

**Zwischenbericht
für das Projekt „Berlin hat Talent“
(Datenbasis 2012/2013)**

Jochen Zinner & Dirk Büsch



Ein gemeinsames Projekt von:

Gliederung

Zusammenfassung

- Teil 1** Einführung und Exploratorische Datenanalyse zur DMT-Untersuchung von 2012 und 2013 in Berlin ($N = 4.986$)
(Zinner & Büsch)
- Teil 2** Zur Problematik der Bildung von Gesamturteilen innerhalb des DMT
(Büsch & Zinner)
- Teil 3** Modellversuch zur Auswertung des DMT mit Hilfe von mehrkriteriellen und FUZZY-Analysen
(Zinner, Büsch & Ester)

Zusammenfassung

Mit Bezug auf die Empfehlung der Sportministerkonferenz (SMK) hat das Land Berlin 2010/11 begonnen, den Deutschen Motorik-Test 6-18 (DMT 6-18, Kurzform DMT) in das Berliner Talentfördermodell zu integrieren und in der dritten Klasse an jährlich – zunächst - repräsentativ ausgewählten Grundschulen durchzuführen. Mit dem DMT sollen sowohl motorisch positiv auffällige als auch motorisch negativ auffällige Schülerinnen und Schüler identifiziert und anschließend gefördert werden. Während die Durchführung des DMT durch die Firma Speed4/Maximum Sport GmbH erfolgt, obliegt die wissenschaftliche Begleitung sowie Evaluation des Projekts der Fakultät Sport der H:G. Es ist beabsichtigt, die Vorgehensweise in Schritten auf alle Schulen in Berlin zu erweitern.

Nachdem 2011 der DMT erstmalig durchgeführt und an die Rahmenbedingungen der Berliner Schulen angepasst wurde, konnte nun begonnen werden, die objektiv und zuverlässig erhobenen umfangreichen Daten aus den Jahren 2012 und 2013 systematisch auszuwerten.

Zunächst wurden dazu die Berliner Daten mit den Normtabellen des DMT verglichen. Die Ergebnisse (Teil 1 und 2 dieses Zwischenberichts) zeigen, dass

- (1) die Leistungen der Berliner Schülerinnen und Schüler von den Normwerten in mehreren Testaufgaben deutlich abweichen,
- (2) ausschließlich über die „Konditions-Items“ ein valider Summenwert zur Einschätzung der motorischen Leistungsfähigkeit berechnet werden darf und
- (3) für die valide Bestimmung eines Koordinations- und Beweglichkeitsfaktor weitere konstruktkonforme Items zu integrieren sind, um eine fundierte Aussage über die umfassende motorische Leistungsfähigkeit respektive motorische Fitness (Kondition, Koordination) formulieren zu können.

Des Weiteren konnte mit einem ausgewählten Datensatz aus dem Jahr 2013 (Teil 3 dieses Zwischenberichts) gezeigt werden, dass mit mehrkriteriellen Fuzzy-Modellen, d. h. Modellen, die neben den validen Konditions-Items weitere Personenmerkmale, z. B. Körperhöhe, Körpergewicht, sowie theoretisch-inhaltliche Überlegungen über die Variablenbeziehungen a priori integrieren, sowohl für jede Person ein Summenwert für die Konditions-Items ermittelt und darauf aufbauend auch Rangreihen gebildet werden können, die eine differenziertere Perspektive zur Beurteilung der motorischen Leistungsfähigkeit gestatten.

*Zwischenbericht für das Projekt „Berlin hat Talent“
(Datenbasis 2012/2013)*

Teil 1:

Exploratorische Datenanalyse zur DMT-Untersuchung von 2012 und
2013 in Berlin (N= 4.986)

(Jochen Zinner & Dirk Büsch)



Ein gemeinsames Projekt von:

Gliederung

- 1.0 Einführung
- 1.1 Perspektiven des Projekts
- 1.2 Exploratorische Datenanalyse
- 1.3 Leistungsunterschiede der Berliner Schüler(innen) auf der Basis der Einordnung in Leistungsklassen
- 1.4 Zusammengefasste Ergebnisse der Datenanalyse
- 1.5 Bestimmung von Leistungsklassen nach unterschiedlichen Vorgehensweisen:
 - 1.5.1 Kategoriengrenzwerte unterschiedlicher Bestimmungsverfahren
 - 1.5.2 Konsequenzen aus den unterschiedlichen Bestimmungsverfahren
 - 1.5.3 Vergleich der unterschiedlich bestimmten Leistungsklassen bzgl. jedes einzelnen Parameters des DMT für die Varianten A und D
 - 1.5.4 Konsequenzen aus dem Vergleich der unterschiedlich bestimmten Leistungsklassen bzgl. jedes einzelnen Parameters für die Varianten A und D
 - 1.5.5 Vergleich der summierten empirischen Leistungsklassen für einen Konditionsfaktor (6 Items, Variante D) mit dem Summenwert der Norm-Quintile (8 Items, Variante A) für die 8- bis 10-Jährigen

1.0 Einführung

Die überwiegende Mehrzahl der Olympischen Fachverbände beklagt eine Abnahme der Zahl der Nachwuchstalente im Bereich des „optimalen“ Einstiegsalters von 5-10 Jahren. Daher besteht ein umfassender Klärungsbedarf für eine zielführende und systematische Talentidentifikation und -förderung bis in den Spitzenbereich der jeweiligen Sportart, wenngleich allgemeine Rahmenkonzepte und Zielkriterien für den langfristigen Leistungsaufbau (LLA) national und international existieren (siehe zusammenfassend Hoffmann & Pfützner, 2013). Insbesondere Verbände mit geringerer Professionalisierungschance sind daher auf sehr frühzeitige Talent-Screenings mit entsprechend langfristiger und noch zu verbessernder Prognosestrategie angewiesen. Vor diesem Hintergrund besteht ein hoher Bedarf an wissenschaftlich gesicherten Modellen einer kombinierten Diagnose- und Interventionskonzeption für einzelne Sportarten oder Sportartengruppen. Der zu untersuchende Entwicklungsrahmen sollte die Frühförderung, d. h. Sportkindergärten und Sportbetonte Grundschulen ebenso wie die Eliteschulen des Sports in Berlin einbeziehen.

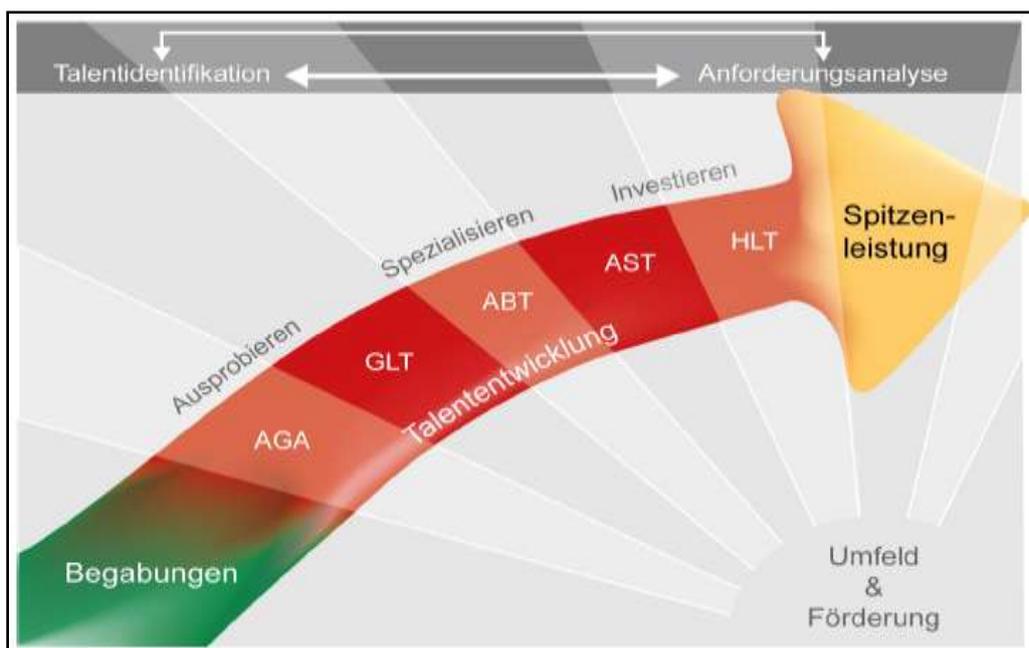
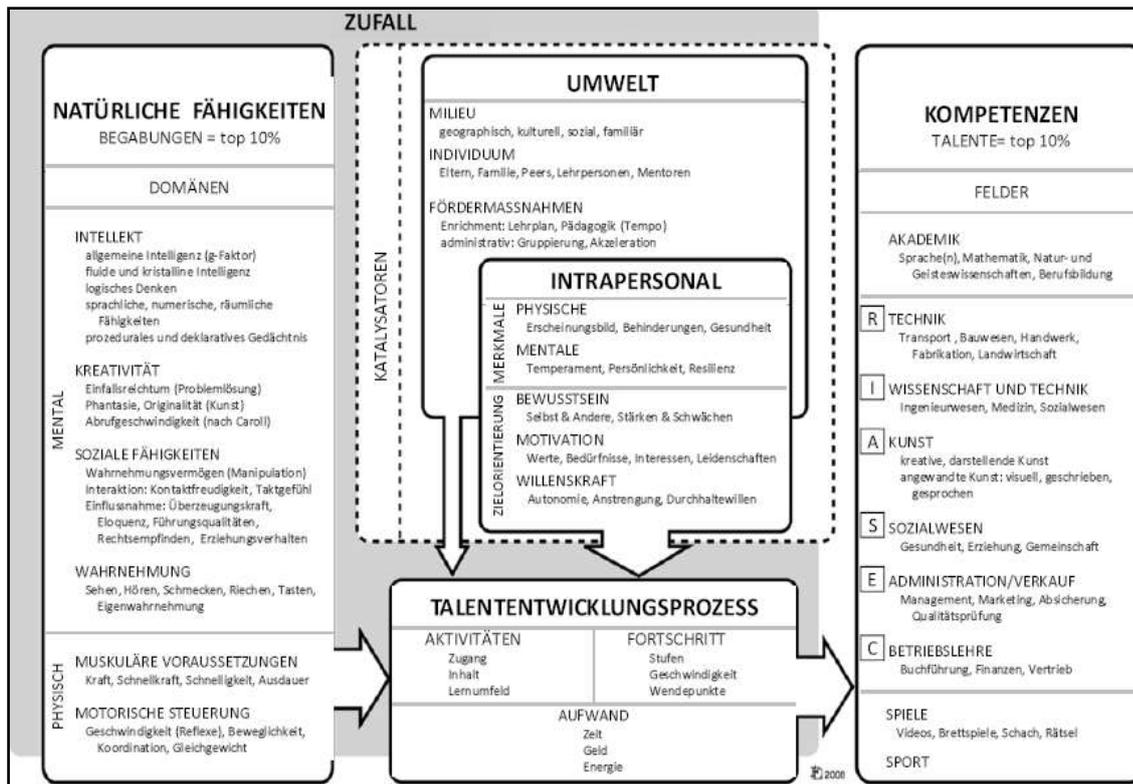


Abbildung 1: Modell der Talentidentifikation und -entwicklung (Hoffmann & Pfützner, 2013, S. 6; in Anlehnung an Gagné, 2010)

Neben der Talentauswahl geht es um die Talententwicklung und -förderung (siehe Abbildung 1) an den sportbetonten und den Sport-Eliteschulen in Berlin sowie in den Landes- und Spitzenfachverbänden, den Bundesstützpunkten und des Olympiastützpunktes in Berlin.

Das Forschungsvorhaben orientiert sich an dem Talententwicklungsmodell von Gagné (2008; 2010; 2013) sowie den Leipziger Positionen zum Nachwuchsleistungssport (Hoffmann & Pfütznern, 2013). Hierin wird der Athlet als bio-psycho-soziale Einheit betrachtet, dessen Entwicklung durch mentale, soziale und physische Merkmale direkt und durch Katalysatoren bzw. Moderator- und Mediatorvariablen erfolgt (siehe Tabelle 1).

Tabelle 1: Differentiated Model of Giftedness and Talent (DMGT, Gagné, 2008)



Dabei differenziert das Modell von Gagné (2008) dezidiert zwischen den Bausteinen Begabung, Talent, Talententwicklungsprozess sowie intrapersonalen und Umweltfaktoren, die entgegen älterer Positionen nicht synonym, sondern in dynamischer Wechselwirkung verstanden werden.

Das Modell von Gagné (2008) soll mit den Voraussetzungen und Randbedingungen in Berlin kontrastiert werden, um das „Berliner“ Talentfördermodell (siehe Abbildung 2) empirisch prüfen und weiter entwickeln zu können.

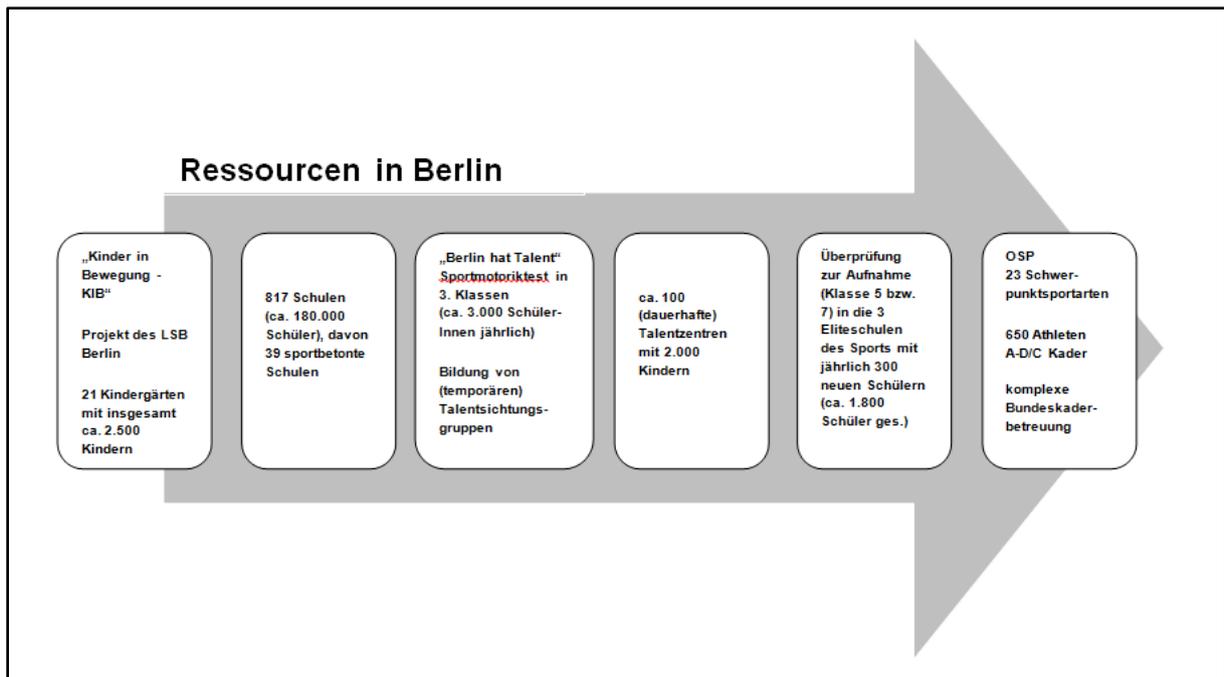


Abbildung 2: Das „Berliner“ Talentfördermodell nach Zinner (2012)

Das „Berliner“ Talentfördermodell berücksichtigt 817 Berliner Schulen, darunter 39 sportbetonte Schulen. An einigen davon wird unter dem Motto „Berlin hat Talent“ seit 2011 der Deutsche Motorik-Test (Bös, Schlenker, Büsch, Lämmle, Müller, Oberger & Tittlbach, 2009) durchgeführt. Zu dem Modell gehören weiterhin 1.800 Schüler und Schülerinnen an den drei Eliteschulen des Sports sowie 700 A-D/C-Kader-Sportler(innen) am OSP Berlin und 1.000 D-Kader-Athlet(inn)en, die vom Landesinstitut für Sportmedizin in Berlin regelmäßig untersucht werden.

Ziel des Projekts ist der Aufbau eines flächendeckenden Sichtungs- und Förderungsnetzwerkes in Berlin durch ergänzende Sportangebote für bewegungsbegabte Kinder. Dabei sollen Mädchen und Jungen an ein leistungssportliches Training herangeführt werden. Geplant – und in Ansätzen bereits realisiert – ist, im Rahmen des Schulsportunterrichts flächendeckend in den 3. Klassen den Deutschen Motorik-Test (DMT) einzuführen. Bewährt hat sich die Übernahme der Erfassung durch Speed4/Maximum Sport GmbH (Rene Orgis, Direktor Maximum Sport GmbH). Neben der Auswertung für die Gestaltung des Schulsportunterrichts kann der organisierte Sport die Ergebnisse des DMT unter anderem für die Förderung von „motorisch positiv auffälligen Kindern“ nutzen. Im Rahmen einer Pilotphase haben bisher rund 6.000 Kinder an dem Test teilgenommen. Im Ergebnis waren von diesen ca. 30 % „motorisch positiv auffällig“ und erhielten eine Einladung, an sportartunabhängigen Übungsstunden teilzunehmen. Nach den bisherigen Erfahrungen nutzen ca. 18 % das Angebot.

Speziell im Jahr 2013 wurden 2.770 Schüler von 54 Grundschulen (aus insgesamt 68 Schulen in zwei untersuchten Bezirken (79 %) getestet, davon 833 (30 %) zu einer „Talentidee“ (Sportartenbörse) eingeladen. 229 (27 %) haben die Einladung angenommen und sich in neun Sportarten vorgestellt. Im Ergebnis dessen werden im Herbst 2013 drei (temporäre) Talentsichtungsgruppen (TSG) (mit jeweils ca. 12 Schülern) aufgebaut, die ab November 2013 jeweils samstags über 12 bis 14 Wochen je eine Trainingseinheit (90 Minuten) absolvieren.

1.1 Perspektiven des Projekts

In Zusammenarbeit zwischen Landessportbund, Senat und H:G ist in diesem Projekt geplant, durch eine wissenschaftliche Begleitung eine Qualifizierung des DMT insbesondere durch die Entwicklung von „Berliner Normtabellen“ vorzunehmen und die Fähigkeitsstruktur sowie Testprofile zu überprüfen, eine Analyse zur Problematik durchzuführen, dass zu wenige der als bewegungsbegabt erkannten Schüler die Sportangebote (z. B. Eingliederung in eine Talentgruppe) annehmen, und die Entwicklung und Validierung eines Curriculums/Trainingsprogramms zur inhaltlichen Gestaltung der „Talentiaden“ und des ca. 14-wöchigen Trainings in den Talentgruppen vorzunehmen. Perspektiven ergeben sich damit wie folgt:

- Entwicklung „Berliner Normtabellen“
- Überprüfung der Fähigkeitsstruktur sowie der Testprofile
- Entwicklung einer Vorgehensweise zur Bildung integrativer Gesamturteile auf der Basis der Einzeltests des DMT
- Entwicklung und Validierung eines Curriculums bzw. Trainingsprogramms zur inhaltlichen Gestaltung der „Talentiaden“ bzw. zum Training der „Auffälligen“ Schüler
- Zwei Anschlussprojekte in Kooperation mit dem LSB/OSP Berlin: Evaluierung und Verbesserung der Aufnahmeprüfungen für die Einschulung in die drei Berliner Eliteschulen
- Sicherung der Anschlussfähigkeit dieser beiden Projekte für die Analyse und Weiterführung vor allem im Kontext einer Dualen Karriere und Vermeidung von Dropouts im Rahmen der D/C- bis A-Kader-Betreuung durch den OSP

Darüberhinaus sollen verschiedene Vorgehensweisen geprüft werden (insbesondere Mixed-Rasch-Modelle, d. h. IRT-Modelle sowie unscharfe mehrkriterielle (Fuzzy-Methoden) mit deren Hilfe sich aus den Teilurteilen der Tests des DMT valide Gesamturteile sowie Rangreihen bilden lassen.

Dieses Basisprojekt soll in den kommenden Jahren durch zwei Anschlussprojekte (und dann in weiterer Kooperation mit dem Olympiastützpunkt Berlin) weitergeführt werden:

Inhaltliche Evaluierung und Verbesserung der Aufnahmeprüfungen für die Einschulung in die drei Berliner Eliteschulen, Sicherung der Anschlussfähigkeit dieser beiden Projekte für die Analyse und Weiterführung vor allem im Kontext einer Dualen Karriere und Vermeidung von Dropouts im Rahmen der D/C- bis A-Kader-Betreuung durch den OSP.

