

**Sibylle Elsäßer**

## **Komponenten von schulischen Leistungen**

Eine Analyse zu Einflussfaktoren auf die Notengebung in der Grundschule

Sibylle Elsässer

Komponenten von schulischen Leistungen. Eine Analyse zu Einflussfaktoren auf die Notengebung in der Grundschule

Dissertationen der LMU München

Band 26

# Komponenten von schulischen Leistungen

Eine Analyse zu Einflussfaktoren auf die Notengebung in der Grundschule

von  
Sibylle Elsässer

Herausgegeben von der  
**Universitätsbibliothek der Ludwig-Maximilians-Universität**  
Geschwister-Scholl-Platz 1  
80539 München

Mit **Open Publishing LMU** unterstützt die Universitätsbibliothek der Ludwig-Maximilians-Universität München alle Wissenschaftlerinnen und Wissenschaftler der LMU dabei, ihre Forschungsergebnisse parallel gedruckt und digital zu veröffentlichen.

Text © Sibylle Elsäßer 2018  
Erstveröffentlichung 2018  
Zugleich Dissertation der Universität zu München 2017

### **Bibliografische Information der Deutschen Nationalbibliothek**

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

Herstellung über:  
readbox unipress  
in der readbox publishing GmbH  
Am Hawerkamp 31  
48155 Münster  
<http://unipress.readbox.net>

Open-Access-Version dieser Publikation verfügbar unter:  
<http://nbn-resolving.de/urn:nbn:de:bvb:19-226439>

978-3-95925-092-4 (Druckausgabe)  
978-3-95925-093-1 (elektronische Version)

# Inhaltsverzeichnis

Abbildungsverzeichnis.....	IX
Tabellenverzeichnis.....	XI
Abkürzungsverzeichnis.....	XIII
1 Einleitung und Problemstellung.....	1
2 Zum Gegenstand schulischer Leistungsbeurteilung – theoretische Grundlagen .....	7
2.1 Der Leistungsbegriff .....	8
2.2 Zum schulischen Leistungsbegriff.....	12
2.3 Die Schulnote als Indikator für Schulleistung .....	17
2.4 Notengebung in der Schule als funktionale Handlung.....	22
2.5 Schulrechtliche Vorgaben der Notengebung.....	32
2.6 Zusammenfassung zentraler Annahmen .....	40
3 Zur messtheoretischen Güte von Schulnoten – theoretische Konstrukte und empirische Evidenzen.....	45
3.1 Objektivität von Schulnoten.....	47
3.2 Reliabilität von Schulnoten .....	54
3.3 Validität von Schulnoten .....	56
3.4 Zur Beziehung von messtheoretischen Gütekriterien und Notengebung in der Schule .....	65
4 Ausgewählte Bestandteile von Schulnoten – theoretische Basis und empirische Evidenzen.....	73
4.1 Ausgewählte Bedingungsmodelle von Schulleistung .....	73
4.1.1 Makro-Modell der Bedingungsfaktoren schulischer Leistungen nach Helmke und Weinert (1997) .....	75
4.1.2 Multikausales Bedingungsmodell schulischer Leistung nach Heller (1995) .....	77
4.2 Individuelle Bedingungsfaktoren von Schulleistung.....	81
4.2.1 Kognitive Prädiktoren von Schulleistung.....	84
4.2.2 Nicht-kognitive Prädiktoren von Schulleistung .....	102
4.2.3 Moderatoreffekte individueller Schülermerkmale ...	109

4.3	Familiäre Bedingungsfaktoren von Schulleistung.....	111
4.3.1	Ausgewählte Erklärungsansätze zu familiär bedingten Ungleichheiten im Schulsystem.....	113
4.3.2	Zum Zusammenhang zwischen familiären Struktur- und Prozessmerkmalen und individuellen Schulleistungskomponenten .....	121
4.4	Schulische Bedingungsfaktoren von Schulleistung.....	133
4.5	Eigenes Gesamtmodell zu Komponenten von Schulleistung	143
5	Ziele und Fragestellungen.....	149
5.1	Individuelle Schülermerkmale und Schulnoten .....	151
5.2	Familiäre Hintergrundmerkmale und Schulnoten.....	152
5.3	Schulische Merkmale und Schulnoten .....	153
6	Datengrundlage .....	155
6.1	Die Koala-S-Studie.....	155
6.2	Durchführung der Untersuchung .....	156
6.3	Stichprobe .....	157
6.4	Instrumente.....	159
6.4.1	Schulnoten und Leistungstests .....	159
6.4.2	Individuelle Schülermerkmale .....	161
6.4.3	Familiäre Hintergrundmerkmale.....	166
6.4.3.1	Strukturmerkmale.....	166
6.4.3.2	Prozessmerkmale .....	167
6.4.4	Übersicht zu den Analysevariablen.....	170
6.5	Datenaufbereitung und statistische Auswertung.....	172
7	Ergebnisse.....	175
7.1	Deskriptive Befunde: Individuelle und familiäre Bedingungsfaktoren von Schulnoten.....	175
7.1.1	Schulnoten in Deutsch und Mathematik am Ende der Grundschulzeit .....	176
7.1.2	Kognitive und nicht-kognitive Schülermerkmale und Schulnoten.....	177
7.1.3	Familiäre Struktur- und Prozessmerkmale und Schulnoten.....	181

---

7.1.4	Beziehung zwischen individuellen und familiären Bedingungsfaktoren .....	184
7.2	Effekte der individuellen Schülermerkmale auf die Deutsch- und Mathematiknote der 4. Klasse .....	189
7.2.1	Effekte kognitiver und nicht-kognitiver Schülermerkmale auf die Deutschnote .....	189
7.2.2	Effekte kognitiver und nicht-kognitiver Schülermerkmale auf die Mathematiknote.....	196
7.2.3	Zusammenfassung der bisherigen Ergebnisse.....	203
7.3	Effekte der familiären und individuellen Hintergrundmerkmale auf die Deutsch- und Mathematiknote der 4. Klasse.....	206
7.3.1	Effekte familiärer und individueller Schülermerkmale auf die Deutschnote .....	207
7.3.2	Effekte familiärer und individueller Schülermerkmale auf die Mathematiknote.....	212
7.3.3	Zusammenfassung der bisherigen Ergebnisse .....	216
7.4	Prüfung des Gesamtmodells.....	218
7.4.1	Prüfung des Gesamtmodells: Effekte auf die Deutschnote .....	220
7.4.2	Prüfung des Gesamtmodells: Effekte auf die Mathematiknote.....	228
7.4.3	Effekte der Klassenkomposition auf die Deutsch- und Mathematiknote .....	235
8	Diskussion .....	241
8.1	Untersuchungsziele der Arbeit .....	241
8.2	Zusammenfassung und Diskussion der Ergebnisse.....	243
8.3	Abschließende Beantwortung der Hypothesen .....	261
8.4	Abschließende Bilanz und zentrale Implikationen.....	273
8.4.1	Bilanz zentraler Befunde .....	273
8.4.2	Kritische Bilanz und zentrale Implikationen.....	276
9	Literaturverzeichnis.....	281
	Anhang.....	313



# Abbildungsverzeichnis

Abbildung 1: Makro-Modell zu Determinanten schulischer Leistung nach Helmke und Weinert (1997, S. 86) .....	75
Abbildung 2: Allg. Bedingungsmodell der Schulleistung nach Heller (1995, S. 984) .....	77
Abbildung 3: Eigenes theoretisches Gesamtmodell zu Komponenten von Schulnoten .....	147
Abbildung 4: Verteilung Zeugnisnoten in Deutsch in Klasse 4.....	176
Abbildung 5: Verteilung Zeugnisnoten in Mathematik in Klasse 4 .....	176
Abbildung 6: Lehrkräfteeinschätzung Schülermerkmale 4. Klasse nach Bildungsstatus der Eltern .....	187
Abbildung 7: Lehrkräfteeinschätzung Schülermerkmale 4. Klasse nach Bildungsaspiration der Eltern .....	187
Abbildung 8: Strukturgleichungsmodell zum Einfluss individueller und familiärer Merkmale auf die Deutschnote .....	221
Abbildung 9: Strukturgleichungsmodell zum Einfluss individueller und familiärer Merkmale auf die Mathematiknote.....	229



# Tabellenverzeichnis

Tabelle 1: Notenstufen KMK (1968, S.1).....	33
Tabelle 2: Auszug der bayerischen und sächsischen Schulordnung zur Bewertung des Sozial-, Arbeits- und Lernverhaltens .....	39
Tabelle 3: Übersicht Erhebungszeitpunkte von Koala-S.....	156
Tabelle 4: Demographische Merkmale der Schülerstichprobe.....	158
Tabelle 5: Zeugnisnoten in Deutsch und Mathematik 4. Klasse Grundschule.....	160
Tabelle 6: Übersicht Testleistungen 4. Klasse.....	161
Tabelle 7: Skala Vorwissen 2. Klasse (LK).....	162
Tabelle 8: Skala Sprachliche Fähigkeiten 4. Klasse (LK) .....	163
Tabelle 9: Begabung in Deutsch und Mathematik 4. Klasse (LK) .....	164
Tabelle 10: Skala Arbeitsverhalten 4. Klasse (LK) .....	164
Tabelle 11: Skala Sozialverhalten 4. Klasse (LK).....	165
Tabelle 12: Übersicht familiäre Strukturmerkmale der Familie .....	167
Tabelle 13: Übersicht Unterstützung Elternhaus 4. Klasse (LK) .....	168
Tabelle 14: Skala Schulkontakt 4. Klasse (LK).....	169
Tabelle 15: Bildungsaspiration Elternhaus 4. Klasse (LK) .....	169
Tabelle 16: Übersicht zu den Analysevariablen.....	171
Tabelle 17: Kognitive Schülermerkmale .....	177
Tabelle 18: Nicht-kognitive Schülermerkmale.....	178
Tabelle 19: Interkorrelationen kognitiver, nicht-kognitiver Schülermerkmale und Schulnoten in Deutsch und Mathematik.....	179
Tabelle 20: Struktur- und Prozessmerkmale der Familie .....	181
Tabelle 21: Interkorrelationen familiärer Struktur- und Prozessmerkmale und Schulnoten in Deutsch und Mathematik.....	182
Tabelle 22: Interkorrelationen individueller und familiärer Bedingungsfaktoren.....	184
Tabelle 23: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch kognitive und nicht-kognitive Schüler- merkmale .....	190
Tabelle 24: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale unter Berücksichtigung von Interaktionseffekten .....	194

---

Tabelle 25: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale .....	197
Tabelle 26: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale unter Berücksichtigung von Interaktionseffekten..	201
Tabelle 27: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch familiäre und individuelle Schülermerkmale.....	208
Tabelle 28: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch familiäre und individuelle Schülermerkmale .....	213
Tabelle 29: Korrelationen der Fehlerterme im Strukturgleichungsmodell für das Fach Deutsch.....	223
Tabelle 30: Standardisierte indirekte Effekte im Strukturgleichungsmodell auf die Deutschnote.....	224
Tabelle 31: Standardisierte totale Effekte im Strukturgleichungsmodell auf die Deutschnote.....	227
Tabelle 32: Korrelationen der Fehlerterme im Strukturgleichungsmodell für das Fach Mathematik .....	231
Tabelle 33: Standardisierte indirekte Effekte im Strukturgleichungsmodell auf die Mathematiknote .....	232
Tabelle 34: Standardisierte totale Effekte im Strukturgleichungsmodell auf die Mathematiknote .....	235
Tabelle 35: Intraklassenkorrelationen für die Deutsch- und Mathematiknote .....	237

# Abkürzungsverzeichnis

CFI	Comparative-Fit-Index
DPC	Data Processing Center
EFB	Elternfragebogen
ELEMENT	Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin
FIML	Full-Information-Maximum-Likelihood
IEA	International Association for the Evaluation of Educational Achievement
IGLU	Internationale Grundschul-Lese-Untersuchung
IQB	Institut für Qualitätsentwicklung im Bildungswesen
ISEI	International Socio-Economic Index of Occupational Status
KFT	Kognitiver Fähigkeitstest
KOALA-S	Kompetenzaufbau und Laufbahnen im Schulsystem
KMK	Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
LFB	Lehrkraftfragebogen
LK	Lehrkräfteeinschätzung
MARKUS	Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext
PISA	Programme for International Student Assessment
RMSEA	Root Mean Square Error of Approximation
SALVE	Systematische Analyse des Lernverhaltens und des Verständnisses in Mathematik: Entwicklungstrends und Fördermöglichkeiten
SBB	Schülerbeurteilungsbogen
SFB	Schülerfragebogen
SRMR	Standardized-Root-Mean-Square-Residual
TIMSS	Trends in International Mathematics and Science Study



# Danke

An dieser Stelle möchte ich mich bei denjenigen Menschen bedanken, die mich auf dem Weg zu meiner Dissertation begleitet und unterstützt haben.

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Hartmut Ditton für die inhaltliche und methodische Unterstützung. Herzlichen Dank für das stets offene Ohr, die anregenden Diskussionen und das Interesse an meinem Projekt, all dies hat maßgeblich zur meiner wissenschaftlichen Entwicklung beigetragen und mich immer wieder ermutigt, weiterzudenken.

Herrn Prof. Dr. Thomas Eckert danke ich für die Übernahme des Zweitgutachtens sowie Herrn Prof. Dr. Markus Paulus für die Teilnahme an meiner Disputation.

Bedanken möchte ich mich außerdem bei Veronika Stahn und Florian Wohlkinger für die inhaltliche und methodische Unterstützung. Ganz besonders möchte ich mich bei Juliane Aulinger bedanken. Danke für Deine inhaltliche und persönliche Unterstützung und den Glauben an das Gelingen dieses Projekts.

Mein großer Dank gilt nicht zuletzt meiner Familie und meinen Freunden und Dir, Kilian. Ihr alle habt mich in den Höhen und Tiefen meines Dissertationsprojekts immer wieder bestärkt und mich bei Korrekturen tatkräftig unterstützt. Vielen Dank für Eure Geduld und Euren Rückhalt.

München, März 2018

Sibylle Elsässer



# 1 Einleitung und Problemstellung

Fragen zu den Bestandteilen von Bildung und der damit zusammenhängenden sozialen Ungleichheit sind bedeutende Themen, welche die öffentliche Debatte und den wissenschaftlichen Diskurs bereits lange Zeit bewegen (Dahrendorf, 1965; Ditton, 1992; Furck, 1961; Picht, 1964).

Während Bildung als Mittel des sozialen Aufstiegs anerkannt und als Bürgerrecht galt, wurde über die Jahre immer deutlicher, dass jener Bildungsaspekt nicht für jedermann gleich zugänglich war und sich daraus schichtspezifische Bildungsmuster entwickelten (Dahrendorf, 1965; Ditton, 1992; Furck, 1961). Ungleiche Bildungsmuster, die bis heute mehr oder weniger eindeutig als illegitim und ungerecht bezeichnet werden und nicht zuletzt durch die internationalen Schulleistungstudien erneut in den Fokus der Öffentlichkeit rücken (Baumert, Artelt et al., 2003).

