

PISA 2006 Skalenhandbuch  
Dokumentation der Erhebungsinstrumente



# PISA 2006

## Skalenhandbuch

Dokumentation der Erhebungsinstrumente

herausgegeben von

Andreas Frey, Päivi Taskinen, Kerstin Schütte,  
Manfred Prenzel, Cordula Artelt, Jürgen Baumert,  
Werner Blum, Marcus Hammann,  
Eckhard Klieme, Reinhard Pekrun

unter Mitarbeit von

Regine Asseburg, Claus H. Carstensen, Barbara Drechsel,  
Timo Ehmke, Birte Harder, Marco Hoffmann, Mareike  
Kobarg, Gráinne Newcombe, Silke Rönnebeck, Katrin  
Schöps, Tina Seidel, Martin Senkbeil, Corinna Steenfatt,  
Oliver Walter, Jörg Wittwer



Waxmann 2009  
Münster / New York  
München / Berlin

### **Bibliografische Informationen der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-8309-2160-8

© Waxmann Verlag GmbH, Münster 2009

[www.waxmann.com](http://www.waxmann.com)

[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Christian Averbeck, Münster

Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,  
säurefrei gemäß ISO 9706

Alle Rechte vorbehalten

Printed in Germany

# Inhalt

1	Einleitung.....	7
1.1	Die Konzeption von PISA 2006 .....	7
1.2	Aufbau des Skalenhandbuchs .....	11
2	Testhefte.....	15
2.1	Lesekompetenz .....	17
2.2	Mathematische Kompetenz .....	18
2.3	Naturwissenschaftliche Kompetenz .....	20
2.4	Motivationale Orientierungen.....	25
2.5	Anstrengungsthermometer.....	29
3	Fragebögen .....	31
3.1	Internationaler Schülerfragebogen.....	33
3.2	Nationaler Schülerfragebogen.....	111
3.3	Internationaler Elternfragebogen.....	157
3.4	Nationaler Elternfragebogen .....	181
3.5	Internationaler Schulfragebogen .....	189
3.6	Nationaler Schulfragebogen.....	215
3.7	Nationaler Lehrerfragebogen.....	267
3.8	Weitere Indizes.....	301
	Literatur .....	369
	Register .....	373
	Abkürzungsverzeichnis.....	383



# 1 Einleitung

Mit dem vorliegenden Skalenhandbuch werden die bei der Hauptstudie des *Programme for International Student Assessment* (PISA) im Jahr 2006 in Deutschland eingesetzten Tests und Fragebögen umfassend dokumentiert. Das Handbuch berichtet statistische Kennwerte auf Item- und Skalenebene, die eine Beurteilung der Verfahren erlauben. Neben diesen statistischen Kennwerten werden die Items der eingesetzten Fragebögen im Wortlaut wiedergegeben. Das Skalenhandbuch kann somit bei der Konzeption neuer empirischer Studien helfen. Darüber hinaus wird die Berechnung der bei PISA 2006 verwendeten Indizes beschrieben (z. B. Indizes der sozialen Herkunft).

Durch die genaue Dokumentation der Instrumente können die in den Berichtsbänden des deutschen PISA 2006-Konsortiums (Prenzel et al., 2007, 2008) publizierten Ergebnisse nachvollzogen werden.

## 1.1 Die Konzeption von PISA 2006

PISA wird von der *Organisation für wirtschaftliche Zusammenarbeit und Entwicklung* (OECD) organisiert. Mit der Studie wird seit dem Jahr 2000 im Abstand von drei Jahren untersucht, wie gut fünfzehnjährige Schülerinnen und Schüler auf Herausforderungen der Wissensgesellschaft vorbereitet sind. Es werden jeweils Bildungsergebnisse in den Kompetenzbereichen Lesen, Mathematik und Naturwissenschaften in einer Vielzahl von Staaten erhoben und im Zusammenhang mit Merkmalen der sozialen und kulturellen Herkunft sowie des schulischen Lernumfeldes interpretiert. Zu jedem Erhebungszeitpunkt liegt der inhaltliche Schwerpunkt auf einem der drei genannten Kompetenzbereiche. Bei PISA 2006 waren dies die Naturwissenschaften. Eine ausführliche Beschreibung der theoretischen Rahmenkonzeption von PISA 2006 findet sich bei OECD (2006) und Prenzel et al. (2007).