1.2 Exploratorische Datenanalyse

Ergebnisse der exploratorischen Datenanalyse für die Items des DMT zeigen die folgenden Abbildungen 3-13:

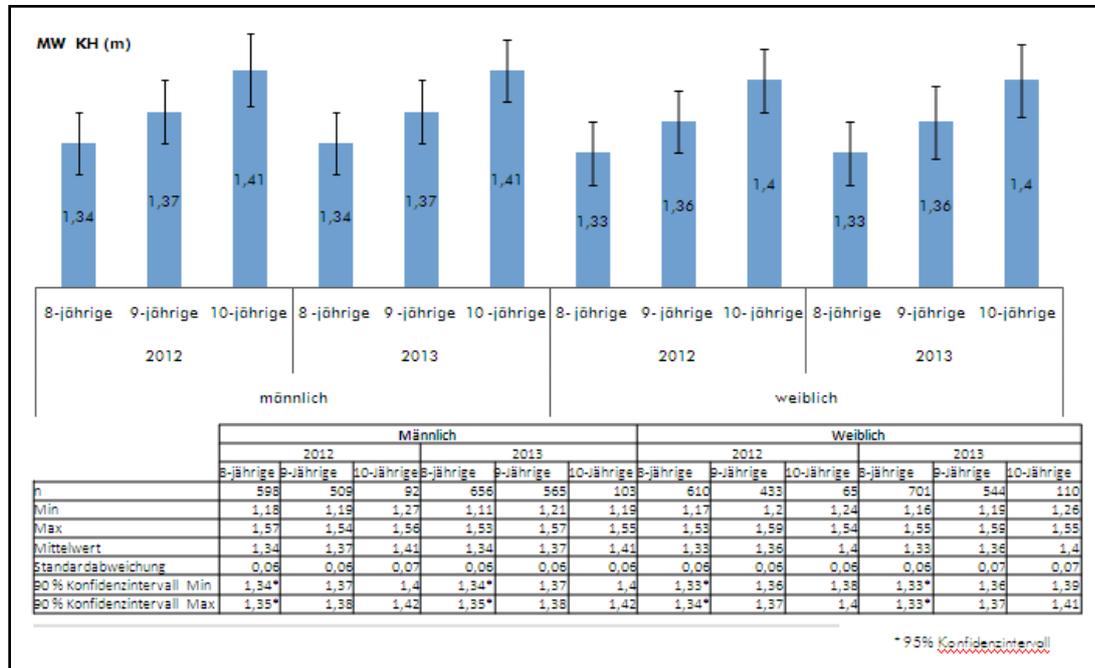


Abbildung 3: Körperhöhe in m

Bei den Untersuchungen im Jahr 2012/13 in Berlin wurden folgende Stichprobenumfänge erreicht:

Männlich: 8-jährige: $n = 1254$
 9-jährige: $n = 1074$
 10-jährige: $n = 195$
 Weiblich: 8-jährige: $n = 1311$
 9-jährige: $n = 977$
 10-jährige: $n = 175$

In der Summe liegen also von $N = 4.986$ Datensätze von 8-, 9- und 10-jährigen Schülerinnen und Schülern vor. Die folgenden Abbildungen zeigen die gemessenen Verhältnisse in den Tests des DMT sowie für die Körperhöhe, das Körpergewicht und den Body-Maß-Index.

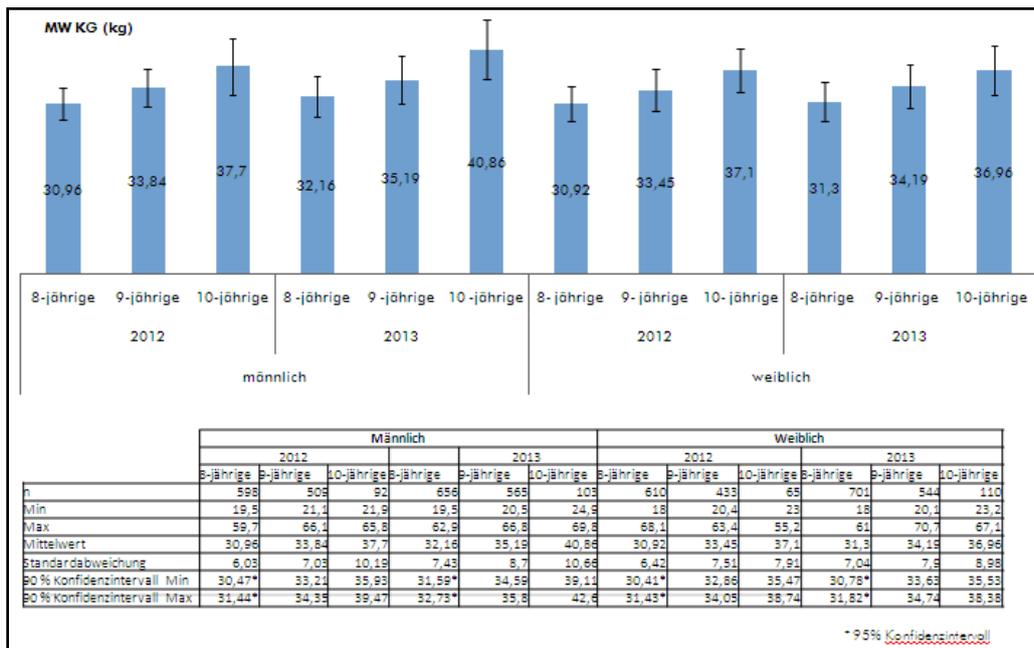


Abbildung 4: Körpergewicht in kg

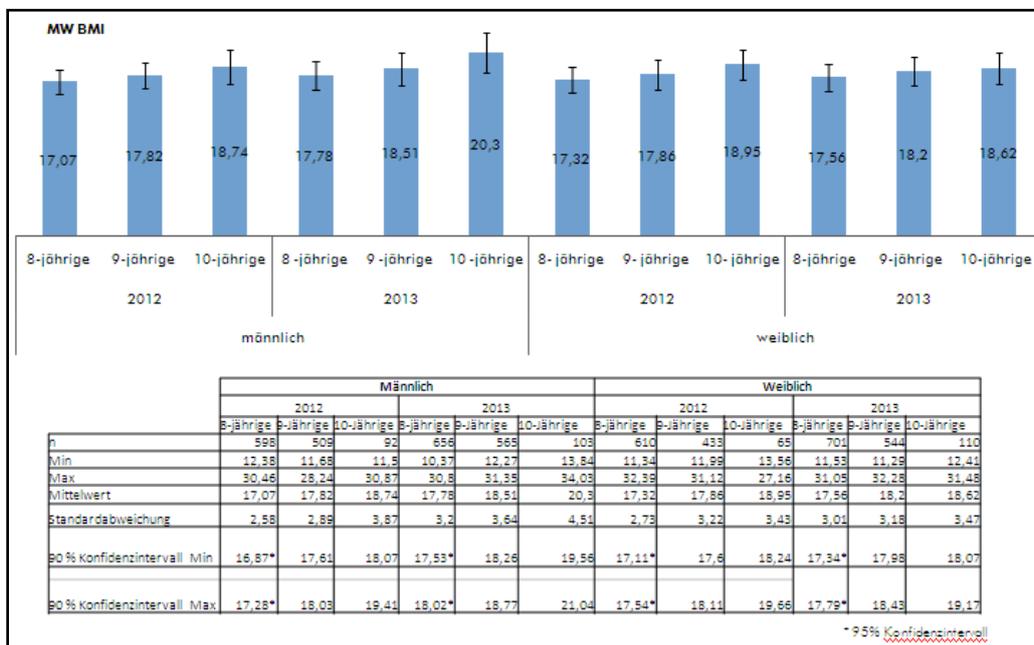


Abbildung 5: BMI

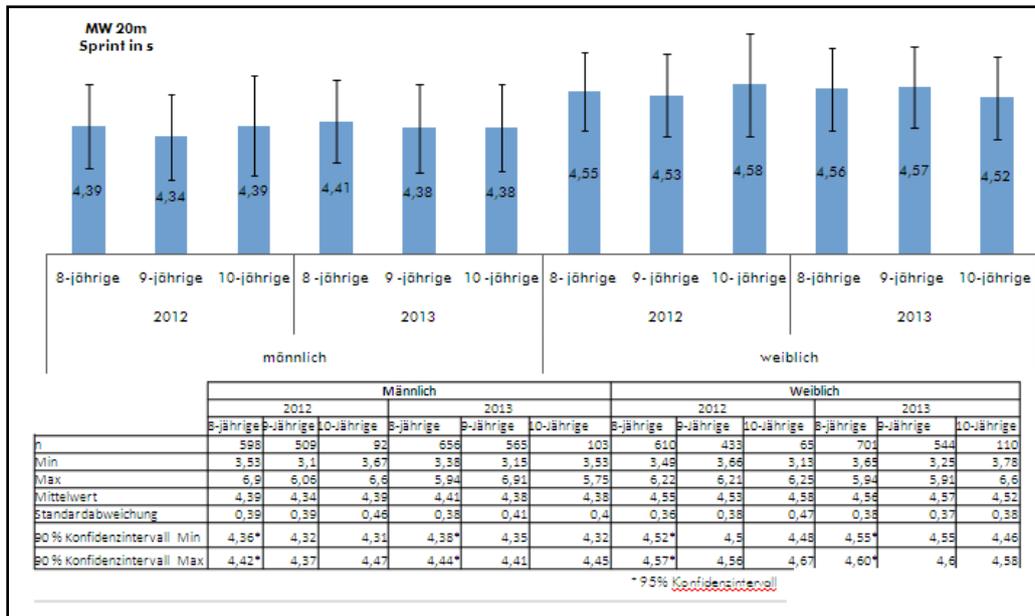


Abbildung 6: 20-m- Sprint in Sekunden

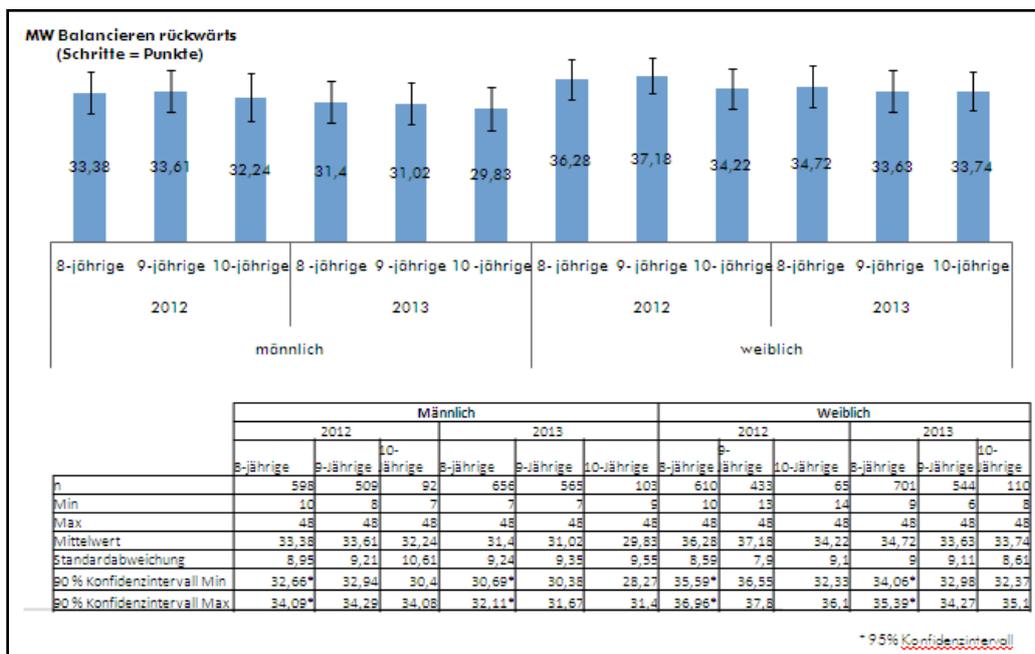


Abbildung 7: Balancieren rückwärts (Schritte = Punkte)

