Utesch, Till and Bardid, Farid and Huyben, Floris and Strauss, Bernd and Tietjens, Maike and De Martelaer, Kristine and Seghers, Jan and Lenoir, Matthieu (2016) Using Rasch modeling to investigate the construct of motor competence in early childhood. Psychology of Sport and Exercise, 24. pp. 179-187. ISSN 1469-0292 , http://dx.doi.org/10.1016/j.psychsport.2016.03.001

Using Rasch modeling to investigate the construct of motor competence in early childhood

Till Utesch[1,*], Farid Bardid[2,*], Floris Huyben[3], Bernd Strauss[1], Maike Tietjens[1], Kristine De Martelaer[3], Jan Seghers[4], Matthieu Lenoir[2]

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

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

[3] *Department of Movement and Sports Training, Vrije Universiteit Brussel, Belgium*

[4] *Department of Kinesiology, KU Leuven – University of Leuven, Belgium*

* These authors equally contributed to this work.

Corresponding author: Till Utesch

University of Münster, Department of Sport Psychology, Horstmarer Landweg 62b, 48419 Münster, Germany

E-mail: Till.Utesch@uni-muenster.de

Abstract

Purpose: The present study investigated the dimensionality and homogeneity of motor competence, which is defined as the ability that underlies the performance of a wide variety of motor skills, in early childhood using a large set of items.

Method: A total of 1467 children (aged 3-6 years) were measured with the Motor Proficiency Test for 4- to 6-Year-old Children (Motoriktest für vier- bis sechsjährige Kinder [MOT 4-6]), which consists of 17 items.

Results: Analyses using the Partial Credit Model and mixed Rasch model revealed a one-dimensional structure (CR = 1.964, p_{CR} = .06; $P\text{-}\chi^2$ = -.227, $p_{P\text{-}\chi^2}$ = .24). Due to unordered threshold parameters, five items were excluded. These items have a scoring system that counts the amount of successful trials (0-2).

Conclusion: The study shows item and person homogeneity within a validated motor score, using 12 items of MOT 4-6. Thus, it provides evidence of a single latent construct (i.e., motor competence), which underlies the performance of motor skills in early childhood. Furthermore, it shows that counting the number of successful trails may be less suitable as a scoring system in motor competence assessment. Present findings also support the use of validated composite scores in motor assessment.

Introduction

Motor competence and related constructs

Motor development is considered an important factor in children's overall health (Hill, 2010; Lubans, Morgan, Cliff, Barnett, & Okely, 2010; Robinson et al., 2015; Stodden et al., 2008). In spite of its significance, a common understanding of the latent construct of motor behavior underlying assessment is lacking. Different hypotheses and concepts have been introduced to explain motor behavior. One popular hypothesis is the classic *General Motor Ability Hypothesis* which states that numerous motor abilities are highly related within a person and form a single general motor ability (Brace, 1927). In their well-known taxonomy, Burton and Miller (1998) defined movement skills, motor abilities and general motor ability in a hierarchical order with movement skills at the top and general motor ability at the bottom. This taxonomy was further elaborated upon by Burton and Rodgerson (2001). Movement skills are defined as a specific group of goal-directed movement patterns, which can be altered through instruction and practice (Burton & Miller, 1998; Burton & Rodgerson, 2001). Motor abilities are described as "general traits or capacities of an individual, that underlie the performance of a variety of movement skills" (Burton & Miller, 1998). This concept has been frequently investigated, e.g. the classification schemes of Fleishman (1964) and Bös (2001). The underlying component in Burton and Miller's (1998) taxonomy is the general motor ability that governs all movement skills. In the research field of motor development, different terminologies are applied to describe the same construct. For instance, movement skills and motor skills are used interchangeably (Gabbard, 2008). Another example is motor competence which refers to the ability to execute a wide variety of motor skills, including both gross (e.g., jumping) and fine motor skills (e.g., manual dexterity) (Haga, 2008). In the context of motor assessment, motor competence can be regarded as general motor ability because – by definition – both are often implicitly measured in assessment tools through a composite score that is built out of a wide range of test items from different motor abilities or motor domains (Burton & Rodgerson, 2001).

Motor competence assessment and underlying theoretical assumptions

Various motor tests have been developed for children and used in both research and educational settings (see for a review Cools et al., 2009). Motor assessment and monitoring are specifically important during early childhood as the preschool years form a sensitive age period for motor development (Gallahue & Cleland-Donnelly, 2007; Haywood & Getchell, 2009). Different aspects need to be considered when selecting an appropriate test, including the total test time (and the relative time amount for each item) and the suitability of the test for the target group (Cools et al., 2009). Another important factor is the purpose of the assessment, which is related to the research or educational question (Mahar & Rowe, 2008). Test instruments are constructed using different theoretical assumptions (e.g., product- or process-oriented approach). These tests should therefore be thoroughly tested on validity and reliability. In general, test instruments can only be as valid as the theoretical construct that is proposed. In turn, the validity of the construct is closely related to the theoretical assumptions, which are specified by a theory of measurement.

The theory of measurement that has generally been used to develop motor test batteries is the classical test theory (CTT). As such, research on the underlying latent trait(s) was mostly conducted using CTT methods like factor analysis and inter-item correlation, which resulted in either hierarchical classification schemes such as muscular strength, endurance, balance and reaction (Bös, 2001; Fleishman, 1964; Rarick, Dobbins, & Broadhead, 1976), or single factor scales (Bruininks, 1978; Calder, 1979; Ulrich, 1985). In spite of limited support for the *General Motor Ability Hypothesis* provided by CTT studies, the concept of a single latent trait (i.e., motor competence) was included in the taxonomy of Burton and Miller (1998) and underpins many widely used assessments (Burton & Rodgerson, 2001). Burton and Rodgerson (2001) argued that the lack of evidence might be due to an inappropriate analysis approach which has dismissed the construct of motor competence due to low correlations between motor composite scores within and between motor tests. For instance, Fransen et al. (2014) found a correlation of .62 between the total scores of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2; Bruininks & Bruininks, 2005) and the Body Coordination Test (Körperkoordinationstest für Kinder [KTK]; Kiphard & Schilling, 1974, 2007) in primary school children (6-12 years). In another example, Cools et al. (2010) reported a correlation of .68 between the Movement Assessment Battery for Children (M-ABC; Henderson & Sugden, 1992; Smits-Engelsman, 1998) and the Motor Proficiency Test for 4- to 6-Year-Old Children (Motoriktest für Vier- bis Sechsjährige Kinder [MOT 4-6]; Zimmer & Volkamer, 1987) in preschool children (4-6 years). Similar results are found in other convergent validity studies (Croce, Horvat, & McCarthy, 2001; Kaplan, Wilson, Dewey, & Crawford, 1998; Smits-Engelsman, Henderson, & Michels, 1998).