### 1.1.1 Stichproben

Im Rahmen der Hauptstudie von PISA 2006 wurden in Deutschland Stichproben von Schülerinnen und Schülern, Eltern, Schulleitungen sowie Lehrkräften untersucht.

Für die *Schülerinnen und Schüler* können mehrere Stichproben unterschieden werden. Erstens wurden 4891 Fünfzehnjährige aus 226 Schulen für den internationalen Vergleich mit Fünfzehnjährigen aus anderen Staaten untersucht. Um klassenbezogene Auswertungen zu ermöglichen, wurden zweitens in den für den internationalen Vergleich gezogenen Schulen jeweils zwei komplette neunte Klassen per Zufall für die Untersuchung ausgewählt. Um Vergleiche zwischen den Ländern Deutschlands zu ermöglichen, wurden mit 37013 Fünfzehnjährigen sowie 40241 Schülerinnen und Schülern neunter Klassen aus 1509 Schulen zwei weitere Stichproben untersucht. In diesen Stichproben sind auch die für den internationalen Vergleich gezogenen Jugendlichen enthalten. Das genaue Vorgehen bei der Stichprobenziehung von PISA 2006 wird für den internationalen Vergleich bei Carstensen, Frey, Walter und Knoll (2007) sowie für den Vergleich der Länder Deutschlands bei Frey, Carstensen, Walter, Rönnebeck und Gomolka (2008) beschrieben.

Neben den Stichproben der Schülerinnen und Schüler gibt es analoge Stichproben für die *Eltern* aller teilnehmenden Jugendlichen, sodass die Größen der Elternstichproben denen entsprechen, die für die Schülerinnen und Schüler angegeben wurden.

Eine weitere Stichprobe umfasste die *Schulleitungen* aller an der Hauptstudie von PISA 2006 teilnehmenden Schulen. Entsprechend wurden für den internationalen Vergleich 226 Schulleiterinnen und Schulleiter und für den Vergleich der Länder Deutschlands 1509 Schulleiterinnen und Schulleiter untersucht.

Schließlich gibt es eine Stichprobe von *Lehrkräften*. In jeder der gezogenen Schulen wurden zufällig maximal sechs Mathematiklehrerinnen oder Mathematiklehrer sowie maximal acht Lehrerinnen und Lehrer ausgewählt, die in der Sekundarstufe I (Klassenstufe 5 bis 10) die Fächer Physik, Biologie, Chemie oder das Fach Naturwissenschaften als kombiniertes Fach unterrichteten. Für den internationalen Vergleich umfasst die Stichprobe 2194 und für den Vergleich der Länder Deutschlands 14660 Lehrerinnen und Lehrer.

Alle Angaben im vorliegenden Skalenhandbuch beziehen sich auf die Stichproben, die an den für den internationalen Vergleich gezogenen Schulen untersucht wurden, und somit auf 4891 Schülerinnen und Schüler, 4891 Eltern, 226 Schulleitungen und 2194 Lehrkräfte.

### 1.1.2 Untersuchungsmaterial

Bei der Hauptstudie von PISA 2006 wurden Testhefte sowie verschiedene Fragebögen eingesetzt. Die Vorgabe aller Instrumente erfolgte im Papier- und Bleistift-Format. Alle verwendeten Testaufgaben und Fragebogen-Items wurden im Jahre 2005 in einem Feldtest oder zu den vorherigen Erhebungszeitpunkten von PISA erprobt.

