Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form: An item response theory perspective

Farid Bardid[1,2,*], Till Utesch[3,*], Matthieu Lenoir[2]

[1] *School of Education, University of Strathclyde, UK*

[2] *Department of Movement and Sports Sciences, Ghent University, Belgium*

[3] *Department of Sport Psychology, University of Münster, Germany*

* These authors contributed equally to this work.

Corresponding author: Farid Bardid
University of Strathclyde, School of Education, 141 St James Road, Glasgow G4 0LT, UK
E-mail: farid.bardid@strath.ac.uk

Abstract

Purpose: Motor assessments generally produce a single motor competence score based on the general motor ability hypothesis, which states that motor competence is a one-dimensional trait underlying a wide range of motor skills. Yet, it is unclear if the general motor ability hypothesis holds true in middle childhood which is marked by an increased participation in sports and other types of physical activity. Therefore, the aim of the study was to evaluate the structure of motor competence in middle childhood using a test battery with a large item set.

Method: A cross-sectional design was used to collect motor competence data of 2,538 children aged 6-11 years. Participants completed the Bruininks-Oseretsky Test of Motor Proficiency – 2nd Edition Short Form (BOT-2 SF), which consists of 14 skill items and covers different motor domains. In accordance with the BOT-2 SF manual, point scores were computed for each item. Polytomous Rasch analyses (i.e., general partial credit model) were carried out to investigate the construct of motor competence.

Results: Rasch analyses revealed different items with unordered threshold parameters, due to ceiling effects. However, after empirically rescaling the category width for each item, follow-up analyses revealed a one-dimensional structure with 12 items.

Conclusion: The study provides some evidence of a one-dimensional construct (i.e., motor competence) underlying motor assessment in middle childhood. Continued efforts should be made to ensure that valid composite scores are used in motor assessment and to better understand the development of motor competence across childhood and into adolescence and adulthood.

Introduction

Motor development refers to continuous changes in motor behavior over the life course, and plays a crucial role in children's overall development.[1,2] One related term is motor competence, which reflects the degree of proficiency in performing various motor skills such as running and catching as well as underlying mechanisms including motor coordination and control.[3] Motor competence is considered a cornerstone for physical activity engagement and has been shown to be related to other health outcomes such as physical fitness, weight status and perceived competence.[4–7] In this respect, it is crucial to evaluate and monitor motor competence during childhood.

Numerous motor assessments have been designed and used in clinical, educational and research contexts.[8,9] Assessment tools can be product-oriented and/or process-oriented. Product-oriented assessment focuses on the outcome of performance (e.g., jump distance, running speed) while process-oriented assessment addresses the quality of performance (e.g., arm movements during jump, arm-leg coordination during running). Assessments are generally used under the assumption that motor competence is a one-dimensional construct, as reflected by the adoption of total scores.[9,10] This assumption stems from the general motor ability hypothesis, which stipulates that there is a general motor ability that underlies numerous motor skills.[11] It should be noted that various terminologies have been adopted in the literature to define the same construct. Within the context of motor assessment, motor competence and general motor ability can be regarded as the same latent trait underlying the performance of a wide range of motor skills.[10] The choice of a motor test depends on different aspects such as test purpose (e.g., identification, program design), group characteristics (e.g., age, health difficulties) and administrative aspects (e.g., user friendliness, test time).[8] Another important aspect to consider in motor assessment is the psychometric quality of the test, i.e., validity and reliability.

The importance of validity is highlighted by Burton and Miller who stated that “an assessment instrument that is not valid is utterly useless”.[12] Moreover, motor assessments can only be as valid as the underlying construct that is being assessed. As previously mentioned, the use of total scores in motor assessment relies on the assumption that the general motor ability hypothesis holds true. Nonetheless, validity research has not provided a clear view on the structure of motor competence. While some studies demonstrated a general factor (e.g., Webster & Ulrich[13]), others found a hierarchical structure with specific factors such as balance and manual dexterity (e.g., Schulz, Henderson, Sugden, & Barnett[14]). It is suggested that the lack of clarity in the literature is (partially) due to the limitations of the measurement theory that underlies the assessment of motor competence.[10]

Classical test theory (CTT) has generally been adopted to develop and evaluate motor tests. CTT methods included in validity research are factor analyses and inter-item correlations.[15,16] Some argue that the limitations of this approach has hindered our understanding of the motor competence construct.[17] For instance, the CTT approach is item-centered, which entails that the items form the latent variable, indicating sample and item dependence of test results.[18] Another limitation of this approach is the requirement of interval-scaled variables in order to calculate correlations due to a lack of relative and absolute interpretation standards. However, raw item scores in assessment tools, such as the Motor Proficiency Test for 4- to 6-year-old Children (Motoriktest für Vier- bis Sechsjährige Kinder, MOT 4-6)[19] and the Movement Assessment Battery for Children, Second Edition (MABC-2)[20], are generally converted into point scores using an ordinal-scaled categorization system before computing a total score.[10] For instance, in the MOT 4-6, raw item scores (e.g., number of dots placed on a sheet in 10 s) are transformed into point scores ranging from 0 to 2, and subsequently summed to calculate a total score. Nonetheless, these scores are often not statistically tested for ordinal scaling, which raises questions about the validity of these measures. An alternative test theory that addresses the aforementioned limitations is item response theory (IRT). The IRT approach is trait-centered and its key feature of parameter invariance entails a sample and item distribution-free calibration of person and item parameters along the continuum of the latent construct. Additionally, ordinal-scaled item scores are transformed into an interval measure through a logit scale by defining the probabilistic relationship between item responses and the measured latent construct.[18] Furthermore, IRT models allow us to connect the content-related definition of a latent construct with assumptions of measurement theory, which can further our understanding of constructs such as motor competence.[10,21]