The argument against the construct of motor competence is mostly based on the general finding of correlations below .70. In a review on individual differences in motor performance, Marteniuk (1974) indicated that a general factor could only be supported if correlations account for ≥ 50% of the variance which has led to the arbitrary cut-off value of .70. However, Cohen (1992) stated that correlation coefficients of .50 are considered to be high in the field of behavioral sciences. In this regard, Burton and Rodgerson (2001) argued that the arbitrary criterion of .70 might not be appropriate to produce valid conclusions about the construct of motor competence. The use of these cut-off values is solely based on human judgment. This shows that using correlations as a criterion to answer those questions implies active choices made by researchers. From a content view one can debate, if a correlation above .4, .5 or .7 would be an indicator of a one-dimensional latent variable. Conducting factorial analysis to investigate the dimensionality of the latent variable comes along with (reasonable) choices such as setting the parameter estimation fixed, free or constrained or including correlated errors to improve model fit (Little, 2013). The CTT approach contains some additional limitations in the context of motor assessment such as sample and item dependence of results (Masters, 2005; Rost, 2004). Additionally, raw item scores are located on different scales. In the process of test construction, these item scores are transformed into an ordinal-scaled categorization system and often summed to a composite score. However, the CTT approach requires interval-scaled variables to conduct correlations. Because composite scores and categorizations are often not statistically verified for ordinal scaling and validity, the lack of a validated theoretical framework hampers the development of meaningful measures in motor assessment.

Rasch modeling in motor competence assessment

Alternative approaches that address the above mentioned limitations are models of item response theory (IRT; also known as probabilistic test theory). IRT models can be valuable when investigating the construct of motor competence, because they address the content related definition of motor competence and link it with test theoretical assumptions (see Strauss, Büsch, and Tenenbaum (2007, 2012) for an overview in the field of sport psychology. IRT models use test and item scores and define the mathematical relationship between the measured latent variable (e.g., motor competence) and the item responses (Sivakumar Alagumalai, Curtis, & Hungi, 2005; Rost, 2004). The major advantage of IRT models is the of parameters, which defines the equality of person and item parameters along different populations (Rost, 2004). This means model conform data imply sample distribution free and indicator distribution free results along the continuum of the measured latent trait. Person ability as well as item raw scores from different measurement units can be measured onto the same scale (logit scale), which is interval-scaled. One of the basic IRT models is the one-parameter Rasch model for dichotomous data (Rasch, 1960), which is based on the concept of fundamental measurement, objectivity and order (Masters, 2005). Since its introduction in 1960, a variety of different Rasch measurement models have been developed.

The use of IRT models in the context of motor assessment has been recommended for decades (Linacre, 2000; Spray, 1987; Strauss et al., 2007; Strauss, 1999; Tenenbaum, Strauss, & Büsch, 2007; Wright & Mok, 2000; Zhu et al., 2011). Beside some work calibrating different test items in the context of motor assessment IRT models can also be used to validate test batteries or to help evaluating, confirming or developing theory. For instance Linacre (2000) applied the Rasch model to the AAHPERD Youth Fitness Test (AAHPERD, 1976) and calibrated the seven items (n = 40). Zhu and Cole (1996) calibrated the Test of Gross Motor Development (Ulrich, 1985) for three to ten year-old children (*n* = 909) and Zhu et al. (2011) calibrated 30 items for children in kindergarten, 2nd and 5th grade. Using the mixed Rasch model Büsch et al. (2009) analyzed two samples of primary school aged children (sample 1: 6-11 years, *M* = 8.4; sample 2: 9-11 years, *M* = 10.28) who completed the six items of the General Sport Motor Test for Children (Allgemeiner Sportmotorischer Test für Kinder [AST]; Bös, 2000). A two-dimensional structure in terms of skill difference between ball handling and locomotion was found in this age group. There are several studies which found one-dimensional scales within a wide range of various item sets. For example, Hands and Larkin (2001) found a separate scale each for five- to six-year-old boys and girls (*n* = 332) out of a wide range of 24 items. Yan and Bond (2011) used the “data fit the model” approach to create a motor scale with four out of nine items for six to twelve year-old children (*n* = 9439). Just recently, Utesch et al. (2015) validated six of the items of the Deutscher Motorik Test 6-18 [German Motor Test 6-18] (Bös et al., 2009) using the mixed Rasch model for nine- to ten-year old children as being one-dimensional.

Summary and study objectives

Currently, the latent trait(s) underlying motor assessment in early childhood is not fully understood. The evidence provided by the CTT approach is inconclusive in validating or rejecting the *General Motor Ability Hypothesis*. CTT neither offers a clear view of this concept nor does it support the current use of composite scores (or linear transformations thereof) in motor assessment. IRT models provide an alternative approach to gain new insights into the latent trait underlying motor assessment on item level. The abovementioned IRT studies support the *General Motor Ability Hypothesis* indicating a one-dimensional structure in early childhood, but only within small item sets.

Using the IRT approach, the aim of this study was to examine the dimensional structure of motor competence in early childhood using a wide variety of motor skills within the large item bank of a motor assessment battery. Based on previous studies, it is expected that the construct of motor competence in this age group will have a one-dimensional structure. Furthermore, the present study demonstrates how the current use of composite scores in motor assessment can be validated.

Methods

Participants

This study is part of a large-scale evaluation of the motor competence of children in Flanders, Belgium ("Multimove for Kids" project; multimove.be). The total sample for this study consisted of 1467 children, aged 3 to 6 years old (see Table 1). Children were recruited from 54 settings (sports clubs, local councils, schools and day care centers) across Flanders, Belgium.

Materials

The MOT 4-6 (Zimmer & Volkamer, 1987) consists of one practice item and 17 test items. The test is easy to use and typically takes 15-20 minutes to administer. According to the authors, different motor domains are represented in the MOT 4-6 test to assess the motor competence of children. Table 2 shows the motor domains and the items representing these domains. In the test manual the original authors describe in detail how to convert each item raw score into a point score ranging from zero (skill not mastered) to two (skill mastered). These point scores are used in practice to interpret test results of children and therefore have to be investigated in terms of empirical validity and order. In line with the test manual, all point scores were summed to attain a sum score.

