscale has a medium influence on school readiness judgement compared to the other subscales, but shows the strongest age dependence. For the gross motor skills subscale, both the influence on school readiness judgement and age dependence are medium compared to the other subscales. The auditory memory subscale is the second weakest subscale in terms of both the effect on school readiness judgement and age dependence. The weakest effect on school readiness judgement (which even failed to reach significance in one analysis) as well as the weakest age dependence has the fine motor skills subscale.

These findings could be used to adapt and improve the PHE at German Steiner schools as follows: Items registering loose teeth and tooth gaps should no longer be collected in future PHEs due to their low discriminatory power. The reason for their small correlation with the second dentition subscale score could be that loose teeth and tooth gaps are temporary and may not be present at the moment of PHE despite ongoing tooth change. However, the other items of the second dentition subscale should remain in the PHE since this subscale has proven to be a good measure of school readiness.

The item of grasping a form also appears to be unsuitable for use in future PHEs because of low discriminatory power. As an item capturing visual perception, it should be relevant for acquiring school skills. Particularly mathematical skills have been demonstrated to be influenced by visual perceptual abilities (Daseking & Petermann, 2011; Werpup & Petermann, 2016), but they are also important for orthographic skills and reading development (Boets, Wouters, van Wieringen, De Smedt, & Ghesquiere, 2008). There are several possible reasons for the poor results of the grasping a form item: One reason could be incomplete standardization – the examiner drew half a pine tree for each child instead of providing it with a printed copy –, another could be that the specific motif is unsuitable for some reason and should be replaced. It is also possible that the type of task is unsuitable. Another frequently used task from the domain of visual perception could be chosen based on other validated instruments measuring school readiness such as GSS (Göppinger sprachfreier Schuleignungstest; Kleiner, 1998) or DVET (Duisburger Vorschul- und Einschulungstest; Meis, 1997), for example identifying two identical images in a series of images.

The unsatisfactory internal consistency of the cognitive/sensory skills subscale should not be given too much weight as the subscale measures school readiness well. The low Cronbach's α in this case is due to the fact that the subscale is heterogeneous in content since it comprises items referring to quantity-number competencies, working memory, visual perception, and visual-spatial integration. These skills should continue to be tested in future PHEs because evidence shows that they are important predictors of school performance (see Introduction section) and they are part of many validated school entrance tests such as the GSS, WTA (Weilburger Testaufgaben für Schulanfänger; Hetzer & Tent, 1994), KST (Kettwiger Schuleingangstest; Meis, 1990), or Schulfähigkeitstest Form C (Seyfried & Karas, 1987).

The fine motor skills subscale could be removed from future PHEs because it measures school readiness poorly. However, fine motor skills would not thereby be omitted, since the two copying a figure items from the cognitive/sensory skills subscale assess visual-spatial integration, which entails both visual processing and fine motor skills. Also in the validated school entrance tests, fine motor skills are assessed almost exclusively by using tasks in which children have to copy figures, for example in the GSS, DVET, or S-ENS (Screening des Entwicklungsstandes bei Einschulungsuntersuchungen; Döpfner, Dietmair, Mersmann, Simon, & Trost-Brinkhues, 2005).

As the second weakest subscale in measuring school readiness, the auditory memory subscale should be subjected to a single item analysis. Pseudoword repetition or other tasks measuring auditory memory are used in some of the validated school entrance tests (e.g., S-ENS, Schulfähigkeitstest Form C), but rhythm skills are not assessed at all. However, there is evidence for a relationship of rhythm skills with phonological awareness and early measures of literacy (Moritz, Yampolsky, Papadelis, Thomson, & Wolf, 2013; Ozernov-Palchik, Wolf, & Patel, 2018), because the processing of certain aspects of auditory temporal structure may be shared (Goswami, 2011). But as phonological awareness appears to be a stronger predictor of reading abilities than rhythm skills (Ozernov-Palchik et al., 2018), it would be worth considering replacing the copying a rhythm pattern items with items measuring phonological awareness in future PHEs developments. These would complement the pseudoword repetition items well since for example Boets et al. (2008) demonstrated that verbal short-term memory measured by a pseudoword repetition test together with phonological awareness predicts reading and spelling development.