Unter Bezug auf das meritokratische Leistungsprinzip wird Bildungsungleichheit dann als ungerecht empfunden, wenn Kinder aufgrund ihrer sozialen Herkunft bei gleicher Schulleistung unterschiedlich hohe Bildungserfolge erzielen und damit die soziale Herkunft als ein maßgeblicher Faktor für Bildungsungleichheiten diskutiert wird (Ditton, 2013a). Im Einklang damit stellt sich Ungerechtigkeit – laut Deutschem Grundrecht – dann ein, wenn eine Person aufgrund ihrer Abstammung, Rasse, Sprache, Geschlecht, Heimat und Herkunft, Glaubens und religiösen oder politischen Anschauungen benachteiligt wird (Bundesministerium der Justiz, 1949). Demnach werden Bildungsunterschiede, die auf einen sogenannten leistungsunabhängigen sozialen Filter zurückgeführt werden, als ungerecht bezeichnet. Offen bleibt in diesem Zusammenhang jedoch oftmals, welche Leistungen als Richtwert für solcherlei Debatten herangezogen werden und inwiefern jene Kriterien ebenfalls einer sozialen Stratifizierung folgen (Solga, 2009).

Als Indikatoren zur Beschreibung der schulischen Situation verschiedener sozialer Gruppen werden häufig deren Bildungsbeteiligung, Bildungserfolg und Schulleistung herangezogen. Jene Schulleistung des Kindes wird dann wiederum über die Punktzahl in Leistungstests, am Übertritt in die Sekundarstufe I über die Grundschulempfehlung oder die Schulnoten von Schülern<sup>1</sup> gemessen (Diefenbach, 2007).

Dabei hat die (kritische) Betrachtung von Schulnoten selbst als Vermittlungsinstanz sozial ungleicher Bildungschancen sowie der sich darin widerspiegelnden Lehrerurteile bereits eine lange Tradition (Hofer, 1986; Ingenkamp, 1976a; Rodax & Spitz, 1978; Sommer, 1983). So werden in der Diskussion um die schulische Leistungsbewertung neben schulrechtlichen und funktionstheoretischen Anforderungen immer wieder Fragen der messtheoretischen Güte der Leistungserhebung thematisiert (Brügelmann, 2001; Ingenkamp, 1976a; Weinert, 2014). Hierzu zeigt die Überprüfung der Leistungsbewertung mittels messtheoretischer Güteverfahren, dass für – subjektiv konnotierte – Schulnoten je nach Studie unterschiedlich starke Zusammenhänge mit parallel eingesetzten – objektiv konnotierten – Schulleistungstests nachgewiesen werden (Bos, Voss et al., 2004; Middendorf, 2012). Neben Fragen des Mehrwerts der Überprüfung von Schulnoten mittels zusätzlich erhobener Schulleistungstests, wird als Folge nicht selten die Subjektivität der Lehrkraftbewertung und der schulrechtlich legitimierte pädagogische Freiraum bei der Notenvergabe diskutiert (Gomolla, 2012; Ingenkamp, 1976a).

Neben diesen schulrechtlichen Bestimmungen, die in Lehrplänen und schulartübergreifenden Richtlinien (fach)spezifische Anforderungen an die Notenkomponenten stellen (z.B. Avenarius & Füssel, 2010), gibt es in Theorie und Empirie zahlreiche Modelle, welche mögliche Determinanten schulischer Leistung in schematischen Einteilungen systematisieren (z.B. Heller, 1995; Helmke & Weinert, 1997; Krapp, 1973, 1976; Sauer & Gattringer, 1985). Gemeinsam ist allen Model-

---

<sup>1</sup> Aus Gründen der Lesbarkeit wird in der vorliegenden Arbeit die männliche Form verwendet und umfasst, sofern nicht eigens angemerkt, alle Geschlechtsformen.

len, dass sie verschiedene Einflüssebenen und Determinantenbereiche benennen und die multikausale Bedingungsstruktur von Schulleistung sichtbar machen; neben individuellen Faktoren der Schülerpersönlichkeit werden meist familiäre Faktoren und schulische Einflussbereiche betrachtet.

Während diese übergreifenden Makromodelle zur multikausalen Bedingungsstruktur auf übergeordneter Ebene die Einflussbereiche und deren Zusammenhänge darstellen, ergeben sich in der empirischen Überprüfung zahlreiche Modelle, welche sich neben direkten Effekten mit den indirekten Vermittlungsmechanismen zwischen individuellen Schülermerkmalen, sozialer Herkunft und Bildungserfolg sowie deren komplexen Wechselwirkungen beschäftigen (Ditton, 2011). Die theoretischen Grundlagen jener Überprüfungen beziehen sich meist mehr oder weniger umfangreich auf Erklärungsansätze zum schichtspezifischen Entscheidungsverhalten und schichtspezifischen Handlungsmustern, die durch unterschiedliche materielle und immaterielle Ausstattungen bestimmte soziale Gruppen in ihren Handlungen begünstigen (Boudon, 1974; Bourdieu & Passeron, 1971; Breen & Goldthorpe, 1997; Coleman, 1988). Auf Ebene der Institution Schule vertritt Solga (2009) die These, dass die meritokratische Bewertungskultur in der Schule selbst einen herkunftsabhängigen Zugang zur Bildung legitimiert, in dem Sinne als dass Lernerfolge im institutionellen Kontext maßgeblich von familialen Ressourcen abhängig sind und damit die familial unterschiedlich ausgeprägten Lernprozesse zu differenten Bildungserfolgen führen.

Obwohl bilanzierend die globalen Tendenzen zur Notenzusammensetzung zunächst relativ eindeutig erscheinen, fehlt bislang eine aktuelle, umfangreiche Aufarbeitung der Komponenten einer Grundschulnote. Neben der unterschiedlichen Operationalisierung des Konstrukts Schulleistung ist wenig eindeutig, welche Informationen auf die Notenzusammensetzung Einfluss nehmen und über welche Strukturen hierbei soziale Ungleichheiten reproduziert werden könnten; welche Komponenten in Noten also vorhanden sein müssen, um Bildungserfolge zu erzielen (Ditton, 2010).

So ist zwar hinlänglich bekannt, dass Kinder in Abhängigkeit von ihrer sozialen Herkunft bei gleicher Testleistung unterschiedliche Schulnoten erhalten, über die Bestandteile der Note – über welche Komponenten diese Ungleichheit also deutlich wird – ist bis heute wenig bekannt. Sollten die Effekte der sozialen Herkunft – auf Struktur- und Prozessebene – durch kognitive und nicht-kognitive Schülermerkmale mediiert werden, könnten durch jene Komponenten ungleichheitsfördernde Bildungsprozesse aufgedeckt werden. Zudem erscheint es erst durch die Kenntnis der Eigenschaften einer guten oder schlechten Note als sinnvoll, die Notengebung auf ihre Prognosefähigkeit zu prüfen. So sind es doch jene Verhaltenseinschätzungen, die im Sinne der Systemkonformität wiederum prognostisch valide den Erfolg im Bildungssystem vorhersagen sollten.

Ziel der vorliegenden Arbeit ist es, jene Komponenten von Schulnoten in der Grundschule aufzudecken. Konkret wird der Frage nachgegangen, welche individuellen, familiären und klassenbezogenen Einflusskomponenten sich auf die Notengebung am Ende der Grundschulzeit auswirken.

Datengrundlage dieser Arbeit ist die von der Deutschen Forschungsgemeinschaft geförderte Längsschnittstudie Koala-S (**K**ompetenzaufbau und **L**aufbahnen im **S**chulsystem) aus den Jahren 2005 bis 2007, durchgeführt an bayerischen und sächsischen Grundschulen.

Im zweiten Kapitel der vorliegenden Arbeit werden die begrifflichen Grundlagen zum Gegenstand schulischer Leistungsbewertung vorgestellt. Nach einer anfänglichen Differenzierung zwischen dem allgemeinen und dem schulischen Leistungsbegriff ([Kap. 2.1](#) – [Kap. 2.2](#)) wird der Fokus auf die Schulnote gelegt ([Kap. 2.3](#)) und funktionale ([Kap. 2.4](#)) und schulrechtliche ([Kap. 2.5](#)) Anforderungen besprochen. Eine abschließende Zusammenfassung ([Kap. 2.6](#)) beendet das zweite Grundlagenkapitel.

Das dritte Kapitel beschäftigt sich mit den messtheoretischen Anforderungen an die Notengebung. Auf theoretischer und empirischer Basis

wird besprochen, welche Bedeutung die Gütekriterien der Objektivität, Reliabilität und Validität für die Schulnoten haben (Kap. 3.1 – Kap. 3.3), um dann auf die Beziehung zwischen messtheoretischen Gütekriterien und Noten einzugehen (Kap. 3.4).

Das vierte Kapitel beschäftigt sich mit den ausgewählten Bestandteilen von Schulnoten. Auf theoretischer und empirischer Basis werden einzelne Bedingungsmodelle von Schulnoten eingeführt (Kap. 4.1), um dann ausgewählte individuelle Bedingungsfaktoren auf Seiten des Schülers – differenziert in kognitive und nicht-kognitive Merkmale – zu betrachten (Kap. 4.2). Kapitel 4.3 widmet sich den familiären Bedingungsfaktoren von Schulleistung, wobei theoretische Erklärungsansätze zur Wirkweise sowie familiäre Struktur- und Prozessmerkmale betrachtet werden. Anschließend werden ausgewählte schulische Bedingungsfaktoren beschrieben, wobei der Fokus auf Faktoren der Lehrkraft und klassenspezifischen Referenzgruppeneffekten liegt (Kap. 4.4). Das vierte Kapitel schließt mit einem bilanzierenden, eigenen theoretischen Gesamtmodell (Kap. 4.5), das in den nachfolgenden Kapiteln einer empirischen Prüfung unterzogen wird.

Nachdem in Kapitel 5 die Ziele und Fragestellungen der Arbeit vorgestellt werden und Kapitel 6 die Datengrundlage der Untersuchung erläutert, folgt in Kapitel 7 die Darstellung der Ergebnisse, untergliedert nach Fragestellung und abhängiger Variable.

Kapitel 8 zeigt nochmals die Untersuchungsziele der Arbeit auf (Kap. 8.1), fasst die gefundenen Ergebnisse zusammen und diskutiert diese (Kap. 8.2), um dann die Hypothesen der vorliegenden Arbeit abschließend zu beantworten (Kap. 8.3), kritisch Bilanz zu ziehen sowie zentrale Implikationen hervorzuheben (Kap. 8.4).



## 2 Zum Gegenstand schulischer Leistungsbeurteilung – theoretische Grundlagen

Die theoretische und empirische Auseinandersetzung mit den Bestandteilen von Noten und der Notengebung in der Grundschule bedarf zunächst – als Grundlage der folgenden Kapitel – eine theoretische Auseinandersetzung mit den Begriffen Leistung, Schulleistung und Schulnoten. Fernab von einer normativen Auseinandersetzung mit der Problematik der schulischen Leistungsbewertung beschäftigt sich der erste Abschnitt dieses Kapitels mit dem vielfach verwendeten Begriff Leistung und seiner Bedeutung für die unterschiedlichen Gesellschafts- und Fachbereiche, um dann in einem zweiten Abschnitt auf den für die Schule relevanten Schulleistungsbegriff einzugehen. Der dritte Teil dieses Kapitels behandelt den – häufig als Indikator für Schulleistung verwendeten – Begriff der Schulnote.

Die Diskussion um die Leistungsbeurteilung<sup>2</sup> im gesamtgesellschaftlichen wie schulischen Kontext umfasst eine Fülle an Funktionen und sowohl kritischer als auch positiver Diskussionsbeiträge, die im vierten Unterkapitel angesprochen werden, um dann die schulrechtlichen Ansprüche an die Notengebung aufzuzeigen. Eine erschöpfende Analyse des theoretischen Kontextes schulischer Leistungsbeurteilung würden den Rahmen dieser Arbeit sicherlich überschreiten. Ziel des Kapitels ist es dennoch, die Komplexität der Thematik mit ihren historischen und aktuellen Bezügen zu beleuchten und auf die für diese Arbeit wesentlichen Grundbezüge zu reduzieren. Im Folgenden wird bewusst auf einen empirischen Bezug der Begrifflichkeiten verzichtet; eine Verknüpfung erfolgt im zweiten Teil der vorliegenden Arbeit, wenn nach den empirisch ermittelten Bestandteilen schulischer Leistung gefragt wird.

---

2 Obwohl in der Literatur durchaus zwischen Leistungsbeurteilung und Leistungsbewertung unterschieden wird (vgl. z.B. Bürgermeister (2014)), wird im Folgenden der Begriff Bewertung synonym für Beurteilung verwendet.

## 2.1 Der Leistungsbegriff

Der Begriff Leistung ist ein vielfach verwendeter Ausdruck, der in den jeweiligen Lebensbereichen und Fachdisziplinen unterschiedliche Bedeutungen annimmt und vor allem in den sechziger und siebziger Jahren des 20. Jahrhunderts ausführlich theoretisch behandelt wurde (Dohse, 1967; Furck, 1961; Heckhausen, 1974; Ingenkamp, 1976a). Geprägt durch die spezifischen gesellschaftlichen, kulturellen und historischen Bezüge bestehen je nach Kontext und Perspektive eine Fülle an Auslegungen und Diskussionen darüber, welche (menschlichen) Anforderungen und Eigenschaften unterschiedliche Ausprägungen von Leistung ausmachen, wie diese erreicht und mittels welcher Maßstäbe Leistung gemessen werden sollte. Eine konkrete Einordnung von Leistung wiederum hängt mit den gesellschaftlich und fachlich geprägten Vorstellungen darüber zusammen, welche Funktionen und Ziele mit dem Leistungskonstrukt erfüllt werden sollen – Vorstellungen, die ebenso vielfältig sind (Sacher, 2001; Sommer, 1983; Ziegenspeck & Lehmann, 1999) und in [Kapitel 2.4](#) weiter vertieft werden.