The MOT 4-6 was constructed using the CTT approach. In the test manual (Zimmer & Volkamer, 1987), the original authors report high test-retest reliability and inter-rater reliability (r = .85 and r = .88 respectively) and a good internal consistency (Cronbach's alpha coefficient = .81). Content and construct validity have been determined through movement skill literature; neither a factor analysis nor cluster analysis demonstrated a valid factor structure (Cools et al., 2009).

Procedure

This study was approved by the ethics committee of the XX University Hospital. For each participant, a written informed consent was obtained from a parent or guardian.

Assessments were conducted by a group of trained assessors in an indoor facility during the period of September-November 2012. The MOT 4-6 was administered to assess the motor competence in young children (Zimmer & Volkamer, 1987). All children completed the tests barefoot in one session, in accordance with the manual guidelines.

Data analysis

Data were analyzed using SPSS 22 for Windows (SPSS Inc., Chicago, IL, USA) and Winmira 2001 (von Davier, 2001). Descriptive statistics were computed for all item scores. To examine the construct of motor competence in the MOT 4-6 data, IRT models were calculated. First, the *Partial Credit Model* (PCM; Masters, 1982) was selected to analyze homogeneity and order within the assumed one-dimensional construct. The PCM is a generalization of the (dichotomous) Rasch model (Rasch, 1960), but for ordinal data (Rost, 2004). It is a test of dimensionality relying on the assumption of equal specificity and sensitivity of indicators (Rost, 2004). Probabilistic threshold parameters between each category as well as item locations are calculated (Strauss et al, 2007, 2012). Model conform data implies invariance of parameters and provides sample distribution free and indicator distribution free results. Furthermore, person ability and item difficulty are measured on the same (logit) scale (Rost, 2004). Second, the mixed Rasch model (MRM) was used, which combines the PCM and Latent Class Analysis (LCA; Rost, 2004) and adds a qualitative aspect to the PCM. This means that possible item difficulty patterns between groups (e. g., boys and girls), are explored and person homogeneity is tested. The latter is shown in case the one-class solution fits best indicating that all persons used the same ability to complete the assessment. In case more-class solutions only differentiate between overall skill level a one-dimensional result indicates that a statistically verified composite score can be constructed with all fitting items.

Applying the PCM the bootstrapping procedure with the recommended 100 bootstrapping samples was executed (Rost, 2004; von Davier, 1997, 2001). The model fit was evaluated in three steps. At first, the global model fit is analyzed checking the statistical values Cressie-Read (CR) and Pearson-χ^2 (P- χ^2). von Davier (1997) recommends checking both values and defines a good model fit at the significance level of 5 % ($p > .05$). Second, local model violations are analyzed. Unordered threshold parameters in form of overlapping item characteristic curves show violations of the order within the ordinal scale. Items showing unordered threshold parameters within the continuum of the latent variable have to be excluded from further analysis. If no valid model was found, the third step would be to analyze local violations in form of item fit statistics. Winmira 2001 (von Davier, 2001) provides the Q-index of each item, which represents likelihood based estimations of the sensitivity. Overfitting items (closer to 0) show significantly better response patterns than the model expects while underfitting (closer to .5) items significantly deviate from it. Using the PCM, reliability is analyzed by Andrich's reliability coefficient (R_A; Andrich, 1988), which is a mean value of the reliability of each step of person test scores.

Conducting the MRM, the fit of two-class solutions is explored in terms of testing the global and local model fits congruently to the PCM. The two-class model is rejected if global or local model fits are violated. In case that both the one- and two-dimensional models fit the data, two types of information criteria are used to select the most appropriate model: Bayes Information Criterion (BIC; Schwarz, 1978) and Consistent Akaikes Information Criterion (CAIC; Bozdogan, 1987). As these criteria indicate the minimum of the global fit function, smaller BIC and CAIC demonstrate a relative better model fit.

Results

Table 3 shows the descriptive statistics of all 17 items; means range from 0.25 to 1.38.

At first, the global model fit regarding the assumed one-dimensional structure was analyzed using PCM. First-step analysis showed a global model fit for all items of the MOT 4-6 (CR = .032, p_{CR} = .43; P- χ^2 = -.356, $p_{P\text{-}\chi^2}$ = .55) and revealed four items with unordered threshold parameters: grasping a tissue with a toe, catching a tennis ring, rolling sideways over the floor and twist jump in/out of a hoop (see Table 4 and Figure 1). These items were excluded from the model because they violated the order within the continuum of the latent variable. The follow-up modeling process using the PCM revealed a global model fit with ordered threshold parameters for the MOT 4-6 (CR = .1.964, p_{CR} = .06; P- χ^2 = -.227, $p_{P\text{-}\chi^2}$ = .24, R_A = .79) and demonstrated good reliability, after the four items with unordered threshold parameters and an additional item (jumping on one leg into a hoop) were excluded. Item locations and threshold parameters of the fitting model with twelve items are presented in Table 5 and Figure 2. The item set for the remaining twelve items conform to the requirements of the PCM and fundamental measurement is attained. Thus, the accumulation of these items to one composite score represents one latent variable. The resulting distribution using the composite score for the remaining twelve items is shown in Figure 3.

To check for better fitting models, the 12 items were analyzed using the MRM. The two-class solution showed acceptable global model fit (CR = 23, p_{CR} = .28; $P\text{-}\chi^2$ = -.53, $p_{P\text{-}\chi^2}$ = .7, $R_{A\_class\ 1}$ = .63; $\mathrm{R}_{\mathrm{A\_class\ 2}}$ = .45). The MRM showed unordered threshold between the classes (class 1: throwing a ball at a target disk, transferring matches; class 2: jumping jacks, balancing forward on a line) rejecting the model (see Table 6). Poor reliability values were reported for both classes. Person homogeneity was shown because the only global fit with ordered threshold parameters was shown in the one class solution, which is identical to the PCM. Person homogeneity means that no qualitative different patterns of item difficulty were found between classes, e.g. boys and girls.

Discussion

Assessment tools are generally as valid as the proposed theoretical construct, which is closely connected to the theoretical assumptions. In the field of motor development, assessments often rely on the *General Motor Ability Hypothesis* as motor competence is implicitly measured as a single latent trait when test scores of a wide range of motor skills are summed up to a composite score (Burton & Rodgerson, 2001). However, these composite scores are often not statistically verified. Prior research has not provided a clear understanding of the latent trait(s) underlying motor assessment, partially due to methodological limitations of the generally adopted CTT approach. Adopting the alternative IRT approach, this study investigated the dimensionality of the construct of motor competence in early childhood using the large item set of a popular motor assessment. This also provided the option to validate the composite score of this assessment tool.