The application of IRT methods in the field of movement sciences has been recommended for decades.[21–25] One popular method that has been used, is the one-parameter IRT model or Rasch model,[18,26] which is based on the principles of fundamental measurement, objectivity and order.[18] These IRT methods have been used to calibrate test items, to validate test batteries and to evaluate theory. Prior IRT research on construct validity in motor assessment has generally revealed a one-dimensional latent trait underlying motor skills.[10,27,28] Although these IRT studies have provided valuable information on the construct validity of motor assessments, few have included a large item set that covers a variety of motor skills to test the general motor ability hypothesis adequately. One such study by Utesch et al.[10] examined the latent construct of motor competence in children aged 3-6 years using the 17 items of the MOT 4-6. While these findings support the general motor ability hypothesis in the early years of childhood, it remains unclear whether this holds true as children grow older. Moreover, middle childhood is characterized by an increased participation in sports, games and other types of physical activity, and the development of motor skills in specific contexts.[1] As such, the current study aimed to investigate the latent structure of motor competence in middle childhood (i.e., age groups 6-8 and 9-11 years) from an IRT perspective and to test the general motor ability hypothesis using a large item set of an existing motor test battery. The study further shows how the use of composite scores in motor assessment practices can be statistically validated as an indicator of motor competence.

Methods

A sample of 2,538 children aged 6-11 years participating in the Flemish Sports Compass project was used in the study. To obtain a representative sample, children were recruited from 29 primary schools, randomly selected across all Flemish provinces and the Brussels Capital Region in Belgium and located in urban and rural areas. Written informed consent was obtained for all participants from a parent or legal guardian. Ghent University Hospital's ethics committee provided approval for this study. Testing took place during the fall of 2007, and was conducted in indoor facilities by trained examiners.

Children's motor competence was assessed with the Bruininks Oseretsky Test of Motor Proficiency-2 Short Form (BOT-2 SF) in accordance with the manual guidelines.[29] The BOT-2 SF is a product-oriented motor test that evaluates the gross and fine motor skills of individuals aged 4–21 years. In its short form, the assessment includes 14 items from eight subtests, reflecting different motor domains: (1) fine motor precision: drawing lines through crooked paths, folding papers; (2) fine motor integration: copying a square, copying a star; (3) manual dexterity: transferring pennies; (4) bilateral coordination: jumping in place–same sides synchronized, tapping feet and fingers–same sides synchronized; (5) balance: walking forward on a line, standing on one leg on a balance beam; (6) speed and agility: stationary hops; (7) upper-limb coordination: dropping and catching a ball with both hands, dribbling a ball with alternating hands; (8) strength: knee push-ups, sit-ups.[29] In accordance with the test manual, we converted item raw scores into point scores with varying category width (i.e., 0–3 to 0–10). These item point scores can then be summed to obtain a total score. With regard to the psychometric quality of the BOT-2 SF, good test-retest reliability ($r \geq 0.80$), excellent inter-rater reliability ($r = 0.98$) and good internal consistency ($r \geq 0.80$) were reported in the manual.[29,30] Content and concurrent validity have also been established.[9,29]

SPSS 23 (IBM Corp., Armonk, NY, USA) was used to generate descriptive statistics for the BOT-2 SF item scores. Using R 3.3.2[31,32], we conducted IRT models to examine the assumed unidimensionality of motor competence across middle childhood. The generalized partial credit model (GPCM[33]) was used to analyze homogeneity and order within the construct of motor competence. The GPCM is an extension of the partial credit model (PCM[34]), which in turn is a (polytomous) generalization of the initial Rasch model for dichotomous data.[26] In contrast to the PCM, the GPCM allows different category widths among items, as present in the BOT-2 SF (e.g., drawing lines [0–7] vs. walking forward on a line [0–4]). Ordinal data is modeled through the estimation of person ability θ on a logit scale. For each item, item difficulty as well as threshold parameters b between categories i were computed; threshold parameters should be ordered as follows: $b_i < b_{i+1}$. Both item difficulty and threshold parameters were estimated on the same scale as person ability.

The GPCM provides on item level χ^2 goodness of fit statistics for each item, with *p*-values above .05 indicating good model fit. Furthermore, ordered threshold parameters show valid ordinal order, which is indicated by each category having a maximum in the continuum of the latent trait. Items with unordered threshold parameters or violating fit statistics need to be excluded from the model. Within the IRT framework, the concept of reliability has been extended from a single index (e.g., Cronbach’s alpha, omega h) —which is helpful to compare average reliability between tests— towards a function measuring the precision of measurement across the continuum. From an IRT perspective, it is common that the standard measurement error will be larger at the ends of the continuum of the measured latent trait. Lower standard measurement error would therefore mean more information[35]. In order to examine possible changes in the latent structure with age, data-analyses were conducted for age groups 6-8 years and 9-11 years separately.