Nähert man sich dem Leistungsbegriff zunächst von der sprachlichen Seite, stellen Ziegenspeck und Lehmann (1999) fest, dass die semantische Einordnung der Wörter leisten und Leistung nicht bei einer eindeutigen Begriffseinordnung helfe, sondern vielmehr die Bedeutung der spezifischen Anwendungsfelder sowie die Mehrdeutigkeit des Leistungsbegriffs deutlich mache. Eine allgemeingültige, kontextunabhängige Definition von Leistung ist in der Literatur nicht zu finden und macht deutlich, dass der Begriff Leistung nicht ohne spezifischen Fachbezug geklärt werden kann (Ziegenspeck & Lehmann, 1999). Dennoch eint die theoretischen Abhandlungen zum Leistungsbegriff das zugrundeliegende Leistungsprinzip, welchem als gesellschaftliches Verteilungsprinzip in der sich entwickelnden Industriegesellschaft eine immer größere Bedeutung bei der Vergabe von Berufs- und Lebenschancen zukam (Sacher, 2009; Sommer, 1983). Anders als in der Ständegesellschaft des Mittelalters erfolgte die Verteilung von Gütern, Zugangschancen und (finanzieller) Anerkennung nicht mehr allein nach Geburt, sondern nach erbrachter Leistung (Furck, 1961; Sacher,

2009). Eine Verteilungsform, die trotz ihres emanzipatorischen Grundgedankens aber bald wegen ihres normativen Charakters als Reproduktionsinstrument bestimmter gesellschaftlicher Verhältnisse kritisiert wurde<sup>3</sup> (Sacher, 2009; Ziegenspeck & Lehmann, 1999). Die begriffliche Grundlage des bis heute vorherrschenden Gesellschaftsprinzips geht allerdings nicht über die sehr allgemein gehaltene Leistungsdefinition hinaus, in der Leistung immer mit einer in einem bestimmten Zeitraum getätigten Handlung in Zusammenhang steht (Hartfiel, 1977). Differenziert nach verschiedenen Handlungsfeldern und Funktionen werden häufig ein physikalischer, soziologischer, psychologischer und pädagogischer Leistungsbegriff unterschieden. Während der *physikalische Leistungsbegriff* als Gütemaßstab für Leistung die (von Mensch oder Maschine) benötigte Zeit für einen Arbeitsvorgang postuliert (Leistung = Arbeit pro Zeiteinheit), behandelt die soziologische Begriffsannäherung die normative Komponente des Leistungsbegriffs und dessen gesamtgesellschaftliche Bedeutung (Fürstenberg, 1995; Sacher, 2001, 2009). Aus *soziologischer Sicht* wird untersucht, welche Maßstäbe zur Bestimmung von Leistung gelten, wer diese Maßstäbe festlegt, welche Instanzen Leistung legitimieren und welche Komponenten Leistung umfassen müssen, um tatsächlich erreicht zu werden (Fürstenberg, 1995). Unter bildungssoziologischer Perspektive werden dann u.a. gesellschaftliche Voraussetzungen des Bildungserwerbs und Folgen schulischer Selektion diskutiert (Terhart, 1999). Während sich die pädagogisch-psychologische Perspektive im Schulkontext mit Fragen der prognostischen und diagnostischen Qualität der Leistungsbeurteilungen beschäftigt, bezeichnet die rein *psychologische Begriffsbestimmung* Leistung als „das Ergebnis einer Untergruppe menschlicher Handlungen“ (Heckhausen, 1974, S.11), bei welcher der handelnde Mensch im Stande sein muss, die geforderte Handlung willentlich auszuführen (Heckhausen, 1974; Terhart, 1999). Differenziert wird zwischen einem anstrengungszentrierten und einem fähigkeits-

---

3 Sommer (1983, S.24) unterscheidet dabei drei wesentliche Kritikstränge an den Begriffen Leistung, Leistungsorientierung und Leistungsprinzip: 1. Die Überbetonung des Leistungsprinzips als alleiniges Verteilungsprinzip mit möglichen inhumanen Folgen, 2. Die Art und Weise der Leistungsabfrage in der Leistungsgesellschaft, 3. Vorherrschende Mechanismen, welche die Leistungsziele und Leistungsmaßstäbe festlegen.

zentrierten Leistungsbegriff, wobei für erfolgreiches Handeln sowohl Können als auch Bemühen vorliegen muss (Heckhausen, 1974). Das Maß für die Fähigkeitsfeststellung wird im psychologischen Kontext wiederum aus der sozialen Bezugsgruppe abgeleitet; die Fähigkeit des Einzelnen beruht damit auf dem interpersonellen Vergleich bei gleicher Handlungsanforderung (Heckhausen, 1974). Aussagen über möglichen Schwierigkeiten der sozialen Bezugsnorm, auch bei ungleicher Ausgangslage, werden in diesem Zusammenhang nicht getätigt. Unerwähnt bleiben soll an dieser Stelle nicht, dass sich auch in psychologischen Disziplinen Abhandlungen zur Schulleistung finden, die in Anlehnung an Heckhausens Leistungsbegriff verschiedene Analyseebenen (hypothetisches Konstrukt und empirische Manifestation) oder aus kognitionspsychologischer Sicht zwischen deklarativem und prozeduralem Wissen unterscheiden (Helmke & Schrader, 2010; Krapp, 1976; Schrader & Helmke, 2008). Diese Einteilung wird allerdings im Folgenden aufgrund ihrer Spezifität nicht weiter ausgeführt.

Analog zu den anderen Handlungsfeldern zeichnet sich auch der *pädagogische Leistungsbegriff* durch seine Heterogenität und die vielfältigen Sinnesausprägungen aus (Ziegenspeck & Lehmann, 1999). Mit der gesellschaftlich steigenden Relevanz des Leistungsprinzips zur Verteilung von Lebenschancen stieg auch die Relevanz von Bildung in der Schule. Schon Furck (1961) wies darauf hin, dass bereits im 18. Jahrhundert der Staat die Pädagogik als Mittel für gesamtgesellschaftliche Notwendigkeiten ansah. Im Fokus standen zu dieser Zeit die Funktionalität der Bildung für künftige, gesellschaftliche Verwendbarkeit mit der einhergehenden Explikation von Eigenschaften, die für diesen Zweck gelehrt werden sollten. Vertreter der Neuhumanisten um Wilhelm von Humboldt entgegneten dieser Zweckgebundenheit und forderten, Menschen – im Gegensatz zu der rein funktionalen Bildung – um ihrer selbst willen zu bilden (Furck, 1961). Furck (1961) betont, dass aus dieser Zeit die Forderung einer allgemeinen Bildung hervorging, die im ganz bewussten Gegensatz zu der Betonung der Funktionalität steht und die reine Berufsbildung in den Schulen ablehnt. Allerdings führte diese Forderung – nach Furck (1961) – zu dem Missverständnis, im Sinne des Prinzips der allgemeinen Bildung junge Men-

schen als zu bildende Objekte zu betrachten und nicht mehr die Person hinter der Schulleistung einzubeziehen. Hauptkritikpunkt aller Abhandlungen zur Bedeutung des Leistungsprinzips für die Schule ist dabei der Tenor, dass es der pädagogische Leistungsbegriff in Vergangenheit und Gegenwart immer weniger schaffe, sich von den gesellschaftlich-wirtschaftlichen auferlegten Funktionen abzugrenzen und den Menschen im Sinne des humanistischen Grundgedanken zu bilden (Ziegenspeck & Lehmann, 1999). Demnach müsse die pädagogische Interpretation des Leistungsbegriffs die Frage nach dem zu bildenden Kind in den Fokus stellen. Aufgabe der Schule sei es, Leistung zu fördern, zu ermöglichen und den Einzelnen auf weitere (gesellschaftliche) Leistungen vorzubereiten (Lichtenstein-Rother, 1976). Diese Bedingungsfaktoren schulischer Sozialisation prägen nach Sommer (1983) das Bild von Schule bis heute maßgeblich.

Zusammenfassend ist festzuhalten, dass die hier verfolgte, ausschnittshafte Behandlung des (fachspezifischen) Leistungsbegriffs einmal mehr die Mehrdeutigkeit des Begriffs widerspiegelt. Allgemein konstatiert werden kann die Tatsache, dass Leistung häufig Arbeit und Zeit bindet, Arbeitsprozesse und Arbeitsergebnisse umfasst und oftmals auf wirtschaftliche und gesellschaftliche Maßstäbe reduziert wird (Ziegenspeck & Lehmann, 1999). Das Fehlen von allgemeingültigen Bewertungsmaßstäben lässt eine objektive Leistungsbewertung nahezu unmöglich erscheinen, beinhaltet die Bewertung doch immer soziale Beziehungen, deren wertende Attribuierung und einen Kontextbezug. Erst wenn man sich bewusst macht, was Leistung im jeweilige Kontext bedeutet, könne man – so argumentiert schon Heid (1992) – das dahinter stehende Leistungsprinzip als Kriterium sozialer Verteilungsgerechtigkeit verwenden. Im nächsten Kapitel wird die Relevanz jener sozialen und kontextuellen Bezüge für den Leistungsbegriff weiter vertieft; es folgt eine Explikation des schulischen Leistungsbegriffs – als spezifische Form des pädagogischen Leistungsbegriffs.

## 2.2 Zum schulischen Leistungsbegriff

Analog zum Leistungsbegriff ist es ebenfalls schwierig, den Schulleistungsbegriff durch eine prägnante Definition zu fassen. Ausgangslage einer begrifflichen Annäherung ist die Annahme, dass es sich bei dem Schulleistungsbegriff um ein inhaltlich leeres Konstrukt handelt, das erst durch spezifische Bewertungsgrundlagen definiert werden kann (Holmeier, 2013; Pache, 1978). Ausgehend von einem pädagogischen Verständnis von Schulleistung dient die Einordnung von Furck (1961) häufig als richtungweisendes Verständnis zum Zusammenhang zwischen Schule und Leistung. Diese umfasst

Leistung als schulische Forderung an den Schüler, Leistung als Tätigkeit des Schülers, Leistung als Ergebnis der Tätigkeit des einzelnen innerhalb der verschiedenen Leistungsbereiche und Leistung als besonderer Beitrag der Schule für Gesellschaft, Staat, Wirtschaft und Wissenschaft. (Furck, 1961, S.118).

Diese weite, nahezu allumfassende Verwendung des schulischen Leistungsbegriffs macht deutlich, dass der pädagogische Leistungsbegriff weder formal noch inhaltlich ausreichend definiert ist, sowie auch der relationale Bezug des Begriffs erkennbar ist (Ziegenspeck & Lehmann, 1999). Häufig werden auch widersprüchliche Anforderungen an den Schulleistungsbegriff moniert, welches von Furck (1961) als pädagogisches Problem der Leistung bezeichnet wird und sich auf den Widerspruch zwischen der Bildung des Menschen um seiner selbst willen gegenüber dem Befähigungsauftrag durch schulische Leistungsbewertungen bezieht. Zunehmend kritisiert wird dazu aktuell die einseitige Ausrichtung des schulischen Leistungsverständnisses an gesellschaftlichen Funktionen und dessen kognitive Verengung (Dietrich & Fricke, 2013).

Die angeführte Problematik zum Schulleistungsbegriff soll nicht das Vorhaben überlagern, das heute vorherrschende pädagogische Verständnis von Schulleistung näher zu beleuchten. Betrachtet werden im Folgenden Bereiche, die für das Verständnis des Konstrukts Schulleis-

tung maßgeblich sind. Fokussiert werden, in Anlehnung an das obenstehende Zitat von Furck (1961), die beteiligten Akteure, die Bedeutung der Bewertungsmaßstäbe für den Begriff, Schwierigkeiten und Kritikpunkte sowie die möglichen Bewertungsdimensionen.

Ausgehend von der Annahme, dass die Schulleistung durch eine Vielzahl von Faktoren beeinflusst wird, ist es der Schüler selbst, der jene zu definierende Leistung erbringt, die von schulischen, individuellen und familiären Hintergrundmerkmalen beeinflusst wird. Beurteilt wird diese Leistung dann wiederum von der beteiligten Lehrkraft, welche sich auf bestimmte Maßstäbe und Bewertungsdimensionen bezieht, die von rechtlichen und schulinternen Vorgaben beeinflusst werden. Diese, sehr verkürzte und allgemein gehaltene Erläuterung beschreibt die unmittelbaren Akteure, die – zumindest sichtbar – auf den Begriff Schulleistung Einfluss nehmen können und in jeder spezifischen Leistungssituation in ihrer individuellen Ausprägung betrachtet werden sollten.

Ebenso relevant für die begriffliche Annäherung an das Konstrukt Schulleistung ist der jeweilige Bewertungsmaßstab, nach welchem die gezeigte Leistung bewertet wird. Nach Furck (1961) wird die Leistung in der Schule an drei Bezugssystemen gemessen, die er als subjektive (am Schüler selbst), intersubjektive (Schüler zur Bezugsgruppe) und objektive (Schüler zum Lerngegenstand) Relationen bezeichnet. Diese, von Rheinberg (2008) auch als individuelle, soziale und sachliche bzw. kriteriale Bezugsnorm bezeichneten Standards werden benötigt, um eine Handlung erst als Leistung bewerten zu können. Der gewählte Bewertungsmaßstab ist auch bei der begrifflichen Einordnung von Schulleistung zu beachten. Aufgrund der schulrechtlichen Unklarheit (vgl. [Kap. 2.5](#)) ist es die jeweilige Schule bzw. die einzelne Lehrkraft, die festlegen muss, ob sie die Leistung des Schülers an dessen eigenem Lernfortschritt (individuell), im Vergleich zu der in der Klasse gezeigten Leistungen (sozial) oder an einem vorher festgelegten Leistungsmaßstab (kriterial) misst. Zusätzlich dazu ist der Zeitpunkt der Leistungsbewertung für das Schulleistungs-konstrukt relevant. Differenziert wird zwischen einem formativen, prozessorientierten und

einem summativen, produktorientierten Bewertungsmaßstab (Jürgens, 2010; Woolfolk, 2008; Ziegenspeck & Lehmann, 1999). Demnach ist es bedeutsam, ob der Prozess zum Handlungsergebnis oder nur das Endresultat selbst bewertet wird. Der zweite Aspekt ist der zeitliche Einbezug der Informationen, die als Leistung bewertet werden. Entscheidend ist, welcher zeitliche Rahmen in die Bewertung miteinfließt: neben der Beurteilung zu einem spezifischen Zeitpunkt ziehen Beurteilungsvorgänge auch Informationen aus der Summe aller implizit und explizit gezeigten, auch zukünftig zu erwartenden Leistungen als Urteilsgrundlage heran (Becker & Birkelbach, 2013; Faber & Billmann-Mahecha, 2010; Furck, 1961; Schrader & Helmke, 2014).

Die am häufigsten (kritisch) diskutierte Größe des Konstrukts Schulleistung sind die Bewertungsdimensionen bzw. Bewertungsbereiche, die bei einer Beurteilung beachtet werden. Gefragt wird, welche Eigenschaften sich in der Leistungsbewertung widerspiegeln sollten, ob ausschließlich kognitive Fähigkeiten oder auch motivationale Aspekte zu bewerten sind und welche kontextuellen Faktoren eine Rolle spielen (dürfen) (Furck, 1961; Heckhausen, 1974; Helmke & Weinert, 1997). Neben einer Fülle an theoretischen Modellen, die sich Fragen der Bedingungsfaktoren von Schulleistung annehmen (z.B. Helmke & Weinert, 1997; Hochweber, 2010; Krapp, 1976; Schrader & Helmke, 2008) muss ebenfalls beachtet werden, welche Funktion und welches Ziel mit der jeweiligen Leistungsbeurteilung erfüllt werden soll (Furck, 1961; Jürgens & Lissmann, 2015). Die hier zu bewertenden Eigenschaften werden im [Kapitel 4](#) weiter vertieft sowie in [Kapitel 2.5](#) die schulrechtlich legitimierten Bewertungsbereiche beleuchtet. Für die begriffliche Einordnung des Konstrukts Schulleistung ist festzuhalten, dass jene festgestellten Dimensionen einen maßgeblichen Anteil an der spezifischen Bestimmung des Begriffs einnehmen. Durch diese Festlegung wird vermutlich erst deutlich, welche Eigenschaften letztendlich als Leistung in der Schule bezeichnet werden. Eigenschaften, die nach Bourdieu (1982) genau jene Merkmale sind, die je nach Legitimitätsgrad mit den vorherrschenden (gesellschaftlichen) Prinzipien in der Schule entsprechend positiv oder negativ bewertet werden.