The gross motor skills subscale measures school readiness moderately well according to our results. For its mediocre importance, this subscale contains relatively many items and could be shortened in future PHEs after a single item analysis has been carried out. In other validated school entrance tests, gross motor skills are rarely assessed (an exception is S-ENS which includes the task of jumping sideways). This is in line with evidence summarized by Cameron, Cottone, Murrah, and Grissmer (2016), showing that gross motor skills are related to the development of social competencies and physical well-being but they are not as strongly associated with academic achievement as fine motor skills.

Limitations

Certain indicators of school readiness mentioned in the Introduction section are not yet covered by the PHE, such as attention, self-regulation abilities, and social-emotional skills. If fine motor skills items would be removed from the PHE, they could be replaced by selective attention tasks or self-regulation tasks. As social-emotional skills are not easily tested within the PHE, Petermann et al. (2008) suggest to include assessments of social and learning behaviour from nursery school teachers.

So far, the school readiness judgement does not result directly from the item scores, but from the subjective ratings of the PHE school team based on their expert knowledge and opinion. This procedure should be made further objective by weighting the PHE items according to their determined relevance for the school readiness judgement and generating a school readiness score.

Conclusions

Overall, the investigated PHE instrument proved to be sufficiently reliable and valid and hence appears suitable for assessing school readiness of preschoolers. Unsatisfactory results were obtained only for the fine motor skills subscale and a few individual items, which should be removed in future updates of the PHE.

Acknowledgements

The authors would like to thank Christian Heckmann, Claudia McKeen, Jan Vagedes, Hanns Ackermann, and Uwe Zickmann for their expert advice and support of the project. We would also like to thank Pilar Maria Urschitz-Duprat for data entry and management and Katherine Taylor for reviewing and proofreading the manuscript. Thanks also go to the principals and teachers of the participating schools for their support and cooperation. Last but not least, we particularly wish to thank all the parents and children for their participation; they made this study possible.

Funding

The IPSUM study was funded by the German Association of Waldorf Schools; the International Federation of Waldorf Kindergartens; the Medical and Pedagogical Section at the Goetheanum, Switzerland; the Educational Research Center of the German Association of Waldorf Schools; the German Physician Association for Anthroposophical Medicine; The Citizen and Patient Association of "Health Active" in Berlin, Germany (Gesundheit Aktiv - Bürger- und Patientenverband), and the Mahle Foundation in Stuttgart, Germany.

References

Bejerot, S., Plenty, S., Humble, A., & Humble, M. B. (2013). Poor motor skills: A risk marker for bully victimization. *Aggressive Behavior, 39*(6), 453-461. doi:10.1002/ab.21489

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology, 99*(4), 288-308. doi:10.1016/j.jecp.2007.12.002

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. In S. T. Fiske (Ed.), *Annual Review of Psychology* (Vol. 66, pp. 711-731). Palo Alto: Annual Reviews.

Boets, B., Wouters, J., van Wieringen, A., De Smedt, B., & Ghesquiere, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language, 106*(1), 29-40. doi:10.1016/j.bandl.2007.12.004

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.

Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403-414.

Büttner, G., Dacheneder, W., Schneider, W., & Weyer, K. (2008). *Frostigs Entwicklungstest der visuellen Wahrnehmung – 2* (1. ed.). Göttingen: Hogrefe.

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development, 83*(4), 1229-1244. doi:10.1111/j.1467-8624.2012.01768.x

Cameron, C. E., Cottone, E. A., Murrah, W. M., & Grissmer, D. W. (2016). How are motor skills linked to children‛s school performance and academic achievement? *Child Development Perspectives, 10*(2), 93-98. doi:10.1111/cdep.12168

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*(1), 77-111. doi:10.1016/s0010-0277(03)00164-1

Daseking, M., & Petermann, F. (2008). Diagnostic of cognitive abilities in pre-school age. Early diagnosis of developmental risks and learning disorders of school age. *Monatsschrift Kinderheilkunde, 156*, 685-693. doi:10.1007/s00112-008-1782-z

Daseking, M., & Petermann, F. (2011). The impact of precursors on reading, spelling, and arithmetic at school 2nd grade. *Gesundheitswesen, 73*(10), 644-649. doi:10.1055/s-0031-1286268

Döpfner, M., Dietmair, I., Mersmann, H., Simon, K., & Trost-Brinkhues, G. (2005). *Screening des Entwicklungsstandes bei Einschulungsuntersuchungen. S-ENS*. Göttingen: Hogrefe.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*. doi:10.1037/0012-1649.43.6.1428

Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*(1), 3-10. doi:10.1016/j.tics.2010.10.001

Griebel, W., & Niesel, R. (2002). *Abschied vom Kindergarten, Start in die Schule*. München: Don Bosco.