Results

Table 1 shows the medians and score distributions of the BOT-2 SF items. Most items displayed skewed distributions. The GPCM, used to test the assumed unidimensional structure of motor competence, showed no global model fit with the 14 items of the BOT-2 SF. For both age groups 6-8 years and 9-11 years, items three and four (i.e., copying a square; copying a star) did not fit the model. However, after excluding both items, unidimensionality was shown for the remaining 12 items covering the motor dimensions fine motor precision, manual dexterity, bilateral coordination, balance, speed and agility, upper-limb coordination, and strength ($19.92 \leq \chi 2 \leq 60.71$; $.06 \leq p \leq .97$, see also Supplementary Table 1). With regard to the order of the categories, most items showed unordered threshold parameters in both 6- to 8-year-old children and 9- to 11-year-old children (see Supplementary Table 2 for violations in $b_i < b_{i+1}$). These unordered threshold parameters are illustrated in Fig. 1 by the overlapping item characteristic curves (see Fig. 1A-B). These indicate violations of order within the ordinal scale. Hence, we empirically merged categories for these items based on (a) unordered thresholds, and (b) low cell counts. For both 6- to 8-year-old and 9- to 11-year-old children, model requirements were met for the recoded categories as indicated by ordered threshold parameters (see Fig. 1C-D) as well as all items fitted into the unidimensional model (all *p* values > 0.05; see also Supplementary Table 2).

Test information and reliability functions showed good sensitivity and reliability across the continuum of motor competence for 6-8 year old children with decreased information at the very end of the scale. However, for 9- to 11-year-old children information and reliability functions indicate decreasing information and reliability for the upper proficiency level of motor competence. This means, the standard measurement error is larger for the upper end of the motor competence continuum in 9- to 11-year-old children, compared to their 6- to 8-year-old counterparts (see Fig. 2).

Discussion

The role of motor development in children's health has underlined the importance to evaluate and monitor motor competence across childhood.[4,5] Motor assessments generally adopt a composite score—based on various skill performances—as an indicator of motor competence. The use of a single score is based on the assumption that motor competence is a one-dimensional construct (i.e., general motor ability hypothesis). In spite of this practice, the dimensionality of the related theoretical underlying construct is not yet fully understood, in part due to the methodological limitations of the widely used CTT approach.[10,17] Thus, the current study aimed to investigate the latent trait motor competence in middle childhood from an IRT perspective with regard to the general motor ability hypothesis.

This study showed a one-dimensional construct of motor competence in middle childhood, which is in line with previous research. For instance, Utesch et al.[27] found a one-dimensional construct in children aged 9-11 years using the German Motor Ability Test 6-18, which consists of eight items (Deutscher Motorik-Test 6-18, DMT[36]). Another study of Utesch et al.[37] showed a single latent trait in children aged 6-9 years, using the same test battery. In contrast to the aforementioned studies, the present research used a large item set (i.e., 14 items) in order to evaluate the general motor ability hypothesis more adequately. Similarly as the investigation of Utesch et al.[10] in early childhood, the present study adopted the large item set of an existing test battery. The BOT-2 SF is a widely adopted assessment tool that reflects different skills and motor domains. Bruininks and Bruininks[29] indicated that the total score (i.e., sum of the 14 item scores) is an indicator of children's motor competence, thus implying a single construct. The Rasch analysis showed a global model fit with 12 BOT-2 SF items. However, due to ceiling effects in the BOT-2 (Brahler et al.[38]; see also Table 1), many items showed unordered threshold parameters indicating that the categories were not related to a child's performance level (see Supplementary Table 1). Nonetheless, after empirically rescaling the category widths, follow-up analyses showed a model fit with ordered threshold parameters. The 12 model conform items of the BOT-2 SF can then be used to compute a valid composite score, reflecting children's motor competence.

Two BOT-2 SF items were omitted because they did not fit the Rasch model (i.e., copying a square; copying a star). Further examination revealed that the excluded items have a process-oriented scoring system in contrast to the product-oriented scoring system for the other 12 BOT-2 SF items. In keeping with prior research[39,40], these findings seem to suggest that product-oriented and process-oriented measures are not equivalent. Still, as noted by Robinson et al.[4], both approaches can be used in motor assessment to provide a more comprehensive view of motor competence.

The present investigation revealed a one-dimensional construct of motor competence for age groups 6-8 years and 9-11 years. Interestingly, the results did show differences in test information and reliability between both age groups, depending on the competence levels. For children aged 6-8 years, the test is suitable when assessing below-average to above-average competence levels. In contrast, for children aged 9-11 years, the suitability of the BOT-2 SF decreases with higher competence levels, as indicated by a less reliable measurement at the higher end of the motor competence continuum. These findings are further supported by differences in item difficulties between younger and older children, in which most items are (much) easier to perform for 9- to 11-year-old children compared to their younger counterparts (see Supplementary Table 2). This may indicate differences in test sensitivity across age and competence levels. Another possible explanation for the age differences in measurement accuracy is that motor competence becomes less general as children grow older. Indeed, Burton and Rodgerson argued that there may be an increased specificity of motor competence or "differentiation of abilities and skills" with age.[17] As children mature, they refine their basic or fundamental motor skills (e.g., jumping; striking) and develop specialized skills (e.g., rope skipping; tennis forehand strike) as they engage in sports, games and other types of physical activity.[1,2] Although there is some evidence for this developmental aspect,[14] the notion of increased specificity over time has generally not been considered in the development and application of motor assessments.[17] A hierarchical structure underlying motor behavior as proposed by different authors[17,41,42] may provide a suitable theoretical framework to capture changes in the construct of motor competence over time, but requires further empirical research. In light of the restrictions of CTT methods such as factor analyses, IRT methods can provide a useful statistical approach to examine the dimensionality and structure of motor competence, and to support the development and adoption of motor assessments in research and practice.