In Anlehnung an die oben eingeführte Relevanz des Bewertungsmaßstabs ist es für den Schulleistungsbegriff unabdingbar, zwei wesentliche Diagnosearten zu unterscheiden. Begreift man die Leistungsbeurteilung als Teil der Pädagogischen Diagnostik<sup>4</sup>, muss zwischen einer Leistungsfeststellung und einer Leistungsbewertung unterschieden werden, wobei erstere der psychometrischen Leistungsmessung und die zweite der Schulnotenvergabe entsprechen könnte (Ingenkamp, 1967; Lehmann, 2001). Je nachdem welche Funktion mit der Schulleistungsdia­gnose verbunden ist und welches Kriterium überprüft werden soll, wird eine andere Form der Leistungserhebung gewählt; in der vorliegenden Arbeit wird zwischen der häufig subjektiv konnotierten Leistungsbeurteilung bzw. Leistungsbewertung und der häufig objektiv konnotierten Leistungsfeststellung bzw. Leistungsmessung unterschieden. Eine Vertiefung dieser messtheoretischen Güteverfahren und deren Bedeutung für die Schulnotenvergabe folgt in [Kapitel 3](#).

Zusammenfassend festgehalten wird an dieser Stelle, dass der Begriff Schulleistung ein ebenso abstraktes wie mehrdeutiges Konstrukt darstellt, das erst vor dem Hintergrund festgelegter (normativer) Bewertungsdimensionen und einbezogener Bewertungsmaßstäbe als Schulleistung bezeichnet werden kann. Probleme der Definition schulischer Leistung ergeben sich vermutlich dadurch, dass man Schulleistung sowohl auf theoretischer als auch empirischer Basis begreifen kann. Jene, auf empirischer Basis ermittelten Schulleistungsvariablen dienen dann wiederum als Indikatoren für das theoretische Konstrukt, das seinerseits wiederum abstraktere Dimensionen bestimmt, die unzählige Operationalisierungen derselben Dimension ermöglichen. Eine Verwendung des Schulleistungsbegriffs ohne jene Differenzierung spricht also beide Seiten des Konstrukts an und kann unzählige Faktoren binden (Krapp, 1976). Fest steht, dass sich der Begriff immer auf das gesamte Leistungsverhalten in der Schule bezieht und

---

<sup>4</sup> Nach Jürgens und Lissmann (2015) umfasst die Pädagogische Diagnostik all jene diagnostischen Tätigkeiten, die Voraussetzungen und Bedingungen der Lehr- und Lernprozesse, die Lernprozesse selbst sowie deren Leistungsergebnisse ermitteln und ggf. optimieren. Jürgens und Sacher (2008) bezeichnen die Zensurengebung gar als die Hauptbeurteilungsform der pädagogischen Diagnostik.

es für theoretisch-normative wie empirische Abhandlungen maßgeblich ist, welche Kriterien man bei der Erfassung schulischer Leistung anlegt. Beachtet werden können sowohl dynamische als auch statische Aspekte des Leistungsprozesses; ebenfalls nimmt die Frage der Perspektive einen maßgeblichen Faktor ein (Ditton, 2013c; Heller, 1984; Helmke & Schrader, 2010). Vor dem Hintergrund, den Schulleistungsbegriff ohne Bezug zu schulischen Inhalten explizit definieren zu können, sind die aktuellen Reformtendenzen in der Leistungsbeurteilung zu beachten, wonach der Kompetenzbegriff immer häufiger an die Stelle des Leistungsbegriffs rückt<sup>5</sup> (Jürgens & Lissmann, 2015; Schrader & Helmke, 2008; Weinert, 2001). Große Schulleistungsstudien messen Schulleistung als Kompetenzen, die wiederum als Fertigkeiten, Kenntnisse oder Fachleistungen bezeichnet werden und häufig als objektive Verfahren der Leistungsmessung gelten (Baumert, Artelt et al., 2002; Kampshoff, 2007; Sauer & Gamsjäger, 1996). Zudem liegen empirischen Studien zu Bedingungsfaktoren von Schulleistung häufig implizite, normative Annahmen zugrunde, was als Leistung verstanden wird. Annahmen, die sich dann wiederum in der entsprechenden Operationalisierung widerspiegeln, wobei häufig die Schulnoten als Indikator für Leistung und Bildungserfolg herangezogen werden (Schrader & Helmke, 2014). Und auch wenn Sommer (1983) bereits darauf hinweist, dass das Verhältnis zwischen dem Konstrukt Schulleistung und Schulnote nicht annähernd vollständig aufzuklären sei, wird im Folgenden auf die Schulnoten eingegangen. Gewählt wird die Perspektive auf die Schulleistung aus institutioneller, schulischer Sicht. Die Bewertung der gezeigten Schulleistung erfolgt damit aus einer Fremdperspektive und klammert die Eigenperspektive des Schü-

---

5 Weinert (2014) definiert Kompetenzen als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösung in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (S.27-28). Jener Begriff rückt vermutlich vornehmlich deswegen in den Fokus, da diese Beschreibung dem Zweck der Leistungsmessung in Kompetenzmodellen der großen Schulleistungsuntersuchungen dient. Der Kompetenzbegriff wird aufgrund seiner Relevanz für die vorliegende Arbeit dennoch nicht weiterverwendet, sollte aber wegen seiner Nähe zum Leistungsbegriff dennoch erwähnt und im Bewusstsein bleiben.

lers selbst bewusst aus, da die Fremdbewertung in Form von Schulnoten letztendlich im Schulsystem richtungsweisend ist.

## 2.3 Die Schulnote als Indikator für Schulleistung

Empirische Studien, die sich mit sozial ungleich verteilten Bildungschancen beschäftigen, sind in Vergangenheit und Gegenwart zahlreich erschienen (z.B. Ditton, 2013b; Ditton & Krüskens, 2009; Maaz, Baeriswyl & Trautwein, 2011; Maaz & Nagy, 2010; Weinert, 2012). Allgemeiner Tenor dieser Analysen ist, dass Leistungsunterschiede zwischen Schülern dann problematisch erscheinen, wenn diese nicht unabhängig von familiären oder schulischen Merkmalen entstanden sind. Die Beurteilung von Schülerleistung, Bildungschancen und Schulerfolg wird in diesem Kontext häufig mithilfe von Schulnoten operationalisiert. Nicht zuletzt aufgrund dieser Tatsache werden die Funktionen und Grenzen der Notengebung bereits seit langer Zeit diskutiert (Dohse, 1967; Ingenkamp, 1976a). Während die [Kapitel 2.4](#) und [2.5](#) die Funktionen der Notengebung und deren schulrechtliche Vorgaben analysieren, geht es in diesem Kapitel vornehmlich um die Frage, welche Strukturmerkmale eine Schulnote auszeichnen, in welcher Form sie vergeben wird und welche Konsequenzen eine Benotung nach sich ziehen kann. An dieser Stelle wird bewusst nicht auf individuelle und institutionelle Einflussfaktoren sowie häufig deklarierte Fehlerquellen in der Notengebung eingegangen; diese Punkte sind Gegenstand eines anderen Kapitels (vgl. [Kap. 4](#)).

Die Schulnote – vornehmlich in älteren Abhandlungen als Zensur bezeichnet – dient in der Schule als Bewertungs- und Klassifikationsmaßstab, welcher die gezeigte Leistung in einem Raster klassifiziert und das gezeigte Verhalten auf einer Rangreihe einstuft (Holmeier, 2013; Ingenkamp, 1995b; Wild & Krapp, 2006). Die Bezeichnung Zensur leitet sich aus dem lateinischen Wort *censere* ab und beschreibt die beurteilende, prüfende und abschätzende Arbeit eines Censors im alten Rom, der entsprechend Tadel durch Zensuren (*censure*) erteilte (Dohse, 1967; Jung, 2013; Lintorf, 2012). Auch der Begriff Note

stammt aus dem Lateinischen und bedeutet nach dem Verb *noscere* so viel wie Kennzeichen oder Merkmal (Jung, 2013). Der Begriff Zensur, im Folgenden synonym verwendet zum Notenbegriff, diente später zur allgemeinen Bewertung von menschlichen Handlungen und zur Feststellung und Einordnung von in der Schule gezeigten Verhaltensweisen; eine Funktion, die sich bis heute fortsetzt (Ziegenspeck & Lehmann, 1999).

Die Vergabe von Ziffernoten erfolgt in einem schulrechtlich vorgegebenen Kodierungssystem, in dem das Ergebnis eines individuellen Lernprozesses auf einer sechsstufigen Skala abgebildet wird (vgl. Kap. 2.5). Betrachtet man diese Einteilung als Ergebnis eines Messvorgangs, können Noten allenfalls auf einem Ordinalskalenniveau gemessen werden, da die Abstände zwischen den einzelnen Notenstufen nicht gleich sind. Würde man Noten auf dem nächsthöheren Intervallskalenniveau messen, mit dem Aussagen über die Abstände zwischen zwei Stufen getroffen werden können, ginge man davon aus, dass der Leistungsabstand zwischen einem Schüler mit der Note 1 zu einem Schüler mit der Note 2 gleich groß sei, wie der Leistungsabstand eines Schülers mit der Note 2 zu einem Schüler mit der Note 3. So müsste beispielsweise eine *sehr gute* Note in Deutsch das Doppelte einer *guten* Note in Deutsch bedeuten. Dieses Ausführung macht deutlich, dass die einzelnen Notenstufen keinesfalls gleiche Leistungsunterschiede deutlich machen und man damit nur Aussagen zu einem besser oder weniger guten Leistungsverhalten tätigen kann sowie nicht angenommen werden kann, dass ein einzelner Benotungsvorgang ausschließlich Informationen aus dem gerade gezeigten Verhalten abbildet (Faber & Billmann-Mahecha, 2010; Sacher, 1985, 2009; Ziegenspeck & Lehmann, 1999). Festgehalten wird an dieser Stelle also, dass die Messung von Leistung durch Schulnoten korrekterweise nur auf dem Ordinalskalenniveau erfolgen kann und eine arithmetische Mittelwertbildung damit streng mathematisch nicht zulässig ist (Sacher, 2009). Dies hat nach Sacher (2009) zur Folge, dass Mittelwerte von Schulnoten – wie sie in Zeugnissen beschrieben stehen – allenfalls eine „grobe Information über die mittlere Leistung“ (S.35) geben und Nachkommastellen keine Beachtung finden sollten. Nach Breidenstein (2014) liefert die

Durchschnittsbildung aus mehreren Einzelnoten sogar eine „Suggestion der Objektivierung qua Mathematisierung“ (Breidenstein, 2014, S.104).

Während die Entstehung der heutigen Zensurenkala und des Zeugnisses eine ganz eigene Entwicklung vollzog, die im Folgenden nicht näher erläutert wird<sup>6</sup>, lohnt ein Blick auf die historische Einteilung zu den Grundformen der schulischen Zensur, als Grundlage für die heutigen Bewertungsformen mittels Ziffern. Die Grundform einer Zensur, so wie sie letztendlich in einem Schulzeugnis geschrieben steht, stammt aus Einzelbenotungen, wodurch die Abhandlung über Einzelzensuren gleichzeitig immer auch eine Abhandlung über das Schulzeugnis beinhaltet (Ziegenspeck & Lehmann, 1999). Ebenfalls bedeutsam für die Notenform ist die dahinterstehende Funktion der Notenvergabe. So war es auch in der Geschichte der Zensurengebung von maßgeblicher Relevanz, welches Ziel die Vergabe von Noten hatte (Dohse, 1967; Sacher, 2009). Die historische Entwicklung der Zensur – als Bestandteil jedes Schulzeugnisses – wird von Dohse (1967) in drei Grundformen unterschieden, die dann wiederum in verschiedene Stufenformen differenziert werden. Die mündliche Zensur – als erste Grundform der Schulzensur – grenzt sich von einer einfachen mündlichen Anerkennung oder Kritik durch die Kategorisierung und Weitergabe an Dritte aus. In der zweiten Grundform, der symbolischen Zensur, wurde die Sitzordnung entsprechend dem Leistungsstand der Klasse justiert (Dohse, 1967; Kleinert, 1951). Die Einteilung in Leistungsstufen nach Zahlen mit entsprechenden Bezeichnungen gilt als dritte Grundform der Zensur. Die verschiedenen Stufenformen Reihenzensur, Polarzensur und Zonenzensur regeln die Struktur der Leistungszuteilung; während die Reihenzensur den Schülerleistungen Ordnungszahlen zuweist, erfolgt bei der polaren Zensur eine Leistungseinstufung nach entsprechenden positiven und negativen Extremwerten, die Zonenzensur klassifiziert dann die gezeigte Leistung und bündelt sie in ebenbürtige Gruppen (Dohse, 1967).

---

6 Bei Dohse (1967) findet sich ab Seite 44 bzw. auf Seite 9 eine differenzierte Abhandlung zur historischen Entwicklung der Zensurenkala und des Schulzeugnisses.

Diese Übersicht der historisch gewachsenen Bewertungsformen lässt einige Übereinstimmungen mit dem heutigen Schulalltag erkennen: So erfolgt eine Benotung meist vor dem Hintergrund eines klasseninternen Bezugsmaßstabs und teilt Leistungen nach Rangfolge in unterschiedliche Leistungsgruppen und Schulformen ein.

Fernab von der historischen Analyse bestehender Benotungsformen sind für den heutigen Schulalltag gängige – schulrechtlich festgelegte – Notenformen festzustellen: die mündliche Note, die schriftliche Einzelnote und die Jahrgangsnote bzw. Zeugnisnote, die wiederum schriftliche und mündliche Noten beinhalten kann und je nach Fach differenziert wird (Birkel, 1984a; Sacher, 2009). Die Form der Note wird wiederum vor dem Hintergrund pädagogischer und gesellschaftlicher Funktionen bestimmt. Gegenstand der folgenden Abhandlung ist die fachspezifische Zeugnisnote, die sowohl im Zwischen- als auch im Jahrgangsstufenzeugnis festgehalten wird und am Ende der Grundschule in einem Übertrittszeugnis mündet. Die Bildung von Noten – egal welcher Form – muss sich wiederum an schulrechtlichen Vorgaben orientieren, die in [Kapitel 2.5](#) expliziert werden. Identisch zu den im vorherigen Kapitel beschriebenen Bewertungsgrundlagen ist es auch für die Kategorisierung von Schulnoten maßgeblich, welche Bewertungsmaßstäbe bei der Notenkategorisierung herangezogen werden, und welche individuellen wie institutionellen Faktoren auf die Note Einfluss nehmen. So wird deutlich, dass die jeweilige Lehrkraft bei dem Leistungsergebnis eines Schülers keine unwesentliche Rolle einnimmt. Neben recht unspezifisch formulierten Rechtsgrundlagen ist es der Lehrkraft bzw. der Einzelschule überlassen, welche Informationen sie bei einer Benotung heranzieht, wie diese erhoben werden und an welchem Maßstab sie das gezeigte Verhalten misst (Faber & Billmann-Mahecha, 2010).

Unmittelbar mit den zugeschriebenen Funktionen verbunden sind die Konsequenzen, die sich aus der Bewertung mittels Noten ergeben können – für die beurteilten Schüler, die beurteilende Lehrkraft, die Schule selbst und nicht zuletzt für die Gesellschaft. Während die Benotung für Schüler u.a. bildungsbiografische- und motivational-emotionale

Konsequenzen nach sich ziehen kann und mögliche Wirkungen von Noten vornehmlich in der Sozialisationsforschung behandelt werden (Fend, 2008; Jachmann, 2003; Lintorf, 2012), wird vermutet, dass die Benotung auch für die Lehrkraft selbst Konsequenzen nach sich zieht. Tent und Birkel (2010) weisen darauf hin, dass die Einschätzung von Schülern mittels Noten für den Lehrer eine Beurteilung seiner eigenen Arbeit sei, obliege es doch ihm, dem Schüler die abgefragten Inhalte ausreichend zu vermitteln.