Die *Testhefte* enthielten verschiedene Elemente. Im Einzelnen handelt es sich dabei um kognitive Testaufgaben, um Fragen zur Messung motivationaler Orientierungen und um ein Instrument zur Messung der subjektiven Anstrengung während des Tests. Den Kern stellen dabei die kognitiven Testaufgaben dar. Sie dienen der Messung der Kompetenzen im Lesen, in der Mathematik und in den Naturwissenschaften. Die Zuordnung der Testaufgaben zu den verschiedenen Testheften wird im Kapitel 2 beschrieben. In allen an PISA 2006 teilnehmenden Staaten wurden die gleichen Testaufgaben in der jeweiligen Landessprache eingesetzt. Um den bei PISA 2006 erstmals als Schwerpunkt untersuchten Kompetenzbereich Naturwissenschaften angemessen abzubilden, konstruierte das internationale PISA-Konsortium im Vorlauf weitere Testaufgaben. Diese Testaufgaben wurden (neben einigen anderen zur Messung der Kompetenzbereiche Lesen und Mathematik) im Rahmen des Feldtests erprobt, um deren Güte zu überprüfen und um eine eindeutige Vergleichbarkeit zwischen den teilnehmenden Staaten zu gewährleisten. Auf Grundlage der dabei beobachteten Antworten wurde eine Itemanalyse auf Basis der Item-Response-Theorie (IRT; z. B. Hambleton & Swaminathan, 1985; Rost, 2004) durchgeführt. Hierbei wurden zentrale Kenngrößen wie Schwierigkeit, Trennschärfe, Itemfit und Häufigkeit fehlender Antworten betrachtet, eine Distraktorenanalyse durchgeführt sowie Differential Item Functioning (DIF) in Abhängigkeit der Variablen *Geschlecht* und *Staat* untersucht. Für die Hauptstudie von PISA 2006 wurden vom internationalen PISA-Konsortium nur Testaufgaben mit unproblematischen Kennwerten ausgewählt. Einzelne Testaufgaben wurden aufgrund der Ergebnisse des Feldtests leicht modifiziert. Bei der Hauptstudie kamen letztendlich für den Kompetenzbereich Naturwissenschaften diese aus dem Feldtest hervorgegangenen Testaufgaben sowie Testaufgaben der vorherigen Erhebungszeitpunkte zum Einsatz. Zur Messung der Kompetenzbereiche Lesen und Mathematik wurden ausschließlich Testaufgaben eingesetzt, die sich bereits bei den vorherigen Erhebungszeitpunkten in den Jahren 2000 und 2003 bewährt hatten. Auf Grundlage der bei der Hauptstudie von PISA 2006 beobachteten Antworten der Schülerinnen und Schüler



auf die kognitiven Testaufgaben wurden im Rahmen einer weiteren IRT-basierten Skalierung Testwertverteilungen berechnet. Eine differenzierte Beschreibung des Prozederes der Aufgabenentwicklung, der Aufgabenadaptation, der Aufgabenkalibrierung und der Skalierung wird bei OECD (2009a) gegeben.

Ferner enthielten die Testhefte Fragen zur Messung der motivationalen Orientierungen *Interesse an den Naturwissenschaften* und *Wertschätzung naturwissenschaftlicher Forschung*, die im Anschluss an ausgewählte Naturwissenschaftsaufgaben präsentiert wurden (vgl. OECD, 2009a). Bei diesen Fragen wurde eine vierstufige Ratingskala als Antwortformat verwendet. Die Auswertung erfolgte wie bei den kognitiven Testaufgaben auf Basis der IRT.

Am Ende aller Testhefte wurde ein Instrument zur Erfassung der subjektiven Anstrengung während des Tests vorgegeben. Dieses Instrument wird nachfolgend *Anstrengungsthermometer* genannt. Es enthält ein grafisches Element und zwei Fragen unter Verwendung einer 10-stufigen Antwortskala.