The main strength of the study is the use of a large item set (see Table 1), which has allowed an adequate evaluation of the general motor ability hypothesis. This study also included a large sample of 2538 children with a wide age range of 6-11 years. Additionally, this study adopted an existing test battery (i.e., BOT-2 SF) and provided information on the valid use of the instrument in research and practice. It should be noted that the BOT-2 SF was developed for a wider age group (i.e., 6-21 years) than was included in the present study. Future investigations are needed to examine the latent structure of motor competence in adolescence and to validate assessment tools such as the BOT-2 SF in older age groups. A limitation of the study is its cross-sectional design, which has not allowed to investigate the structural development of children's motor competence over time. As noted by other authors, in order to examine changes in motor competence across childhood and into adolescence and adulthood, it is important that future studies adopt a longitudinal design.[4,10]

Perspective

The importance of assessing and monitoring motor competence for children's health[4] necessitates valid measurements to enable appropriate interpretation. For this, a clear understanding of the latent trait underlying motor assessment is needed. This study provides insights into the theoretical definition of motor competence in middle childhood. Using the IRT approach, the study revealed a one-dimensional structure underlying multiple motor domains and skills, providing support for the general motor ability hypothesis in children aged 6-11 years. These findings are in line with previous IRT research on motor assessment in early childhood[10] and support the use of composite scores in practice. The decrease in test reliability for older children scoring above average indicates that motor competence becomes less general as children grow older and develop skills in contexts of sports, games and other types of physical activity.[1,17] However, further longitudinal research in children and adolescents is needed to explore the construct of motor competence and potential changes over time. The IRT approach provides a unique perspective into motor assessment due to its methodological advantages[22] and could be used in conjunction with other measurement approaches such as CTT to better understand motor competence, and to validate tests for use in research and practice.

Acknowledgements

We wish to thank all children, parents and schools that participated in the Flemish Sports Compass project, funded by the Flemish Government.

References

1. Gallahue DL, Ozmun JC, Goodway JD. *Understanding Motor Development: Infants, Children, Adolescents, Adults*. 7th ed. New York, NY: McGraw-Hill; 2012.
2. Haywood KM, Getchell N. *Life Span Motor Development*. 6th ed. Champaign, IL: Human Kinetics; 2014.
3. Utesch T, Bardid F. Motor competence. In: Hackfort D, Schinke R, Strauss B, eds. *Dictionary of Sport Psychology: Sport, Exercise, and Performing Arts*. Amsterdam: Elsevier; 2019:186.
4. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. *Sport Med*. 2015;45(9):1273-1284.
5. Barnett LM, Lai SK, Veldman SLC, et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sport Med*. 2016;46(11):1663-1688. doi:10.1007/s40279-016-0495-z
6. Utesch T, Dreiskämper D, Naul R, Geukes & K. Understanding physical (in-) activity, overweight, and obesity in childhood: Effects of congruence between physical self-concept and motor competence OPEN. *Sci Rep*. 2018;8(5908):1-10. doi:10.1038/s41598-018-24139-y
7. Utesch T, Bardid F, Büsch D, Strauss B. The relationship between motor competence and physical fitness from early childhood to early adulthood: A meta-analysis. *Sport Med*. 2019;49(4):541-551. doi:10.1007/s40279-019-01068-y
8. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. *J Sci Med Sport*. 2019;22(3):311-318. doi:10.1016/j.jsams.2018.08.007
9. Cools W, De Martelaer K, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. *J*

Sports Sci Med. 2009;8(2):154-168.

10. Utesch T, Bardid F, Huyben F, et al. Using Rasch modeling to investigate the construct of motor competence in early childhood. *Psychol Sport Exerc*. 2016;24:179-187. doi:10.1016/j.psychsport.2016.03.001
11. Brace DK. *Measuring Motor Ability: A Scale of Motor Ability Tests*. New York: A. S. Barnes; 1927.
12. Burton AW, Miller DE. *Movement Skill Assessment*. Champaign, IL: Human Kinetics; 1998.
13. Webster EK, Ulrich DA. Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition. *J Mot Learn Dev*. 2017;5(1):45-58. doi:10.1123/jmld.2016-0003
14. Schulz J, Henderson SE, Sugden DA, Barnett AL. Structural validity of the Movement ABC-2 test: Factor structure comparisons across three age groups. *Res Dev Disabil*. 2011;32(4):1361-1369.
15. Bös K. Motorische Leistungsfähigkeit von Kindern und Jugendlichen. In: Schmidt W, Hartmann-Tews I, Brettschneider W-D, eds. *Erster Deutscher Kinder Und Jugendsportbericht*. Schorndorf: Verlag Karl Hofmann; 2003:1-23.
16. Ulrich DA. *Test of Gross Motor Development*. Austin, TX: PRO-ED; 1985.
17. Burton AW, Rodgerson RW. New perspectives on the assessment of movement skills and motor abilities. *Adapt Phys Act Q*. 2001;18:347-365.
18. Masters GN. Objective measurement. In: Alagumalai S, Curtis D, Hungi N, eds. *Applied Rasch Measurement: A Book of Exemplars*. The Netherlands: Springer; 2005:15-25.
19. Zimmer R, Volkamer M. *Motoriktest Für Vier- Bis Sechsjährige Kinder [Motor Proficiency Test for 4- to 6-Year-Old Children]*. Weinheim: Beltz Test; 1987.
20. Henderson SE, Sugden DA, Barnett A. *Movement Assessment Battery for Children*. 2nd ed. London: Harcourt Assessment; 2007.