Zusammenfassend zeigt die Übersicht zu dem Begriff der schulischen Notengebung, dass sich neben schulrechtlichen und gesellschaftlichen Rahmenbedingungen auch historische und strukturelle Grundlagen für die Notengebung ergeben. Die Schulnote dient vornehmlich als Klassifikationsschema für Verhaltensbewertungen – in schulischen und gesellschaftlichen Kontexten – und wird entsprechend häufig als Indikator für gezeigte Schulleistungen verwendet. Als Ausgangspunkt aller folgenden Abhandlungen wird die Zeugnisnote – als eine spezifische Form der Schulnote – verwendet. Unerwähnt bleiben soll an dieser Stelle nicht, dass diese Note häufig Gegenstand kritischer theoretischer wie empirischer Abhandlungen ist (Ditton, 2010; Faber & Billmann-Mahecha, 2010; Ingenkamp, 1976a; Ziegenspeck & Lehmann, 1999). Während Kritiker damals wie heute den mangelnden objektiven, reliablen wie validen Informationsgehalt von Noten kritisieren, werden auch kritische Stimmen an der psychologisch-diagnostischen Überprüfung schulischer Noten lauter, zeichne sich beispielsweise die zu einem Zeitpunkt gemessene Intelligenz eines Schülers doch nur in geringem Maße in einer Schulnote ab (Furck, 1961; Lintorf, 2012; Weinert, 2014). Welche Komponenten in einer Schulnote Beachtung finden (sollten), hängt wiederum mit den besagten Funktionen und Zielen zusammen, die mit dem Leistungskonstrukt erfüllt werden sollen – Vorstellungen, die ebenso vielfältig sind und im nachfolgenden Kapitel weiter vertieft werden.

## 2.4 Notengebung in der Schule als funktionale Handlung

Die Auseinandersetzung mit dem schulischen Leistungsbegriff und der daraus resultierenden Notengebung kommt nicht umhin, sich mit den Funktionen von Schule und Leistungsbewertung zu beschäftigen. Funktionen, die es im Folgenden zu benennen gilt, um sich bewusst zu machen, welche mannigfachen, wechselnden Aufgaben – neben rechtlichen Belangen – Noten zugeschrieben werden, zumal sich diese Zuschreibung wieder mit den daraus resultierenden Folgen ändert. In der Literatur sind dazu bereits zahlreiche aktuelle wie zeitgeschichtliche Abhandlungen zu Einteilung und Folgen funktionaler Schulhandlungen erschienen (z.B. Dohse, 1976; Fend, 2008; Weiss, 1976a). Auffällig ist dabei sowohl die differente Betitelung und Ausformung zugeschriebener Funktionen als auch die vermutlich – in pädagogischer Literatur – gleichgesetzte Bedeutung mit schulischen Aufgaben und Zielen (z.B. Fend, 2008; Jachmann, 2003; Tillmann & Vollstädt, 2009; Titze, 2000). Im vorliegenden Kapitel wird eine Differenzierung zwischen den zugeschriebenen Funktionen für Schule und Notengebung vorgenommen. Nach einer kurzen Einführung in die der Institution Schule selbst zugeschriebenen Funktionen folgt eine Analyse der angesprochenen Funktionen an die schulinterne Notengebung<sup>7</sup>; so werden diese in der Literatur auch häufig nicht voneinander unterschieden (Hochweber, 2010; Tillmann & Vollstädt, 2009). Angemerkt wird außerdem, dass die Betitelung Funktionen den Aufgaben und Zielen von Schule und Notengebung gleichgesetzt werden, da diese Differenzierung zwischen den Begrifflichkeiten eine eigene Abhandlung benötigen würde. Ebenso kann auch hier festgehalten werden, dass die zitierten Arbeiten häufig älteren Datums sind, die Zuschreibungen aber aufgrund der seitdem wenig veränderten institutionellen

---

7 Um die Komplexität des Sachverhalts zu reduzieren, werden im Folgenden die Begriffe Notengebung, Leistungsbeurteilung, Leistungsfeststellung, Zeugniszensuren und Zeugnisse synonym verwendet. Obwohl eine Differenzierung in Anlehnung an die vorangegangenen Kapitel für die spezifischen Bewertungsformen notwendig ist, geht es im folgenden Kapitel um die übergestellten Anforderungen an die schulische Bewertung. Diese ist für alle Ausformungen ähnlich.

Rahmenbedingungen – mit Ausnahmen einiger neu hinzugekommenen Anforderungen – immer noch zutreffen.

Ziel des Kapitels ist es, einen Überblick über die den Noten zugeschriebenen Funktionen – in Abgrenzung zur Schule – zu geben. Damit soll der Frage nach dem *Warum von Leistungsbewertungen* nachgegangen werden, ist diese Frage doch häufig der zentrale Ausgangspunkt kontroverser Diskussionen um die Spannungsfelder zu Notengebung, Alternativen und deren Bedeutung für die Schulpraxis (Tillmann & Vollstädt, 2009). In Anlehnung an den betrachteten (schulischen) Leistungsbegriff sind es jene teils differenten gesellschaftlichen, historischen und fachlichen Vorstellungen darüber, welche Funktionen mit dem Leistungskonstrukt erfüllt werden sollen, die es im vorliegenden Kapitel zu behandeln gilt. Die Darstellung bezieht sich dabei allerdings nur auf Funktionen, die dem *deutschen* Schulwesen zugesprochen werden, würde eine kulturübergreifende Betrachtung doch das Ausmaß der Arbeit übersteigen.

### **Funktionen von Schule**

Begreift man die Schule als eine Institution, die sich tagtäglich mit den verschiedensten kulturellen, gesellschaftlichen, teils wechselnden und neuartigen Anforderungen konfrontiert sieht, müssen einige Grundannahmen der strukturellen, sozialisationstheoretischen Begebenheiten jener gesellschaftlichen Funktionszuschreibungen an die Schule betrachtet werden, wobei hier keinesfalls der Anspruch auf Vollständigkeit besteht. Aus strukturfunktionalistischer Systemperspektive kommt der Schule im Sozialisationsprozess die Aufgabe zu, die soziale Ordnung der Gesellschaft aufrecht zu erhalten (Niederbacher & Zimmermann, 2011). Ausgangspunkt ist die Annahme, schulische Praktiken nicht vornehmlich unter pädagogischer Perspektive zu beleuchten, sondern die Schule im Kontext der Gesellschaft zu verorten. Zentrale Fragestellungen beziehen sich hier nicht auf die Suche nach der einen richtigen und guten Schule, sondern fragen nach dem Beitrag von gewinnbringenden Strukturen für die Gesellschaft (Rademacher & Wernet, 2015). Nach Parsons (1979) kommen der Schulklasse zwei zentrale Aufgaben zu: die Verinnerlichung von gesellschaftlich aner-

kannten Rollen und die Verteilung dieser Rollen in der Gesellschaft. Die Einübung dieser hat hiernach eine zentrale schulische Funktion, die sich durch die Schüler-Lehrer Interaktion selbstregulierend entwickelt, um das gesamtgesellschaftliche System zu stabilisieren. Im Sinne dieser Rollenverinnerlichung kommt der stetigen Leistungsbewertung eine entscheidende, entwicklungsbestimmende Rolle zu. Während die Vergabe von Noten eine funktionale Notwendigkeit für die Differenzierung in Bildungs- und Berufslaufbahnen ist, erfolgt durch die hierarchische Struktur der Lehrer-Schüler Beziehung im besten Falle eine Identifikation des Schülers mit der ihm zugewiesenen Position, die auf zukünftige Rollen in der Gesellschaft vorbereiten soll (Niederbacher & Zimmermann, 2011; Parsons, 1979; Rademacher & Wernet, 2015).

Die zweite ausgewählte Perspektive zu schulischen Funktionen stammt von Fend (1981), der sich in seiner Theorie kritisch mit dem Parson'schen Erklärungsansatz auseinandersetzt und die theoretischen Annahmen zur Rolle der Schule bzw. Schulklasse im Sozialisationsprozess zu einer Theorie der Schule weiterentwickelt (Rademacher & Wernet, 2015). Ausgehend von einer übergeordneten Sozialisationsfunktion der Schule und der Erörterung über das Darf und Soll von Schule, ist Fends Einteilung Ausgangspunkt vieler Abhandlungen zum schulischen Funktionsbegriff (Fend, 2008; Herzog, 2011; Rademacher & Wernet, 2015; Rolff, 1997; Sacher, 2009; Tillmann & Vollstädt, 2009). Obgleich die einzelnen Funktionen der Schule im Folgenden nicht weiter vertieft werden, geht Fend (2008) – anders als Parsons – von einer Doppelfunktion des Bildungswesen aus. Dabei ist es Auftrag der Schule, einerseits die Stabilität der Gesellschaft durch die Formung des Individuums zu sichern, andererseits aber auch dem Einzelnen zu eigenen Handlungsweisen und Wertvorstellungen zu verhelfen (Fend, 1977). Nach Furck (1961) dürfe man aber nicht vergessen, den einzelnen Schüler und seine individuelle Bildsamkeit zu fördern. Jener von Furck (1961) befürchtete funktionale Bildungsgedanke überlagert nach Grundmann (2011) den emanzipatorischen Charakter von Bildung bis heute und benutzt die schulischen Handlungen als gesellschaftliches Instrument. Allerdings weisen Rademacher und Wernet (2015) darauf hin, dass erst die Benennung von Funktionen im Sinne

Führt eine differenzierte Analyse des funktionalen Zusammenhangs zwischen Schule und Gesellschaft möglich macht.

Betrachtet man die Institution Schule abschließend noch aus einer dritten, organisationstheoretischen Perspektive, ist es demnach nicht möglich, eine (soziale) Organisation in all ihren Abläufen funktionell zu durchschauen (Plake, 2010). Anders als die vorherigen Theorien geht es in diesem Ansatz nicht darum, der Schule spezifische Funktionen zuzuschreiben, sondern die Institution Schule in ihren Organisationsvorgängen mit all ihren internen Abläufen und Glaubenssätzen zu verstehen (Gomolla, 2012). Gemäß diesem neo-institutionalistischen Ansatz wird die Einzelschule als lose gekoppeltes System verstanden, deren alltägliche Abläufe im Einzelnen relativ autonom verlaufen und somit wenig Wissen über die eigentlichen Aktivitätsstrukturen existiert (Gomolla, 2012). Dennoch wird sie durch die Übernahme von gesellschaftlich anerkannten, für effizient erachteten Strukturen in ihren Handlungen legitimiert. Verkürzt betrachtet besteht nach Plake (2010) allerdings die Vermutung, dass jene Strukturen im alltäglichen Schulbetrieb nicht umgesetzt werden, da diese nicht zu den alltäglichen Anforderungen der Organisation Schule passten. Folge dieser fehlenden Übernahme sei damit eine schulinterne Entkopplung der Alltagsaktivitäten von den von außen auferlegten Formalstrukturen, legitimiert durch die Umsetzung struktureller Vorgaben; zum Beispiel durch die Vergabe von Noten (Gomolla, 2012; Plake, 2010; Tacke, 2006).

Bezieht man diesen theoretischen Ansatz zur Organisation von Schule auf die eingangs thematisierten Ansätze, könnte man davon ausgehen, dass eine noch so differenzierte Zusprechung von Funktionen an die Schule die realen Schulaktivitäten wenig berühren. In eine ähnliche Richtung diskutiert auch Breidenstein (2014) auf Grundlage von Befunden aus einer teilnehmenden Beobachtungsstudie (Zaborowski, Meier & Breidenstein, 2011). Demnach zeichne sich in der schulischen Notengebungspraxis eine ausgeprägte Eigenlogik ab, die von subtilen Zuschreibungen, Stigmatisierungen und einer häufig erst rückwirkenden Reflexion geprägt sei (Breidenstein, 2014). Theorieübergreifend

wird der Notengebung in allen drei Ansätzen eine institutionell notwendige Tätigkeit zugesprochen. Ausgehend von der Annahme, dass die Notengebung ein greifbarer Maßstab ist, der als Produkt schulischer Handlungen tatsächlich feststellbar ist, werden im Folgenden die nicht minder differenzierten funktionalen Zuschreibungen an die schulische Notengebung thematisiert; so sind diese doch neben den schulrechtlichen Grundlagen eine strukturelle Vorgabe, die bei der Notenvergabe explizit befolgt werden soll.

### Funktionen von Noten

Obgleich die Funktionen der Schule nach Furck (1961) nicht auf die Funktion der Leistungsfeststellung reduziert werden kann, ist für die Beantwortung der Frage nach dem *Warum der Leistungsbewertung* und ihren Bestandteilen ein differenzierter, historischer Blick auf die Funktionszuschreibungen an die Notengebung notwendig. So ist es für die explizierten Begriffe Leistung, Schulleistung und Schulnoten maßgeblich, welche gesellschaftlich und kulturell differenten Funktionen mit der Leistungsfeststellung erfüllt werden sollen. In Anlehnung an die in [Kapitel 2.1](#) erfolgte (historische) Beleuchtung des Leistungsprinzips war und ist es ein grundlegendes Ziel der Leistungsfeststellung, eine Einstufung und Bewertung gezeigter Handlungen vorzunehmen, um auf dessen Grundlage eine – leistungsbasierte – Bildungsselektion vorzunehmen. Während in den Anfängen des Berechtigungswesens mit der Einführung des Jahrgangsklassensystems<sup>8</sup> die Vergleichbarkeit von Jahrgangsklassen notwendig wurde, wurde die Zensur zum Instrument der Berechtigungsvergabe (Ingenkamp, 1995b). Ausgehend von einem auf Wunsch ausgestellten Benefizienzeugnis wurde den Zensuren – verkürzt beschrieben – mit der Weimarer Verfassung differenzierte Aufgaben zugesprochen, die nun für alle Leistungsnachweise gleichermaßen berücksichtigt werden mussten (Gomolla, 2012; Ingenkamp, 1995b). Ingenkamp (1995b) summiert zur historischen Betrachtung der Notengebung, dass die formalen Funktionen den pädagogischen Funktionen damals wie heute in ihrer Bedeutung

---

<sup>8</sup> Eine ausführliche Übersicht zu der Entwicklung des Schulzeugnisses ist bei Dohse (1976) zu finden.

allerdings weit überlegen waren. Während in der Vergangenheit auf Grundlage der Leistungsbewertung eine rangmäßig verteilte Sitzordnung hergestellt wurde, erfolgt auch heute noch durch die Bewertung eine personale Einstufung des Schülers (Dohse, 1976; Ziegenspeck & Lehmann, 1999). Während primär ältere Quellen Zensuren häufig fünf Hauptfunktionen zuschreiben<sup>9</sup>, werden in neueren Quellen neben unzähligen anderen Einteilungen inhaltlich häufig zwei große Funktionskategorien unterschieden: die gesellschaftlichen und pädagogischen Funktionsbereiche (Jürgens & Sacher, 2008; Langfeldt & Tent, 1999; Lintorf, 2012; Tillmann & Vollstädt, 2009).