Auch bei den *Fragebögen* wurden seitens der OECD einheitliche Instrumente bereitgestellt, die in die jeweilige Landessprache übersetzt wurden. Mit diesen wurden umfangreiche Hintergrundinformationen zu den untersuchten Jugendlichen erhoben. Den Schwerpunkt bildeten dabei Informationen, die für das Lernen in den Naturwissenschaften bedeutsam sind. An der Entwicklung der internationalen Fragebögen waren Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Staaten beteiligt. In Zusammenarbeit mit dem internationalen PISA-Konsortium verfasste eine international zusammengesetzte Expertengruppe eine im Englischen *Contextual Framework* genannte Rahmenkonzeption (s. OECD, 2005c), die den theoretischen Hintergrund für die Konstruktion der Fragebögen bildete. Weitere Hintergrundinformationen zu den internationalen Fragebögen finden sich in OECD (2009a). Im Unterschied zu den kognitiven Tests konnten die internationalen Fragebögen auf nationaler Ebene durch zusätzliche Items oder durch komplette Instrumente ergänzt werden. International bereitgestellt wurden ein Fragebogen für Schülerinnen und Schüler, ein Fragebogen für Schulleiterinnen und Schulleiter (im Weiteren: Schulfragebogen) sowie ein Fragebogen für Eltern (der nicht in allen Staaten eingesetzt wurde). Alle drei Instrumente wurden in Deutschland durch zusätzliche Items ergänzt. Darüber hinaus kamen in Deutschland eigenständige nationale Fragebögen für die Schülerinnen und Schüler, die Eltern, die Schulleitungen und die Lehrkräfte (im Weiteren: Lehrerfragebogen) zum Einsatz. Mit diesen zusätzlichen Instrumenten wurden speziell auf das deutsche Bildungssystem bezogene Informationen erhoben, die nicht durch die internationalen Fragebögen abgedeckt waren. Alle bei der Hauptstudie von PISA 2006 verwendeten Items waren vorab im Feldtest oder zu den vorherigen Erhebungszeitpunkten erprobt worden. Items, die sich dabei nicht bewährt hatten, wurden in der Hauptstudie nicht verwendet oder in Einzelfällen in modifizierter Form vorgegeben. Bei Items, die gemeinsam eine Skala bilden, erfolgte eine Skalierung auf Basis der IRT mit dem Computerprogramm ConQuest (Wu, Adams, Wilson & Haldane, 2007). Die in den Berichtsbänden zu PISA 2006 für solche Skalen angegebenen Ergebnisse basieren auf Weighted Likelihood Estimates (WLEs; Warm, 1989) als Schätzer individueller Merkmalsausprägungen.

Zusammenfassend sind die folgenden sieben Fragebögen zu unterscheiden:

- Internationaler Schülerfragebogen
- Nationaler Schülerfragebogen
- Internationaler Elternfragebogen
- Nationaler Elternfragebogen
- Internationaler Schulfragebogen
- Nationaler Schulfragebogen
- Nationaler Lehrerfragebogen

### 1.1.3 Durchführung

Hinsichtlich der Durchführung ist zwischen Schülerinnen und Schülern, Eltern, Schulleitungen und Lehrkräften zu differenzieren.

Die Testungen der Schülerinnen und Schüler im Rahmen der Hauptstudie von PISA 2006 fanden in der Zeitspanne vom 18. April 2006 bis zum 29. Mai 2006 statt. Die für den internationalen Vergleich gezogenen Jugendlichen wurden an zwei aufeinander folgenden Tagen untersucht. Die Testung der für den Ländervergleich zusätzlich gezogenen Jugendlichen beschränkte sich auf den ersten Testtag. An diesem wurden den Schülerinnen und Schülern die kognitiven Tests und der internationale Schülerfragebogen zur Bearbeitung vorgelegt. Es nahmen damit alle bei PISA 2006 untersuchten Jugendlichen am Standardverfahren von PISA teil. Darüber hinaus wurde die Infrastruktur der PISA-Erhebung genutzt, um Testaufgaben zur Überprüfung der Bildungsstandards in Mathematik für den Mittleren Schulabschluss zu erproben. Diese Testaufgaben hatten die für den internationalen Vergleich gezogenen Jugendlichen an einem zweiten Testtag zu bearbeiten. Die hieraus resultierenden Ergebnisse finden sich bei Prenzel und Blum (2007). Ferner wurden am zweiten Testtag der nationale Schülerfragebogen sowie die Subskala *Figurale Analogien* des Kognitiven Fähigkeits-tests (KFT; Heller & Perleth, 2000) vorgegeben. Der Ablauf der Testungen an den beiden Testtagen ist in Tabelle 1.1.3-1 dargestellt. Die Testsitzungen dauerten inklusive zweier Pausen jeweils ungefähr 3,5 Zeitstunden.