21. Strauss B, Büsch D, Tenenbaum G. Rasch modeling in sports. In: Tenenbaum G, Kamata A, Bar-Eli M, eds. *Handbook of Measurement in Sport and Exercise Psychology*. New York: Human Kinetics; 2012:75-81.

22. Linacre JM. New approaches to determining reliability and validity. *Res Q Exerc Sport*. 2000;71(2):129-136.

23. Hands B, Larkin D, Sheridan B. Rasch Measurement Applied to Young Children. *Aust Educ Dev Psychol*. 1997;14(01):11-22. doi:10.1017/S0816512200027619

24. Hands B, Sheridan B, Larkin D. Creating performance categories from continuous motor skill data using a Rasch measurement model. *J Outcome Meas*. 1999;3(3):216-232.

25. Spray JA. Recent developments in measurement and possible applications to the measurement of psychomotor behavior. *Res Q Exerc Sport*. 1987;58(3):203-209.

26. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danisch Institute for Educational Research; 1960.

27. Utesch T, Strauß B, Tietjens M, Büsch D, Ghanbari M-C, Seidel I. Die Überprüfung der Konstruktvalidität des Deutschen Motorik-Tests 6-18 für 9- bis 10-Jährige [Assessment of the construct validity of the German Motor Test 6-18 for 9- to 10-Year-Olds]. *Zeitschrift für Sport*. 2015;22(2):77-90. doi:10.1026/1612-5010/a000143

28. Hands B, Larkin D. Using the Rasch measurement model to investigate the construct of motor abilities in young children. *J Appl Meas*. 2001;2(2):101-120.

29. Bruininks R, Bruininks B. *Bruininks-Oseretsky Test of Motor Proficiency*. 2nd ed. Minneapolis, MN: Pearson Assessment; 2005.

30. Deitz JC, Kartin D, Kopp K. Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2). *Phys Occup Ther Pediatr*. 2007;27(4):87-102. doi:10.1300/J006v27n04_06

31. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016. https://www.r-project.org.

32. Chalmers RP. mirt : A Multidimensional Item Response Theory Package for the R Environment. *J Stat Softw*. 2012;48(6):1-29. doi:10.18637/jss.v048.i06

33. Muraki E. A generalized partial credit model: Application of an EM algorithm. *Appl Psychol Meas*. 1992;16:159-176. doi:10.1002/j.2333-8504.1992.tb01436.x

34. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149-174.

35. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

36. Bös K, Schlenker L, Büsch D, et al. *Deutscher Motorik Test 6-18 [German Motor Ability Test 6-18]*. Hamburg: Czwalina; 2009.

37. Utesch T, Dreiskämper D, Strauss B, Naul R. The development of the physical fitness construct across childhood. *Scand J Med Sci Sports*. April 2017:1-8. doi:10.1111/sms.12889

38. Brahler CJ, Donahoe-Fillmore B, Mrowzinski S, Aebker S, Kreill M. Numerous Test Items in the Complete and Short Forms of the BOT-2 Do Not Contribute Substantially to Motor Performance Assessments in Typically Developing Children Six to Ten Years Old. *J Occup Ther Sch Early Interv*. 2012;5(1):73-84. doi:10.1080/19411243.2012.674746

39. Logan SW, Barnett LM, Goodway JD, Stodden DF. Comparison of performance on process- and product-oriented assessments of fundamental motor skills across childhood. *J Sports Sci*. May 2016. doi:10.1080/02640414.2016.1183803

40. Ré AHN, Logan SW, Cattuzzo MT, Henrique RS, Tudela MC, Stodden DF. Comparison of motor competence levels on two assessments across childhood. *J Sports Sci*. January 2017:1-6. doi:10.1080/02640414.2016.1276294

41. Cratty BJ. A Three Level Theory of Perceptual-Motor Behavior. *Quest*. 1966;6(1):3-9. doi:10.1080/00336297.1966.10519605

42. Hands B, McIntyre F, Parker H. The General Motor Ability Hypothesis: An Old Idea

Revisited. *Percept Mot Skills*. January 2018. doi:10.1177/0031512517751750

Table 1. Descriptive statistics for the BOT-2 SF items, stratified by age group.