### **Funktion von Noten: Gesellschaftliche Funktionszuschreibungen**

Zu den gesellschaftlichen Funktionen werden in diesem Kontext jene Bereiche gezählt, in denen Schülerbewertungen die augenscheinliche Grundlage für gesellschaftsrelevante Entscheidungen sind (Lintorf, 2012). Sie umfassen die Selektions- oder Allokationsfunktion, die Sozialisations- oder Qualifikationsfunktion, die Legitimationsfunktion und die Kontrollfunktion, wobei diese Einteilung sicherlich kontrovers diskutiert werden könnte.

Die wohl am häufigsten diskutierte *Selektions- oder Allokationsfunktion* bezieht sich auf die Aufgabe, Schüler anhand ihrer Noten entsprechenden Bildungsgängen zuzuordnen. Demnach dient die Zensur der Funktion der Auslese und Berechtigung; Schüler werden z.B. zurückgestellt oder für den Besuch einer weiterführenden Schulform berechtigt (Dohse, 1976; Tillmann & Vollstädt, 2009; Weiss, 1976a). Während dieser Prozess einen innerschulischen Ablauf umfasst, wird das entsprechende Ergebnis von außerschulischen Erwartungen gesteuert; so ist es die direkte Aufgabe des Schulsystems, Schüler auf Grundlage ihrer individuellen Leistung für zukünftige Positionen im Beschäftigungssystem zu berechtigen. Indirekt erfolgt vermutlich eine Verinnerlichung auf Seiten des Schülers, das Leistungskriterium als Maßstab zur Verteilung unterschiedlicher Chancen anzusehen (Tillmann & Vollstädt, 2009).

---

9 Dohse (1976) unterscheidet z.B. zwischen Auslesefunktion, Kontrollfunktion, rechtlicher Funktion, Anreizfunktion und pädagogischer Funktion.

Ungeklärt bleibt jedoch auch hier, welche Bestandteile Leistung umfasst, *nach was* also selektiert werden soll. So sollte nach Furck (1961) nicht nur die – auch von ihm undeutlich definierte – Leistung, sondern auch der Grad der Bildsamkeit des Schülers berücksichtigt werden. An jenen Internalisierungsgedanken schließt der Grundsatz der *Sozialisations- und Qualifikationsfunktion* an. Ausgehend von der Annahme, dass die Sozialisationsfunktion von Noten unbewusst erfolgt, wird konstatiert, dass durch die Vergabe von Noten ein verengtes, schulisches Leistungsverständnis internalisiert wird, im Zuge dessen die Einübung berufsrelevanter Fähigkeiten den größten Stellenwert einnimmt (Sacher, 2009). Jene Einübung qualifiziert dann den Heranwachsenden, dem entsprechenden *Qualifikationsniveau* im Beschäftigungssystem zugeteilt zu werden (Tillmann & Vollstädt, 2009). Mittels dieser zweckdienlichen Lernstandserhebung erfolgt nach Sacher (2009) eine *Legitimation* administrativer und unterrichtlicher Handlungen. So können Noten beispielsweise finanzielle Unterstützung für bestimmte Bildungsbereiche oder strukturelle Neuorientierungen rechtfertigen. Ebenso wird dieser Legitimationsfunktion auch ein Nutzen für Schule und Lehrer zugesprochen, da durch die entsprechende Notenbasis indirekt immer eine Legitimation der entsprechenden (internen) Schulmaßnahmen erfolgt (Sacher, 2009). An jene Legitimationsfunktion schließt die *Kontrollfunktion* der Notengebung an, wobei diese sowohl für gesellschaftliche als auch pädagogische Zwecke dienlich sein kann. Dienlich sind die Noteninformationen demnach für die beteiligten Lehrkräfte und ihren Unterricht. Sie informieren über den Leistungsstand der Schüler und überprüfen die zugrundeliegenden Lehrpläne auf ihre Stimmigkeit. Ebenso dienen jene Daten u.a. als Grundlage für das im internationalen Vergleich durchgeführte Bildungsmonitoring (Prenzel, Sälzer, Klieme & Köller, 2013; Sacher, 2009).

Zusätzlich zu jenen, bereits lange Zeit diskutierten Funktionsbereichen sind in neuerer Zeit unzählige Funktionszuschreibungen hinzugekommen. Als logische Folge der Selektionsfunktion wird in diesem Zusammenhang immer häufiger von einer Prognosefunktion gesprochen, indem davon ausgegangen wird, dass die zensierte Leistung des

Schülers ein konstantes Verhalten widerspiegeln soll, so dass eine prognostisch valide Notenvergabe erfolgt (Hochweber, 2010). Ebenso häufig wird die diagnostische Funktion der Schulleistungsdiagnostik diskutiert, in der mittels einer validen und prognostisch korrekten Überprüfung eine treffsichere Beurteilung des Schülerverhaltens, der Lernziele und Lehr-Lernprozesse erfolgen soll (Heller, 1984; Hesse & Latzko, 2012).

### **Funktion von Noten: Pädagogische Funktionszuschreibungen**

Unter den pädagogischen Funktionen der Notengebung werden jene Vorgaben an die Leistungsbewertung zusammengefasst, die zum Fortschritt des individuellen Lernprozesses beitragen – sie werden in der Literatur allerdings bezüglich ihrer pädagogischen Intentionen kontrovers diskutiert. Differenziert werden häufig die Rückmeldungs- und Optimierungsfunktion und die Berichtsfunktion für Lehrkräfte, Eltern und Schüler (Tillmann & Vollstädt, 2009). Vorrangig dienlich sein soll die Bewertung durch Noten als *Rückmeldung und Bericht* über gezeigte Leistungen an die besagten Personengruppen, zudem wird der Schulnote im Sinne einer Anreizfunktion eine motivierende Funktion zugesprochen. Diese wird allerdings aufgrund von Befunden zu demotivierenden Folgen von Noten und spezifischen Unterformen motivationaler Wirkungen<sup>10</sup> kontrovers diskutiert (Arnold & Jürgens, 2001; Breidenstein, Meier & Zaborowski, 2011; Heckhausen, 1974; Lintorf, 2012; Rieder, 1990; Weiss, 1976a). Eine disziplinierende Funktion von Noten sollte wiederum mit einer realistischeren Selbsteinschätzung, positiv konnotiertem Schulverhalten und einer größeren Anstrengungsbereitschaft einhergehen (Sacher, 2009).

Die kritische Beleuchtung der – gesellschaftlichen wie pädagogischen – funktionalen Zuschreibungen an die Notengebung und deren sozialschichtspezifischer Legitimationscharakter wird bereits lange Zeit diskutiert (Bourdieu, 1976; Dohse, 1967, 1976; Ingenkamp, 1976a; Jürgens & Lissmann, 2015; Kalthoff, 1996; Weiss, 1976a). Während Sacher (2009) von einer funktionalen Überfrachtung der schulischen

---

<sup>10</sup> Weiss (1976a) unterscheidet dabei zwischen sozialen und sachlichen Motivationsanreizen.

Leistungsbeurteilung spricht, diskutiert bereits Furck (1961) das Spannungsverhältnis zwischen Bildungswert und Funktionalitätsgedanke, welches die Leistungsbeurteilung nur unzureichend kläre. Zudem sind die widersprüchlichen Anforderungen an die Leistungsbewertung Gegenstand zahlreicher Abhandlungen (Arnold & Jürgens, 2001; Sacher, 2009; Tillmann & Vollstädt, 2009; Ziegenspeck & Lehmann, 1999). Während Ingenkamp (1995b) die Funktionszuschreibungen wegen ihrer Fülle an differenten, teils widersprüchlichen Aussagen angreift und ihm sich der fortwährende Einsatz von Noten nur aufgrund ihrer vielseitigen, einfachen Handhabung erschließt, weist schon Sommer (1983) trotz aller Kritik auf die Vorteile der Funktionszuschreibungen an die Leistungsbewertung hin. Demnach hätten diese im historischen Kontext letztlich zu einer gegliederten Schulstruktur geführt, die wiederum den Schülern Rückmeldung geben könnte, in welcher gesellschaftlichen Funktion sie sich befänden (Sommer, 1983). Die Frage der Bedeutung der Erhebungsperspektive und der fremdbestimmten Funktionszuweisung wird allerdings weiterhin nicht gelöst. Ebenso widersprüchlich ist der zugleich fördernde und selektive Funktionsgedanke und die hohe diagnostische Anforderung an die Benotung bei gleichzeitig unzureichender diagnostischer Kompetenzermittlung (Tillmann & Vollstädt, 2009). Kritisch beleuchtet wird also vor allem die große Spannweite der Anforderungen, welche durch eine einzige Note erfüllt werden soll.

Auch wenn die dargelegten Befunde theoretischer Art sind und deutlich von empirischen Evidenzen zur Notengebung getrennt werden müssen, sind es doch diese Bestimmungen, welche die Art der Fragen an die Praxis der Notengebung bestimmen (Breidenstein et al., 2011; Titze, 2000). Betrachtet man beispielsweise die Selektions- und Allokationsfunktion der Notengebung, ergibt sich erstens die empirisch zu lösende Frage, welcher diagnostische und prognostische Wert von einer Benotung ausgeht, wie *valide* die Selektion also ist. Zweitens gilt zu beantworten, mittels welcher Beurteilungsform und welchem Maßstab die Benotung zustande kam und drittens, wie vergleichbar die Noten letztlich über Schulen und Schulformen hinweg sein können (Breidenstein et al., 2011; Terhart, 1999; Ziegenspeck & Lehmann, 1999).

So besteht bei all den vielfältigen Anforderungen an die Notengebung immer die Frage, welche Funktionen eine einzelne Note primär erfüllt und ob spezifische Noten auch differente Funktionen erfüllen können. Bei den pädagogischen Funktionen wird von einer gänzlich fehlerhaften Zuschreibung pädagogischer Aspekte an die Notengebung gesprochen, könne man doch mit einer einzigen Note niemals alle möglichen Folgen und pädagogisch intendierten Belange widerspiegeln und aus einem vornehmlich einstufigen Instrument kein pädagogisch wirksames Fazit ziehen (Arnold & Jürgens, 2001; Jürgens & Sacher, 2008; Saldern, 1997). Empirische Fragen an die pädagogische Funktion der Notengebung ergeben sich z.B. aus dem Forschungsstand zur Attribuierung von schulischen Noten, die von der Bezugsgruppe beeinflusst werden und Folgen für die Selbsteinschätzung, Selbstkonzeptentwicklung und die gesamte Persönlichkeitsentwicklung von Schülern haben können (Jerusalem & Schwarzer, 1991; Möller & Jerusalem, 1997). Zudem sind es vor allem ältere Studien, die sich mit der Bedeutung von Funktionszuschreibungen für beteiligte Personengruppen beschäftigen.

Festgehalten wird, dass die Frage nach dem *Warum von Leistungsbeurteilung* mit einer theoretischen Analyse gesellschaftlicher Rahmenbedingungen von Schule und Notengebung beantwortet werden muss. Die kritische Reflexion der der Notengebung zugeschriebenen Funktionen zeigt auf, dass die Voraussetzungen, unter denen die Funktionen überhaupt die zugeschriebenen Erwartungen erfüllen können, teils divergierend sind (Trautwein, Maaz & Baeriswyl, 2014). Obwohl oftmals nicht klar ist, ob eine Funktion tatsächlich umgesetzt wird, existieren bislang wenige Abhandlungen darüber, über welche Eigenschaften Schulnoten verfügen müssen, um die ihr zugesprochenen Funktionen zu erfüllen. Auch wenn institutionelle Einflüsse im Rahmen dieser Arbeit nur indirekt geprüft werden, ist die Untersuchung von Noten und ihrer Einflussfaktoren bis heute das einzig valide Instrument, um empirisch Informationen darüber zu erhalten, welche Funktionen sich in den Bestandteilen einer Note widerspiegeln. Und auch wenn für Ingenkamp (1976a) die empirische Prüfung von Funktionen der Notengebung nahezu unmöglich erscheint, erfolgt die

Überprüfung bislang häufig über messtheoretische Gütekriterien, die wiederum Aussagen über die Tauglichkeit der Notengebung bezüglich der ihr zugeschriebenen Funktionen machen (Ingenkamp & Lissmann, 2005; Jürgens & Sacher, 2008). Nach der begrifflichen Klärung und kritischen Reflexion der Funktionen von Schule und Leistungsbewertung wird im Folgenden darauf eingegangen, welche konkreten Anforderungen von Seiten des Schulrechts an die schulische Notengebung gestellt werden.

## 2.5 Schulrechtliche Vorgaben der Notengebung

Das deutsche Prüfungsrecht ist durch eine Vielzahl von Einzelfallentscheidungen geprägt (Rademacher, 2009). Im Folgenden werden jene Rechtsgrundlagen vorgestellt, die formal für alle schulischen Beurteilungsvorgänge gelten. Die Darstellung bezieht sich zunächst auf länder- und schulartübergreifende Rechtsgrundlagen, die unabhängig von den Landesschulgesetzen für alle Bundesländer gleichermaßen gelten. Anschließend daran wird in Anbetracht der dieser Arbeit zugrunde gelegten Untersuchungsstichprobe aus Koala-S auf die aktuelle Gesetzeslage in der Grundschule eingegangen, wobei die Bundesländer Bayern und Sachsen und deren fachspezifische Regelungen vertieft betrachtet werden. Diese Ausführung bezieht sich immer auf Rechtsgrundlagen, die auch zum Zeitpunkt der vorliegenden Koala-S-Studie (2005-2007) bis auf kleine aktuelle Änderungen geltend waren. Generell gilt außerdem, dass die Verbindlichkeit der gesetzlichen Vorgaben von den durch die Benotung ausgelösten Folgen abhängt; während pädagogisch intendierte Bewertungen weniger starr und differenziert festgeschrieben sind, müssen für den Schüler richtungsweisende Benotungen zwingend rechtlichen Grundlagen folgen (Füssel, 2010). Angemerkt wird zudem, dass die beschriebenen rechtlichen Anforderungen vornehmlich bei berufsqualifizierenden und übergangsrelevanten Benotungen Anwendung finden, diese allerdings auch für andere schulische Leistungsbewertungen gelten (Füssel, 2010; Rademacher, 2009).

Die aktuelle, länderübergreifende Rechtsgrundlage der Notengebung beruht auf einem Beschluss der KMK aus dem Jahr 1968, der eine lernzielbezogene Leistungsbeurteilung durch Noten deklariert (KMK, 1968). Ein Überblick zu den sechs Notenstufen in Ziffern und zugrundeliegenden Definitionen gibt die [Tabelle 1](#).

Tabelle 1: Notenstufen KMK (1968, S.1)

Bezeichnung	Ziffer	Definition
sehr gut	1	Die Note „sehr gut“ soll erteilt werden, wenn die Leistung den Anforderungen in besonderem Maße entspricht.
gut	2	Die Note „gut“ soll erteilt werden, wenn die Leistung den Anforderungen voll entspricht.
befriedigend	3	Die Note „befriedigend“ soll erteilt werden, wenn die Leistung im allgemeinen den Anforderungen entspricht.
ausreichend	4	Die Note „ausreichend“ soll erteilt werden, wenn die Leistung zwar Mängel aufweist, aber im ganzen den Anforderungen noch entspricht.
mangelhaft	5	Die Note „mangelhaft“ soll erteilt werden, wenn die Leistung den Anforderungen nicht entspricht, jedoch erkennen läßt, daß die notwendigen Grundkenntnisse vorhanden sind und die Mängel in absehbarer Zeit behoben werden könnten.
ungenügend	6	Die Note „ungenügend“ soll erteilt werden, wenn die Leistung den Anforderungen nicht entspricht und selbst die Grundkenntnisse so lückenhaft sind, daß die Mängel in absehbarer Zeit nicht behoben werden könnten.