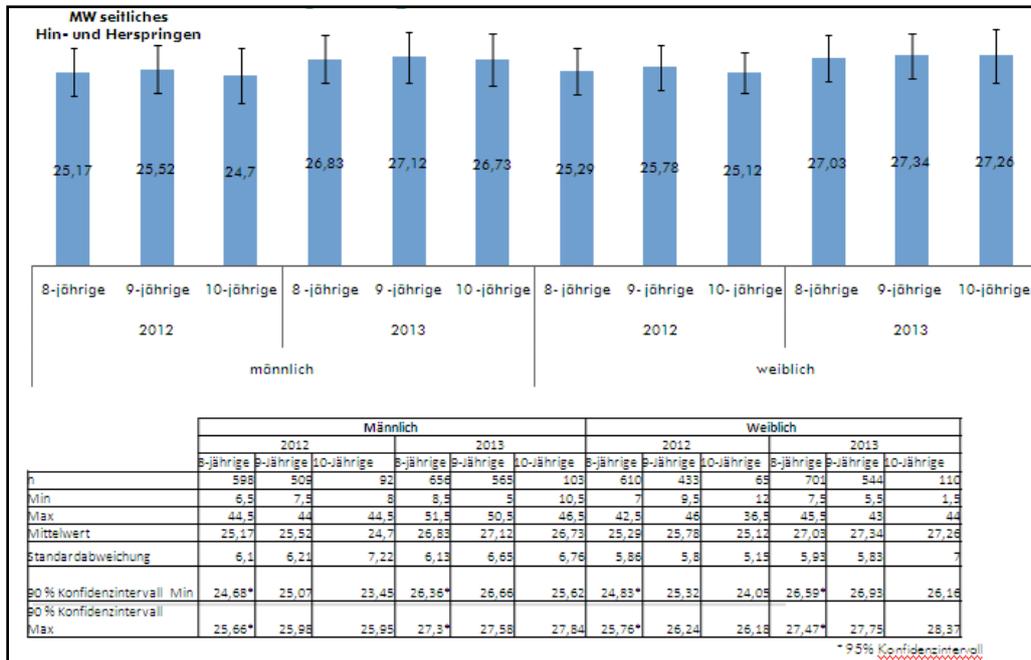


Abbildung 8: Seitliches Hin- und Herspringen

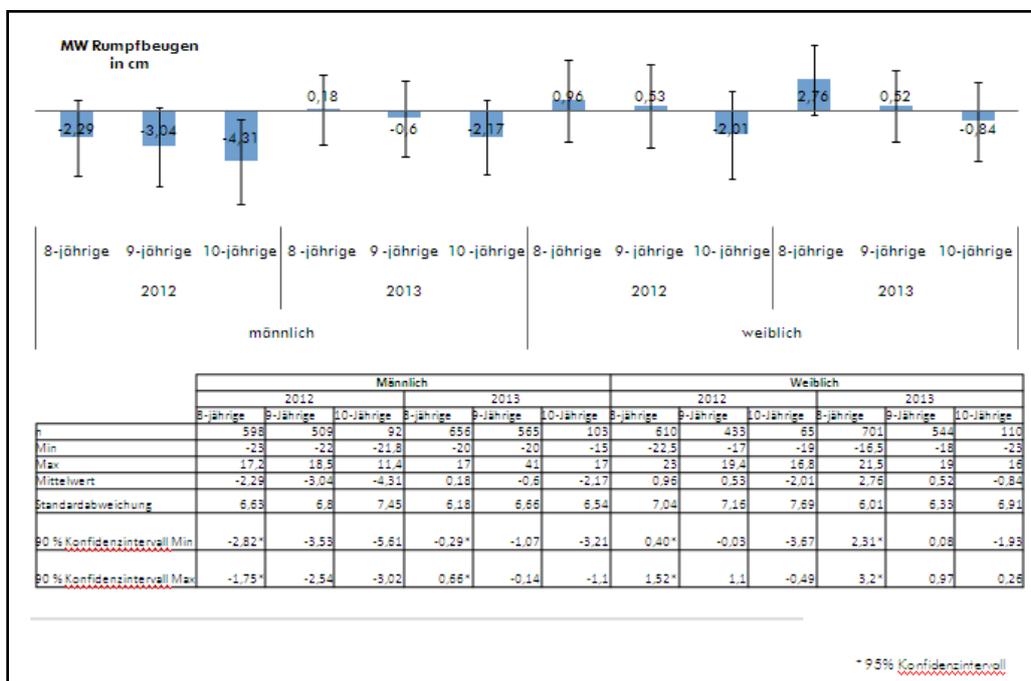


Abbildung 9: Rumpfbeugen in Zentimeter

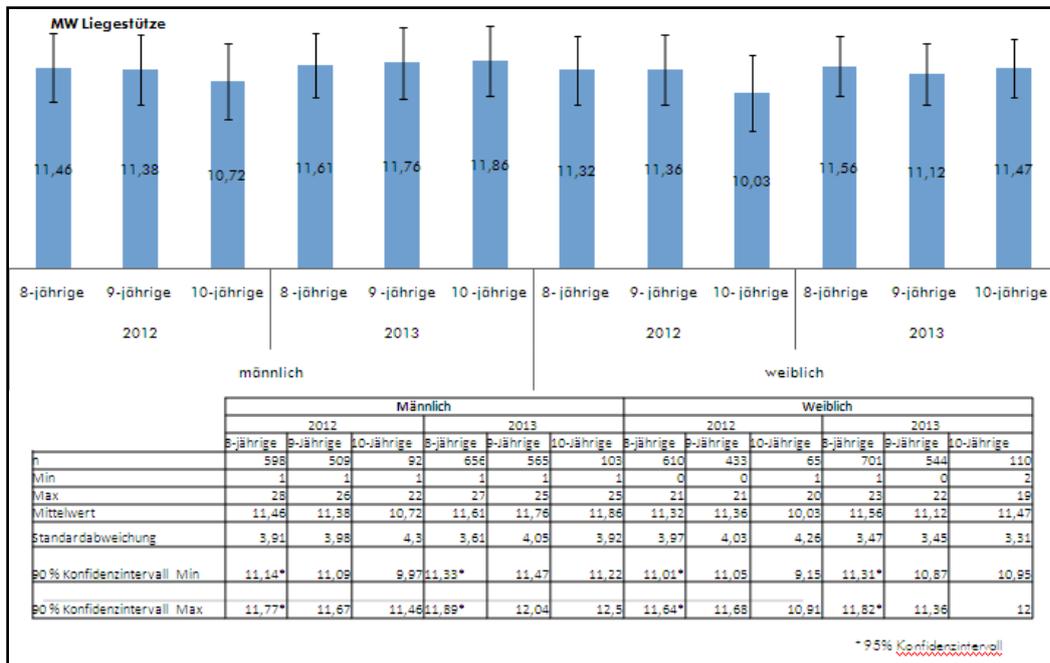


Abbildung 10: Liegestütze

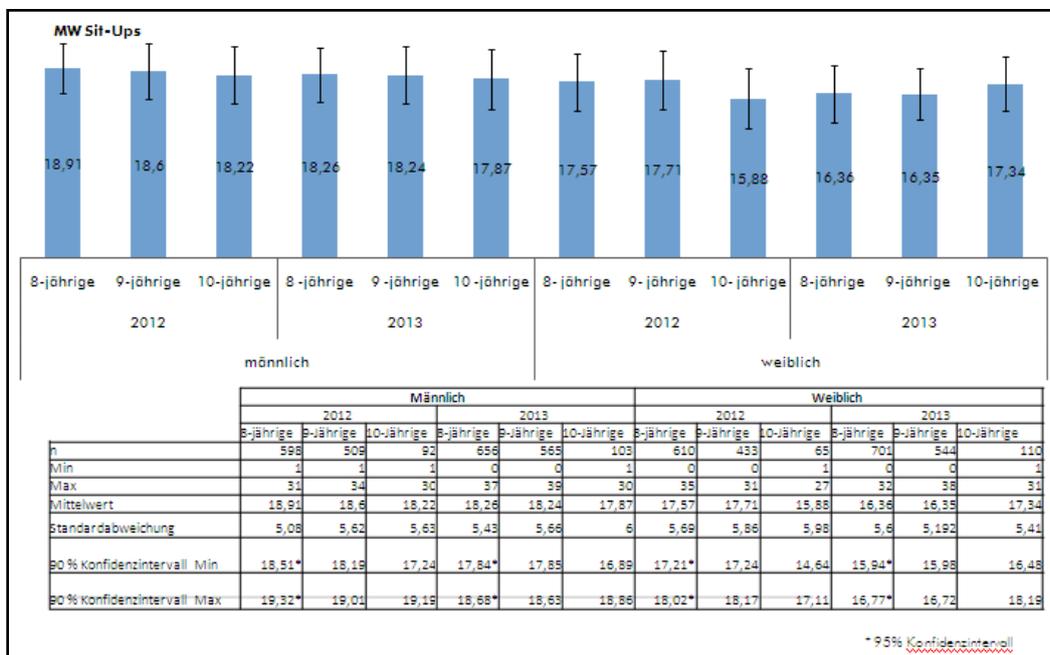


Abbildung 11: Sit-Ups

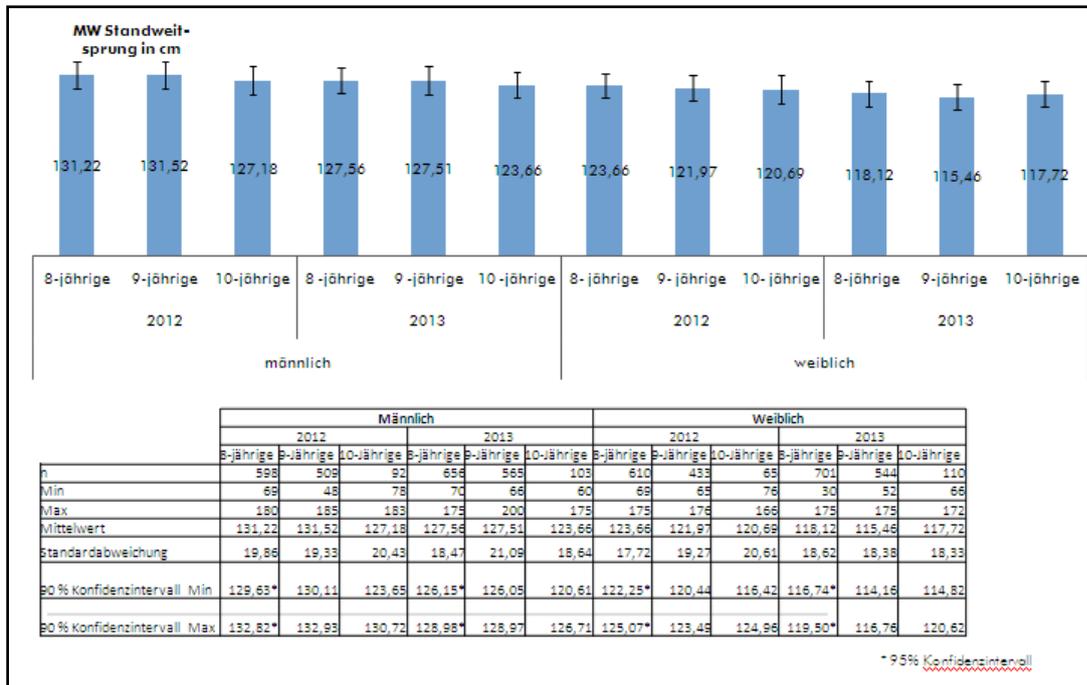


Abbildung 12: Standweitsprung in cm

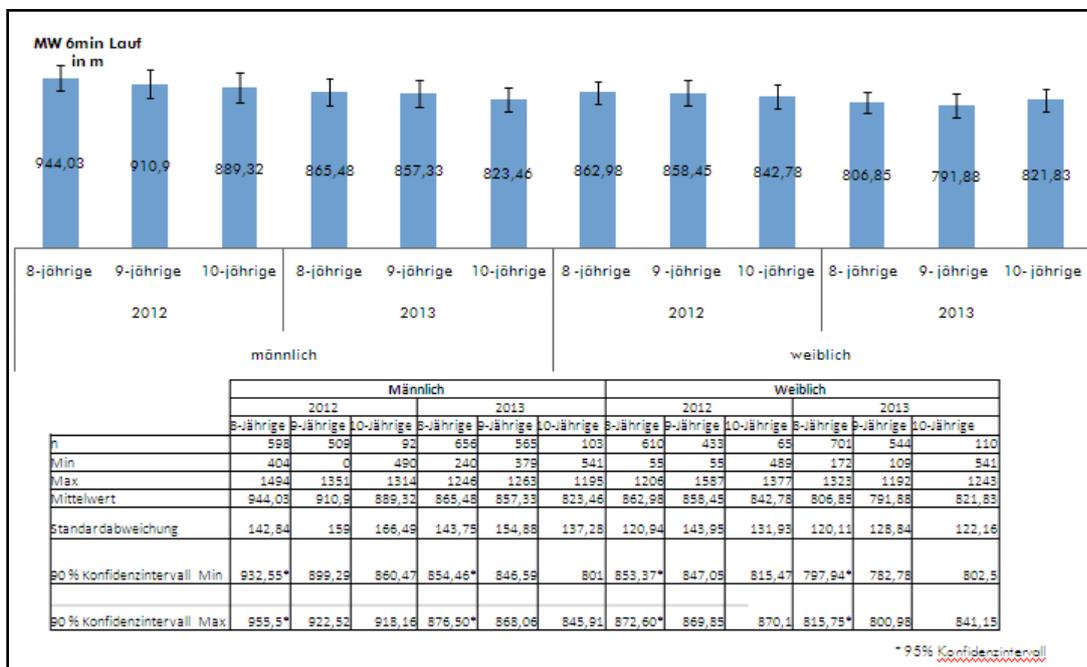


Abbildung 13: 6-Minuten-Lauf in Meter

1.3 Leistungsunterschiede von Berliner Schüler(inne)n zwischen 2012 und 2013

In den folgenden Tabellen ist eine Einteilung der 2012/13 untersuchten 2.458 Berliner Mädchen und Jungen im Alter von 9 und 10 Jahren in die Leistungsklassen (auf der Grundlage von so genannten Norm-Quintilen, die in die Auswertungs-Software implementiert ist, die vom Forschungszentrum für den Schulsport und den Sport von Kindern und Jugendlichen (FoSS) über die Website <http://www.sport.kit.edu/dmt/> erworben werden kann) für die Parameter des DMT erfolgt. Daraus ersichtlich sind marginale Leistungsunterschiede zwischen den beiden Erfassungsjahren.

Tabelle 2: 20m Sprint

			$\chi^2 (FG = 4, N = 2458) = 9,99, p = 0,04$					Gesamt
			Leistungsklassen (Quintile) für den 20-m-Sprint in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					
Durchführungsjahr der Untersuchung	2012	Anzahl	1	2	3	4	5	
		% innerhalb von Durchführungsjahr der Untersuchung	231	185	198	239	253	1106
		% der Gesamtzahl	20,9%	16,7%	17,9%	21,6%	22,9%	100,0%
	2013	Anzahl	9,4%	7,5%	8,1%	9,7%	10,3%	45,0%
		% innerhalb von Durchführungsjahr der Untersuchung	327	232	244	304	245	1352
		% der Gesamtzahl	24,2%	17,2%	18,0%	22,5%	18,1%	100,0%
		% der Gesamtzahl	13,3%	9,4%	9,9%	12,4%	10,0%	55,0%
Gesamt		Anzahl	558	417	442	543	498	2458
		% innerhalb von Durchführungsjahr der Untersuchung	22,7%	17,0%	18,0%	22,1%	20,3%	100,0%
		% der Gesamtzahl	22,7%	17,0%	18,0%	22,1%	20,3%	100,0%

Vorläufige Interpretation
 Die Häufigkeitsverteilung über die Leistungsklassen unterscheidet sich geringfügig zwischen 2012 und 2013 ($w = 0,17$)!
 Innerhalb der beiden Jahre bestehen ebenfalls nur geringfügige unterschiedliche Häufigkeitsverteilungen ($0,12 \leq w \leq 0,14$)!

Tabelle 3: Balancieren rückwärts

			$\chi^2 (FG = 4, N = 2454) = 61,49, p < 0,001$					Gesamt
			Leistungsklassen (Quintile) für das Balancieren rückwärts in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					
Durchführungsjahr der Untersuchung	2012	Anzahl	1	2	3	4	5	
		% innerhalb von Durchführungsjahr der Untersuchung	132	102	153	251	468	1106
		% der Gesamtzahl	11,9%	9,2%	13,8%	22,7%	42,3%	100,0%
	2013	Anzahl	5,4%	4,2%	6,2%	10,2%	19,1%	45,1%
		% innerhalb von Durchführungsjahr der Untersuchung	262	166	247	265	408	1348
		% der Gesamtzahl	19,4%	12,3%	18,3%	19,7%	30,3%	100,0%
		% der Gesamtzahl	10,7%	6,8%	10,1%	10,8%	16,6%	54,9%
Gesamt		Anzahl	394	268	400	516	876	2454
		% innerhalb von Durchführungsjahr der Untersuchung	16,1%	10,9%	16,3%	21,0%	35,7%	100,0%
		% der Gesamtzahl	16,1%	10,9%	16,3%	21,0%	35,7%	100,0%

Tabelle 4: Seitliches Hin- und Herspringen

• **Seitl. Hin-/Herspringen** $\chi^2 (FG = 4, N = 2455) = 37,75, p < 0,001$

			Leistungsklassen (Quintile) für das seitliche Hin- und Herspringen in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	247	145	185	291	238	1106
		% innerhalb von Durchführungsjahr der Untersuchung	22,3%	13,1%	16,7%	26,3%	21,5%	100,0%
		% der Gesamtzahl	10,1%	5,9%	7,5%	11,9%	9,7%	45,1%
	2013	Anzahl	224	140	217	341	427	1349
		% innerhalb von Durchführungsjahr der Untersuchung	16,6%	10,4%	16,1%	25,3%	31,7%	100,0%
		% der Gesamtzahl	9,1%	5,7%	8,8%	13,9%	17,4%	54,9%
Gesamt	Anzahl	471	285	402	632	665	2455	
	% innerhalb von Durchführungsjahr der Untersuchung	19,2%	11,6%	16,4%	25,7%	27,1%	100,0%	
	% der Gesamtzahl	19,2%	11,6%	16,4%	25,7%	27,1%	100,0%	

Tabelle 5: Rumpfbeugen

• **Rumpfbeugen** $\chi^2 (FG = 4, N = 2457) = 22,44, p < 0,001$

			Leistungsklassen (Quintile) für die Rumpfbeuge in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	310	224	186	218	168	1106
		% innerhalb von Durchführungsjahr der Untersuchung	28,0%	20,3%	16,8%	19,7%	15,2%	100,0%
		% der Gesamtzahl	12,6%	9,1%	7,6%	8,9%	6,8%	45,0%
	2013	Anzahl	280	255	266	302	248	1351
		% innerhalb von Durchführungsjahr der Untersuchung	20,7%	18,9%	19,7%	22,4%	18,4%	100,0%
		% der Gesamtzahl	11,4%	10,4%	10,8%	12,3%	10,1%	55,0%
Gesamt	Anzahl	590	479	452	520	416	2457	
	% innerhalb von Durchführungsjahr der Untersuchung	24,0%	19,5%	18,4%	21,2%	16,9%	100,0%	
	% der Gesamtzahl	24,0%	19,5%	18,4%	21,2%	16,9%	100,0%	

Tabelle 6: Liegestütz

• **Liegestütz** $\chi^2 (FG = 4, N = 2448) = 12,49, p = 0,01$

			Leistungsklassen (Quintile) für die Liegestütz in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	204	180	214	197	310	1105
		% innerhalb von Durchführungsjahr der Untersuchung	18,5%	16,3%	19,4%	17,8%	28,1%	100,0%
		% der Gesamtzahl	8,3%	7,4%	8,7%	8,0%	12,7%	45,1%
	2013	Anzahl	211	205	248	314	365	1343
		% innerhalb von Durchführungsjahr der Untersuchung	15,7%	15,3%	18,5%	23,4%	27,2%	100,0%
		% der Gesamtzahl	8,6%	8,4%	10,1%	12,8%	14,9%	54,9%
Gesamt	Anzahl	415	385	462	511	675	2448	
	% innerhalb von Durchführungsjahr der Untersuchung	17,0%	15,7%	18,9%	20,9%	27,6%	100,0%	
	% der Gesamtzahl	17,0%	15,7%	18,9%	20,9%	27,6%	100,0%	

Tabelle 7: 6-Minuten-Lauf

• **6-min-Lauf** $\chi^2 (FG = 4, N = 2461) = 14,01, p = 0,007$

			Leistungsklassen (Quintile) für den Gesamtwert in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	74	271	510	225	30	1110
		% innerhalb von Durchführungsjahr der Untersuchung	6,7%	24,4%	45,9%	20,3%	2,7%	100,0%
		% der Gesamtzahl	3,0%	11,0%	20,7%	9,1%	1,2%	45,1%
	2013	Anzahl	110	384	610	227	20	1351
		% innerhalb von Durchführungsjahr der Untersuchung	8,1%	28,4%	45,2%	16,8%	1,5%	100,0%
		% der Gesamtzahl	4,5%	15,6%	24,8%	9,2%	0,8%	54,9%
Gesamt	Anzahl	184	655	1120	452	50	2461	
	% innerhalb von Durchführungsjahr der Untersuchung	7,5%	26,6%	45,5%	18,4%	2,0%	100,0%	
	% der Gesamtzahl	7,5%	26,6%	45,5%	18,4%	2,0%	100,0%	

Tabelle 8: Sit-Ups

• **Sit-Ups** $\chi^2 (FG = 4, N = 2453) = 15,78, p = 0,003$

			Leistungsklassen (Quintile) für die Sit-Ups in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	365	269	286	136	50	1106
		% innerhalb von Durchführungsjahr der Untersuchung	33,0%	24,3%	25,9%	12,3%	4,5%	100,0%
		% der Gesamtzahl	14,9%	11,0%	11,7%	5,5%	2,0%	45,1%
	2013	Anzahl	520	313	353	118	43	1347
		% innerhalb von Durchführungsjahr der Untersuchung	38,6%	23,2%	26,2%	8,8%	3,2%	100,0%
		% der Gesamtzahl	21,2%	12,8%	14,4%	4,8%	1,8%	54,9%
Gesamt	Anzahl	885	582	639	254	93	2453	
	% innerhalb von Durchführungsjahr der Untersuchung	36,1%	23,7%	26,0%	10,4%	3,8%	100,0%	
	% der Gesamtzahl	36,1%	23,7%	26,0%	10,4%	3,8%	100,0%	

Tabelle 9: Standweitsprung

• **Standweitsprung** $\chi^2 (FG = 4, N = 2455) = 49,49, p < 0,001$

			Leistungsklassen (Quintile) für den Standweitsprung in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Durchführungsjahr der Untersuchung	2012	Anzahl	332	282	197	190	105	1106
		% innerhalb von Durchführungsjahr der Untersuchung	30,0%	25,5%	17,8%	17,2%	9,5%	100,0%
		% der Gesamtzahl	13,5%	11,5%	8,0%	7,7%	4,3%	45,1%
	2013	Anzahl	554	338	232	142	83	1349
		% innerhalb von Durchführungsjahr der Untersuchung	41,1%	25,1%	17,2%	10,5%	6,2%	100,0%
		% der Gesamtzahl	22,6%	13,8%	9,5%	5,8%	3,4%	54,9%
Gesamt	Anzahl	886	620	429	332	188	2455	
	% innerhalb von Durchführungsjahr der Untersuchung	36,1%	25,3%	17,5%	13,5%	7,7%	100,0%	
	% der Gesamtzahl	36,1%	25,3%	17,5%	13,5%	7,7%	100,0%	

Die nachfolgenden Tabellen zeigen nun die Leistungsunterschiede zwischen männlichen und weiblichen Schülern im Alter von 9 und 10 Jahren ($N = 2.458$) für alle Tests des DMT aus den Untersuchungsjahren 2012 und 2013 auf der Grundlage der Norm-Quintile (Bös et al., 2009).

Tabelle 10: 20-m-Sprint

• **20-m-Sprint** $\chi^2 (FG = 4, N = 2458) = 48,64, p < 0,001$

			Leistungsklassen (Quintile) für den 20-m-Sprint in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	266	179	230	293	320	1288
		% innerhalb von Geschlecht	20,7%	13,9%	17,9%	22,7%	24,8%	100,0%
		% der Gesamtzahl	10,8%	7,3%	9,4%	11,9%	13,0%	52,4%
	weiblich	Anzahl	292	238	212	250	178	1170
		% innerhalb von Geschlecht	25,0%	20,3%	18,1%	21,4%	15,2%	100,0%
		% der Gesamtzahl	11,9%	9,7%	8,6%	10,2%	7,2%	47,6%
Gesamt		Anzahl	558	417	442	543	498	2458
		% innerhalb von Geschlecht	22,7%	17,0%	18,0%	22,1%	20,3%	100,0%
		% der Gesamtzahl	22,7%	17,0%	18,0%	22,1%	20,3%	100,0%

Tabelle 11: Balancieren rückwärts

• **Balancieren rückwärts** $\chi^2 (FG = 4, N = 2454) = 11,50, p = 0,021$

			Leistungsklassen (Quintile) für das Balancieren rückwärts in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	233	150	204	265	436	1288
		% innerhalb von Geschlecht	18,1%	11,6%	15,8%	20,6%	33,9%	100,0%
		% der Gesamtzahl	9,5%	6,1%	8,3%	10,8%	17,8%	52,5%
	weiblich	Anzahl	161	118	196	251	440	1166
		% innerhalb von Geschlecht	13,8%	10,1%	16,8%	21,5%	37,7%	100,0%
		% der Gesamtzahl	6,6%	4,8%	8,0%	10,2%	17,9%	47,5%
Gesamt		Anzahl	394	268	400	516	876	2454
		% innerhalb von Geschlecht	16,1%	10,9%	16,3%	21,0%	35,7%	100,0%
		% der Gesamtzahl	16,1%	10,9%	16,3%	21,0%	35,7%	100,0%

Tabelle 12: Seitliches Hin- und Herspringen

• **Seitl. Hin-/Herspringen** $\chi^2 (FG = 4, N = 2457) = 157,89, p < 0,001$

			Leistungsklassen (Quintile) für das seitliche Hin- und Herspringen in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	186	240	216	291	355	1288
		% innerhalb von Geschlecht	14,4%	18,6%	16,8%	22,6%	27,6%	100,0%
		% der Gesamtzahl	7,6%	9,8%	8,8%	11,9%	14,5%	52,5%
	weiblich	Anzahl	285	45	186	341	310	1167
		% innerhalb von Geschlecht	24,4%	3,9%	15,9%	29,2%	26,6%	100,0%
		% der Gesamtzahl	11,6%	1,8%	7,6%	13,9%	12,6%	47,5%
Gesamt		Anzahl	471	285	402	632	665	2455
		% innerhalb von Geschlecht	19,2%	11,6%	16,4%	25,7%	27,1%	100,0%
		% der Gesamtzahl	19,2%	11,6%	16,4%	25,7%	27,1%	100,0%

Tabelle 13: Rumpfbeugen

• **Rumpfbeugen** $\chi^2 (FG = 4, N = 2457) = 37,77, p < 0,001$

			Leistungsklassen (Quintile) für die Rumpfbeuge in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	270	247	214	295	262	1288
		% innerhalb von Geschlecht	21,0%	19,2%	16,6%	22,9%	20,3%	100,0%
		% der Gesamtzahl	11,0%	10,1%	8,7%	12,0%	10,7%	52,4%
	weiblich	Anzahl	320	232	238	225	154	1169
		% innerhalb von Geschlecht	27,4%	19,8%	20,4%	19,2%	13,2%	100,0%
		% der Gesamtzahl	13,0%	9,4%	9,7%	9,2%	6,3%	47,6%
Gesamt		Anzahl	590	479	452	520	416	2457
		% innerhalb von Geschlecht	24,0%	19,5%	18,4%	21,2%	16,9%	100,0%
		% der Gesamtzahl	24,0%	19,5%	18,4%	21,2%	16,9%	100,0%

Tabelle 14: Liegestütz

• **Liegestütz** $\chi^2 (FG = 4, N = 2448) = 13,06, p = 0,011$

			Leistungsklassen (Quintile) für die Liegestütz in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	228	186	228	254	386	1282
		% innerhalb von Geschlecht	17,8%	14,5%	17,8%	19,8%	30,1%	100,0%
		% der Gesamtzahl	9,3%	7,6%	9,3%	10,4%	15,8%	52,4%
	weiblich	Anzahl	187	199	234	257	289	1166
		% innerhalb von Geschlecht	16,0%	17,1%	20,1%	22,0%	24,8%	100,0%
		% der Gesamtzahl	7,6%	8,1%	9,6%	10,5%	11,8%	47,6%
Gesamt		Anzahl	415	385	462	511	675	2448
		% innerhalb von Geschlecht	17,0%	15,7%	18,9%	20,9%	27,6%	100,0%
		% der Gesamtzahl	17,0%	15,7%	18,9%	20,9%	27,6%	100,0%

Tabelle 15: Sit-Ups

• **Sit-Ups** $\chi^2 (FG = 4, N = 2453) = 57,92, p < 0,001$

			Leistungsklassen (Quintile) für die Sit-Ups in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	530	305	308	120	22	1285
		% innerhalb von Geschlecht	41,2%	23,7%	24,0%	9,3%	1,7%	100,0%
		% der Gesamtzahl	21,6%	12,4%	12,6%	4,9%	0,9%	52,4%
	weiblich	Anzahl	355	277	331	134	71	1168
		% innerhalb von Geschlecht	30,4%	23,7%	28,3%	11,5%	6,1%	100,0%
		% der Gesamtzahl	14,5%	11,3%	13,5%	5,5%	2,9%	47,6%
Gesamt		Anzahl	885	582	639	254	93	2453
		% innerhalb von Geschlecht	36,1%	23,7%	26,0%	10,4%	3,8%	100,0%
		% der Gesamtzahl	36,1%	23,7%	26,0%	10,4%	3,8%	100,0%