Der internationale und der nationale Elternfragebogen wurden zu einem gemeinsamen Dokument zusammengefügt und den Eltern der gezogenen Schülerinnen und Schüler durch ihre Kinder zugestellt. Die Jugendlichen bekamen den Fragebogen am ersten Testtag mit der Bitte um Weitergabe an ihre Eltern ausgehändigt. Die ausgefüllten Elternfragebögen waren von den Schülerinnen und Schülern entweder bei der Testung am zweiten Testtag bei der Testleiterin oder beim Testleiter beziehungsweise bei der PISA-Koordinatorin oder dem PISA-Koordinator der Schule in einem verschlossenen Umschlag abzugeben.

Den Schulleitungen wurden bis spätestens 11 Tage vor Beginn der Testungen der internationale und der nationale Schulfragebogen zur Bearbeitung übergeben. Die Fragebögen wurden von der PISA-Koordinatorin oder dem PISA-Koordinator der betreffenden Schule in einem verschlossenen Umschlag in Empfang genommen und den Testleiterinnen oder Testleitern am Testtag übergeben.

Den für die Untersuchung ausgewählten Lehrkräften wurde bis spätestens 11 Tage vor Beginn der Testungen der nationale Lehrerfragebogen ausgehändigt. Auch hier wurden die ausgefüllten

Fragebögen von der PISA-Koordinatorin oder dem PISA-Koordinator der Schule in verschlossenen Umschlägen empfangen und den Testleiterinnen und Testleitern am Testtag übergeben.

*Tabelle 1.1.3-1:* Ablaufplan der Testungen von PISA 2006 an Testtag 1 und 2

<b>1. Testtag</b>		<b>2. Testtag</b>	
<b>Aktivität</b>	<b>Dauer</b>	<b>Aktivität</b>	<b>Dauer</b>
Verteilung des Materials, Einweisung	15-20 min	Verteilung des Materials, Einweisung	15-20 min
Bearbeitung des kognitiven Tests (PISA)	60 min	Bearbeitung des kognitiven Tests (Bildungsstandards Mathematik)	60 min
PAUSE	10 min	PAUSE	10 min
Bearbeitung des kognitiven Tests (PISA)	60 min	Bearbeitung des kognitiven Tests (Bildungsstandards Mathematik)	60 min
PAUSE und Austeilen der Elternfragebögen	15 min	Bearbeitung des kognitiven Fähigkeitstests (KFT)	ca. 10 min
Einweisung in die Bearbeitung des internationalen Schülerfragebogens	ca. 5 min	PAUSE	15 min
Bearbeitung des internationalen Schülerfragebogens	50 min	Einweisung in die Bearbeitung des nationalen Schülerfragebogens	ca. 5 min
Beenden der Testsitzung, Einsammeln der Materialien	3-5 min	Bearbeitung des nationalen Schülerfragebogens	30 min
		Beenden der Testsitzung, Einsammeln der Materialien	3-5 min

*Anmerkung.* min = Minuten.

## 1.2 Aufbau des Skalenhandbuchs

Der Hauptteil des vorliegenden Skalenhandbuchs gliedert sich in zwei Teile. Der erste widmet sich den Testheften, der zweite den Fragebögen.

### 1.2.1 Testhefte

Im Kapitel 2 werden Informationen zu den Testaufgaben zur Messung der Lesekompetenz, der mathematischen Kompetenz und der naturwissenschaftlichen Kompetenz, Fragen zur Messung der motivationalen Orientierungen Interesse an den Naturwissenschaften und Wertschätzung naturwissenschaftlicher Forschung und dem Anstrengungsthermometer dargestellt.