Der Maßstab für die Zuordnung zu einer der sechs Notenstufen soll je nach *Anforderung* in Schulart, Jahrgangsstufe und Fach festgelegt werden, wobei eine niedrigere Note eine höhere Leistung bedeutet (Füssel, 2010; Ingenkamp, 1976a; KMK, 1968). Die spezifische Differenzierung des Begriffs *Anforderungen* bezieht sich laut KMK (1968) auf den Umfang und die selbstständige, richtige Anwendung von Kenntnissen und deren Darstellung. Die fehlende vertiefte Erläuterung jener Anforderungen sowie die mangelnde Ausdifferenzierung der Notenstufen wird vielfach wegen ihrer Interpretationsspielräume kritisiert (z.B. Ingenkamp, 1976a; Ziegenspeck & Lehmann, 1999). Auf der

anderen Seite wird die Einführung von Bildungsstandards als Orientierungsrahmen für die Lehrkräfte diskutiert (KMK, 2004).

Länderübergreifende Einheitlichkeit besteht bezüglich der Bewertung einzelner Schulleistungen. Differenziert wird in Bewertungen von Einzelleistungen in schriftlicher, mündlicher und praktischer Form und in Bewertungen durch Zeugnisse, in welchen die gesamten Kenntnisse und Fähigkeiten der Schüler aufgeführt werden (Füssel, 2010). Neben unzähligen Funktionen, welche die Benotung durch Zeugnisse nach sich zieht (vgl. [Kap. 2.4](#)), ist es Zeugnissen vorbehalten, zum Ende jedes Schulhalbjahres zusammenfassend aus mehreren Einzelnoten fachspezifische Bewertungen abzugeben<sup>11</sup> (Füssel, 2010). Bei der Bildung der Zeugnisnote muss nach aktueller Rechtslage Umfang und Schwierigkeitsgrad der erbrachten Leistungsnachweise berücksichtigt werden. Klassenarbeiten fallen stärker als mündliche Leistungen ins Gewicht. Die Art und Weise der mündlichen Leistungsprüfung obliegt dabei dem pädagogischen Ermessensspielraum der jeweiligen Lehrkraft, sowie auch das Standardwerk für Schulrechtsfragen von Avenarius und Füssel (2010) festhält, dass eine mündliche Nachprüfung bei Schülern mit unklarem Notenbild ebenfalls nach Ermessen möglich ist. Im Rahmen des schulischen Erziehungsauftrags ist es der Schule rechtlich außerdem erlaubt, das gezeigte Sozialverhalten in Form von Kopfnoten oder verbalisierter Form separat im Zeugnis festzuhalten, nicht aber den elterlichen Erziehungsvorstellungen entgegen zu wirken. Ebenfalls ist es – mit entsprechenden länderspezifischen Ausformungen – erlaubt, Mitarbeit und Verhalten nach Vorschlägen einzelner Lehrer im Zeugnis aufzuführen sowie die Leistungsentwicklung und das Lernverhalten bei der Notenvergabe zu berücksichtigen (Füssel, 2010).

Neben länderspezifischen Ausformungen besteht in Rechtsprechung und Fachliteratur Einigkeit darüber, dass es sich bei Leistungsbewertungen immer um einen individuellen Beurteilungsvorgang handelt.

---

11 Zu den möglichen Fehlerquellen, die sich aus der Berechnung eines Notendurchschnitts ergeben können, findet sich bei Sacher (1985) eine hilfreiche Übersicht sowie [Kapitel 2.3](#) die Durchschnittsbildung in der Notengebung diskutiert.

Lehrkräften wird rechtlich explizit ein Beurteilungs- und Bewertungsspielraum zugesprochen (Füssel, 2010; Rademacher, 2009; Ziegen-speck & Lehmann, 1999). Begründet wird dies durch die Tatsache,

[...] dass Prüfungsnoten nicht isoliert gesehen werden dürfen, sondern in einem Bezugssystem zu finden sind, das durch die persönlichen Erfahrungen und Vorstellungen der Prüfer beeinflusst wird; es widerspräche dem das Prüfungsverfahren prägenden Grundsatz der Chancengleichheit [...], wenn einzelne Kandidaten bei einer rechtlichen Überprüfung die Chance einer vom Vergleichsrahmen unabhängigen Bewertung erhielten. (Füssel, 2010, S. 443).

Abgeleitet aus der pädagogischen Funktion der Leistungsbewertung ist es inhaltlich untersagt, rein individuelle Bewertungsmaßstäbe für einzelne Schüler anzuwenden, vor allem bei Abschlussprüfungen: „Maßstab der Leistungsbewertung sind vielmehr die verbindlichen Anforderungen, wie sie sich insbesondere aus den Lehrplänen bzw. Bildungsstandards ergeben [...], auch wenn vergleichende Betrachtungen innerhalb der Lerngruppe zulässig sind.“ (Füssel, 2010, S.446-447). Gleichzeitig wird bezüglich der Bewertungsmaßstäbe untersagt, durch eine vollkommene Relativierung der Leistungsbeurteilung das Gebot der Chancengleichheit zu missachten. So sei es die Pflicht jeder Schule, gerade bei qualifizierenden Bewertungen generelle Standards heranzuziehen, um eine Benachteiligung durch absolute und relative Bewertungsmaßstäbe zu verhindern (Füssel, 2010).

Die inhaltlichen Anforderungen an schulische Bewertungsverfahren geben u.a. vor, dass bei einer Einzelbenotung nur die aktuell gezeigte Leistung bewertet werden darf und weder ehemalige noch zukünftig erwartete Leistungen Beachtung finden sollten; andererseits ist es zulässig, dass sich die Lehrkraft über vorausgegangene Bewertungen anderer Prüfer informiert. Ebenfalls wird bundeslandübergreifend festgehalten, dass formale Schwächen schriftlicher Arbeiten nur dann bewertet werden dürfen, wenn dies explizit Gegenstand der Prüfung war (Füssel, 2010). Rechtlich vorangestellt werden dem erheblichen Entscheidungsspielraum der Lehrkräfte allerdings einige Grenzen.

Gemäß der Rechtsprechung sei der Bewertungsspielraum dann überschritten, wenn die allgemeingültigen Bewertungsmaßstäbe überschritten oder sachfremde Erwägungen mit in die Bewertung aufgenommen werden<sup>12</sup>.

Neben dieser landesübergreifenden Rechtsgrundlage liegen je nach Schulart, Fach und Bundesland weitere, teils differente Bewertungsrichtlinien vor (Lintorf, 2012). Grundlage dieser länderspezifischen Ausformungen bilden die jeweils durch die KMK erlassenen, schulartspezifischen Bildungsstandards, die für die jeweiligen Schulen als Leitziel deklariert werden (KMK, 2004). Für die Grundschule – jene Schulart, aus der die Daten der vorliegenden Studie stammen – wird von Seiten des Schulrechts vorgegeben, in den Jahrgangsstufen 1 und 2, anstelle einer Benotung durch Ziffern, einen Bericht über individuelle Stärken, Schwächen und Fortschritte des Schülers zu verfassen. Frühestens am Ende der zweiten Jahrgangsstufe erfolgt dann eine Bewertung durch (Halb-)Jahreszeugnisse. Neben den oben beschriebenen, allgemeingültigen Bewertungsmaßstäben werden in Grundschulen individuelle Zeugnisnoten mit dem Leistungsniveau der Bezugsgruppe verglichen (KMK, 2014). Zusätzlich zu der expliziten Befürwortung der sozialen Bezugsnorm wird es den Bundesländern von Rechts wegen freigestellt, eine Beurteilung des Lern- und Sozialverhaltens mit in das Zeugnis aufzunehmen – einer Empfehlung, der laut KMK (2014) etwa die Hälfte aller Bundesländer nachkommt, während diese Beurteilung in den übrigen Ländern zum Teil kontroverse Diskussionen auslöst.

In ihren Benotungs- bzw. Zeugnisregelungen für den Primarbereich unterscheiden sich die Bundesländer Sachsen und Bayern organisatorisch nur insofern, dass in Sachsen bereits in der zweiten Jahrgangsstufe im Halbjahreszeugnis eine Bewertung durch Noten vor-

---

12 Gemäß der Rechtslage liegen sachfremde Erwägungen dann vor, wenn das Gebot der Fairness und Sachlichkeit z.B. durch voreingenommene Randbemerkungen des Prüfers eingeschränkt werden. Allerdings läge nach Rademacher (2009) eine solche Überschreitung nur dann vor, wenn man die fehlende Distanz und mangelnde Sachlichkeit explizit nachweisen könne.

genommen wird; für bayerische Grundschulen gilt dies erst Ende der Jahrgangsstufe 2. In beiden Ländern bilden die erlassenen Lehrpläne, Stundentafeln und definierten Bildungsstandards die Grundlage für die Leistungsanforderungen, welche anhand schriftlicher, mündlicher und praktischer Leistungen erhoben werden. Auch führen beide Schulordnungen an, bei Einzel- und Zeugnisnoten zusätzlich eine Bewertung des Schülerverhaltens vorzunehmen, wobei die Ausformung dieser Bewertung unterschiedlich gehandhabt wird (Bayerisches Staatsministerium für Unterricht und Kultus, 2008; Sächsische Staatskanzlei, 2013).

In Bayern kann nach § 38 GrSo bei der Bewertung schriftlicher Arbeiten die äußere Form berücksichtigt werden, sprachliche Verstöße und schwere Ausdrucksmängel seien zu kennzeichnen, wobei in Einzelfällen (z.B. bei Schülern mit sonderpädagogischem Förderbedarf oder nicht deutscher Muttersprache) davon abgesehen werden soll. Während die Einzelbenotung in Anlehnung an die von der KMK (1968) verfassten Notenstufen erfolgt, wird in den bayerischen Zwischen- und Jahreszeugnissen neben Noten für die Pflichtfächer Deutsch, Mathematik und Heimat- und Sachkunde eine Bewertung des Sozial-, Lern- und Arbeitsverhaltens nach vier Stufen vorgenommen (sehr gut, gut, befriedigend, nicht-befriedigend). Das Arbeits- und Lernverhalten wird wiederum in die Oberbegriffe Interesse und Motivation, Lern- und Arbeitsweise sowie Konzentration und Ausdauer des Schülers differenziert. Das Sozialverhalten umfasst dagegen die soziale Verantwortung, Kooperation, Kommunikation und das Konfliktverhalten des Schülers (Staatsinstitut für Schulqualität und Bildungsforschung München, 2015). Zeugnisnote und Bewertung des Schülerverhaltens werden vom Klassenlehrer nach Rücksprache mit den in der Klasse unterrichtenden Lehrkräften angegeben, wobei die Bildung der Einzelnoten „in pädagogischer Verantwortung“ erfolgt (GrSO, § 43, Abs. 7). In der vierten Jahrgangsstufe wird in Bayern das Zwischenzeugnis durch ein Übertrittszeugnis ersetzt, das sich in Anlehnung an die Empfehlung der KMK (2015) zusätzlich zu den anderen Bewertungsmaßstäben auf die für den Schulerfolg allgemein wichti-

gen Fähigkeiten stützen soll, die wiederum in diesem Zusammenhang nicht spezifisch definiert werden.

In Sachsen wird als Maßstab für die Leistungsbewertung in der Primarstufe der individuelle Lernfortschritt des Schülers angeführt, welcher bei der Benotung der Fächer Deutsch, Mathematik und Sachunterricht berücksichtigt werden soll (Sächsische Staatskanzlei, 2013). Während die Rechtslage in diesem Bundesland ebenfalls vorgibt, Einzelnoten in Anlehnung an die von der KMK (1968) deklarierten Notenstufen zu vergeben, wird zusätzlich festgehalten, dass sich die Ausrichtung der Notenstufen immer an der „[...]selbst-ständigen und richtigen Anwendung der Kenntnisse, Fähigkeiten und Fertigkeiten, einschließlich der Art der Darstellung“ (SOGS, § 15, Abs. 4) orientieren soll. Über mögliche Nachprüfungen oder den Umgang mit nicht erbrachten Leistungen habe die Lehrkraft je nach Alter und Entwicklungsstand des Schülers zu entscheiden. In sächsischen Grundschulen werden Mitarbeit, Fleiß, Ordnung und Betragen – anders als in Bayern – in fünf Stufen (sehr gut 1 – mangelhaft 5) benotet. Die Charakterisierung dieser Verhaltensmerkmale werden in § 15 SOGS Abs. 6 differenziert aufgeführt und nachfolgend in [Tabelle 2](#) der bayerischen Schulordnung zur Bewertung des Sozial-, Arbeits- und Lernverhaltens gegenübergestellt (vgl. Sächsische Staatskanzlei, 2013; Staatsinstitut für Schulqualität und Bildungsforschung München, 2005). Unschwer zu erkennen ist, dass die einzelnen Inhaltskategorien wiederum durch eine Vielzahl an Untereigenschaften charakterisiert werden.

Die Bewertung in den Fächern Deutsch und Mathematik erfolgt auf Grundlage der fachspezifischen Lehrpläne. Für die Fächer Deutsch und Mathematik liegen für Bayern und Sachsen von der KMK Bildungsstandards vor, die im Folgenden aufgrund ihrer Differenziertheit nicht ausführlich expliziert werden (KMK, 2005a, 2005b). Festgehalten wird lediglich, dass darin die inhaltlich zu vermittelnden Kompetenzbereiche in ihren Ausformungen beschrieben und voneinander unterschieden werden (z.B. Sprechen und Zuhören, Zahlen und Operationen etc.) (KMK, 2005a, 2005b). Als Grundlage für die fachspezifische Bewertung dienen wieder mündliche, schriftliche und praktische Leistungen (Sächsische Staatskanzlei, 2013).