Tabelle 16: Standweitsprung

• **Standweitsprung** $\chi^2 (FG = 4, N = 2455) = 22,16, p < 0,001$

			Leistungsklassen (Quintile) für den Standweitsprung in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	424	314	244	203	103	1288
		% innerhalb von Geschlecht	32,9%	24,4%	18,9%	15,8%	8,0%	100,0%
		% der Gesamtzahl	17,3%	12,8%	9,9%	8,3%	4,2%	52,5%
	weiblich	Anzahl	462	306	185	129	85	1167
		% innerhalb von Geschlecht	39,6%	26,2%	15,9%	11,1%	7,3%	100,0%
		% der Gesamtzahl	18,8%	12,5%	7,5%	5,3%	3,5%	47,5%
Gesamt		Anzahl	886	620	429	332	188	2455
		% innerhalb von Geschlecht	36,1%	25,3%	17,5%	13,5%	7,7%	100,0%
		% der Gesamtzahl	36,1%	25,3%	17,5%	13,5%	7,7%	100,0%

Tabelle 17: 6-min-Lauf

• **6-min-Lauf** $\chi^2 (FG = 4, N = 2443) = 5,88, p = 0,21$

			Leistungsklassen (Quintile) für den 6-Min-Lauf in Anlehnung an den DMT in Abhängigkeit des Alters, Geschlechts und Erfassungsjahres					Gesamt
			1	2	3	4	5	
Geschlecht	männlich	Anzahl	622	252	180	159	66	1279
		% innerhalb von Geschlecht	48,6%	19,7%	14,1%	12,4%	5,2%	100,0%
		% der Gesamtzahl	25,5%	10,3%	7,4%	6,5%	2,7%	52,4%
	weiblich	Anzahl	521	256	173	138	76	1164
		% innerhalb von Geschlecht	44,8%	22,0%	14,9%	11,9%	6,5%	100,0%
		% der Gesamtzahl	21,3%	10,5%	7,1%	5,6%	3,1%	47,6%
Gesamt		Anzahl	1143	508	353	297	142	2443
		% innerhalb von Geschlecht	46,8%	20,8%	14,4%	12,2%	5,8%	100,0%
		% der Gesamtzahl	46,8%	20,8%	14,4%	12,2%	5,8%	100,0%

1.4 Zusammengefasste Ergebnisse der Datenanalyse

Aus den zurückliegenden Tabellen lässt sich eine Vielzahl von Schlussfolgerungen ableiten. So fällt in der deskriptiven Globalanalyse für den Gesamtdatensatz, d. h. die 8-, 9- und 10 Jährigen Jungen und Mädchen (Tabelle 18) zum Beispiel auf, dass bei den Tests Sit-Ups, Standweitsprung und 6-min-Lauf vergleichsweise wenige Schüler(innen) in der besten Leistungsklasse (Normtabelle), wohingegen beim Balancieren rückwärts und dem seitlichen Hin- und Herspringen mehr Schüler(innen) in der besten Leistungsklasse vertreten sind! Mit anderen Worten: Die Berliner sind im Vergleich zu den Stichproben in den Normtabellen von Bös et al. (2009) in diesen Parametern, insbesondere im Spitzenbereich offenbar weniger gut. Ersichtlich anders ist die Situation im Parameter Balancieren rückwärts und im seitlichen Hin- und Herspringen

Tabelle 18: Deskriptive Globalanalyse

item labels and sample frequencies:								
no.	label	n of cats	categories					N
			1	2	3	4	5	
1	LKQ_20M	5	807	790	992	1311	1095	4995
2	LKQ_BAL	5	607	505	751	1019	2113	4995
3	LKQ_SHH	5	493	647	704	1050	2101	4995
4	LKQ_RB	5	1073	912	903	1107	1000	4995
5	LKQ_LS	5	770	763	825	1074	1563	4995
6	LKQ_SU	5	1439	1088	1385	786	297	4995
7	LKQ_SW	5	1394	1098	939	896	668	4995
8	LKQ_6M	5	1921	1046	791	725	512	4995

Weitere – in diesem Sinne - ausgewählte Ergebnisse der deskriptiven Datenanalyse der in Berlin untersuchten Drittklässler (unabhängig vom kalendarischen Alter) in den Jahren 2012 und 2013 ($N = 4.986$) zeigt die Tabelle 19.

Aus ihr lässt sich beispielsweise ableiten:

- In den Sit-Ups, dem Standweitsprung und dem 6-Minuten-Lauf sind die Berliner Jungen und auch die Berliner Mädchen im bundesweiten Vergleich auf der Grundlage der Stichproben von Bös deutlich schlechter: Nur 1,7% der Jungen und nur 6,1% der Mädchen schaffen es bei den Sit-Ups in die Leistungsklasse 5 (analog 8,0%/7,3% im Standweitsprung bzw. 5,2%/6,5% im 6-Minuten-Lauf [Spalte A]). Diese Situation wird von 2012 zu 2013 noch kritischer (siehe Spalte B) und das wiederum gilt gleichermaßen für die Mädchen (siehe Spalte C) und die Jungen (siehe Spalte D).
- Die (eventuelle) Erwartung, dass die Älteren in den motorischen Parametern leistungsstärker sind als die Jüngeren, gilt komplett für keinen der Parameter. Im Gegenteil: Die Mädchen werden beim Rumpfbeugen mit zunehmendem Alter sogar (gleichmäßig) schlechter (siehe Spalte E), für die Jungen gilt das für drei Parameter (Rumpfbeugen, Sit-Ups und 6-Minuten-Lauf [siehe Spalte F]).
- Die sehr geringen Unterschiede in den Mittelwerten praktisch aller Parameter von 2012 und 2013 kann Anlass sein, beide Jahrgänge zu **einer** (homogenen) Stichprobe zu vereinen und als „Berliner Normtabelle“ weiter zu entwickeln bzw. sie gegebenenfalls auch mit den Bös'schen Tabellen zusammenzuführen.

Anmerkung: Da wir hier nur zwei Jahrgänge (auch noch aus einem abgegrenzten Gebiet) betrachtet werden, kann das natürlich nur im Sinne einer Tendenz gesehen werden, die in den kommenden Jahren weiter zu replizieren und zu konkretisieren ist.

Tabelle 19: Ausgewählte, zusammengefasste Ergebnisse der deskriptiven Datenanalyse der in Berlin untersuchten Drittklässler in den Jahren 2012 und 2013 (N = 4.986)

	Unterschiede der Mädchen und Jungen (Mädchen n=2.463, Jungen n=2.523)	Körpergröße	Gewicht	BMI	20m Lauf	Balancieren rw	Seitliches Hin-	Rumpfbeugen	Liegestütze	Sit-Ups	Standweitsprung	6-Minuten-Lauf
Gesamt (n=4.986)	Spalte A	Jungen sind im Durchschnitt 1cm größer als die Mädchen	Jungen sind schwerer als Mädchen (besonders 10-Jährige)	2012: annähernd gleich (8-Jährige Mädchen liegen über Jungen), 2013: Jungen haben höheren BMI als Mädchen (vor allem 10-Jährige)	In Leistungsklassen sind die Jungen besser als die Mädchen J: 20,7%; LK1 M: 25%; LK1 J: 24,8%; LK5 M: 15,2%; LK5 2012: 20,9% in LK1 & 22,9% in LK5; 2013: 24,2% in LK1 & 18,1% in LK5	In Leistungsklassen sind die Mädchen besser als die Jungen J: 18,1%; LK1 M: 13,8%; LK1 J: 27,6%; LK5 M: 37,7%; LK5 2012: 11,9% in LK1 & 42,3% in LK5; 2013: 19,4% in LK1 & 30,3% in LK5	In Leistungsklassen sind die Mädchen besser als die Jungen J: 14,4%; LK1 M: 24,4%; LK1 J: 27,6%; LK5 M: 26,6%; LK5 2012: 22,3% in LK1 & 21,5% in LK5; 2013: 16,6% in LK1 & 31,7% in LK5	In Leistungsklassen sind die Jungen besser als die Mädchen J: 21%; LK1 M: 27,4%; LK1 J: 20,3%; LK5 M: 13,2%; LK5 2012: 28% in LK1 & 15,2% in LK5; 2013: 20,7% in LK1 & 18,4% in LK5	In Leistungsklassen sind die Jungen besser als die Mädchen J: 17,8%; LK1 M: 16%; LK1 J: 30,1%; LK5 M: 24,8%; LK5 2012: 18,5% in LK1 & 28,1% in LK5; 2013: 15,7% in LK1 & 27,2% in LK5	In Leistungsklassen sind die Mädchen besser als die Jungen J: 41,2%; LK1 M: 30,4%; LK1 J: 1,7%; LK5 M: 6,1%; LK5 2012: 33% in LK1 & 4,5% in LK5; 2013: 38,6% in LK1 & 3,2% in LK5	In Leistungsklassen sind die Jungen besser als die Mädchen J: 48,8%; LK1 M: 44,8%; LK1 J: 5,2%; LK5 M: 6,5%; LK5 2012: 30% in LK1 & 9,5% in LK5; 2013: 41,1% in LK1 & 6,2% in LK5	In Leistungsklassen sind die Mädchen besser als die Jungen J: 56,1m; 9,66,6m; 10,20,95m 2012: 8-9-10 2013: 10-8-9 Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m)
	Spalte B	Unterschiede 2012-2013 (Mädchen n=1.108, Jungen n=1.199, Mädchen n=1.108, Jungen n=1.199, 2013)	keine	Werden minimal schwerer (400-800g) 2012: 8-9-10 2013: 8-9-10	BMI steigt an, außer bei den 10-Jährigen (sinkt leicht) 2012: 8-9-10 2013: 8-9-10	8/9 werden langsamer (0,01-0,04s), 10-Jährigen besser (0,06s) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 1,56; 9: 3,5; 10: 0,48) 2012: 9-8-10 2013: 8-10-9	Werden besser 2012: 9-8-10 2013: 9-10-8	Werden besser außer 9-Jährige 2012: 8-9-10 2013: 8-9-10	Werden minimal besser (10-Jährige am stärksten) 2012: 9-8-10 2013: 8-10-9	Werden schlechter (außer 10-Jährige) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9
Mädchen (n=2.463)	Spalte C	keine	Werden minimal schwerer (400-800g) 2012: 8-9-10 2013: 8-9-10	BMI steigt an, vor allem bei 10-Jährigen 2012: 8-9-10 2013: 8-9-10	8/9 werden langsamer (0,02-0,04s), 10-Jährige minimal besser (0,01s) 2012: 9-10-8 2013: 9-10-8	Werden schlechter (8: 1,98; 9: 2,59; 10: 3,59) 2012: 9-8-10 2013: 8-9-10	Werden besser 2012: 9-8-10 2013: 9-8-10	Werden besser 2012: 8-9-10 2013: 8-9-10	Werden minimal besser (10-Jährige am stärksten) 2012: 8-9-10 2013: 10-9-8	Werden leicht schlechter (10-Jährige am stärksten) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9
	Spalte E	keine	Werden schwerer (8/9 ca. 1kg, 10 ca. 3kg) 2012: 8-9-10 2013: 8-9-10	BMI steigt an, vor allem bei 10-Jährigen 2012: 8-9-10 2013: 8-9-10	8/9 werden langsamer (0,02-0,04s), 10-Jährige minimal besser (0,01s) 2012: 9-10-8 2013: 9-10-8	Werden schlechter (8: 1,98; 9: 2,59; 10: 3,59) 2012: 9-8-10 2013: 8-9-10	Werden besser 2012: 9-8-10 2013: 9-8-10	Werden besser 2012: 8-9-10 2013: 8-9-10	Werden minimal besser (10-Jährige am stärksten) 2012: 8-9-10 2013: 10-9-8	Werden leicht schlechter (10-Jährige am stärksten) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9
Jungen (n=2.523)	Spalte D	keine	Werden schwerer (8/9 ca. 1kg, 10 ca. 3kg) 2012: 8-9-10 2013: 8-9-10	BMI steigt an, vor allem bei 10-Jährigen 2012: 8-9-10 2013: 8-9-10	8/9 werden langsamer (0,02-0,04s), 10-Jährige minimal besser (0,01s) 2012: 9-10-8 2013: 9-10-8	Werden schlechter (8: 1,98; 9: 2,59; 10: 3,59) 2012: 9-8-10 2013: 8-9-10	Werden besser 2012: 9-8-10 2013: 9-8-10	Werden besser 2012: 8-9-10 2013: 8-9-10	Werden minimal besser (10-Jährige am stärksten) 2012: 8-9-10 2013: 10-9-8	Werden leicht schlechter (10-Jährige am stärksten) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9
	Spalte F	keine	Werden schwerer (8/9 ca. 1kg, 10 ca. 3kg) 2012: 8-9-10 2013: 8-9-10	BMI steigt an, vor allem bei 10-Jährigen 2012: 8-9-10 2013: 8-9-10	8/9 werden langsamer (0,02-0,04s), 10-Jährige minimal besser (0,01s) 2012: 9-10-8 2013: 9-10-8	Werden schlechter (8: 1,98; 9: 2,59; 10: 3,59) 2012: 9-8-10 2013: 8-9-10	Werden besser 2012: 9-8-10 2013: 9-8-10	Werden besser 2012: 8-9-10 2013: 8-9-10	Werden minimal besser (10-Jährige am stärksten) 2012: 8-9-10 2013: 10-9-8	Werden leicht schlechter (10-Jährige am stärksten) 2012: 9-8-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9	Werden schlechter (8: 56,1m; 9,66,6m; 10,20,95m) 2012: 8-9-10 2013: 10-8-9

1.5 Bestimmung von Leistungsklassen nach unterschiedlichen Vorgehensweisen

In Anlehnung an das Schulnotensystem werden generell fünf Leistungsklassen unterschieden!

Vier unterschiedliche Vorgehensweisen zur Bestimmung der Leistungsklassen wurden realisiert:

- **Variante A:** Normkategorien, die auf der Basis unterschiedlich großer Stichproben von Bös et al. (2009) über Quintile festgelegt wurden (Norm-Quintile oder nG-Kategorien)
- **Variante B:** Quintile, die auf der Basis der Berliner Daten berechnet wurden (empirische Quintile oder eG-Kategorien)
- **Variante C:** Normleistungsklassen, die auf der Basis unterschiedlich großer Stichproben von Bös et al. (2009) über eine Z-Transformation festgelegt wurden (nZ-Kategorien)
- **Variante D:** Standardnoten, die auf der Basis der Berliner Daten nach einer T-Transformation berechnet wurden (empirische Leistungsklassen oder eT-Kategorien)

1.5.1 Kategoriengrenzwerte unterschiedlicher Bestimmungsverfahren

Die nachfolgenden Tabellen zeigen die bei den jeweiligen Verfahren entstehenden Grenzwerte für die fünf Leistungsklassen (Schulnoten).

Tabelle 20: Kategoriengrenzwerte für den 20m-Sprint am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

20m-Sprint [Sekunden]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	4,89-7,13	4,63-4,88	4,41-4,62	4,15-4,40	3,38-4,14
eG-Kategorien (qg)	4,72-7,13	4,45-4,71	4,26-4,44	4,10-4,25	3,38-4,09
nZ-Kategorien (lkz)	5,20-7,13	4,79-5,19	4,30-4,78	3,89-4,29	3,38-3,88
eT-Kategorien (tg)	5,31-7,13	4,77-5,29	4,04-4,75	3,78-4,03	3,38-3,77

Tabelle 21: Kategoriengrenzwerte für das Balancieren rückwärts am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Balancieren rückwärts [Punkte]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	7-19	20-23	24-28	29-33	34-48
eG-Kategorien (qg)	7-24	25-30	31-35	36-41	42-48
nZ-Kategorien (lkz)	7-13	14-21	22-31	32-39	40-48
eT-Kategorien (tg)	7-1	12-21	22-40	41-46	47-48

Tabelle 22: Kategoriengrenzwerte für das Seitliche Hin- und Herspringen am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Seitlich hin und her springen [Anzahl]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	6,5-16,0	16,5-19,0	19,5-21,5	22,0-25,0	25,5-51,5
eG-Kategorien (qg)	6,5-21,0	21,5-24,5	25,0-27,5	28,0-31,5	32,0-51,5
nZ-Kategorien (lkz)	6,5-12,0	12,5-17,0	17,5-23,5	24,0-28,5	29,0-51,5
eT-Kategorien (tg)	6,5-13,5	14,0-19,0	19,5-31,5	32,0-37,5	38,0-51,5

Tabelle 23: Kategoriengrenzwerte für das Rumpfbeugen am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Rumpfbeuge [Zentimeter]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	-23,0 - -7,5	-7,3 - -3,7	-3,6 - -0,5	-0,4 - 3,3	3,4 - 17,2
eG-Kategorien (qg)	-23,0 - -7,0	-6,9 - -2,2	-2,0 - 1,0	1,1 - 4,5	4,6 - 17,2
nZ-Kategorien (lkz)	-23,0 - -11,9	-11,8 - -6,0	-5,8 - 1,2	1,3 - 7,0	7,5 - 17,2
eT-Kategorien (tg)	-23,0 - -15,2	-15,0 - -7,9	-7,8 - 5,3	5,4 - 10,5	10,6 - 17,2

Tabelle 24: Kategoriengrenzwerte für Liegestütze am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Liegestütze [Anzahl]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	0-7	8-9	10-11	12-13	14-28
eG-Kategorien (qg)	0-8	9-11	12-13	14-14	15-28
nZ-Kategorien (lkz)	0-5	6-8	9-12	13-15	16-28
eT-Kategorien (tg)	0-2	3-6	7-13	14-18	19-28

Tabelle 25: Kategoriengrenzwerte für Sit-Ups am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Sit-Ups [Anzahl]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	0-15	16-18	19-22	23-26	27-37
eG-Kategorien (qg)	0-15	16-18	19-20	21-23	24-37
nZ-Kategorien (lkz)	0-10	11-16	17-24	25-30	31-37
eT-Kategorien (tg)	0-5	6-12	13-22	23-27	28-37

Tabelle 26: Kategoriengrenzwerte für den Standweitsprung am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

Standweitsprung [Zentimeter]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	69-111	112-123	124-133	134-145	146-180
eG-Kategorien (qg)	69-113	114-125	126-135	136-145	146-180
nZ-Kategorien (lkz)	69-97	98-115	116-138	139-156	157-180
eT-Kategorien (tg)	69-88	89-109	110-147	148-167	168-180

Tabelle 27: Kategoriengrenzwerte für den 6-Min-Lauf am Beispiel der 8-jährigen Jungen, Anmerkung: Die Daten liegen für die 9- und 10-jährigen Jungen und Mädchen in allen Tests des DMT (siehe die nachfolgenden Tabellen) separat vor!

6-Min-Lauf [Meter]	Leistungsklassen				
	1	2	3	4	5
nG-Kategorien (lkq)	240-838	839-914	915-980	981-1056	1057-1494
eG-Kategorien (qg)	240-779	780-870	871-945	946-1033	1034-1494
nZ-Kategorien (lkz)	240-749	752-868	869-1013	1014-1130	1133-1494
eT-Kategorien (tg)	240-595	601-757	758-1050	1051-1168	1172-1494

1.5.2 Konsequenzen aus den unterschiedlichen Bestimmungsverfahren für die Grenzwerte

- Unterschiedliche Bestimmungsverfahren für die Leistungsklassen führen zu unterschiedlichen Ergebnissen bzw. Grenzwerten, die wiederum mit gravierenden Konsequenzen (bspw. Fehlentscheidungen zu Gunsten oder zu Lasten von Talenten) für die Betroffenen verbunden sein können!
- Aus formal-empirischen Gründen wird die Variante D, d. h. aus den Berliner Daten empirisch berechnete Standardnoten im Gegensatz zu Bös et al. (2009) favorisiert, da zumindest mit dieser Variante für die sechs Konditionstests ein (homogener) Summenwert berechnet werden darf, der eine valide Leistungsbeurteilung der Kondition gestattet! Eine ausführliche Begründung für dieses Vorgehen liefert der Teil 2, insbesondere 2.4 (Konsequenzen aus den dimensionsanalytischen Prüfungen des Strukturmodells)

Und zwar aus folgendem Grund: Die Einordnung der Probanden in die Leistungsklassen nach den Normwerten ist nicht gleichverteilt. Deshalb verbietet sich z.B. der Einsatz von Faktorenanalysen, um die Dimensionalität der Tests des DMT zu prüfen. Das geschieht mit Hilfe von Mixed-Rasch-Modellen (auch als Item-Response-Theory-Modelle, IRT-Modelle bezeichnet). In unserem Fall zeigen diese, dass die Tests RB und Balrw nicht zu den 6 weiteren – vorwiegend konditionell verursachten – Tests „passen“. In einer – additiv – gestalteten Gesamteinschätzung dürfen diese Tests somit nicht verwendet werden.