			Range	*Mdn*	Count										
Age group		Test item			**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
6-8 years (*N* = 1243)	1	Draw lines through crooked paths	0-7	7	10	2	8	47	79	215	223	659			
	2	Folding paper	0-7	5	29	50	72	106	161	260	170	395			
	3	Copying a square	0-5	3	6	64	300	543	267	63					
	4	Copying a star	0-5	2	145	216	380	364	120	18					
	5	Transferring pennies	0-9	5		1	18	130	337	414	258	79	6		
	6	Jumping in place	0-3	3	17	23	165	1038							
	7	Tapping feet and fingers	0-4	4	12	10	51	198	972						
	8	Walking forward on a line	0-4	4		2	18	45	1178						
	9	Standing on one leg on a balance beam	0-4	4	1	87	179	194	782						
	10	Stationary hop	0-10	9				1	0	3	8	19	315	604	293
	11	Dropping and catching a ball	0-5	5	36	45	66	113	175	808					
	12	Dribbling a ball	0-7	5	11	48	101	172	278	173	94	366			
	13	Knee push-ups	0-9	5	5	4	12	51	173	410	350	178	47	13	
	14	Sit-ups	0-9	5	32	24	85	132	266	327	253	90	26	8	
9-11 years (*N* = 1295)	1	Draw lines through crooked paths	0-7	7				11	19	78	133	1054			
	2	Folding paper	0-7	7	5	3	14	28	57	119	164	905			
	3	Copying a square	0-5	4	1	21	101	514	540	118					
	4	Copying a star	0-5	3	35	84	264	499	363	50					
	5	Transferring pennies	0-9	6				9	92	288	441	344	104	17	
	6	Jumping in place	0-3	3	6	6	71	1212							
	7	Tapping feet and fingers	0-4	4			26	90	1179						
	8	Walking forward on a line	0-4	4		1	9	20	1265						
	9	Standing on one leg on a balance beam	0-4	4	1	22	70	133	1069						
	10	Stationary hop	0-10	10					1	1	1	5	149	415	723
	11	Dropping and catching a ball	0-5	5	8	9	14	42	119	1103					
	12	Dribbling a ball	0-7	7	1	2	15	44	119	117	119	878			
	13	Knee push-ups	0-9	6		2	9	21	82	208	380	361	168	64	
	14	Sit-ups	0-9	6	2	5	10	48	105	311	402	281	100	31	

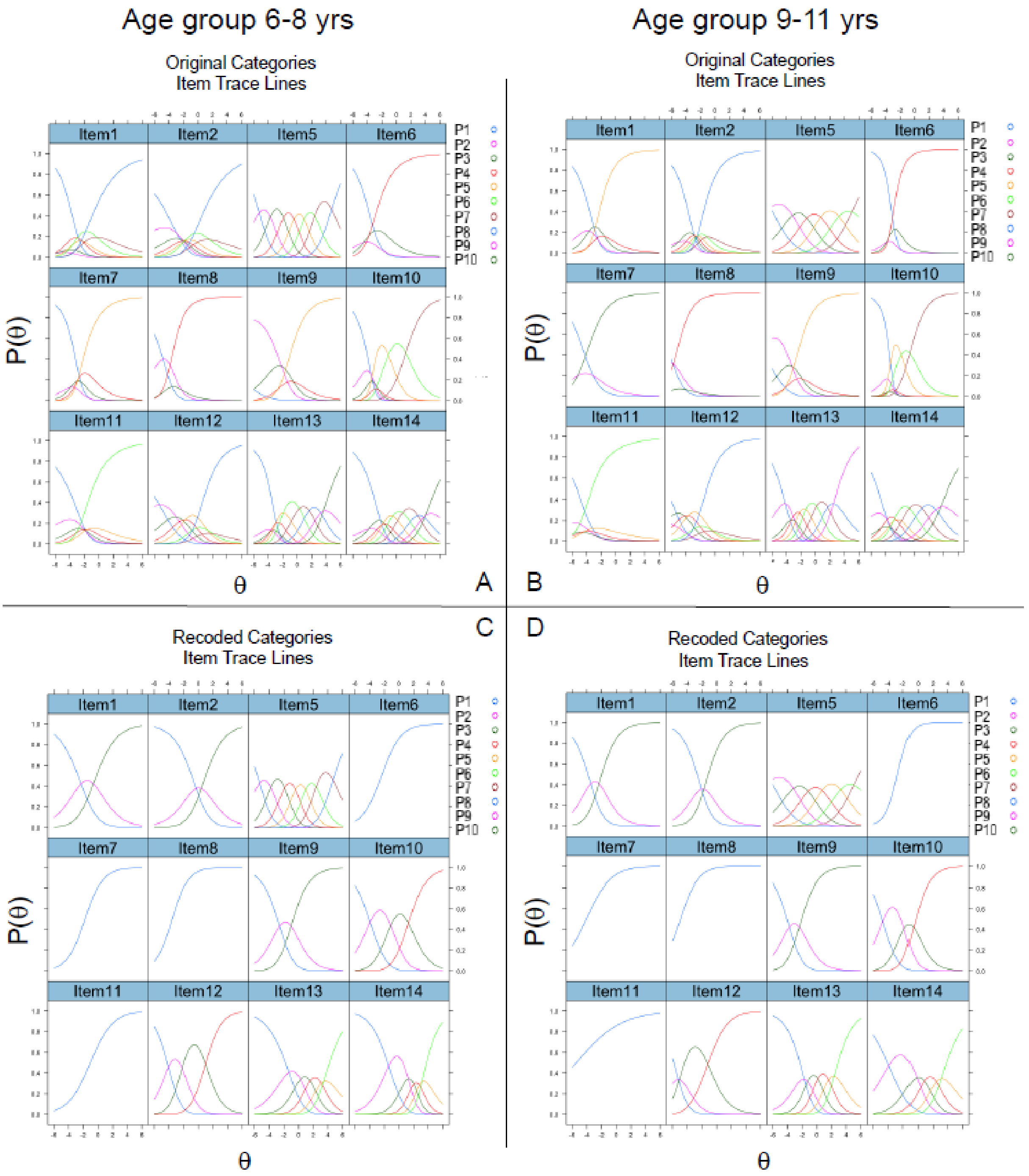

Fig. 1. Category trace lines for the 12 BOT-2 SF items meeting the model requirements. P*i* represents the probability of category *i* at a certain skill level θ (e.g., P1 represents the probability of category 1 for each item). Original categories for 6-8-year-olds are shown in A (top left) and for 9-11-year-olds in B (top right). Statistical categories for 6-8-year-olds are shown in C (bottom left) and for 9-11-year-olds in D (bottom right).