Bei PISA wird nur eine begrenzte Anzahl kognitiver Testaufgaben zu Illustrationszwecken öffentlich zugänglich gemacht. Aufgabenbeispiele in deutscher Sprache sind auf der Internetseite ‚<http://pisa.ipn.uni-kiel.de/>‘ sowie in englischer Sprache in OECD (2006, 2009b) zu finden. Der Großteil der Testaufgaben wird indes nicht veröffentlicht, um eine hohe Vergleichbarkeit über die Erhebungszeitpunkte für Veränderungsaussagen zu gewährleisten.

Für die kognitiven Testaufgaben und die Fragen zur Messung der motivationalen Orientierungen werden Ergebnisse der für den internationalen Vergleich von PISA 2006 durchgeführten IRT-basierten Skalierung berichtet. Die Ergebnisse sind dem technischen Bericht zu PISA 2006 entnommen (OECD, 2009a). Auf Aufgabenebene werden die in den OECD-Staaten beobachteten Lösungshäufigkeiten in Prozent (% korrekt) mit zugehörigen Standardfehlern ( $SE$ ) sowie die Itemschwierigkeiten ( $Delta$ ) angegeben. Bei Testaufgaben mit mehrkategoriellem Antwortformat wird als Itemschwierigkeit der mittlere Schwellenparameter berichtet (vgl. Wu et al., 2007). Auf Skalenebene wird die Anzahl der Testaufgaben ( $N_{Items}$ ), der Mittelwert ( $M$ ), die Standardabweichung ( $SD$ ) sowie die EAP/PV-Reliabilität der Skala berichtet. Für das Anstrengungsthermometer werden das grafische Element, der Wortlaut der Fragen sowie die Häufigkeiten der Antworten je Antwortkategorie angegeben.

## 1.2.2 Fragebögen

Im Kapitel 3 werden Informationen zu den eingesetzten Fragebögen sowie zu Indizes berichtet. Die Reihenfolge der Items im vorliegenden Skalenhandbuch orientiert sich an der Reihenfolge, in der sie in den Fragebögen präsentiert wurden. Einige Items aus international vorgegebenen Instrumenten wurden an die deutschen Gegebenheiten angepasst. So wurden beispielsweise die schulischen und beruflichen Abschlüsse der Eltern differenzierter als im internationalen Vergleich erfasst. Bei solchen Items berichtet das vorliegende Skalenhandbuch die nationalen Ergänzungen. Die berichteten Ergebnisse wurden ungewichtet berechnet.

Alle Items werden im vollen Wortlaut dargestellt. Während bei Items mit offenem Antwortformat darüber hinaus keine weiteren Angaben gemacht werden, werden bei Items mit anderen Antwortformaten zusätzliche Ergebnisse berichtet. Für Items mit dichotomem Antwortformat (zweistufig, z. B. „ja“, „nein“), im Mehrfachwahl-Antwortformat (multiple choice) und im Ratingskalen-Antwortformat (mehrstufig mit ordinaler Ordnung z. B. „stimme ganz zu“, „stimme eher zu“, „stimme eher nicht zu“, „stimme gar nicht zu“) werden die absoluten und die relativen Häufigkeiten der einzelnen Antwortkategorien in geeigneten Tabellen angegeben. Die jeweiligen Antwortmöglichkeiten werden unter den Tabellen genannt. Neben den jeweils gültigen Antwortalternativen werden drei Arten von fehlenden Werten unterschieden. Wenn auf Itemebene einzelne Antworten fehlen, gelten die Fragen als „nicht bearbeitet“. Derart fehlende Werte wurden mit der Ziffer 9 kodiert (bzw. bei mehrstellig kodierten Testaufgaben mit 99, 999 usw.). Fehlen hingegen für eine Person die Angaben aus einem Fragebogen vollständig, so gilt der Fragebogen als nicht vorgelegt und alle zugehörigen Fragen wurden als „nicht bearbeitbar“ kategorisiert (mit der Ziffer 7 bzw. bei mehrstellig kodierten Testaufgaben mit 97, 997 usw. kodiert). Schließlich wurden nicht zulässige Antworten, zum Beispiel Mehrfachantworten auf Ratingskalen, als „nicht gültig“ kategorisiert (mit der Ziffer 8 bzw. 98, 998 usw. kodiert).