„Unsere“ Variante D bezieht deshalb nur die verbleibenden 6 (konditionellen) Tests in die Gesamteinschätzung ein, transformiert die entsprechenden Messwerte in eine T-Skala (die dann normalverteilt ist) und ermittelt sodann den Mittelwert über die 6 T-Werte (Rasch-bereinigter T-transformierter Mittelwert).

Eine weitere Vorgehensweise wäre, mehrere Tests zur Koordination/Beweglichkeit hinzu zu nehmen, um damit auch formal-statistisch die Voraussetzungen für den Nachweis eines zweiten Faktors zu schaffen, da mit einem oder zwei Tests zum Konstrukt Koordination/Beweglichkeit der Nachweis nicht valide erbracht werden kann.

1.5.3 Vergleich der unterschiedlich bestimmten Leistungsklassen bzgl. jedes einzelnen Parameters des DMT für die Varianten A und D

Die folgenden Tabellen zeigen die Leistungsklassen auf Basis von empirisch berechneten Standardnoten für die Berliner Stichprobe (alle 8- bis 10-jährigen Jungen und Mädchen aus den Jahren 2012 und 2013, $N = 5038$), die mit den Leistungsklassen auf der Basis der Normquintile nach Bös et al. (2009) verglichen wurden. Bei allen Berechnungen mit $N = 5038$ wurde zuerst die geschlechts- und altersspezifische Einteilung in die Leistungsklassen vorgenommen, bevor die einzelnen Datensätze zusammengefügt wurden.

Tabelle 28: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim 20m-Sprint

• **20-m-Sprint** $\chi^2 (FG = 16, N = 5038) = 6176,06, p < 0,001, CI = 0,55$

Leistungsklassen (Quintile) für den 20-m-Sprint (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) * Leistungsklassen für den 20-m-Sprint auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für den 20-m-Sprint auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für den 20-m-Sprint (DMT in	1	128	543	147	0	0	818
Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	2	0	176	615	0	0	791
	3	0	0	1000	0	0	1000
	4	0	0	1314	11	0	1325
	5	0	0	362	642	100	1104
Gesamt		128	719	3438	653	100	5038

Ergebnis

- (1) Nach den Norm-Daten von Bös et al. (2009) erreichen 1104 die höchste Leistungsklasse und nach den empirisch bestimmten Leistungsklassen (Standardnoten) nur 100 Schüler!
- (2) Gleichzeitig erhöht sich das Risiko, ein Talent nicht zu identifizieren und zu fördern, weil die Grenzwerte strenger formuliert sind!

Erläuterung Tabelle 28 (gilt analog für die Tabellen 29-35):

Da die Variationsbreite für die empirisch bestimmten Leistungsklassen (T-transformierte Standardnoten) qua Definition in den Randbereichen 1 und 5 kleiner und in der mittleren Leistungsklasse größer ist, lassen sich die Ergebnisse wie folgt interpretieren:

- (1) Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 1104 die höchste Leistungsklasse und nach den empirisch bestimmten Leistungsklassen nur 100 Schüler (das gilt nur für den 20-m-Sprint)!

(2) Gleichzeitig erhöht sich der Beta-Fehler, d. h. die Nicht-Berücksichtigung eines potenziellen Talents, weil die Grenzwerte strenger formuliert sind.

Tabelle 29: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim Balancieren rückwärts

• **Bal. rückw.** $\lambda^2 (FG = 16, N = 5033) = 6176,06, p < 0,001, CI = 0,55$

Leistungsklassen (Quintile) für das Balancieren rückwärts (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) *
 Leistungsklassen für Balancieren auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für Balancieren auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für das Balancieren rückwärts	1	85	481	50	0	0	616
	2	0	113	397	0	0	510
	3	0	0	754	0	0	754
	4	0	0	1023	0	0	1023
	5	0	0	1146	563	421	2130
Gesamt		85	594	3370	563	421	5033

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 2130 die höchste Leistungsklasse im Balrw und nach den empirisch bestimmten Leistungsklassen nur 421 Schüler.

Tabelle 30: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim Seitlichen Hin- und Herspringen

• **Seilt. H&H** $\lambda^2 (FG = 16, N = 5034) = 4898,47, p < 0,001, CI = 0,49$

Leistungsklassen (Quintile) für das seitliche Hin- und Herspringen (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) *
 Leistungsklassen für Seitliches Hin- und Herspringen auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für Seitliches Hin- und Herspringen auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für das seitliche Hin- und Herspringen	1	91	354	55	0	0	500
	2	0	263	389	0	0	652
	3	0	0	711	0	0	711
	4	0	0	1054	0	0	1054
	5	0	0	1220	738	159	2117
Gesamt		91	617	3429	738	159	5034

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 2117 die höchste Leistungsklasse im SHH und nach den empirisch bestimmten Leistungsklassen nur 159 Schüler.

Tabelle 31: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim Rumpfbeugen

• **Rumpfbeuge** $\lambda^2 (FG = 16, N = 5035) = 6525,01, p < 0,001, CI = 0,57$

Leistungsklassen (Quintile) für die Rumpfbeuge (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) *
 Leistungsklassen für Rumpfbeuge auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für Rumpfbeuge auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für die Rumpfbeuge (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	1	101	644	336	0	0	1081
	2	0	0	921	0	0	921
	3	0	0	912	0	0	912
	4	0	0	1063	51	0	1114
	5	0	0	220	648	139	1007
Gesamt		101	644	3452	699	139	5035

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 1007 die höchste Leistungsklasse im RB und nach den empirisch bestimmten Leistungsklassen nur 139 Schüler.

Tabelle 32: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen bei den Liegestützen

• **Liegestütz** $\lambda^2 (FG = 16, N = 5023) = 4898,47, p < 0,001, CI = 0,49$

Leistungsklassen (Quintile) für die Liegestütz (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) * Leistungsklassen für Liegestütz auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für Liegestütz auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für die Liegestütz (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	1	63	431	285	0	0	779
	2	0	0	766	0	0	766
	3	0	0	832	0	0	832
	4	0	0	1058	19	0	1077
	5	0	0	354	1016	199	1569
Gesamt		63	431	3295	1035	199	5023

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 1569 die höchste Leistungsklasse im LS und nach den empirisch bestimmten Leistungsklassen nur 199 Schüler.

Tabelle 33: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen bei den Sit-Ups

• **Sit-Ups** $\lambda^2 (FG = 16, N = 5032) = 6252,82, p < 0,001, CI = 0,56$

Leistungsklassen (Quintile) für die Sit-Ups (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) * Leistungsklassen für Sit-ups auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren
Kreuztabelle

Anzahl

		Leistungsklassen für Sit-ups auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für die Sit-Ups (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	1	83	504	862	0	0	1449
	2	0	0	1097	0	0	1097
	3	0	0	1227	171	0	1398
	4	0	0	217	550	22	789
	5	0	0	0	135	164	299
Gesamt		83	504	3403	856	186	5032

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 299 die höchste Leistungsklasse in den SU und nach den empirisch bestimmten Leistungsklassen nur 164 Schüler.

Tabelle 34: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim Standweitsprung

• **Standweitspr.** $\lambda^2 (FG = 16, N = 5031) = 5562,81, p < 0,001, CI = 0,53$

Leistungsklassen (Quintile) für den Standweitsprung (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) * Leistungsklassen für Standweitsprung auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren
Kreuztabelle

Anzahl

		Leistungsklassen für Standweitsprung auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für den Standweitsprung (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	1	99	642	667	0	0	1408
	2	0	0	1105	0	0	1105
	3	0	0	917	26	0	943
	4	0	0	692	207	2	901
	5	0	0	53	494	127	674
Gesamt		99	642	3434	727	129	5031

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 674 die höchste Leistungsklasse im Standweitsprung und nach den empirisch bestimmten Leistungsklassen nur 127 Schüler.

Tabelle 35: Vergleich der empirischen Leistungsklassen (Standardnoten) mit den Norm-Quintilen für die 8- bis 10-Jährigen beim 6-Min-Lauf

• **6-Min-Lauf** $\chi^2 (FG = 16, N = 5016) = 5133,38, p < 0,001, CI = 0,51$

Leistungsklassen (Quintile) für den 6-Min-Lauf (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren) * Leistungsklassen für 6min-Lauf auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren Kreuztabelle

Anzahl

		Leistungsklassen für 6min-Lauf auf der Grundlage T-transformierter Daten (Standardnoten) in Abhängigkeit des Alters u. Geschlechts in beiden Jahren					Gesamt
		1	2	3	4	5	
Leistungsklassen (Quintile) für den 6-Min-Lauf (DMT in Abhängigkeit von Alter u. Geschlecht in beiden Jahren)	1	107	676	1151	0	0	1934
	2	0	0	1048	0	0	1048
	3	0	0	766	26	0	792
	4	0	0	454	274	2	730
	5	0	0	0	395	117	512
Gesamt		107	676	3419	695	119	5016

Ergebnis

Die empirisch bestimmten Leistungsklassen (Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Item weniger Personen als Talente eingeschätzt bzw. gefördert werden. Nach den Norm-Daten erreichen 512 die höchste Leistungsklasse im 6-Minuten-Lauf und nach den empirisch bestimmten Leistungsklassen nur 117 Schüler.

1.5.4 Konsequenzen aus dem Vergleich der unterschiedlich bestimmten Leistungsklassen bzgl. jedes einzelnen Parameters für die Varianten A und D

Die Variationsbreite für die empirisch bestimmten Leistungsklassen (T-transformierte Standardnoten) ist in den Randbereichen 1 und 5 kleiner und in der mittleren Leistungsklasse größer als bei den Norm-Quintilen. Daraus ergibt sich:

- (1) Die empirisch bestimmten Leistungsklassen (Standardnoten auf der Basis der Berliner Daten) reduzieren insbesondere die Personen in der höchsten Leistungsklasse, so dass je Testaufgabe teilweise deutlich weniger Personen als Talente eingeschätzt bzw. gefördert werden!
- (2) Die Unterschiede variieren zwischen den Testaufgaben!
- (3) Gleichzeitig erhöht sich der Beta-Fehler, d. h. das Risiko, ein potenzielles Talent nicht zu berücksichtigen erhöht sich, weil die Grenzwerte bei den empirisch bestimmten Leistungsklassen bessere Leistungen, insbesondere in der oberen Leistungsklasse verlangen!

Anmerkung: In dem Excel-Programm, das Bös für den DMT anbietet, sind die Leistungsklassen nach Bös identisch mit den Norm-Quintilen nach Bös!

1.5.5 Vergleich der summierten empirischen Leistungsklassen für einen Konditionsfaktor (6 Items, Variante D) mit dem Summenwert der Norm-Quintile (8 Items, Variante A) für die 8- bis 10-Jährigen

Den Vergleich der summierten empirischen Leistungsklassen für einen Konditionsfaktor (6 Items, Variante D) mit dem Summenwert der Norm-Quintile (8 Items, Variante A) für die 8- bis 10-Jährigen zeigt die Tabelle 36:

Tabelle 36: Vergleich der summierten empirischen Leistungsklassen für einen Konditionsfaktor (6 Items, Variante D) mit dem Summenwert der Norm-Quintile (8 Items, Variante A) für die 8- bis 10-Jährigen

• **Summenwert** $\chi^2 (FG = 16, N = 4995) = 3369,70, p < 0,001, CI = 0,47$

Kategorisierte Summenwerte auf der Grundlage der Quintile nach Bös * Kategorisierte Summenwerte auf der Grundlage der empirisch berechneten Standardnoten für den Konditionsfaktor Kreuztabelle

Anzahl

Anzahl		Kategorisierte Summenwerte auf der Grundlage der empirisch berechneten Standardnoten für den Konditionsfaktor				Gesamt
		2,00	3,00	4,00	5,00	
Kategorisierte Summenwerte auf der Grundlage der Quintile nach Bös	1,00	15	6	0	0	21
	2,00	87	571	3	0	661
	3,00	5	1410	293	0	1708
	4,00	0	607	1341	8	1956
	5,00	0	14	567	68	649
Gesamt		107	2608	2204	76	4995

Was für die Einzelwerte gilt, gilt auch für den Gesamtwert: Statt 649 Mädchen in der höchsten Leistungsklasse gelangen nach dem „Rasch-bereinigten Summenwert“ nur 68 Mädchen in diese höchste Klasse. Kein Mädchen verbleibt dagegen in der Klasse 1, also der Klasse der Schlechtesten.

*Zwischenbericht für das Projekt „Berlin hat Talent“
(Datenbasis 2012/2013)*

Teil 2:

Zur Problematik der Bildung von Gesamturteilen innerhalb des DMT

(Dirk Büsch & Jochen Zinner)



Ein gemeinsames Projekt von:

Gliederung

- 2.1 Das Strukturmodell motorischer Fähigkeiten als Grundlage für den DMT
- 2.2 Prüfung des Strukturmodells mit Verfahren der klassischen Testtheorie
- 2.3 Prüfung des Strukturmodells mit Verfahren der probabilistischen Testtheorie (IRT-Modelle)
- 2.4. Konsequenzen aus den dimensionsanalytischen Prüfungen des Strukturmodells
- 2.5 Das Strukturmodell motorischer Fähigkeiten als Grundlage für den „Berliner Konditionsfaktor“

2.1 Das Strukturmodell motorischer Fähigkeiten als Grundlage für den DMT

Das grundlegende Strukturmodell der motorischen Fähigkeiten im Kontext zum DMT zeigt Abbildung 1.

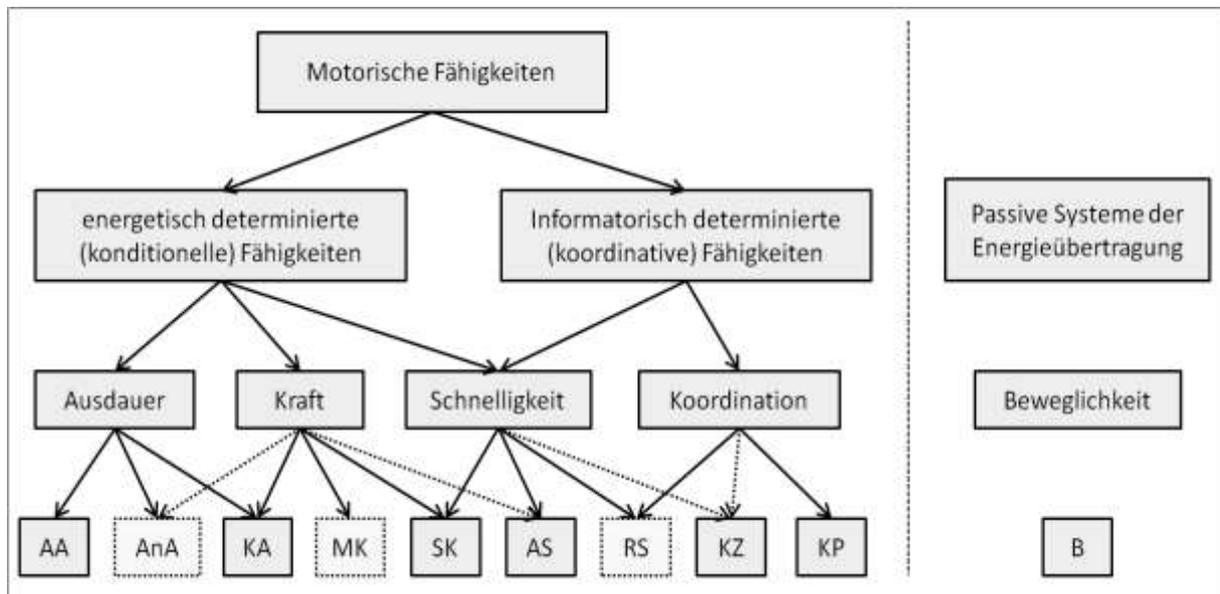


Abbildung 1: Strukturmodell motorischer Fähigkeiten (nach Bös, 2001, S. 2).

Anmerkung: Die markierten Haupterscheinungsformen werden durch die Items des DMT 6-18 abgedeckt. (AA, Aerobe Ausdauer; AnA, Anaerobe Ausdauer; KA, Kraftausdauer; MK, Maximalkraft; SK, Schnellkraft; AS, Aktionsschnelligkeit; RS, Reaktionsschnelligkeit; KZ, Koordination unter Zeitdruck; KP, Koordination bei Präzisionsaufgaben; B, Beweglichkeit).

2.2 Prüfung des Strukturmodells mit Verfahren der klassischen Testtheorie

Bei der Prüfung des Strukturmodells mit Verfahren der klassischen Testtheorie zeigt sich:

- Die konfirmatorischen Faktorenanalysen zeigen unzureichende Modellgeltungsprüfungen. Selbst mit den implementierten Modifikationsindizes ist keine suffiziente Modellgeltung herbeizuführen.
- Die exploratorischen Faktorenanalysen zeigen zumindest für die „Konditionsitems“ ca. 40 % Varianzaufklärung.
- Die Reliabilitätsanalysen (interne Konsistenz) zeigen unzureichende Werte für die Gesamt- ebenso für die „Konditionsskala“ (Chronbachs $\alpha_{\text{Jungen}} \leq 0,17$ und Chronbachs $\alpha_{\text{Mädchen}} \leq 0,15$).

Ein wesentlicher Grund dürfte in der unzureichenden Anzahl manifester Variablen für die beiden latenten Variablen „koordinative Fähigkeiten“ und „Beweglichkeit“ liegen, die zunächst zum Ausschluss der Koordinations- und Beweglichkeits-Items führen müssen oder

durch eine größere Anzahl manifester Variablen (mindestens vier für eine latente Variable) perspektivisch geprüft werden können!

2.3 Prüfung des Strukturmodells mit Verfahren der probabilistischen Testtheorie (IRT-Modelle)

Bei der Prüfung des Strukturmodells mit Verfahren der probabilistischen Testtheorie (IRT-Modelle) zeigt sich:

- Die Berechnung und Interpretation eines Summenwerts ist nur dann gerechtfertigt, wenn die Homogenität der Items (Item-Homogenität), d. h. die Eindimensionalität gegeben ist. Diese Voraussetzung kann nur mit IRT-Modellen, d. h. konkret mit dem (ordinalen) Rasch-Modell geprüft werden.
- Die Prüfung mit dem ordinalen Rasch-Modell bestätigt die Eindimensionalität für die Items 20-m-Sprint, Liegestütz, Sit-Ups, Standweitsprung, Seitliches Hin- und Herspringen sowie den 6-min-Lauf. Das Gemeinsame dieser Items ergibt sich aus dem Strukturmodell von Bös (2001) und dokumentiert einen „Konditionsfaktor“ (siehe Abbildung 2).