Present study provided evidence of a one-dimensional construct of motor competence in early childhood using a large number of items. These findings are in agreement with a previous IRT study on preschool children (Hands & Larkin, 2001); the authors found a one-dimensional latent structure for five- and six-year-old children using a set of 24 items. In contrast to the present findings, an IRT study of Büsch et al. (2009), that evaluated three locomotor and three object control skills in children aged 9 and 10, revealed qualitatively different item difficulties for the AST (Bös, 2000). One group showed higher item difficulties in object control skills and the other in locomotor skills. However, in our study no differentiation between object control and locomotor skills was found for the preschool age group. One possible explanation is that the latent trait underlying motor assessment might divide in multiple motor domains due to an interaction of maturation and environmental experiences as found in other studies (e.g., Schulz, Henderson, Sugden, & Barnett, 2011).

Compared to the findings of Hands and Larkin (2001), who also found a one-dimensional construct of motor competence, the MRM did not reveal different item properties between groups. Hands and Larkin (2001) analyzed boys and girls separately and found descriptive differences between these groups. However, the MRM conducted in this study did not reveal differences between groups or classes, because only the one-dimensional model fitted the data. Instead, this study revealed a one-dimensional structure for all 17 items of the MOT 4-6 in early childhood. Furthermore, homogeneity was shown for 12 items, which can be used for both boys and girls.

Zimmer and Volkamer (1987) constructed the MOT 4-6 and selected 17 items to cover multiple motor domains and a wide range of motor skills (see Table 2). In addition, the authors built a composite score with all items based on the implicit assumption that a single latent trait underlies the MOT 4-6 (Burton & Rodgerson, 2001). Our results support that implicit assumption of a single latent trait from a measurement-theoretical perspective; 12 items of the

MOT 4-6 met the Rasch model requirements and therefore provided a valid measurement of motor competence through a composite score. Five items violated the assumption of order in the ordinal scale indicating that the categorization of one or two points is not related to the person's skill level but is random. Upon inspection of these items violating the model assumption, we found no similar content between these items; the items, grasping a tissue with a toe, catching a tennis ring, rolling sideways over the floor, twist jump in/out of a hoop and jumping on one leg into a hoop, represent different motor dimensions. However, the scoring system was equal for of all these items: zero successful trials giving zero points, one successful trial giving one point and two or more successful trials giving two points. Thus, the results indicate that this scoring system seems inadequate under certain circumstances. With regard to this finding, categorization systems should be taken into account in the construction and analysis of motor assessments.

The IRT approach provides a solution for the limitations of generally used CTT methods and contributes to a better understanding of the latent trait(s) underlying motor assessment. In view of limited IRT studies in the field of motor assessment, present study examined the motor competence in early childhood using the IRT approach and provides evidence for the *General Motor Ability Hypothesis* in that age group, which states that numerous motor abilities are highly related within a person and form a single general motor ability (i.e., motor competence). The main strength of our study is the use of a large set of 17 items, which covers a wide range of motor skills, and a large sample of 1476 children aged three to six years. In addition, this study investigated the items of an existing test battery (i.e., MOT 4-6) which provides information on the validation of the assessment. However, this study is not without limitations. One limitation to this study is the small amount of object control skills in the MOT 4-6. Since other test batteries include more object control items, this might restrict the generalizability of present findings. Future IRT research should evaluate motor assessments that include a wide item-set with a larger proportion of object control skills. Another limitation relates to the product-oriented approach of the MOT 4-6 where motor skills are scored based on the outcome of the performance (such as speed and frequency). However, qualitative factors such as arm-leg coordination are also important for motor performance. Future IRT studies should include item sets with process- and product-oriented approaches to better encompass motor competence. Finally, current IRT research – including the present study – has analyzed the construct of motor competence using a cross-sectional design. However, there is a need for longitudinal studies to investigate how the construct of motor competence might change across childhood.

Conclusion

In view of the importance of motor development in children's overall health, it is imperative to have valid measurements in order to make sound interpretations and decisions (Mahar & Rowe, 2008). This study gives insights into the latent trait(s) underlying motor assessment in early childhood. Rasch measurement provided support for the theoretical definition of motor competence (or general motor ability) and evidence for the *General Motor Ability Hypothesis*, which could expand to older age groups. Whereas previous research investigating the taxonomy underlying motor assessment used the CTT approach and arbitrary cut-off values (correlations) based on random human judgment, the IRT approach provides models with goodness of fit statistics to address that limitation. Furthermore, this study shows the capacity of IRT models in the context of motor development research. It provides an alternative approach to test theories, to validate test instruments and detect non-fitting items. IRT models are specifically valuable to evaluate test instruments that use composite scores to describe motor behavior and should be included in the evaluation of the methodological standard for those test instruments.

The present study does not imply that only IRT models should be used in motor test and construct validation. Rather, a combination of appropriate psychometric approaches can further enrich scientific discourse and provide a deeper understanding of the underlying latent trait(s) of motor assessment.

Acknowledgments

This study was financed by the Flemish Government. The authors would like to acknowledge Brecht Ameys, Wout Chielens, Sara Dauwe, Hanne De Ruyck, Eline Sarrazyn and Anouk Vanblaere for their help in the data collection. The authors also wish to thank the participants and the staff of all participating organizations.

Appendix

Motor Proficiency Test for 4- to 6-Year-Old Children – Short Form (MOT 4-6 SF)

The twelve model conform items can be used as a short form of the MOT 4-6 and can be called the Motor Proficiency Test for 4- to 6-Year-Old Children – Short Form (*Motoriktest für vier- bis sechsjährige Kinder* [MOT 4-6 SF]). Compared to the MOT 4-6, the MOT 4-6 SF is a more economic test from a practical perspective. It consists of following items using the original three point (0 to 2) scale system: (1) balancing forward on a line, (2) placing dots on a sheet, (3) jumping sideways over a rope, (4) catching a stick, (5) moving balls from box to box, (6) balancing backwards on a line, (7) throwing a ball at a target disk, (8) transferring matches, (9) passing through a hoop, (10) jumping jacks, (11) jumping over a cord, and (12) standing and sitting while holding a ball on the head.