Hetzer, H., & Tent, L. (1994). *Weilburger Testaufgaben für Schulanfänger. WTA.* Weinheim: Beltz.

Horstschräer, J., & Muehler, G. (2014). School entrance recommendation: A question of age or development? *Education Economics, 22*(3), 270-292. doi:10.1080/09645292.2011.645126

Kammermeyer, G. (2000). *Schulfähigkeit - Kriterien und diagnostische / prognostische Kompetenz von Lehrerinnen, Lehrern und Erzieherinnen.* Bad Heilbrunn: Klinkhardt.

Kinsey, K., Rose, M., Hansen, P., Richardson, A., & Stein, J. (2004). Magnocellular mediated visual-spatial attention and reading ability. *Neuroreport, 15*(14), 2215-2218. doi:10.1097/00001756-200410050-00014

Kleiner, A. (1998). *GSS: Göppinger sprachfreier Schuleignungstest.* Göttingen: Beltz.

Krajewski, K., & Schneider, W. (2006). Predicting mathematical school achievement by mathematical precursors assessed in kindergarten. *Psychologie in Erziehung und Unterricht, 53*(4), 246-262.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*(4), 516-531. doi:10.1016/j.jecp.2009.03.009

Linder, M., & Grissemann, H. (2000). *Zürcher Lesetest (ZLT)* (6. ed.). Bern: Huber.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471-490. doi:10.1016/j.ecresq.2006.09.003

Meis, R. (1990). *Kettwiger Schuleingangstest. KST.* Weinheim: Beltz.

Meis, R. (1997). *Duisburger Vorschul- und Einschulungstest. DVET.* Weinheim: Beltz.

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Reading and Writing, 26*(5), 739-769. doi:10.1007/s11145-012-9389-0

Oberer, N., Gashaj, V., & Roebers, C. M. (2018). Executive functions, visual-motor coordination, physical fitness and academic achievement: Longitudinal relations in typically developing children. *Human Movement Science, 58*, 69-79. doi:10.1016/j.humov.2018.01.003

Oldenhage, M., Daseking, M., & Petermann, F. (2009). Assessment of developmental status considering current requirements of the health examination for school entry. *Gesundheitswesen, 71*(10), 638-647. doi:10.1055/s-0029-1239514

Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *Journal of Experimental Child Psychology, 167*, 354-368. doi:10.1016/j.jecp.2017.11.009

Patzlaff, R., Boeddecker, D., & Schmidt, M. (2006). Einschulungsalter und Gesundheitsentwicklung. Ein Forschungsprojekt des IPSUM-Instituts. *Erziehungskunst, 5*, 531-543.

Petermann, F., Petermann, U., & Krummrich, M. Z. (2008). Erfassung des Sozial- und Lernverhaltens vor dem Schuleintritt. *Psychologie in Erziehung und Unterricht, 55,* 114-122.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353. doi:10.1037/1040-3590.8.4.350

Schneider, W., & Näslund, J. C. (1999). Impact of early phonological processing skills on reading and spelling in school: Evidence from the Munich longitudinal study. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 126-147). New York, NY: Cambridge University Press.

Seyfried, H., & Karas, E. (1987). *Schulfähigkeitstest Form C* Wien: Ketterl.

Statistisches Bundesamt, W. (2013). *Schulanfänger zum Schuljahr 2008/09 nach Einschulungsart und Schulart.* Available at http://www.destatis.de [Retrieved 12 Oct 2015]

Werpup, L., & Petermann, F. (2016). Cognitive abilities of primary school children with dyscalculia: Focussing visual perception. *Kindheit und Entwicklung, 25*(4), 238-249. doi:10.1026/0942-5403/a000203

Wirtz, M. (2004). Determining the quality of rater judgements using intraclass correlation, and enhancing rater judgements. *Rehabilitation, 43*(6), 384-389. doi:10.1055/s-2003-814935

Woll, A., Kurth, B. M., Opper, E., Worth, A., & Bos, K. (2011). The ‚Motorik-Modul' (MoMo): physical fitness and physical activity in German children and adolescents. *European Journal of Pediatrics, 170*(9), 1129-1142. doi:10.1007/s00431-010-1391-4