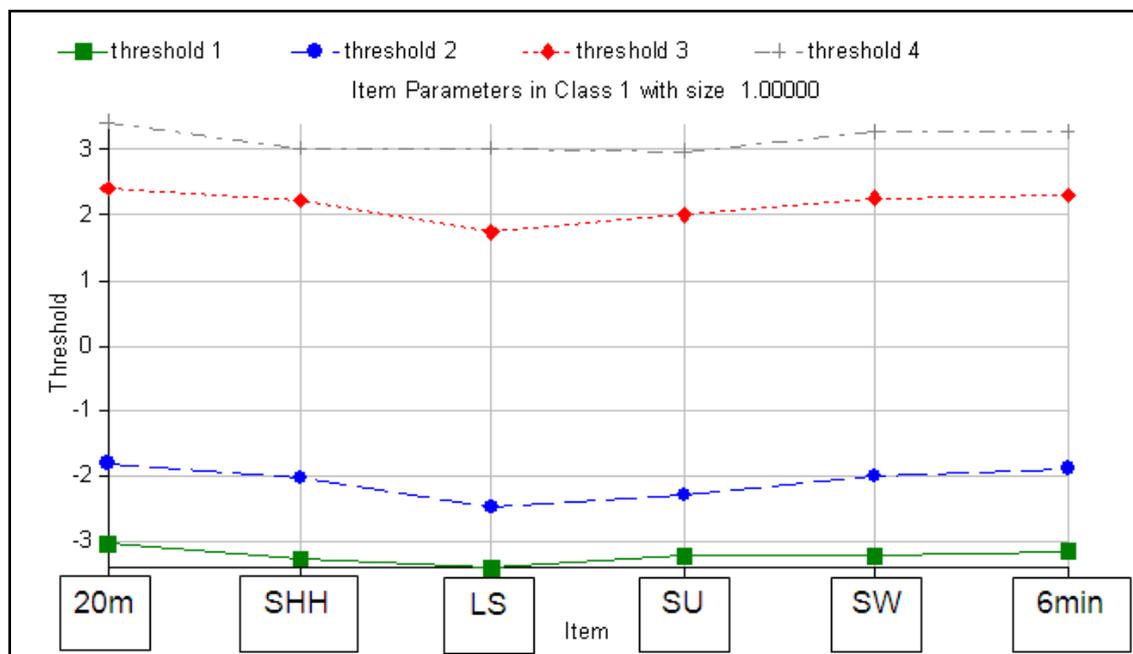


Abbildung 2: Die geordneten Schwellenparameter für das Ein-Klassen-Modell für die Leistungskategorien auf der Basis t-transformierter Standardnoten

Add 1) Da es sich um ordinal skalierte Modelle handelt, können nicht die Rohdaten, sondern die (Norm-) Quintile oder (Norm-Leistungsklassen) für die Prüfung verwendet werden. In den von Bös et al. bereit gestellten Excel-Eingabe- und -Ausgabe-Blättern werden irreführenderweise die Norm-Quintile als Leistungsklassen verwendet. Die Rohdaten wurden also im ersten Schritt diesen definierten Leistungsklassen zugeordnet. Dabei fällt auf, dass die Berliner Daten über die Leistungsklassen nicht gleichverteilt sind (siehe Abbildung 3).

```

item labels and sample frequencies:

```

no.	label	n of cats	0	1	2	3	4	N
1	LK_20MTG	5	123	715	3415	645	100	4998
2	LK_SHHTG	5	89	612	3404	734	159	4998
3	LK_LSTG	5	62	429	3278	1030	199	4998
4	LK_SUTG	5	82	501	3379	851	185	4998
5	LK_SWTG	5	96	637	3416	723	126	4998
6	LK_6MTG	5	105	672	3409	693	119	4998


```

expected category frequencies and item scores:

```

Item label	Item's Score	Item's Stdev	relative category frequencies				
			0	1	2	3	4
LK_20MTG	1.98	0.67	0.025	0.143	0.683	0.129	0.020
LK_SHHTG	2.05	0.68	0.018	0.122	0.681	0.147	0.032
LK_LSTG	2.18	0.69	0.012	0.086	0.656	0.206	0.040
LK_SUTG	2.11	0.69	0.016	0.100	0.676	0.170	0.037
LK_SWTG	2.03	0.67	0.019	0.127	0.683	0.145	0.025
LK_6MTG	2.01	0.67	0.021	0.134	0.682	0.139	0.024

Sum: | 12.35

```

threshold parameters: ordinal (partial credit) model

```

item label	item location	threshold parameters			
		1	2	3	4
LK_20MTG	0.24697	-3.033	-1.811	2.408	3.423
LK_SHHTG	-0.02556	-3.285	-2.038	2.210	3.011
LK_LSTG	-0.28785	-3.422	-2.483	1.723	3.030
LK_SUTG	-0.13876	-3.225	-2.291	2.001	2.960
LK_SWTG	0.07453	-3.229	-1.982	2.250	3.260
LK_6MTG	0.13068	-3.168	-1.905	2.306	3.289

Abbildung 3: Item-Parameter für das Ein-Klassen-Modell für die Leistungskategorien auf der Basis t-transformierter Standardnoten

Die Rasch-Analysen mit diesen Leistungsklassen bestätigen keine Modellgeltung in Anlehnung an das Strukturmodell von Bös et al. (2001).

Aufgrund von zahlenmäßigen Unterschieden zwischen den Norm-Quintilen und den empirischen Quintilen der Berliner Daten wurden die Rasch-Analysen mit den empirischen Quintilen wiederholt. Auch hierfür konnte keine Modellgeltung nachgewiesen werden.

Entsprechend dem Vorschlag von Büsch et al. (2009), Holzweg et al. (2012) und Utesch et al. (2014) wurden die empirischen Daten in Standardnoten (empirische Leistungsklassen von 1 bis 5) flächentransformiert (T-transformierte Standardnoten) und mit diesen Daten die Mixed-Rasch-Analyse wiederholt. Bei diesen Daten ergibt sich eine valide Ein-Klassen-Lösung für die „Konditions-Items“, wenn aufgrund nicht modellkonformer Antwortmuster die Items „Balancieren rückwärts“ und „Rumpfbeugen“ eliminiert werden.

Dabei ist generell zu berücksichtigen, dass die Modelle geschlechtsspezifisch und altersspezifisch geprüft und repliziert werden müssen, bevor eine abschließende Aussage formuliert werden kann!

2.4 Konsequenzen aus den dimensionsanalytischen Prüfungen des Strukturmodells

Die Unabhängigkeit der Beweglichkeit sowie die zu geringe Schwierigkeit des Balancierens rückwärts konnten bereits mehrfach bestätigt werden (u. a. Büsch et al., 2009; Holzweg et al., 2012; Utesch et al., 2014) und spiegeln sich auch in den Berliner Daten wider.

Für die weiterführenden Analysen im Teil 1 der Zwischeneinschätzung „Berlin hat Talent“ wurden daher die Leistungsklassen (Norm-Quintile) aus den Tabellen von Bös et al. (2009) (Variante A) mit den empirischen Leistungsklassen (T-transformierte Standardnoten) (Variante D) verglichen.

2.5 Das Strukturmodell motorischer Fähigkeiten als Grundlage für den „Berliner Konditionsfaktor“

Das Strukturmodell motorischer Fähigkeiten als Grundlage für den „Berliner Konditionsfaktor“ zeigt Abbildung 5. Die grün markierten Haupterscheinungsformen werden durch den Konditionsfaktor (6 Items des DMT 6-18) abgedeckt (AA, Aerobe Ausdauer; AnA, Anaerobe Ausdauer; KA, Kraftausdauer; MK, Maximalkraft; SK, Schnellkraft; AS, Aktionsschnelligkeit; KZ, Koordination unter Zeitdruck). Die rot markierten Haupterscheinungsformen (KP, Koordination bei Präzisionsaufgaben und B, Beweglichkeit) dürfen in einem Summenwert für den Konditionsfaktor keine Berücksichtigung finden.

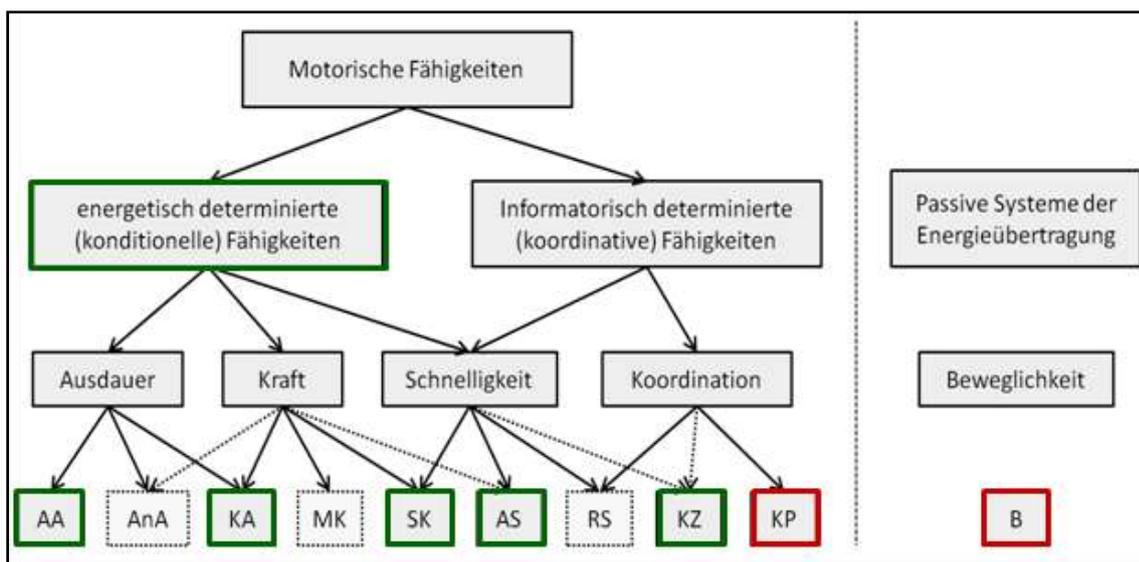


Abbildung 4: Strukturmodell motorischer Fähigkeiten (nach Bös, 2001, S. 2).

*Zwischenbericht für das Projekt „Berlin hat Talent“
(Datenbasis 2012/2013)*

Teil 3:

Modellversuch zur Auswertung des DMT mit Hilfe von mehrkriteriellen und FUZZY-Analysen. Datenbasis Lichtenberg 2014 (Mädchen)

(Jochen Zinner, Dirk Büsch & Jochen Ester)



Ein gemeinsames Projekt von:

Gliederung

- 3.1 Zielstellung
- 3.2 Mehrkriterielles und FUZZY-Vorgehen
- 3.3 Modellversuch Lichtenberger Mädchen
- 3.4 Vergleich der verschiedenen Klassifizierungsverfahren mit den unterschiedlichen FUZZY-Rangreihen
- 3.5 Schlussfolgerungen

3.1 Zielstellung

Ein wesentliches Problem in der Auswertung des Deutschen Motorik Test (DMT) besteht darin, für jeden Schüler letztlich ein (quantitatives) Gesamturteil sozusagen als globales Maß für die allgemeine motorische Leistungsfähigkeit zu erhalten. Bös empfiehlt wegen der Multidisziplinarität der Tests auf ein solches Gesamturteil zu verzichten. Wenn dennoch ein solches Urteil im Sinne einer ersten Orientierung getroffen werden soll, dann schlägt er vor, in sieben der acht Testaufgaben des DMT (außer Rumpfbeugen) mit Hilfe der Normtabellen die Z-Werte zu berechnen und dafür einen mittleren Wert zu bestimmen (Bös et al., S. 61). In unseren bisherigen Untersuchungen konnte gezeigt werden, dass man nicht nur *Rumpfbeugen [RB]*, sondern auch den Test *Balancieren rückwärts [Balrw]* heraus lassen sollte und dann einen T-Wert (transformiert in Standardnoten) auf der Basis der verbleibenden sechs Testwerte bildet („Rasch-bereinigter Summenwert“, siehe Büsch & Zinner, Zwischeneinschätzung „Berlin hat Talent“, Teil 2).

Mehrkriterielle, darunter auch unscharfe FUZZY-Analysen, sind gegenüber solchen „additiven, starren“ Methoden eventuell besser geeignet, für die mit dem DMT untersuchten Schüler einen Gesamtwert abzuschätzen. Das soll im vorliegenden Modellversuch geprüft werden.

3.2 Mehrkriterielles und FUZZY-Vorgehen

Charakteristisch für multikriterielle und FUZZY-Analysen ist die Konstruktion von Zugehörigkeitsfunktionen unter Berücksichtigung von Expertenwissen und das Rechnen mit solchen Funktionen (siehe Ester & Zinner, 1996; Zinner & Ester, 1994). Die Zugehörigkeitsfunktion gibt einen Wert dafür an, dass ein Schüler bezüglich aller Parameter besser ist als alle anderen Schüler. Auf diese Weise wird für die Menge der Schüler eine unscharfe (FUZZY-)Menge erzeugt, in der jeder Schüler einen Zugehörigkeitswert dafür bekommt, dass er besser als alle anderen ist. Der Sportler mit dem größten Zugehörigkeitswert kann dann im Sinne eines Gesamturteils über alle Parameter als „Bester“ gelten usw.

Ein derartige Verfahrensweise (Methode der unscharfen Dominanzmengen) wurde mit den Daten der 2014 untersuchten Schülerinnen aus dem Stadtbezirk Lichtenberg ($n = 533$) durchgeführt.

Im Sinne eines „Expertenratings“ (diskursive Validierung) wurden dazu zunächst die Präferenzrichtungen der einzelnen Parameter festgelegt, danach waren die Messfehlerbereiche und schließlich die paarweisen Gewichtungen zu schätzen (siehe Abbildung 1).

	Opt	Δ	Fkt.	KH	KG	20m	Bal rw	SHH	RB	LS	SU	SW	6 min
KH	max	0,5	lin	x	2	1/10	1/4	1/4	1/2	1/8	1/8	1/8	1/6
KG	min	0,5	lin		x	1/12	1/6	1/6	1/4	1/10	1/10	1/10	1/8
20m	min	0,05	lin			x	6	6	8	2	2	2	4
Bal rw	max	1	lin				x	1	2	1/4	1/4	1/4	1/2
SHH	max	1	lin					x	2	1/6	1/6	1/6	1/4
RB	max	0,5							x	1/6	1/6	1/6	1/4
LS	max	1	lin							x	1	1	2
SU	max	1	lin								x	1	2
SW	max	1	lin									x	2
6 min	max												x

Abbildung 1: Präferenzrichtungen, Fehlerbereiche, Parameterfunktionen und paarweise Gewichtungen entsprechend „Expertenrating“

Es ist ein großer Vorteil der FUZZY-Vorgehensweise, dass auf diese Weise subjektive Urteile in die mathematischen Algorithmen einbezogen werden können.

Die paarweisen Gewichtungen werden nun mit Hilfe von Lagrange'schen Multiplikatoren in „echte“ (konsistente) Gewichtskoeffizienten und sogenannte Substitutionsratentransformiert (siehe Abb. 2, 3, 4, 5).

Wichtigungen - Substitutionsraten - Wichtigkeiten für das Problem LBW01n

Zurück Dateien Normierung Wichtigungen

Umrechnen ← variable Substitutionsraten → Umrechnen

Kriterium	Wichtung	Substitutionsraten		Wieviel ist Zeile wichtiger als Spalte?									
		KHVar	KG	Gr/max	Gw/min	@20m/min	Bal/max	SHH/max	RB/max	LS/max	SU/max	SW/max	@6Min/max
Gr/max	0,0210267	13,202425	122	1	2	0,1	0,25	0,25	0,5	0,125	0,125	0,125	0,1666667
Gw/min	0,0160138	17,466581		0,5	1	0,0333333	0,1666667	0,1666667	0,25	0,1	0,1	0,1	0,125
@20m/min	0,2797098		1	10	12	1	6	6	8	2	2	2	4
Bal/max	0,0448028	6,2290218		4	6	0,1666667	1	1	2	0,25	0,25	0,25	0,5
SHH/max	0,0324743	8,8131702		4	6	0,1666667	1	1	2	0,1666667	0,1666667	0,1666667	0,25
RB/max	0,0294284	9,5046534		2	4	0,125	0,5	0,5	1	0,1666667	0,1666667	0,1666667	0,25
LS/max	0,1643818	1,7015675		8	10	0,5	4	6	8	1	1	1	2
SU/max	0,1643818	1,7015675		8	10	0,5	4	6	8	1	1	1	2
SW/max	0,1643818	1,7015675		8	10	0,5	4	6	8	1	1	1	2
@6Min/max	0,0933007	3,2677957		6	6	0,25	2	4	4	0,5	0,5	0,5	1

Normierung auf Null-Eins
schlechteste Ausprägung auf Null, beste Ausprägung auf Eins bei Gr, Gw, 20m, Bal, SHH, LS, SU, SW

Konsistente Gewichte

Bei RB: Wunschwert bei 1,5, Ausprägung = 1, kleinster Wert mit Ausprägung 0,5 größter Wert mit Ausprägung 0,5

Bei 6Min: Wunschwert bei 775, Ausprägung = 1, kleinster Wert mit Ausprägung 0,7 größter Wert mit Ausprägung 0,3

Kriterium	Toleranz-/Fehlerangaben
Gr/max	0,005
Gw/min	0,5
@20m/min	0,05
Bal/max	1
SHH/max	1
RB/max	0,5
LS/max	1
SU/max	1
SW/max	1
@6Min/max	10

Abbildung 2: Wichtungskoeffizienten für die Tests des DMT sowie Körpergröße und Körpergewicht

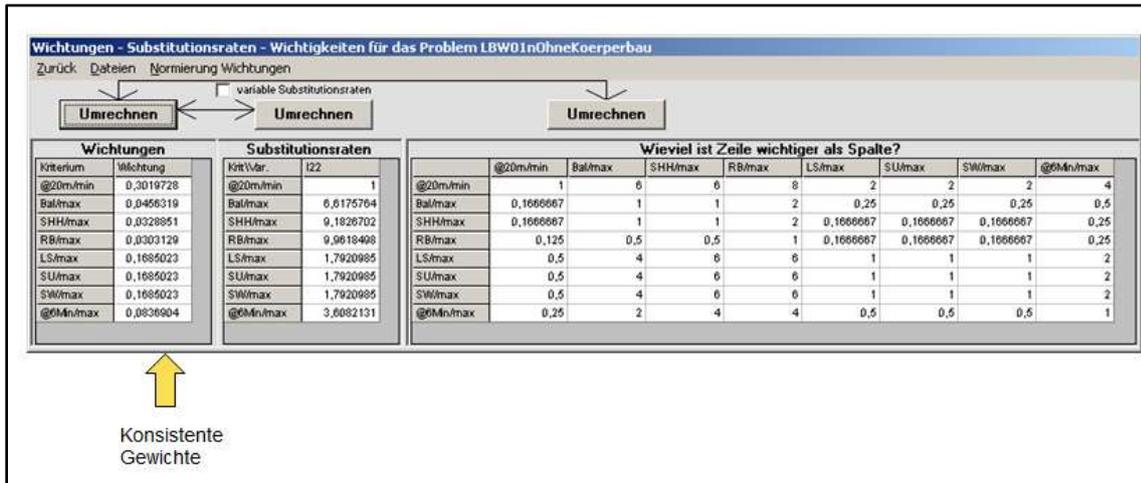


Abbildung 3: Wichtungskoeffizienten für die Tests des DMT

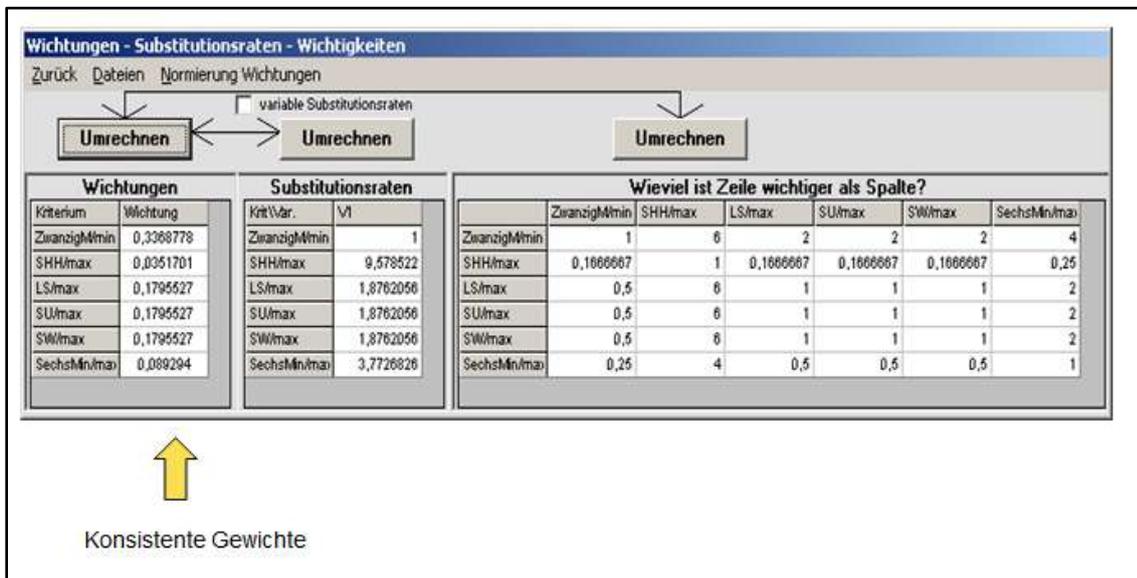


Abbildung 4: Wichtungskoeffizienten für die „konditionellen“ Parameter

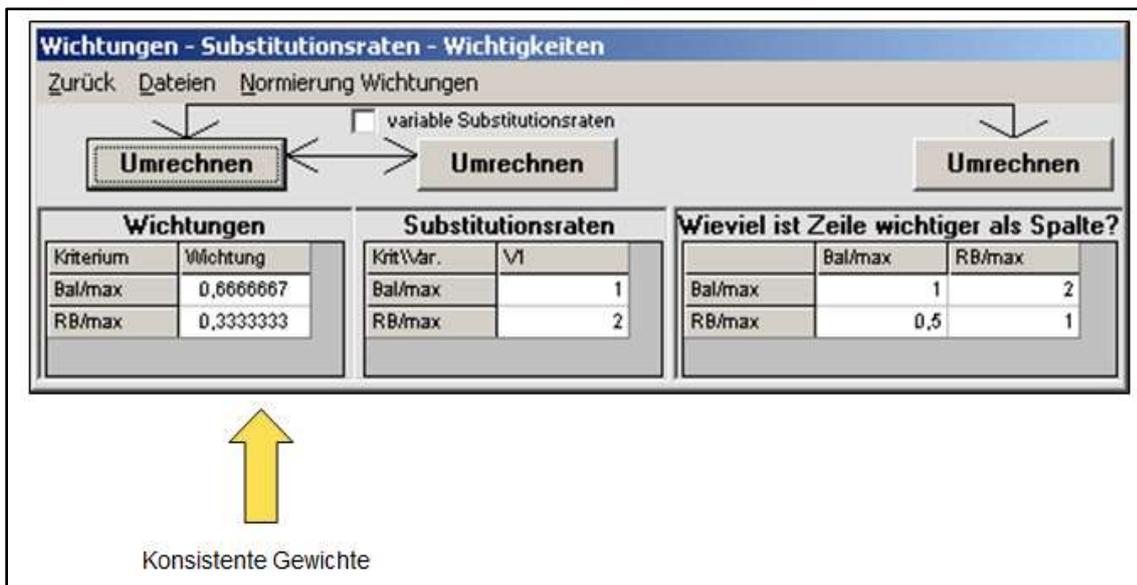


Abbildung 5: Wichtungskoeffizienten für die „koordinativen“ Parametern

Exkurs 1

Die Matrix der Gewichtungsverhältnisse zwischen den einzelnen Kriterien muss unabhängig von den gewählten Maßeinheiten sein. Wenn der Parameter „20-Meter-Lauf“ z. B. 6 mal so wichtig ist wie der Parameter „Seitliches Hin- und Herspringen“, dann bezieht sich das auf den Inhalt der beiden Merkmale. Man kann also die berechneten Gewichtungen aus der o. g. Matrix nur benutzen, wenn die Kriteriumsausprägungen dimensionslos dargestellt werden, also auf das Intervall von 0 bis 1 normiert bzw. standardisiert werden (0 für die schlechteste und 1 für die beste Ausprägung).