All items of the MOT 4-6 SF can be accumulated to a single valid composite score (or a linear transformation thereof) which represents children's motor competence. As shown in this study the age range of the MOT 4-6 SF can be extended to three-year-old children, since preschools or kindergartens in Europe are often offered to three to six year-old children.

Table 1

Age and sex distribution of the study sample

Age	Sex	*N*	%
3 years	Girls	137	46.8
	Boys	156	53.2
	Total	293	100
4 years	Girls	180	40.8
	Boys	261	59.2
	Total	441	100
5 years	Girls	191	47.9
	Boys	208	52.1
	Total	399	100
6 years	Girls	164	49.1
	Boys	170	50.9
	Total	334	100
Total	Girls	672	45.8
	Boys	795	54.2
	Total	1467	100

Table 2

Description of the motor domains and corresponding items used for the Motor Proficiency Test for 4- to 6-year-old Children (Motoriktest für Vier- bis Sechsjährige Kinder [MOT 4-6]).

Domains	Items
Fine motor skill	Placing dots on a sheet*
	Grasping a tissue with toes
	Transferring matches*
Agility and coordination	Moving balls from box to box*
	Passing through a hoop
	Jumping jacks
	Rolling sideways over the floor
	Twist jump in/out of a hoop*
Balance	Balancing forward on a line
	Balancing backwards on a line
	Jumping on one leg into a hoop
	Standing and sitting while holding a ball on the head
	Twist jump in/out of a hoop*
Reaction time	Catching a stick
	Catching a tennis ring
Power	Jumping over a cord
	Twist jump in/out of a hoop
Speed	Placing dots on a sheet*
	Jumping sideways over a rope
	Moving balls from box to box*
Accuracy	Throwing a ball at a target disk
	Transferring matches*

* Items that represent two motor domains.

Table 3
Descriptive statistics of the MOT 4-6 test items and the score distribution for each test item.

	Test item*	*M*	*SD*	0		1		2	
				Count	%	Count	%	Count	%
2	Balancing forward on a line	1.00	.81	479	32.7	514	35.0	474	32.3
3	Placing dots on a sheet	.92	.82	558	38.0	467	31.8	442	30.1
4	Grasping a tissue with toes	1.25	.84	380	25.9	338	23.0	749	51.1
5	Jumping sideways over a rope	.91	.80	545	37.2	516	35.2	406	27.7
6	Catching a stick	.82	.47	316	21.5	1100	75.0	51	3.5
7	Moving balls from box to box	.75	.71	598	40.8	634	43.2	235	16.0
8	Balancing backwards on a line	.25	.53	1169	79.7	231	15.7	67	4.6
9	Throwing a ball at a target disk	.52	.72	906	61.8	365	24.9	196	13.4
10	Transferring matches	.68	.79	766	52.2	406	27.7	295	20.1
11	Passing through a hoop	1.38	.74	233	15.9	442	30.1	792	54.0
12	Jumping on one leg into a hoop	.84	.82	635	43.3	433	29.5	399	27.2
13	Catching a tennis ring	.87	.89	685	46.7	282	19.2	500	34.1
14	Jumping jacks	.77	.81	687	46.8	429	29.2	351	23.9
15	Jumping over a cord	.94	.82	548	37.4	465	31.7	454	30.9
16	Rolling sideways over the floor	1.40	.78	273	18.6	332	22.6	862	58.8
17	Standing and sitting while holding a ball on the head	.95	.78	484	33.0	576	39.3	407	27.7
18	Twist jump in/out of a hoop	1.04	.87	521	35.5	361	24.6	585	39.9

Item 1 (jumping forward into a hoop) is a practice item and therefore not rated.

Table 4

Item location and threshold parameters for all test items of the MOT 4-6.

	Test item*	Location	Threshold parameters	
			1	2
2	Balancing forward on a line	-.27	-.69	.14
3	Placing dots on a sheet	-.11	-.38	.16
4	Grasping a tissue with toes **	-.80	-.68	-.92
5	Jumping sideways over a rope	-.07	-.50	.36
6	Catching a stick	.74	-1.80	3.27
7	Moving balls from box to box	.35	-.51	1.22
8	Balancing backwards on a line	1.65	1.55	1.75
9	Throwing a ball at a target disk	.81	.65	.98
10	Transferring matches	.42	.26	.58
11	Passing through a hoop	-1.17	-1.55	-.79
12	Jumping on one leg into a hoop **	.07	-.11	.24
13	Catching a tennis ring **	-.03	.38	-.44
14	Jumping jacks	.22	.03	.40
15	Jumping over a cord	-.14	-.40	.12
16	Rolling sideways over the floor **	-1.15	-1.11	-1.18
17	Standing and sitting while holding a ball on the head	-.16	-.76	.44
18	Twist jump in/out of a hoop **	-.36	-.27	-.46

* Item 1 (Jumping forward into a hoop) is a practice item and therefore not included.

** Items with unordered threshold parameters.

Table 5

Item location and threshold parameters for the 12 test items of the MOT 4-6 meeting the model requirements

	Test item	Location	Q-Index	Threshold parameters	
				1	2
2	Balancing forward on a line	-.46	.20	-.87	-.04
3	Placing dots on a sheet	-.29	.09	-.56	-.02
5	Jumping sideways over a rope	-.25	.22	-.67	-.17
6	Catching a stick	.54	.07	-2.00	3.07
7	Moving balls from box to box	.16	.22	-.69	1.02
8	Balancing backwards on a line	1.44	.11	1.36	1.52
9	Throwing a ball at a target disk	.62	.18	.47	.78
10	Transferring matches	.23	.19	.08	.38
11	Passing through a hoop	-1.36	.32	-1.75	-.97
14	Jumping jacks	.03	.16	-.15	.21
15	Jumping over a cord	-.32	.09	-.58	-.07
17	Standing and sitting while holding a ball on the head	-.34	.18	-.94	.26

Table 6

Item location and threshold parameters for the 12 test items of the MOT 4-6 2-class solution.