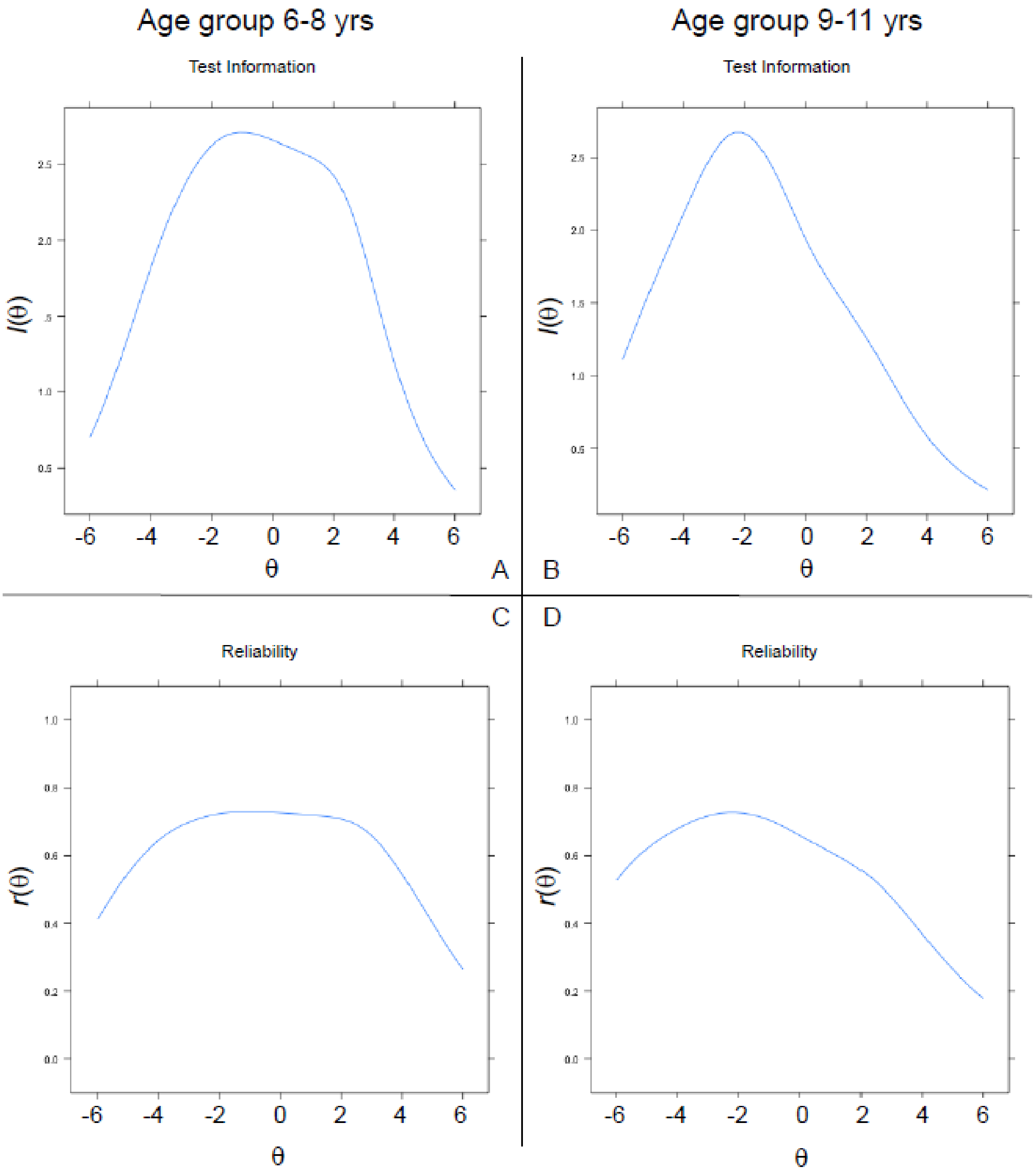

Fig. 2. Test information and reliability curves of the 12-item IRT models. The curves represent the quality of information $I(\theta)$ as well as the reliablity $r(\theta)$ of the measurement the test provides at a certain skill level θ. Curves for 6-8-year-olds are shown in A and C (left) and for 9-11-year-olds in B and D (right).

Supplementary Table 1.

Item fit statistics.

		6-8 years			9-11 years		
	Test item	χ^2	*df*	*p*	χ^2	*df*	*p*
1	Draw lines through crooked paths	40.81	33	0.16	13.63	25	0.97
2	Folding paper	40.07	32	0.15	36.19	27	0.11
5	Transferring pennies	46.19	55	0.80	55.97	46	0.15
6	Jumping in place	21.11	18	0.27	9.25	14	0.81
7	Tapping feet and fingers	18.92	19	0.46	16.58	16	0.41
8	Walking forward on a line	24.92	17	0.10	12.28	13	0.51
9	Standing on one leg on a balance beam	41.55	31	0.10	21.13	23	0.57
10	Stationary hop	45.48	40	0.25	38.44	26	0.06
11	Dropping and catching a ball	23.88	19	0.20	13.68	17	0.69
12	Dribbling a ball	31.86	31	0.42	22.45	23	0.49
13	Knee push-ups	60.71	57	0.34	39.04	44	0.68
14	Sit-ups	53.99	50	0.32	46.66	46	0.44

Note: There is item fit if $p > 0.05$.