Exkurs 2

Als schlechteste resp. beste Ausprägung gilt per definitionem: Die schlechteste/beste innerhalb der Stichprobe oder auch die **denkbar** schlechteste/beste. Dadurch ist eine Möglichkeit gegeben, Erfahrungen aus Untersuchungen früherer Jahre oder anderer Stichproben „mitzunehmen“: Der beste zum Beispiel im 6-Minuten-Lauf des aktuellen Jahrgangs muss also nicht den Wert 1 erhalten, sondern wird abgestuft, weil es in anderen Jahren Bessere gab.

Aus den konsistenten Gewichtungen werden auf folgende Weise die notwendigen Substitutionsraten berechnet:

- die Kriterien werden auf Null-Eins normiert (dabei werden die Messfehler ebenfalls transformiert!),
- auf die normierten Ausprägungen werden die Gewichtungsverhältnisse angewendet - daraus die Gewichtungen und daraus wiederum die Substitutionsraten berechnet. Eine Substitutionsrate gibt an, um wie viel Einheiten ein Merkmal besser sein muss, um einen definierten Verlust eines anderen Merkmals zu kompensieren („Naturalpreis“), wenn die gleiche „Güte“ erhalten werden soll.

Für die weiter unten – neben der Methode der unscharfen Dominanzmengen – zusätzlich genutzten Verfahren ist ergänzendes, teilweise auch anderes Expertenwissen notwendig – z. B. Angaben zu VETO- bzw. INDIFFERENZSCHWELLEN (Ester, 2014; Zinner, 1997).

Selbstverständlich kann das von uns bisher sehr pragmatisch vorgenommene und in Abbildung 1 dokumentierte Expertenrating sozusagen nur als „Anfangsiteration“ gesehen werden und muss künftig im Sinne von erweiterten und empirisch begründeten Iterationsschritten konkretisiert werden. Die entsprechenden Auswirkungen auf die Rangfolgen können dann umgehend verifiziert werden.

Exkurs 3

Die Erzeugung der Zugehörigkeitsfunktionen für die Probanden und damit deren Einordnung in die Rangfolge basiert auf (exakten) mathematischen/geometrischen Verfahren und stellt deshalb eine „scharfe“ Lösung dar. Die Unschärfe resultiert aus unserer Unsicherheit, weil wir nur schwer abschätzen können, auf welche Weise sich die „Gesamt-Güte“ aus den „Einzel-Güten“ ergibt. Ausdruck dieser Unsicherheit sind die im Expertenrating (diskursive Validierung) zu schätzenden Fehlerbereiche, Wichtungsverhältnisse usw. Je realitätsnäher diese Abschätzungen getroffen werden können, desto schärfer gilt die Rangfolge! Die Bildung des Gesamturteils nach den

Prinzipien des DMT (Bös et al., 2009, S. 60/61) und auch unsere in Teil 1 und 2 beschriebene Vorgehensweise zur Ermittlung einer Gesamteinschätzung verzichten völlig auf solche Zusatzinformationen und bilden „lediglich“ den Mittelwert (die lineare Transformation) der entsprechenden Standardnoten oder Z-Werte der Einzeltests.

3.3 Modellversuch Lichtenberger Mädchen

Auf der Basis dieser mehrkriteriellen FUZZY-Vorgehensweise wurden auf der Datenbasis der Lichtenberger Mädchen von 2014 ($n = 532$) vier Analysen durchgeführt und damit auch vier Rangfolgen ermittelt:

- Analyse mit 10 Parametern (alle Tests des DMT plus Körperhöhe [KH] und Körpergewicht [KG]),
- Analyse mit 8 Parametern (alle Tests des DMT)
- Analyse nur mit den sechs ermittelten „konditionellen“ Parametern (DMT außer „Balancieren rückwärts“ [Balrw] und „Rumpfbeugen“ [RB])
- Analyse nur mit den 2 „koordinativen“ Parametern (Balrw und RB).

Andererseits wurden folgende Klassifizierungsverfahren auf der Grundlage der herkömmlichen Methoden der Bildung von Leistungsklassen (Bös et al., 2009) durchgeführt:

- Klassifizierung nach den Leistungsklassen von Bös (auf der Stichproben-Datenbasis von Bös – 8 Items gemittelt) - im Weiteren bezeichnet mit **LK**
- Klassifizierung nach der im Teil 2 von Büsch und Zinner favorisierten Bildung der Leistungsklassen auf der Basis der „Rasch-bereinigten Mittelwerte“ (6 Items auf der Datenbasis Berlin 2012/13) – im Weiteren bezeichnet mit **Lktgesk.**
- Klassifizierung nach den Leistungsklassen von Bös (8 Items gemittelt, aber Datenbasis von Berlin) – im Weiteren bezeichnet **Lktgesb.**

Hinweis: Bei den beiden letzten Klassifizierungsverfahren schafft es kein Mädchen aus der Lichtenberger Stichprobe in die auf der Basis der $N = 2.288$ Berliner Mädchen 2012/2013 ermittelten Leistungsklasse 1 bzw. Leistungsklasse 5 zu kommen (siehe Anmerkungen am Schluss des Textes).

Tabelle 1: Rangreihen und Leistungsklassen für die jeweils „besten“ und „schlechtesten“ 20 Probandinnen

		FUZZY-Rangfolgen			„Bös“-Leistungsklassen		
RF 8	ID	RF 10	RF 6	RF 2	LK	Lktgesk	Lktgeskb
1	I184	1	3	118	4	4	4
2	I117	3	1	236	5	4	4
3	I1004	2	2	147	4	4	4
4	I1017	5	11	3	5	4	4
5	I118	4	4	323	4	4	4
6	I1092	6	7	164	4	4	4
7	I1023	7	5	459	5	4	4
8	I185	8	6	481	5	4	4
9	I22	11	10	322	5	4	4
10	I1094	9	8	373	5	4	4
11	I1006	10	20	13	4	4	4
12	I341	17	16	60	4	4	4
13	I495	12	17	69	5	4	4
14	I394	16	15	120	4	4	4
15	I596	20	19	233	5	4	3
16	I41	13	12	490	4	4	4
17	I1022	15	14	156	4	4	3
18	I1178	14	24	177	5	4	4
19	I546	19	18	264	4	3	3
20	I261	21	21	306	5	4	3

		FUZZY-Rangfolgen			„Bös“-Leistungsklassen		
RF 8	ID	RF 10	RF 6	RF 2	LK	Lktgesk	Lktgeskb
513	I793	486	488	469	2	3	3
514	I292	518	525	295	3	3	3
515	I1024	524	515	289	3	3	3
516	I178	479	467	466	3	3	3
517	I962	520	504	238	1	2	2
518	I755	516	511	336	2	3	3
519	I1152	505	526	342	2	3	3
520	I74	515	432	496	2	3	3
521	I458	513	497	448	2	3	3
522	I476	519	516	413	2	2	2
523	I561	529	496	491	2	2	3
524	I472	522	523	172	1	2	2
525	I615	517	430	530	2	3	3
526	I490	527	503	485	1	2	2
527	I794	525	494	516	1	2	2
528	I644	526	482	521	1	3	2
529	I477	528	518	529	1	2	2
530	I475	530	529	522	1	2	2
531	I83	531	524	528	1	2	2
532	I119	532	530	532	1	2	2

3.4 Vergleich der verschiedenen Klassifizierungsverfahren mit den unterschiedlichen FUZZY-Rangreihen

Um den Vergleich bei der relativ großen Datenbasis anschaulich zu gestalten, wurden (zunächst) jeweils nur die 20 „besten“ bzw. 20 „schlechtesten“ Probandinnen bezüglich der verschiedenen Rangfolgen und Klassifizierungsverfahren verglichen. Eine Übersicht dazu zeigt die Tabelle 1, bei der die Rangfolge über alle zehn Parameter als Basis verwendet wurde. Die Tabelle lässt erkennen:

- dass die Rangfolge auf der Grundlage aller zehn Parameter (8 motorische Parameter plus KH und KG – Spalte RF 10) sehr gut mit der Rangfolge auf der Grundlage der im DMT verwendeten 8 alleinigen motorischen Parametern übereinstimmt (Spalte LK). Für die Ermittlung der Rangfolge auf diese unscharfe Weise sollten also auch in Zukunft alle 10 Parameter genutzt werden, weil damit eben auch ein „Mehr“ an Informationen in die Analysen einbezogen wird,
- dass die in der Rangfolge „besten“ 20 Mädchen auch allesamt (Ausnahme: ID 546) in die höchste Leistungsklasse LK4 bzgl. des „Rasch-bereinigten“ Summenwertes eingeordnet werden (Spalte Lktgesk),
- dass die in der Rangfolge „schlechtesten“ Mädchen jeweils auch in die schlechteste (10x) bzw. in die zweitschlechteste (in diesem Fall mittlere) Leistungsklasse eingeordnet werden (Spalte Lktgesk),
- dass die Rangfolge nur auf der Grundlage der beiden (koordinativen) Parameter (RB, Balrw - Spalte RF 2) etwas „ganz anderes“ als alle anderen Rangfolgen ist. Dieses ist verständlich – drücken diese beiden Merkmale ja auch ganz andere sportliche Fähigkeiten aus. Das ist dann also ein weiterer Beleg für die Herausnahme dieser beiden Testwerte aus der Vorgehensweise zur Ermittlung eines Gesamtwertes.

Den Vergleich der **absoluten Messwerte** der in der Rangfolge mit allen zehn Parametern drei „besten“ bzw. „schlechtesten“ Schülerinnen zeigt die Tabelle 2. Auch diese bestätigt schließlich die Plausibilität des Fuzzy-Vorgehens.

Tabelle 2: Vergleich der 3 „Besten“ mit den 3 „Schlechtesten“

ID		Gr/max	Gw/min	20m/min	Bal/max	SHH/max	RB/max	LS/max	SU/max	SW/max	6Min/max
184,00	1184	1,35	27,50	4,05	43,00	36,00	4,00	16,00	27,00	144,00	766,00
117,00	1117	1,34	34,20	4,07	32,00	41,50	4,00	17,00	25,00	160,00	853,00
1004,00	11004	1,42	36,50	4,19	35,00	24,00	0,10	19,00	32,00	143,00	748,00
475,00	1475	1,26	22,00	5,09	17,00	20,00	-11,00	7,00	10,00	79,00	549,00
83,00	183	1,39	35,30	5,83	5,00	11,00	-11,50	6,00	13,00	90,00	658,00
119,00	1119	1,42	55,10	6,01	7,00	9,50	-20,00	9,00	9,00	61,00	654,00

Da das Grundproblem der Rangreihe ein multiattributives Entscheidungsproblem ist, kann man natürlich auch andere bewährte Methoden der Multiattributiven Entscheidungstheorie anwenden. Deren Grundgedanken lassen es problemlos zu, auch diese als unscharfe Methoden zu bezeichnen. Im Laufe der Zeit sind eine Menge solcher Methoden entwickelt worden, die z. B. von unterschiedlichen Ausgangsinformationen (Gewichtungen, Gewichtungsverhältnisse, Substitutionsraten, Unsicherheiten, Mess- und Bewertungsfehler, Empfindlichkeiten usw.) ausgehen und somit auch unterschiedliche Seiten des Entscheidungsprozesses berücksichtigen. Da es nicht sinnvoll ist, ein Ranking dieser Methoden untereinander durchzuführen, wird hier der folgende Standpunkt vertreten:

Man benutzt die gängigsten, verständlichsten Methoden und natürlich nur solche, für die man die Ausgangsinformationen hat, berechnet jeweils die Rangfolgen, um diese danach zu vergleichen und auszuwerten.

Unter dieser Prämisse wurden folgende weitere Methoden auf den vollständigen Datensatz (10 Parameter) angewendet:

- der Analytische Hierarchieprozess,
- die Rückwärtsfilterung (mit $p = 1$, $p = 2$, $p = 20$),
- das Promethee-Verfahren (mit Kerngrad, mit net-flow),
- die unscharfe Güte (ohne und mit faktorieller Bewertung).

Diese Rangfolgen lassen sich mit der Auswertungssoftware MAOE (Modellieren, Analysieren, Optimieren, Entscheiden) (siehe ESTER, 2013) relativ schnell miteinander vergleichen (Abb. 6).

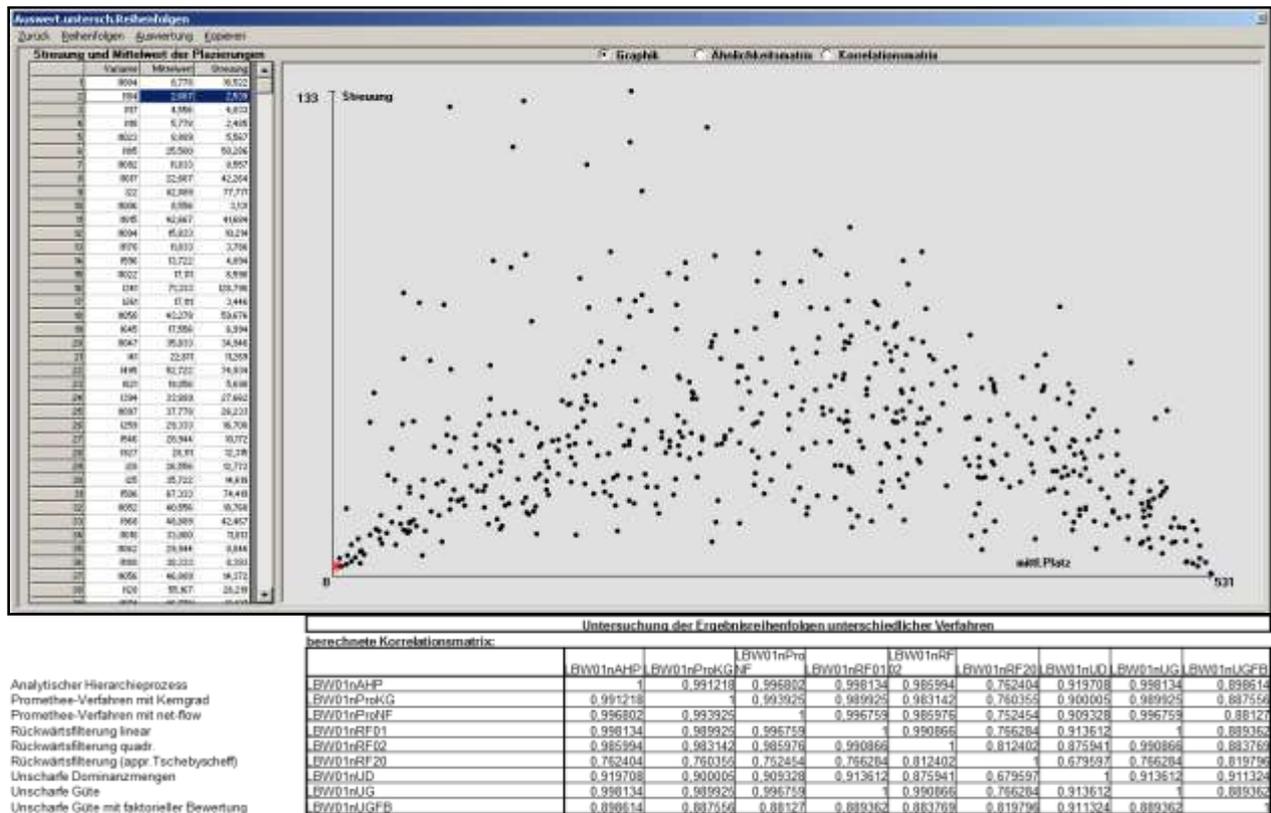


Abbildung 6: Vergleich verschiedener Verfahren/Ähnlichkeitsmatrix

Die Ähnlichkeitsmaße liegen dabei zwischen 0,9 und 0,7 (1 bedeutet exakt die gleiche Reihenfolge - 0 bedeutet die genau umgekehrte Reihenfolge).

Es zeigt sich, dass die verschiedenen Rangfolgen unserer Modellrechnungen sehr ähnlich sind. Zur Nutzung des Informationsgehaltes aller unterschiedlichen Rangfolgen berechnet man nun den mittleren Platz eines jeden Kandidaten bzgl. aller Rangfolgen und die Streuung seiner Platzierung.

Die Grafik Streuung/Mittlerer Platz (in Abbildung 6) zeigt, dass sich für niedrige und hohe Platzziffern geringe Streuungen ergeben. Mittlere Platzziffern sind dagegen mit hohen Streuungen verbunden. Da wir aber besonders mit der Gruppe der Besten bzw. der Schlechtesten arbeiten wollen, kommt uns diese Situation sehr entgegen. Alle Verfahren ordnen also gute und schlechte Kandidaten als solche auch ein - bei mittleren Kandidaten ist die Einordnung unsicherer. Ursache dafür kann sein, dass alle Kandidaten mit guten Ausprägungen bei vielen Merkmalen vorn liegen, alle mit schlechten Ausprägungen bei vielen Merkmalen hinten liegen, während bei Kandidaten im Mittelfeld sich gute und schlechte Ausprägungen die Waage halten. Da die unterschiedlichen Methoden die einzelnen Merkmalsunterschiede auch unterschiedlich stark berücksichtigen, kommt es dort auch zu den größeren Streuungen – was aus o. g. Gründen aber nicht wesentlich stört.

Man kann nun die mittlere Platzierung als Ergebnis für eine weitere Rangfolgenbildung benutzen. Damit lässt sich dann der Kandidat bestimmen, der unter den **verschiedensten** multikriteriellen Gesichtspunkten am besten abschneidet, ebenso der zweitbeste und so weiter. Den Vergleich der so entstandenen Rangfolge auf der Basis der mittleren

Platzierungen in allen Rangfolgen mit der Rangfolge bzgl. der zehn Parameter für die ersten 20 Probanden zeigt Tabelle 3.

Tabelle 3: Rangfolge mittlere Platzierungen

Alle Merkmale					
Unsch.Dom.mengen			Mittlere Platzierung		
Platz	Variante	Bewertung	Platz	Variante	Bewertung
1	I184	1	1	I184	1
2	I1004	0,8904473	2	I117	0,9964308
3	I117	0,8565215	3	I118	0,9941214
4	I118	0,8125951	4	I1006	0,9888726
5	I1017	0,8037129	5	I1004	0,9884527
6	I1092	0,741695	6	I1023	0,9882427
7	I1023	0,7392524	7,5	I1178	0,982679
8	I185	0,7223457	7,5	I1092	0,982679
9	I1094	0,6950548	9	I596	0,9791098
10	I1006	0,6645828	10	I1094	0,9751207
11	I22	0,6384205	11,5	I261	0,9727063
12	I495	0,5985121	11,5	I1022	0,9727063
13	I41	0,5891997	13	I645	0,9718665
14	I1178	0,586849	14	I821	0,9690321
15	I1022	0,5820312	15	I41	0,9623137
16	I394	0,5759453	16	I1017	0,9622087
17	I341	0,565884	17	I185	0,9568549
18	I927	0,5607076	18	I28	0,9548604
19	I546	0,5606284	19	I927	0,9519211
20	I596	0,5603668	20	I546	0,9503464

Man erkennt, dass die Ergebnisse in diesem Vergleich sehr ähnlich sind, nur zwei Probanden fallen aus der Rangfolge raus, zwei kommen rein. Auf diese Weise kann man die Sicherheit der Rangreihe-Aussagen immer stärker stabilisieren! Bezieht man nach diesem Schema auch die Erkenntnisse und Ergebnisse der herkömmlichen Verfahren nach Bös et al. (2009) ein, kann das nur zu einer weiteren Erhöhung der Sicherheit der Klassifizierung führen.