	Test item	Location	Q-Index	Threshold parameters 1	Threshold parameters 2
	Class 1				
2	Balancing forward on a line	-.61	.24	-1.08	-.16
3	Placing dots on a sheet	.99	.19	.46	1.52
5	Jumping sideways over a rope	-1.03	.17	-1.21	-.85
6	Catching a stick	.98	.15	-.65	2.6
7	Moving balls from box to box	.06	.29	-2	2.13
8	Balancing backwards on a line	1.41	.19	-.62	3.45
9	Throwing a ball at a target disk*	.84	.15	1.21	.47
10	Transferring matches*	.18	.2	.28	.08
11	Passing through a hoop	-2.08	.21	-2.34	-1.83
14	Jumping jacks	-.14	.18	-.55	.28
15	Jumping over a cord	.03	.15	-.55	.61
17	Standing and sitting while holding a ball on the head	-.63	.21	-1.24	-.04
	Class 2				
2	Balancing forward on a line*	.02	.27	.14	-.1
3	Placing dots on a sheet	-.17	.28	-.79	.45
5	Jumping sideways over a rope	.04	.27	-.34	.43
6	Catching a stick	-1.19	.23	-2.3	-.09
7	Moving balls from box to box	.4	.38	-2.73	3.54
8	Balancing backwards on a line	.25	.3	-1.41	.91
9	Throwing a ball at a target disk	1.62	.28	1.43	1.81

10	Transferring matches	.79	.28	.58	1.01
11	Passing through a hoop	-.34	.29	-.46	-.21
14	Jumping jacks*	.17	.27	.22	.12
15	Jumping over a cord	-.96	.29	-1.77	-.17
17	Standing and sitting while holding a ball on the head	-.14	.26	-.64	.37

* Items with unordered threshold parameters

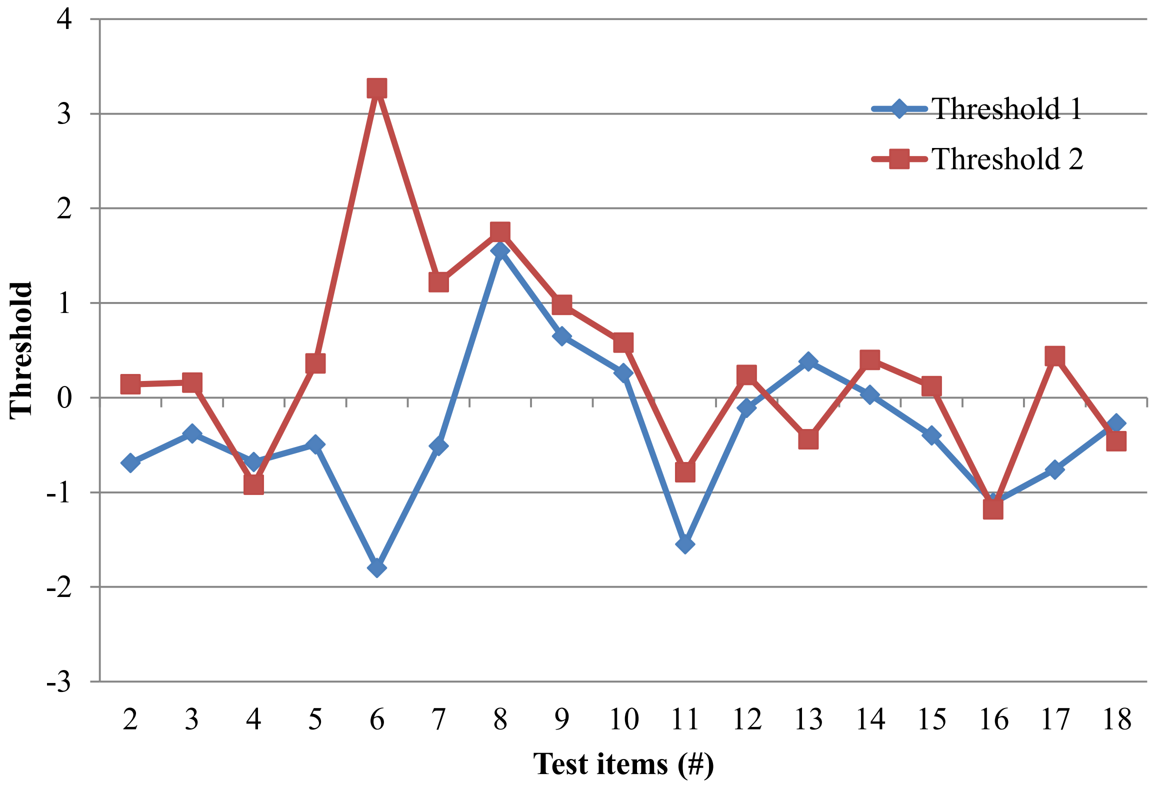

Fig 1. Threshold parameter profile of the *Partial Credit Model* for all MOT 4-6 items.

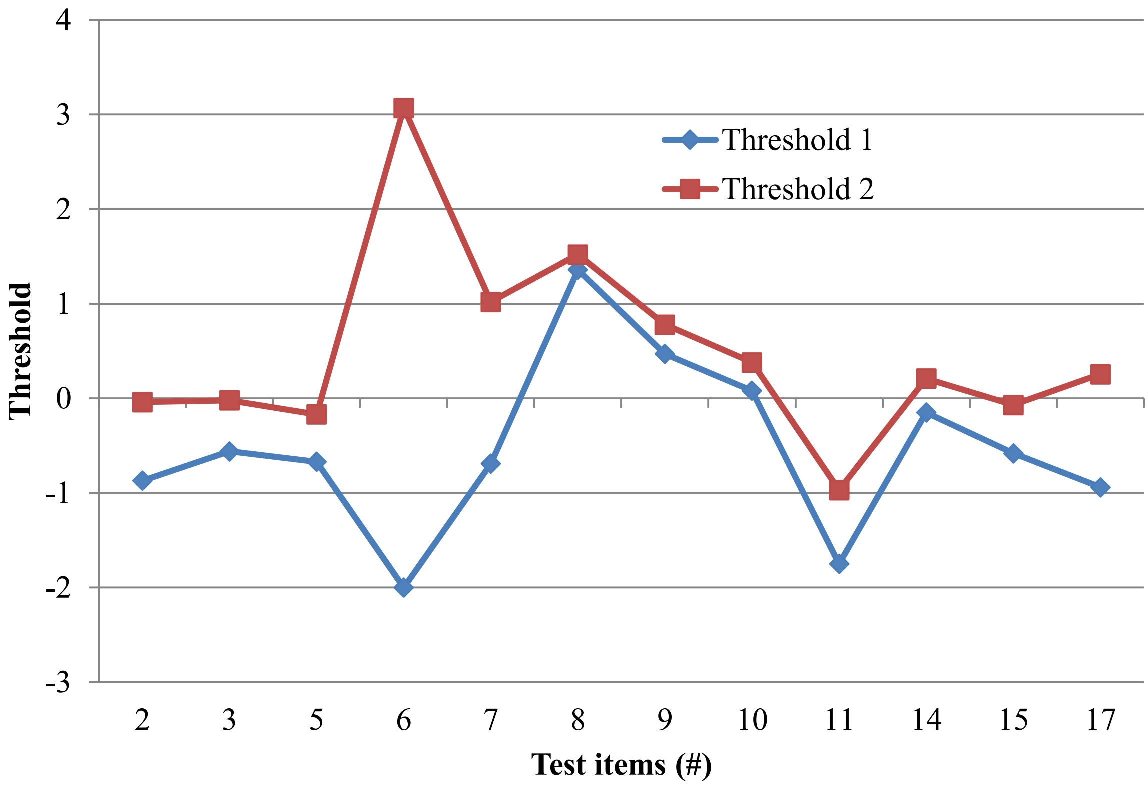

Fig 2. Threshold parameter profile of the *Partial Credit Model* for the 12 MOT 4-6 items meeting the model requirements.