Supplementary Table 2

Item difficulty and threshold parameters for the BOT-2 SF items meeting the model requirements.

				Threshold parameters								
Age group		Test item	Item difficulty	**b1**	**b2**	**b3**	**b4**	**b5**	**b6**	**b7**	**b8**	**b9**
6-8 years (original categories)	1	Draw lines through crooked paths	-2.57	3.77	-5.59	-6.58	-2.33	-3.62	-0.27	-3.35		
	2	Folding paper	-1.67	-2.91	-2.08	-1.95	-1.84	-1.84	1.62	-2.68		
	5	Transferring pennies	-0.45	-5.22	-3.70	-1.90	-0.44	0.96	2.40	4.75		
	6	Jumping in place	-3.30	-1.62	-4.49	-3.78						
	7	Tapping feet and fingers	-2.73	-1.12	-3.90	-2.97	-2.91					
	8	Walking forward on a line	-4.17	-4.83	-2.28	-5.41						
	9	Standing on one leg on a balance beam	-3.81	-9.86	-2.17	-0.52	-2.67					
	10	Stationary hop	-2.23	-3.47	-3.10	-2.45	-4.52	-1.11	1.30			
	11	Dropping and catching a ball	-2.39	-1.67	-1.92	-2.14	-1.56	-4.66				
	12	Dribbling a ball	-1.76	-5.33	-2.98	-2.12	-1.72	1.32	1.97	-3.49		
	13	Knee push-ups	-0.52	-1.41	-3.44	-3.73	-2.96	-1.94	0.30	1.65	3.25	3.60
	14	Sit-ups	0.35	-0.56	-3.81	-1.65	-1.94	-0.51	0.86	2.96	3.77	3.99
6-8 years (recoded categories)	1	Draw lines through crooked paths	-1.47	-2.30	-0.63							
	2	Folding paper	0.06	-0.29	0.40							
	5	Transferring pennies	-0.48	-5.29	-3.78	-1.92	-0.43	0.97	2.39	4.70		
	6	Jumping in place	-2.36	-2.36								
	7	Tapping feet and fingers	-1.76	-1.76								
	8	Walking forward on a line	-3.55	-3.55								
	9	Standing on one leg on a balance beam	-1.81	-2.59	-1.03							
	10	Stationary hop	-1.25	-3.92	-1.12	1.29						
	11	Dropping and catching a ball	-1.03	-1.03								
	12	Dribbling a ball	-1.76	-4.06	-2.20	0.99						
	13	Knee push-ups	1.40	-1.32	0.26	1.57	3.08	3.41				
	14	Sit-ups	1.57	-1.40	1.28	2.04	2.79	3.16				

Supplementary Table 2 (continued).

				Threshold parameters								
Age group		Test item	Item difficulty	**b1**	**b2**	**b3**	**b4**	**b5**	**b6**	**b7**	**b8**	**b9**
9-11 years (original categories)	1	Draw lines through crooked paths	-2.90	-2.39	-3.64	-1.56	-4.01					
	2	Folding paper	-2.64	-0.62	-5.08	-2.85	-2.60	-2.35	-1.06	-3.89		
	5	Transferring pennies	-0.26	-6.18	-3.15	-1.18	0.71	3.23	5.01			
	6	Jumping in place	-3.05	-1.70	-3.88	-3.57						
	7	Tapping feet and fingers	-4.18	-3.10	-5.25							
	8	Walking forward on a line	-5.76	-5.80	-2.37	-9.10						
	9	Standing on one leg on a balance beam	-4.18	-7.45	-3.25	-1.82	-4.21					
	10	Stationary hop	-2.64	-2.27	-2.04	-3.64	-5.37	-1.86	-0.66			
	11	Dropping and catching a ball	-4.28	-1.47	-2.47	-4.71	-4.25	-8.48				
	12	Dribbling a ball	-3.58	-3.48	-7.15	-4.19	-3.69	-0.46	-0.27	-5.79		
	13	Knee push-ups	-1.18	-4.15	-2.83	-3.28	-2.22	-1.30	0.15	1.71	2.47	
	14	Sit-ups	-1.12	-3.75	-2.97	-4.65	-2.56	-2.93	-0.72	1.02	2.91	3.58
9-11 years (recoded categories)	1	Draw lines through crooked paths	-2.92	-3.47	-2.36							
	2	Folding paper	-1.89	-2.05	-1.73							
	5	Transferring pennies	-0.28	-6.28	-3.18	-1.18	0.72	3.25	5.02			
	6	Jumping in place	-2.72	-2.72								
	7	Tapping feet and fingers	-4.09	-4.09								
	8	Walking forward on a line	-4.92	-4.92								
	9	Standing on one leg on a balance beam	-3.07	-3.74	-2.39							
	10	Stationary hop	-2.38	-4.70	-1.82	-0.62						
	11	Dropping and catching a ball	-5.44	-5.44								
	12	Dribbling a ball	-3.82	-5.19	-5.08	-1.19						
	13	Knee push-ups	0.23	-1.70	-1.24	0.14	1.60	2.35				
	14	Sit-ups	0.46	-3.87	-0.17	0.83	2.42	3.09				