Noch eine Bemerkung zum Verfahren „Analytischer Hierarchieprozess (AHP):

Hierarchische Methoden bieten die Möglichkeit, die Merkmale zu strukturieren, auf jeder Ebene eine gesonderte Bewertung durchzuführen und die Gesamtbewertung besser an das mögliche Ziel anzupassen.

In der folgenden Abbildung 7 ist eine solche „Bewertungshierarchie“ dargestellt.

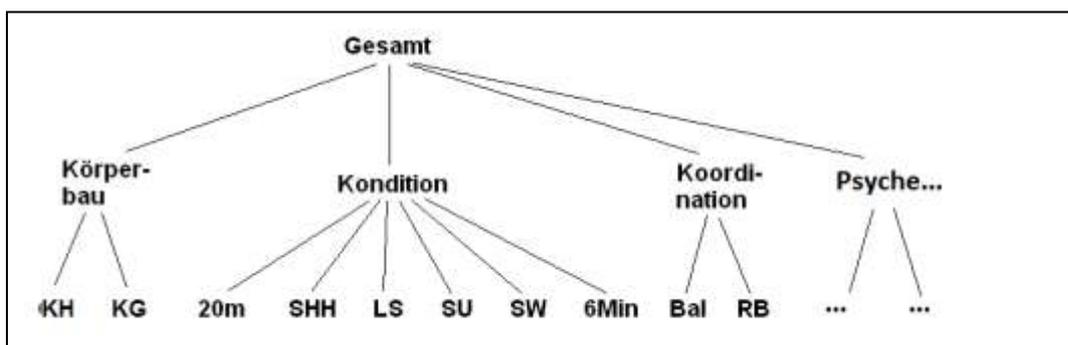


Abbildung 7: Bewertungshierarchie

Man kann also nach jedem der drei „Zwischenmerkmale“ bewerten, ordnen und klassifizieren (wie geschehen). Die Bewertungen der Zwischenmerkmale können dann herangezogen werden, um eine Gesamtbewertung, Ordnung und Klassifikation durchzuführen (das heißt, jede der Zwischenbewertungen wird wieder gewichtet). Dabei kann diese Zusammenfassung sogar mit unterschiedlichen Prioritäten geschehen, da ja in den unterschiedlichen Sportarten Kondition und Koordination unterschiedlich in Anspruch genommen werden. Falls es inhaltlich notwendig ist, kann die Zahl der Bewertungsebenen auch größer als drei sein und es ist auch nicht zwingend, dass jedes Merkmal nur genau einem übergeordneten Merkmal zugeordnet werden muss. Auch kann ein Merkmal die nächste Ebene überspringen, weil es direkt der übernächsten usw. Ebene zuzuordnen ist.

3.5 Schlussfolgerungen

Eine abschließende Beurteilung der Effizienz von Rangreihen nach unscharfen FUZZY-Methoden für die Auswertung des DMT wäre sicher verfrüht. Es gelingt jedoch mit diesem Vorgehen, Testprofile zu analysieren, subjektive Erfahrungen und Bewertungen einzubringen und integrale Urteile zu bilden. Sie bieten darüberhinaus den Vorteil, interaktiv zu arbeiten. Das heißt, die Auswirkungen von Veränderungen der Gewichte oder Bewertungsfunktionen auf das Ergebnis in der Rangfolge lassen sich (unter Nutzung eines entsprechenden Programmsystems, siehe Ester, 2014) schnell „durchspielen“. Das hat den positiven Nebeneffekt, dass die „Experten“ sich gründlich mit den Daten auseinandersetzen, oft deutliche Differenzen zwischen ihnen diskutieren müssen und deshalb meist eine viel bewusstere Position zu den letztlichen Analyseergebnissen einnehmen als bei der Anwendung von „starr“ Verfahren, bei denen man die Ergebnisse entweder annimmt oder verwirft. Auch die Möglichkeit zur Nutzung einer Vielzahl unterschiedlicher Vorgehensweisen zur Rangreihenbildung, die alle ja auch unterschiedliche Ergebnisse liefern, sollte nicht Grund zu Unbehagen sein, sondern zeigt ein reales Bild von der Wirklichkeit. Sie ist beliebig komplex und für komplexe, komplizierte Geschehnisse gibt es immer eine Vielzahl von Modellen, die dann auch widersprüchlich sein können. Dabei lohnt es sich nicht, zwischen den Modellen zu schwanken und sich für genau eines entscheiden zu wollen. Vielmehr sollte man die in der Modellvielfalt steckenden Informationen zur Findung neuer, modifizierter Ansichten über die Realität nutzen und konstruktiv diskutieren.

In diesem Sinne soll die charakterisierte FUZZY-Vorgehensweise natürlich nicht als Alternative zu den bisherigen Analyseschritten im DMT, sondern als Erweiterung und Ergänzung dieser Vorgehensweisen gesehen werden.

Exkurs 4

Im „Zwischenbericht für das Projekt „Berlin hat Talent (Datenbasis 2012/2013)“, Teil 2 wurde eine nach spezifische Vorgehensweisen zur Ermittlung von Summenwerten aus den Items des Deutschen Motoriktests (Rasch-bereinigte Summenwerte auf der Datenbasis Berlin 2012 und 2013) vorgestellt. Der Vergleich dieser Methode (gemittelte Berliner (empirische) Leistungsklassen für einen Konditionsfaktor (6 Items)) mit dem standardisierten Vorgehen (mittlere Kategorie der Normquintile von 8 Items) zeigt Tabelle 4.

Tabelle 4: Vergleich der gemittelten Berliner (empirischen) Leistungsklassen für einen Konditionsfaktor (6 Items) mit der mittleren Kategorie der Norm-Quintile (8 Items) für die 8- bis 10-Jährigen Mädchen (Lichtenberg 2014)

$\lambda^2 (FG = 8, N = 532) = 339,66, p < 0,001, CI = 0,57$

• **Summenwert**

Zusammengefasste mittlere Kategorie nach Bös * Gemittelte Kategorie für die 6 Konditions-Items Kreuztabelle^a

Anzahl

		Gemittelte Kategorie für die 6 Konditions-Items			Gesamt
		2	3	4	
Zusammengefasste	1	10	7	0	17
mittlere Kategorie nach	2	20	80	1	101
Bös	3	3	241	4	248
	4	0	89	49	138
	5	0	4	24	28
Gesamt		33	421	78	532

a. Geschlecht = weiblich

Während nach der mittleren Kategorie der Normquintile 17 Probandinnen der Lichtenberger Mädchen in die LK 1 und 28 Mädchen in die LK5 kommen, schafft es nach der Methode der Rasch-bereinigten Summenwerte auf der Datenbasis Berlin 2012/2013 keine Schülerin in die höchste bzw. niedrigste Leistungsklasse. Diese beiden Leistungsklassen bleiben also leer. Somit ist die von den Mädchen erreichte höchste (beste) Leistungsklasse die Leistungsklasse 4, die niedrigste (schlechteste) die Leistungsklasse 2.

Das ist zunächst zwar sehr eigenartig, aber evtl. zumindest insoweit gegeben, dass eben kein Mädchen öfter als drei Mal nur die (schlechteste) Leistungsklasse 1 erreicht. Das wiederum ist zunächst positiv, weil sich damit das Gesamtniveau gegenüber 2012/13 als verbessert zeigt. Andererseits ist jedoch auch kein Mädchen so gut, dass es öfter als drei Mal die (beste) Leistungsklasse 5 erreicht. Es gibt also auch keine wirklichen Spitzenleistungen bei den Lichtenberger Mädchen in diesem Jahrgang. Vielleicht liegt das am Stadtteil, vielleicht liegt es an Ausreißern (z. B. hat das Körpergewicht einen Gipfel bei 35 kg und einen bei 50kg), vielleicht ist es auch ein Artefakt, da wir uns an den Leistungsklassen 2012/13 orientieren – da muss demnach die Auswertung weiterer Bezirke Aufklärung bringen.

Bei der von uns favorisierten Vorgehensweise der Rasch-bereinigten Summenwerte landen die Schüler(innen) nur dann in der Leistungsklasse 1 (schlechteste Klasse), wenn sie einen Summenwert < 9 (bei 6 Items bedeutet dies, dass man bei mindestens 3 Items die Leistungsklasse 2 erreichen muss, um auch insgesamt in die Leistungsklasse 2 zu kommen) und in der besten Leistungsklasse einen Summenwert > 26 (bei 6 Items bedeutet dies, dass man bei mindestens 3 Items die Leistungsklasse 5 erreichen muss, um auch insgesamt in die Leistungsklasse 5 zu kommen) erreichen.

Zu berücksichtigen ist, dass es sich bei der Einteilung der Leistungsklassen um eine restriktive Festlegung und zugleich erste Näherung handelt, da im Gegensatz zu den Verfahren des DMT nicht die Rohdaten in Prozentränge umgerechnet wurden, sondern nur die Leistungsklassen mit Bezug auf die Berliner Daten (über die Grenzwerte der

Leistungsklassen) kategorisiert wurden. Wenn die Berliner Daten 2012/2013 als Bezugsgrundlage dauerhaft verwendet werden sollen (was man in Abstimmung mit Klaus Bös noch einmal diskutieren sollte), dann müssten noch aktualisierte Normwert-Tabellen, die auch die Ergebnisse aus anderen Studien mit dem DMT einbeziehen (z. B. Utesch et al., 2014) erstellt werden.

Tabelle 5: Vergleich der gemittelten Berliner (empirischen) Leistungsklassen für die 8 DMT-Items mit der mittleren Kategorie der Norm-Quintile (8 Items) für die 8- bis 10-jährigen Mädchen (Lichtenberg 2014)

$\lambda^2 (FG = 8, N = 532) = 390,04, p < 0,001, CI = 0,61$

• **Summenwert**
Zusammengefasste mittlere Kategorie nach Bös * Gemittelte Kategorie für die 8 DMT-Items nach den Berliner Kategorien Kreuztabelle^a

Anzahl

		Gemittelte Kategorie für die 8 DMT-Items nach den Berliner Kategorien			Gesamt
		2	3	4	
Zusammengefasste mittlere Kategorie nach Bös	1	12	5	0	17
	2	12	89	0	101
	3	0	244	4	248
	4	0	104	34	138
	5	0	6	22	28
Gesamt		24	448	60	532

a. Geschlecht = weiblich

Hinweis

Für die gemittelte Kategorie der 6 Konditions-Items mussten $n = 77$ 7-jährige, $n = 1$ 6-jährige und $n = 3$ 11-jährige gelöscht werden, da für diese keine Referenzdaten aus der Stichprobe 2012/2013 vorliegen.

Wenn anstelle der 6 Konditions-Items die 8-DMT-Items nach der Vorgabe von Bös gemittelt werden (siehe Tabelle 5), dann verändert sich das Gesamtbild aus der vorherigen Folie nur marginal. Dennoch empfiehlt es sich, mit den 6 Konditions-Items zu rechnen, da für diese die Rasch-homogene Eindimensionalität und nur dann die Verwendung eines Summenwerts angezeigt ist.

Literatur (Teil 1-3)

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

Büsch, D. & Strauß, B. (2005, August). *Qualitative differences in coordination tasks*. Vortrag. 11th World Congress of Sport Psychology der International Society of Sport Psychology (ISSP), Australian Psychological Society (APS) and College of Sport Psychologists (CoSP), Sydney, Australien.

Büsch, D. & Strauss, B. (2005). *Qualitative differences in performing coordination tasks*. *Measurement in Physical Education and Exercise Science*, 9 (3), 161-180.

Büsch, D., Strauß, B., Seidel, I. & Pabst, J. (2010, September). *Konfirmatorische Prüfung des Strukturmodells sportmotorischer Fähigkeiten mit einem mehrdimensionalen Rasch-Modell*. Vortrag. 8. gemeinsames Symposium der drei dvs-Sektionen Biomechanik, Sportmotorik und Trainingswissenschaft zum Thema "Bewegung und Leistung - Sport, Gesundheit & Alter", Hamburg.

Büsch, D., Strauß, B., Seidel, I., Pabst, J., Müller, L., Kretschmer, J., et al. (2008, Mai). *Die Konstruktvalidität auf Itemebene des allgemeinen sportmotorischen Tests für Kinder (AST 6-11)*. Vortrag. 40. Tagung der Arbeitsgemeinschaft für Sportpsychologie, Bern.

Büsch, D., Strauß, B., Seidel, I., Pabst, J., Tietjens, M., Müller, L., et al. (2009). *Die Konstruktvalidität des Allgemeinen Sportmotorischen Tests für Kinder*. *Sportwissenschaft*, 39 (2), 95-103.

Büsch, D., Strauß, B., Seidel, I., Pabst, J., Tietjens, M., Ziroli, S., et al. (2009, September). *Welche Konstrukte messen unterschiedliche motorische Testbatterien für das frühe Schulkindalter?* Vortrag. 19. Sportwissenschaftlicher Hochschultag der Deutschen Vereinigung für Sportwissenschaft, Münster.

Ester, J. (2014). *Ausgewählte Methoden des multiattributiven Variantenvergleichs*. Arbeitspapier.

Ester, J. & Zinner, J. (1996). *Multiattributive und FUZZY-Bewertungsmethoden in der Trainingswissenschaft*. In Krug, J. (Hrsg.), *Zeitreihenanalyse und „multiple“ statistische Verfahren in der Trainingswissenschaft*, Band 4. S. 119-136. Bundesinstitut für Sportwissenschaft.

Gagné, F. (2008). *Building gifts into talents. Brief overview of the DMGT 2.0*: Université du Québec.

Gagné, F. (2010). *Begabungen in Talente umsetzen. Kurze Übersicht über das differenzierte Modell von Begabung und Talent DMGT 2.0*. *Swiss Gifted*, 3 (1), 14-19.

Gagné, F. (2013). *The DMGT: Changes within, beneath, and beyond*. *Talent Development and Excellence*, 5 (1), 5-19.

- Hands, B. & Larkin, D. (2001). Using the Rasch measurement model to investigate the construct of motor ability in young children. *Journal of Applied Measurement*, 2 (2), 101-120.
- Hoffmann, A. & Pfützner, A. (2013). Leipziger Positionen zum Nachwuchsleistungssport in Deutschland. *Leistungssport* 43 (5), 5-9.
- Holzweg, M., Ketelhut, K. & Brandt, S. (2012). Der Einfluss des gewählten Bewertungsverfahrens auf die Interpretation der Ergebnisse des Deutschen Motorik Tests (DMT 6-18). *Sportunterricht*, 61 (8), 239-243.
- Linacre, J. M. (2000). New approaches to determining reliability and validity. *Research Quarterly for Exercise and Sport*, 71 (2), 129-136.
- 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.
- Poller, T. (2012). *Berlin hat Talent*. Vortrag. Runder Tisch des Spitzensports. Berlin.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2. Aufl.). Bern: Huber.
- Rowe, D. A. & Mahar, M. T. (2006). Validity. In T.M. Wood & W. Zhu (Hrsg.), *Measurement theory and practice in kinesiology* (S.9-26). Champaign, IL: Human Kinetics.
- Schlizio, F. & Kugel, R. (2013). *Berlin hat Talent – Das Netzwerk*. Vortrag. Runder Tisch des Spitzensports. Berlin.
- Strauß, B., Büsch, D. & Tenenbaum, G. (2006). Anwendungen probabilistischer Modelle in der Sportpsychologie. *Zeitschrift für Sportpsychologie*, 13 (2), 68-74.
- Strauss, B., Büsch, D. & Tenenbaum, G. (2007). New perspectives on measurement and testing in sport psychology. In G. Tenenbaum & R.C. Eklund (Hrsg.), *Handbook of sportpsychology* (3. Aufl., S.737-756). Boston, MA: Wiley.
- Strauss, B., Büsch, D. & Tenenbaum, G. (2012). Rasch modeling in sports. In G. Tenenbaum, R.C. Eklund & A. Kamata (Hrsg.), *Measurement in sport and exercise psychology* (S.75-80). Champaign IL: Human Kinetics.
- Strauss, B., Büsch, D., Tietjens, M., Seidel, I., Pabst, J. & Kretschmer, J. (2009, Juni). *IRT models and motor abilities*. Vortrag. Conference of the North American Society for the Psychology of Sport and Physical Activity (NASPSPA), Austin, Texas.
- Strauss, B., Tietjens, M. & Büsch, D. (2010, Oktober). *Diagnostik motorischer Fähigkeiten im Kindes- und Jugendalter*. Vortrag. Conference "Deutschland und Japan in der Globalisierung des Sports", Tokyo.
- Strauss, B., Tietjens, M. & Büsch, D. (2011b). Diagnostik motorischer Fähigkeiten im Kindes- und Jugendalter. In K. Hottenrott, O. Stoll & R. Wollny (Hrsg.), *Kreativität-Innovation-Leistung. Wissenschaft bewegt Sportbewegt Wissenschaft. 20. dvs-Hochschultag. Halle. 21-23. September 2011. Abstracts* (S.109). Hamburg: Czwalina.

- Tenenbaum, G., Strauss, B. & Büsch, D. (2007). Applications of the generalized Rasch models in sport, exercise and the motor domain. In M. Von Davier & C. Carstensen (Hrsg.), *Generalized Rasch models* (S.347-356). New York: Springer.
- Tietjens, M., Strauss, B., Utesch, T., Büsch, D., Ghanbari, M. & Seidel, I. (2013a, June). *Testing the general fitness of children: Construct validity of the German Motor Ability Test*. Vortrag. Symposium of the North American Society for the Psychology of Sport and Physical Activity: Sport and Exercise Psychology – Effects of acute exercise on cognitive performance: Understanding the role of potential moderators, New Orleans, LA.
- Tietjens, M., Strauss, B., Utesch, T., Büsch, D., Ghanbari, M. & Seidel, I. (2013b). Testing the general fitness of children: Construct validity of the German Motor Ability Test. *Journal of Sport & Exercise Psychology*, 35, S72-S73.
- von Davier, M. & Strauss, B. (2003). New developments in testing probabilistic models. *International Journal of Sport and Exercise Psychology*, 1, 61-81.
- Wood, T. M. & Zhu, W. (Hrsg.). (2006). *Measurement theory and practice in kinesiology*. Champaign: Human Kinetics.
- Yan, Z. & Bond, T. G. (2011). Developing a Rasch Measurement Physical Fitness Scale for Hong Kong Primary School-Aged Students. [doi: 10.1080/1091367X.2011.590772]. *Measurement in Physical Education and Exercise Science*, 15 (3), 182-203. doi: 10.1080/1091367x.2011.590772
- Zhu, W. (2001). An empirical investigation of Rasch equating of motor function tasks. *Adapted Physical Activity Quarterly*, 18, 72-89.
- Zhu, W. (2006). Constructing test using item response theory. In T.M. Wood & W. Zhu (Hrsg.), *Measurement theory and practice in kinesiology* (S.53-76). Champaign, IL: Human Kinetics.
- 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.
- Zinner, J. & Ester, J. (1996). *Fuzzy in der Datenanalyse. Berichte und Materialien des Bundesinstituts für Sportwissenschaft 1996/1*. Köln: Sport und Buch Strauß.
- Zinner, J., Ester, J., Pansold, B., Wolff, R. (1994) Zur Nutzung unscharfer (Fuzzy-) Bewertungsmethoden für die Auswertung leistungsdiagnostischer Untersuchungen. *Leistungssport* 24(4), 22-26
- Zinner, J. (1997). *Multiattributive und unscharfe (FUZZY-) Bewertungsmethoden zur Datenanalyse im Sport*. OSP Berlin.
- Zinner, J., Westphal, S. & Malz, J. (1999). Positionen des Olympiastützpunkts zur Entwicklung der Schwerpunktsportarten am Standort Berlin bis zum Jahr 2000. *Leistungssport* 29(1), S. 18-25.
- Zinner, J., Schwarz, P., Malz, J., Westphal, S. & Ledig, R. (2004). *Offensive um Qualität im Leistungssport*. *Leistungssport* 34(5), 18-22.

Zinner, J. (2007). *Neue Modelle braucht das Land*. Leistungssport 37(3), 25-28.

Zinner, J. (2013). *Langfristige Leistungsentwicklung in Berlin*. Vortrag. Präsidiumsklausur Landessportbund Berlin. Berlin.