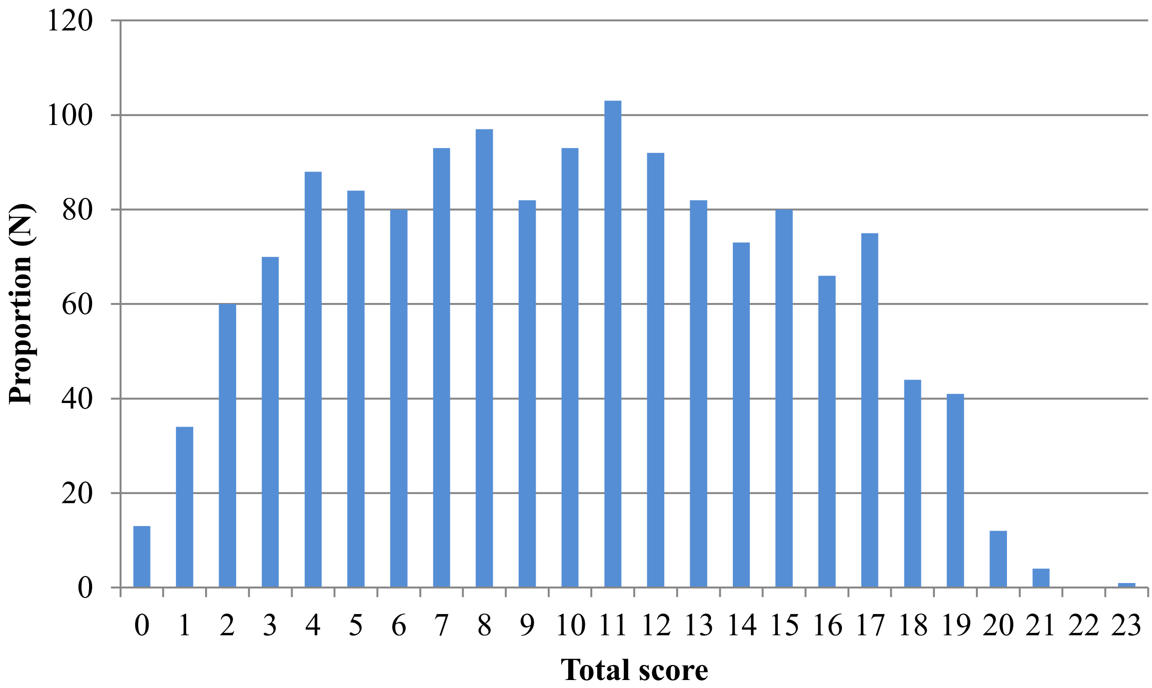

Fig 3. Distribution of the corrected total score* in the study sample.

** Based on the Rasch analysis, following items were not included in the total score: 4 (grasping a tissue with toes), 12 (jumping on 1 leg into a hoop), 13 (catching a tennis ring), 16 (rolling sideways over the floor) and 18 (twist jump in/out of a hoop).*

References

AAHPERD. (1976). *AAHPERD Youth Fitness Test Manual* (19th ed.). Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance.

Andrich, D. (1988). *Rasch models for measurement*. London: Sage.

Bös, K. (2000). AST 6-11: Allgemeiner Sportmotorischer Test für Kinder von 6-11 Jahren. *Haltung Und Bewegung*, *20*(2), 5–9.

Bös, K. (2001). *Handbuch motorische tests* (2nd ed.). Göttingen: Verlag für Psychologie.

Bös, K., Schlenker, L., Büsch, D., Lämmle, L., Müller, H., Oberger, J., … Tittlbach, S. (2009). *Deutscher Motorik Test 6-18*. Hamburg: Czwalina.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. doi:10.1007/BF02294361

Brace, D. K. (1927). *Measuring motor ability: A scale of motor ability tests*. New York: A. S. Barnes.

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

Burton, A. W., & Miller, D. E. (1998). *Movement Skill Assessment*. Champaign, IL: Human Kinetics.

Burton, A. W., & Rodgerson, R. W. (2001). New Perspectives on the Assessment of Movement Skills and Motor Abilities. *Adapted Physical Activity Quarterly*, *18*, 347–365.

Büsch, D., Strauß, B., Seidel, I., Pabst, J., Tietjens, M., Müller, L., … Wirszing, D. (2009). Die Konstruktvalidität des Allgemeinen Sportmotorischen Tests für Kinder. *Sportwissenschaft*, *39*, 95–103. doi:10.1007/s12662-009-0030-6

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159.

Cools, W., De Martelaer, K., Samaey, C., & Andries, C. (2009). Movement Skill Assessment of Typically Developing Preschool Children: A Review of Seven Movement Skill Assessment Tools. *Journal of Sports Science & Medicine*, *8*(2), 154–168.

Cools, W., De Martelaer, K., Vandaele, B., Samaey, C., & Andries, C. (2010). Assessment of Movement Skill Performance in Preschool Children: Convergent Validity Between MOT 4-6 and M-ABC. *Journal of Sports Science & Medicine*, *9*(4), 597–604.

Croce, R. V, Horvat, M., & McCarthy, E. (2001). Reliability and concurrent validity of the movement assessment battery for children. *Perceptual and Motor Skills*, *93*(1), 275–80. doi:10.2466/pms.2001.93.1.275

Fleishman, E. A. (1964). *The structure and measurement of physical fitness*. Englewood Cliffs, NJ: Prentice-Hall.