Appendix

Table
Description of the preschool health examination items and their rating scales

Item	Brief description	Rating
I. Physical form		
Form changes	Limb extension, visibility of the joints, flattening of the ribcage, forming of the waist	Yes / No / In transition
Second dentition	Look into the mouth to see if change of teeth has started	No / In transition
Loose teeth		Yes / No
Tooth gaps		Yes / No
New incisors		Yes / No
New molars (6er)		Yes / No
Body length	Measured without shoes in upright posture	Number (in cm)
Span	Measured from the longest finger tip of one side to the longest finger tip of the other side with arms stretched out horizontally	Number (in cm)
Head circumference	Measured above the eyebrows	Number (in cm)
Body weight	Weighed without shoes	Number (in kg)
II. Motor skills.		
Backward tight-rope walk 1	The way the child balances backwards on a 3 meter long strip, frequency of stepping off	Number
Backward tight-rope walk 2	The way the child balances backwards on a 3 meter long strip, qualitative	5-point scale
Hopping on the right leg 1	Hopping on a marked cross 10 times with the right leg, number of hops	9-10 / 7-8 / 5-6 / 3-4 / <3
Hopping on the left leg 1	Hopping on a marked cross 10 times with the left leg, number of hops	9-10 / 7-8 / 5-6 / 3-4 / <3
Hopping on the right leg 2	Quality of the hopping on the right leg	5-point scale
Hopping on the left leg 2	Quality of the hopping on the left leg	5-point scale
Jumping sideways 1	Jumping from side to side between two small fields using both feet, as many jumps as possible in 10 seconds	Number
Jumping sideways 2	Quality of the side-to-side jumping	5-point scale
Sequential finger-thumb opposition 1	All fingers of the left hand touch the left thumb one after the other, running the complete sequence back and forth	5-point scale
Sequential finger-thumb opposition 2	All fingers of the right hand touch the right thumb one after the other, running the complete sequence back and forth	5-point scale
Rapid hand-turning 1	Child turns the palm of the left hand up and down about 4 times per second, standing in front of the examiner	5-point scale

Rapid hand-turning 2	Child turns the palm of the right hand up and down about 4 times per second, standing in front of the examiner	5-point scale
III. Sensory and cognitive skills		
Copying a rhythm pattern 1	Examiner taps the 1st rhythm on the table, child taps back: v v – v v – (v short, – long) one repetition allowed	3-point scale (1st attempt / 2nd attempt / not successful)
Copying a rhythm pattern 2	Examiner taps the 2nd rhythm on the table, child taps back: – v v – – (v short, – long), one repetition allowed	3-point scale (as above)
Copying a rhythm pattern 3	Examiner taps the 3rd rhythm on the table, child taps back: v v v v – – (v short, – long), one repetition allowed	3-point scale (as above)
Repeating a sequence of syllables 1	The child repeats the 1st sequence of syllables (pseudoword): to-pa-mo-ki, one repetition allowed	3-point scale (1st attempt / 2nd attempt / not successful)
Repeating a sequence of syllables 2	The child repeats the 2nd pseudoword: ka-to-pi-na-fe, one repetition allowed	3-point scale (as above)
Repeating a sequence of syllables 3	The child repeats the 3rd pseudoword: ga-li-no-ma-re-se, one repetition allowed	3-point scale (as above)
Recognition of quantities	Instant recognition (without counting) of the quantities of differently coloured beads (6 red, 4 blue)	3-point scale (6 were recognized / 4 were recognized / neither of the quantities was recognized)
Continuing counting	The child should continue counting from 10 to 20: with no errors / the child began counting again from 1 and then counted to 20 with no errors / the child counted with errors	3-point scale
Number memory	The child should remember that a total of 10 beads are hidden under the examiner's hand	Yes / No (the correct number was remembered or not)
Remembering the coloured quantities	The child should remember that 4 blue and 6 red beads are hidden under the examiner's hand	3-point scale (both quantities were remembered / one of the two quantities was remembered / neither of the quantities was remembered)
Grasping a form	Half a pine tree is dawn for the child without naming it, the child should recognize the pine tree	3-point scale (pine tree is recognized when the sheet is on the table / recognized after the sheet is raised / not recognized)
Copying a figure (cross)	The cross is drawn for the child without naming it, the child copies it	5-point scale
Copying a figure (fish)	The fish is drawn for the child without naming it, the child copies it	5-point scale