Bei Ratingskalen, deren Fragen ein gemeinsames Konstrukt erfassen, werden darüber hinaus statistische Kennwerte berichtet. Vor der Berechnung dieser Kennwerte wurden invertierte Items entsprechend umkodiert. Es werden mit der Itemschwierigkeit ( $Delta$ ) und der Itemtrennschärfe ( $r_i$ ) zum einen zwei zentrale Itemkennwerte der Klassischen Testtheorie (KTT; z. B. Moosbrugger, 2007) angegeben. Um einschätzen zu können, wie gut einzelne Fragen den Annahmen des Skalierungsmodells entsprechen, werden zum anderen mit dem  $MNSQ$  und dem  $WMNSQ$  zwei IRT-basierte Statistiken für den Itemfit berichtet. Der Erwartungswert beider Statistiken beträgt 1, sodass geringe

Abweichungen von 1 auf eine gute Modellpassung des betreffenden Items hinweisen. Werte  $< 1$  weisen darauf hin, dass die Diskrimination des Items größer ist als vom Modell erwartet; Werte  $> 1$  weisen darauf hin, dass die Diskrimination des Items geringer ist als vom Modell erwartet. Somit sind insbesondere sehr hohe Itemfit-Werte als problematisch anzusehen. Weitere Angaben zu  $MNSQ$  und  $WMNSQ$  sind beispielsweise bei Wu et al. (2007) zu finden. Ferner werden für Ratingskalen, deren Items ein gemeinsames Konstrukt erfassen, Ergebnisse auf Skalenebene angegeben. Es werden der Mittelwert ( $M$ ) und die Standardabweichung ( $SD$ ) der Skalenrohwerte berichtet, wobei die Skalenrohwerte anhand des Mittelwerts über alle der Skala zugeordneten Items berechnet wurden (ggf. nach entsprechender Umkodierung invertierter Items). Zur Beurteilung der Zuverlässigkeit der Ratingskalen werden sowohl Cronbachs  $\alpha$  als häufig verwendeter KTT-basierter Reliabilitätsindex als auch die EAP/PV-Reliabilität (s. Wu et al., 2007) als IRT-basierter Reliabilitätsindex berichtet. Um die Belastbarkeit der berichteten Skalenkennwerte einschätzen zu können, werden außerdem die Anzahl der Items einer Skala ( $N_{\text{Items}}$ ), die Anzahl von Jugendlichen mit fehlenden Skalenwerten ( $N_{\text{fehlend}}$ ) und die Anzahl von Jugendlichen mit gültigen Werten ( $N_{\text{gültig}}$ ) genannt. Hierbei wurde ein gültiger Skalenwert berechnet, wenn für ein Individuum bei allen Items der Skala gültige Antworten vorlagen.

Im Anschluss an die Kapitel zu den einzelnen Fragebögen werden in einem gesonderten Kapitel 3.8 weitere Indizes berichtet, die sich thematisch zu Indikatoren der sozialen Herkunft, der Migration, zu naturwissenschaftsbezogenen Berufen, zum Unterricht, zu strukturellen Merkmalen der Schule und Lehrerkooperation gruppieren lassen. Diese Indizes werden aus einem oder mehreren Items durch Transformationen oder Verrechnungen gebildet. Neben der Beschreibung der Art der Berechnung wird der Wortlaut der in einen Index eingehenden Variablen oder Skalen und wenn möglich, Häufigkeitstabellen sowie Skalenkennwerte angegeben. Um die Nutzung des vorliegenden Skalenhandbuchs möglichst komfortabel zu gestalten, werden im Kapitel 3.8 zu den Indizes alle jeweils zugehörigen Items erneut mit vollständigen Informationen angeführt, auch wenn diese bereits zuvor in einem der Kapitel 3.1 bis 3.7 beschrieben wurden.