Fransen, J., D'Hondt, E., Bourgois, J., Vaeyens, R., Philippaerts, R. M., & Lenoir, M. (2014). Motor competence assessment in children: Convergent and discriminant validity between the BOT-2 Short Form and KTK testing batteries. *Research in Developmental Disabilities*, *35*(6), 1375–1383. doi:10.1016/j.ridd.2014.03.011

Gabbard, C. P. (2008). *Lifelong Motor Development* (5th ed.). San Francisco, CA: Pearson Benjamin Cummings.

Gallahue, D. L., & Cleland-Donnelly, F. (2007). *Developmental physical education for all children* (5th ed.). Champaign, IL: Human Kinetics.

Haga, M. (2008). The relationship between physical fitness and motor competence in children. *Child: Care, Health and Development*, *34*(3), 329–34. doi:10.1111/j.1365-2214.2008.00814.x

Hands, B., & Larkin, D. (2001). Using the Rasch Measurement MOdel to Investigate the Construct of Motor Abilities in Young Children. *Journal of Applied Measurement*, *2*(2), 101–120.

Haywood, K. M., & Getchell, N. (2009). *Life span motor development* (5th ed.). Champaign, IL: Human Kinetics.

Hill, E. L. (2010). The importance of motor skill in general development. *Developmental Medicine and Child Neurology*, *52*(10), 888. doi:10.1111/j.1469-8749.2010.03700.x

Kaplan, B. J., Wilson, B., Dewey, D., & Crawford, S. G. (1998). DCD may not be a discrete disorder. *Human Movement Science*, *17*(4-5), 471–490. doi:10.1016/S0167-9457(98)00010-4

Kiphard, E. J., & Schilling, F. (1974). *Körperkoordinationstest für Kinder*. Weinheim: Beltz Test GmbH.

Kiphard, E. J., & Schilling, F. (2007). *Körperkoordinationstest für Kinder*. Weinheim: Beltz Test GmbH.

Linacre, J. M. (2000). New approaches to determining reliability and validity. *Research Quarterly for Exercise and Sport*, *71*(2), 129–136.

Lubans, D. R., Morgan, P. J., Cliff, D. P., Barnett, L. M., & Okely, A. D. (2010). Fundamental movement skills in children and adolescents: Review of associated health benefits. *Sports Medicine*, *40*(12), 1019–35. doi:10.2165/11536850-000000000-00000

Mahar, M. T., & Rowe, D. A. (2008). Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise Science, 12*, 126–145. DOI: 10.1080/10913670802216106.

Marteniuk, R. G. (1974). Individual differences in motor performance and learning. *Exercise & Sport Sciences Reviews*, *2*(1), 103–130.

Masters, G. N. (2005). Objective Measurement. In S. Alagumalai, D. Curtis, & N. Hungi (Eds.), *Applied Rasch Measurement: A book of exemplars* (pp. 15–25). The Netherlands: Springer.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danisch Institute for Educational Research.

Robinson, L. E., Stodden, D. F., Barnett, L. M., Lopes, V. P., Logan, S. W., Rodrigues, L. P., & D'Hondt, E. (2015). Motor Competence and its Effect on Positive Developmental Trajectories of Health. *Sports Medicine*, (September). doi:10.1007/s40279-015-0351-6

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2nd ed.). Bern: Huber.

Schulz, J., Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2011). Structural validity of the Movement ABC-2 test: factor structure comparisons across three age groups. *Research in Developmental Disabilities*, *32*(4), 1361–9. doi:10.1016/j.ridd.2011.01.032

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Smits-Engelsman, B. C. M., Henderson, S. E., & Michels, C. G. J. (1998). The assessment of children with Developmental Coordination Disorders in the Netherlands: The relationship between the Movement Assessment Battery for Children and the Korperkoordinations Test für Kinder. *Human Movement Science*, *17*, 699–709.

Spray, J. A. (1987). Recent Developments in Measurement and Possible Applications to the Measurement of Psychomotor Behavior. *Research Quarterly for Exercise and Sport*, *58*(3), 203–209. doi:10.1080/02701367.1987.10605449

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity: An Emergent Relationship. *Quest*, *60*(2), 290–306. doi:10.1080/00336297.2008.10483582

Strauss, B. (1999). Latent trait and latent class models. *International Journal of Sport Psychology*, *30*(1), 17–40.

Strauss, B., Büsch, D., & Tenenbaum, G. (2007). New developments in measurement and testing. In G. Tenenbaum & R. Eklund (Eds.), *Handbook of Sport Psychology* (3rd ed., pp. 737–756). Boston, MA: Wiley.

Tenenbaum, G., Strauss, B., & Büsch, D. (2007). Applications of Generalized Rasch Models in the Sport, Exercise, and the Motor Domains. *Statistics for Social and Behavioral Sciences*, (1). Retrieved from http://www.springerlink.com/content/x8680k3782k08052/

Ulrich, D. A. (1985). *Test of Gross Motor Development*. Austin, TX: PRO-ED.

Utesch, T., Strauß, B., Tietjens, M., Büsch, D., Ghanbari, M.-C., & Seidel, I. (2015). Die Konstruktvalidität des Deutschen Motorik-Tests 6-18 für 9- bis 10-Jährige. *Zeitschrift Für Sportpsychologie*, *22*(2), 77–90. doi:10.1026/1612-5010/a000143

von Davier, M. (1997). *Methoden zur Prüfung probabilistischer Testmodelle [Methods to evaluate probabilistic test models]*. Kiel: IPN.

von Davier, M. (2001). Winmira 2001. St. Paul, MN: Assessment Systems Corporation.

Wright, B. D., & Mok, M. (2000). Rasch models overview. *Journal of Applied Measurement*, *1*(1), 83–106.

Yan, Z., & Bond, T. G. (2011). Developing a Rasch Measurement Physical Fitness Scale for Hong Kong Primary School-Aged Students. *Measurement in Physical Education and Exercise Science*, *15*(3), 182–203. doi:10.1080/1091367X.2011.590772

Zhu, W., & Cole, E. L. (1996). Many-Faceted Rasch Calibration of a Gross Motor Instrument. *Research Quarterly for Exercise and Sport*, *67*(1), 24–34.

Zhu, W., Rink, J., Placek, J. H., Graber, K. C., Fox, C., Fisette, J. L., … Raynes, D. (2011). PE Metrics: Background, Testing Theory, and Methods. *Measurement in Physical Education and Exercise Science*, *15*(2), 87–99. doi:10.1080/1091367X.2011.568363

Zimmer, R., & Volkamer, M. (1987). *Motoriktest für vier- bis sechsjährige Kinder*. Weinheim: Beltz Test.