Tabelle 2: Auszug der bayerischen und sächsischen Schulordnung zur Bewertung des Sozial-, Arbeits- und Lernverhaltens

<b>Schulordnung Sachsen</b>		<b>Schulordnung Bayern</b>	
Betragen	Aufmerksamkeit, Hilfsbereitschaft, Zivilcourage Angemessener Umgang mit Konflikten, Rücksichtnahme, Toleranz/Gemeinsinn sowie Selbsteinschätzung	Sozialverhalten	Soziale Verantwortung Kooperation Konfliktverhalten Kommunikation
Fleiß	Lernbereitschaft, Ausdauer und Regelmäßigkeit beim Erfüllen von Aufgaben	Lern- und Arbeitsverhalten	Interesse und Motivation Lern- und Arbeitsweise Konzentration und Ausdauer
Mitarbeit	Initiative, Kooperationsbereitschaft, Teamfähigkeit, Beteiligung im Unterricht, Selbstständigkeit, Kreativität sowie Verantwortungsbewusstsein		
Ordnung	Sorgfalt, Pünktlichkeit, Zuverlässigkeit, Einhalten von Regeln/Absprachen sowie Bereithalten notwendiger Unterrichtsmaterialien		
Notenbereich	1-5		1-4

Zusammenfassend ist festzuhalten, dass im deutschen Schulrecht den Einzelschulen und den darin tätigen Lehrkräften mit ihren Unterrichtsanforderungen ein gewichtiger Stellenwert bei der schulischen Leistungsbewertung zukommt. Neben schulart-, fach- und länderspezifischen Unterschieden in den Regelungen für die Zusammensetzung von Einzel- und Zeugnisnoten, hält die Rechtsprechung übergreifend fest, dass die Auslegung bzw. Zuordnung zu den jeweiligen Notestufen einen verantwortungsvollen – individuellen – pädagogischen Bewertungsvorgang voraussetzt. Ein festgeschriebener Freiraum bei dem offen bleibt, ob sich der Bewertungsvorgang je nach Lehrkraft und schulischem Sozialkontext voneinander unterscheidet und in unterschiedlichen Noten für dieselbe Leistung mündet. Unterschiede, die dann häufig mithilfe von unabhängigen Leistungstests überprüft

werden. Die Bedeutung dieser Überprüfung von Noten mittels messtheoretischer Gütekriterien wird nach einer Zusammenfassung der zentralen Grundgedanken dieses Kapitels in [Kapitel 3](#) weiter vertieft.

## 2.6 Zusammenfassung zentraler Annahmen

Ziel des zweiten Kapitels war es, die Komplexität des schulischen Leistungsbegriffs mit seinen historischen und aktuellen Bezügen zu beleuchten und auf die, für diese Arbeit wesentlichen, begrifflichen Grundbezüge zu reduzieren.

Ausgehend von dem allgemeinen Leistungsbegriff wird deutlich, dass dieser je nach Fachbereich unterschiedliche Leistungsaspekte anspricht, wodurch sich die Mehrdeutigkeit des Begriffs widerspiegelt. Allgemein bindet Leistung meist Arbeit und Zeit, umfasst Arbeitsprozesse und Arbeitsergebnisse und wird häufig auf wirtschaftliche und gesellschaftliche Maßstäbe reduziert (Ziegenspeck & Lehmann, 1999). Fehlende allgemeingültige Bewertungsmaßstäbe und kontextuelle Einflüsse lassen eine Leistungsbewertung – unabhängig vom Schulkontext – als subjektive Handlung erscheinen, deren zugrundeliegendes Leistungsprinzip als Kriterium sozialer Verteilungsgerechtigkeit gilt.

Betrachtet man den Leistungsbegriff im schulischen Kontext, werden für eine pädagogische Begriffsannäherung beteiligte Akteure, Bewertungsmaßstäbe, Bewertungsdimensionen und Arten der Erhebung relevant. Ausgehend von der Annahme Furcks (1961) Leistung als Anforderung an den Schüler, dessen Tätigkeit und Handlungsergebnis sowie gesellschaftlichen Nutzen zu beschreiben, ist für die begriffliche Einordnung maßgeblich, welche Zeitpunkte Leistung umfasst (Lernprozess, Lernergebnis, zukünftige Leistung) und durch welchen Maßstab diese bewertet wird. Differenziert nach *Leistungsfeststellung* und *-bewertung*, muss für die Bewertung der soziale, kriteriale oder individuelle Maßstab herangezogen werden, woraus sich unterschiedliche

Konsequenzen für die zu bewertende Schulleistung und die daraus abgeleiteten Folgen ergeben (Ditton, 2016).

Versteht man Schulleistung als ein abstraktes, mehrdeutiges Konstrukt, das man sowohl theoretisch als auch empirisch fassen kann, wird Schulleistung im Schulkontext durch Noten festgehalten, die wiederum als Indikator für Bildungserfolg herangezogen werden. Gemessen auf Ordinalskalenniveau existieren unterschiedliche Arten von Schulnoten: Einzel- und Zeugnisnoten können schriftliche oder mündliche Leistungen beinhalten, für die vorliegende Arbeit fokussiert werden fachspezifische Zeugnisnoten am Ende der Grundschulzeit.

Für die begriffliche Einordnung von Schulleistung ist auch die Frage nach den einzubeziehenden Schülermerkmalen entscheidend. Gefragt wird z.B., ob ausschließlich kognitive Fähigkeiten oder auch motivationale Aspekte zu berücksichtigen sind und welche kontextuelle Faktoren eine Rolle spielen; welche Eigenschaften also letztlich als Schulleistung bezeichnet werden. Die Frage nach den einbezogenen Schülereigenschaften hängt unmittelbar mit der Funktionszuschreibung an die Leistungsbeurteilung zusammen, welche Funktionen also durch die Notengebung erfüllt werden sollen.

Als Klassifikationsschema von gezeigtem Schülerverhalten gilt für Schulnoten ebenfalls die Frage, welchen Einflüssen diese unterliegen (dürfen) sowie die Form der Note je nach Anforderung gewählt wird.

Bei den allgemeinen Funktionen der Schule und der darin verorteten Notengebung werden pädagogische und gesellschaftliche Funktionen unterschieden, die vornehmlich theoretisch deklariert und häufig nur implizit empirisch geprüft werden. Geleitet durch die Frage nach dem *Warum von Leistungsbewertung* macht die Abhandlung aus [Kapitel 2.4](#) deutlich, dass die pädagogischen und gesellschaftlichen Erwartungen an die Notengebung teils widersprüchlich sind. Ebenso existieren bislang wenige Abhandlungen, die sich mit der Passung theoretischer

Funktionszuschreibung und empirisch gefundenen Bestandteilen von Noten beschäftigen.

Im Schulrecht sollten dann alle jene im Schulalltag zu entscheidenden Fragen geklärt werden, die sich aus dem Leistungsbegriff ergeben (Bewertungsmaßstab, Bewertungsdimensionen, Bewertungsfunktion, Art und Zeitpunkt der Erhebung etc.). Von Seiten des Schulrechts beruht die Notengebung auf einem Beschluss der KMK aus dem Jahr 1968 zur lernzielbezogenen Leistungsbeurteilung. Differenziert in sechs Notenstufen erfolgt eine Bewertung anhand der bewältigten, wenig transparent beschriebenen *Anforderungen* sowie die länderübergreifende Rechtsgrundlage festhält, dass Einzelleistungen in schriftlicher, mündlicher und praktischer Form in das Zeugnis einfließen. Während eine mündliche Nachprüfung je nach pädagogischem Ermessen erfolgt, darf Mitarbeit und Lernverhalten bei der Notenvergabe berücksichtigt werden, das Sozialverhalten soll dagegen nur separat aufgeführt werden. Neben der Befürwortung eines pädagogischen Bewertungsspielraums wird aus der deutschen Schulrechtsabhandlung nicht eindeutig ersichtlich, welche Bezugsnorm angewendet werden soll, wohl aber sind Vergleiche mit der Bezugsgruppe zulässig. Bezüglich der zeitlich einbezogenen Informationen sollten nur aktuell gezeigte Leistungen bewertet werden, Nachfragen zu vorausgegangenen Bewertungen sind aber legitimiert.

Während die länderübergreifende Rechtsprechung allgemeine Grundlagen klärt, liegen je nach Schulart, Fach und Bundesland weitere Bewertungsrichtlinien vor; für die Grundschule sind z.B. soziale Bezugsnormvergleiche zulässig. Die Grundschulverordnungen aus Bayern und Sachsen vergeben Schulnoten ab der zweiten Jahrgangsstufe sowie Sozial-, Lern- und Arbeitsverhalten separat im Zeugnis mittels Noten aufgeführt werden. Auch für den Grundschulbereich obliegt die Notenbildung der pädagogischen Verantwortung sowie die Lehrkraft in Bayern in der vierten Jahrgangsstufe auf die für den Schulerfolg allgemein wichtigen Fähigkeiten achten und in Sachsen als Maßstab für die Leistungsbewertung individuelle Lernfortschritte beachtet werden sollen. Inhaltliche Grundlage fachspezifischer Noten

sind die jeweiligen Lehrpläne und Bildungsstandards. Inwiefern der festgeschriebene Bewertungsspielraum in unterschiedlichen Noten derselben Leistung mündet, wird häufig mittels messtheoretischer Gütekriterien überprüft, die im Folgenden betrachtet werden.



### 3 Zur messtheoretischen Güte von Schulnoten – theoretische Konstrukte und empirische Evidenzen

Ähnlich zur Beschäftigung mit den Funktionen schulischer Notengebung, kommt man bei der Analyse von Noten nicht umhin, sich mit Fragen zur messtheoretischen Güte auseinander zu setzen. So werden in der (kritischen) Diskussion um die schulische Leistungsbewertung, neben schulrechtlichen und funktionstheoretischen Anforderungen, immer wieder Fragen der messtheoretischen Güte thematisiert (Brügelmann, 2001; Ingenkamp, 1976a; Weinert, 2014). Aussagen zur Güte der Leistungsbewertung werden häufig aus dem Vergleich zwischen Schulnoten und parallel erhobenen Testleistungen der Schüler gewonnen (Schrader & Helmke, 2014). Tenor jener Vergleiche ist, dass bei der schulischen Notengebung, häufig erhoben am Übertritt in die Sekundarstufe I, immer wieder eine schwache bis mangelnde Korrelation mit parallel eingesetzten Schulleistungstests nachgewiesen wird (Bos, Voss et al., 2004; Middendorf, 2012). Als Folge wird dann nicht selten über den Einfluss sachfremder Faktoren auf die Notengebung diskutiert, häufig resümierend in einem Plädoyer für die Einhaltung messtheoretischer Gütekriterien in der Schule, um eine höhere Transparenz und Objektivität zu erhalten (Gomolla, 2012; Ingenkamp, 1976a; Schrader & Helmke, 2014). Dabei ist die Diskussion um die mangelnde Güte von Noten bereits lange Zeit ein viel diskutiertes Thema, das vor allem mit Ingenkamps Grundlagenarbeiten zum Gegenstandsbereich in den Fokus fachlicher und gesamtgesellschaftlicher Diskussion rückte (z.B. Ingenkamp, 1976a, 1981, 1993). Die Überprüfung psychometrischer Güteaspekte in der Notengebung erfolgt heute vornehmlich zu einzelnen Themenbereichen (Lintorf, 2012). Häufig diskutiert werden beispielsweise Fragen zur prognostischen und diagnostischen Qualität der Leistungserhebung und zum Einfluss klassenspezifischer Kompositionsmerkmale (z.B. Anders, McElvany & Baumert, 2010; Böhnelt, 1996; Scharenberg, 2014). Weniger untersucht wird aus themenübergreifender Perspektive, welche Bestandteile eine Note konkret vorweist, um zu fra-

gen, welche Eigenschaften es überhaupt sind, die schulrechtskonform oder -widersetzend einfließen und durch Gütemaßstäbe aufgedeckt werden müssten. Während die Erhebung von Leistungsdimensionen in Anlehnung an [Kapitel 2.2](#) sowohl beurteilende als auch rein feststellende, messende Komponenten beinhaltet, umfasst die Leistungsmessung einige messtheoretische Grundlagen, die es aufgrund ihrer Relevanz für die nachgestellten Unterkapitel knapp zu explizieren gilt.

Die methodischen Kriterien zur Überprüfung pädagogischer Diagnostikinstrumente werden meist auf Grundlage der psychologischen Testtheorie formuliert (Ingenkamp, 1976a). Ausgehend von der empirisch-sozialwissenschaftlichen Annahme, dass Merkmale einer Person oder Gruppe mittels Testerhebungen gemessen bzw. sichtbar gemacht werden können, werden ausschnittshafte, empirisch gemessene Merkmale in ein numerisches Relativ überführt, deren Zahlenwerte dann wiederum mehr oder weniger explizit die Verhältnisse der empirisch erfassten Merkmale widerspiegeln. Ein Messvorgang kann dabei von unterschiedlicher Qualität und Informationsgehalt sein (Bühner, 2011; Ingenkamp & Lissmann, 2005; Langfeldt, 1984; Ziegenspeck & Lehmann, 1999). Die Information der Messung ist vom gewählten Skalenniveau abhängig und lässt bereits hier einen aktiven Eingriff des Messenden erkennen. Die Qualität der Messung wiederum wird mithilfe von messtheoretischen Güteverfahren überprüft. Dabei wird im Rahmen psychologischer Tests und Analysen der empirischen Bildungsforschung im engeren Sinne zwischen psychometrischen Persönlichkeitstests und Leistungstests unterschieden. Erfasst werden können unter bestimmten Voraussetzungen Fähigkeiten, Fertigkeiten, Eigenschaften und Zustände der Befragten (Bühner, 2011; Gniewosz, 2015). Differenziert man die Leistungstests für den schulischen Kontext nochmals in Leistungserfassung mittels Testleistung oder Noten, sind wiederum unterschiedliche Formen und Konstruktionsmöglichkeiten zu unterscheiden, die sich als Gegenstand sozialer Bewertungsprozesse ebenfalls mit Fragen der Qualität konfrontiert sehen (Woolfolk, 2008). Betrachtet man die Bewertung durch Noten als zu messendes Objekt, wird deren Qualität neben einigen Nebenkriterien ebenfalls an den drei Hauptgütekriterien Objektivität, Reliabilität und

Validität gemessen, obgleich nach der klassischen Testtheorie diese primär bei der Konstruktion von standardisierten Schulleistungstests Anwendung finden (Langfeldt, 1984). In den folgenden Kapiteln wird darauf eingegangen, welche Annahmen diese Gütekriterien beinhalten und welche Anforderungen sich daraus für die Notengebung ergeben, um dann ausgewählte Befunde für die psychometrische Güte von Noten anzuführen. Abschließend wird in [Kapitel 3.4](#) beleuchtet, in welchem Verhältnis die in der pädagogischen Leistungsdiagnostik vertorten Leistungstests und Schulnoten zueinanderstehen. Die dargestellten Befunde beziehen sich hauptsächlich auf experimentelle Studien zur schulischen Leistungsbeurteilung, bei welchen Noten häufig mittels zusätzlich erhobener Leistungstests überprüft werden. Befunde aus den internationalen Schulleistungsstudien werden nur am Rande behandelt, da sich diese nur als Nebenfragestellungen mit Schulnoten beschäftigen (Hochweber, 2010; Valtin, Wäagner & Schwippert, 2005).

### 3.1 Objektivität von Schulnoten

Die Objektivität einer Messung liegt dann vor, wenn das Messergebnis unabhängig vom Messenden festgehalten wird und damit gewährleistet ist, dass unterschiedliche Untersucher bei der Messung desselben Gegenstands zum gleichen Ergebnis kommen (Gniewosz, 2015; Ingenkamp & Lissmann, 2005; Ziegenspeck & Lehmann, 1999). Jene Forderung nach der Ausschaltung und Kontrolle subjektiver Einflüsse der beteiligten Personen auf das Messergebnis ist nicht nur für die empirische Forschung ein hohes Gut, werden doch auf Grundlage von Messergebnissen häufig richtungsweisende Entscheidungen getroffen. Wäre das Messergebnis desselben Tests bearbeitet von derselben Person je nach Testleiter ein anderes, könnte man diesem Maßstab zur Folge aus diesem Ergebnis keine Folgerungen ableiten (Ingenkamp & Lissmann, 2005). Der Ausschluss intersubjektiver Einflüsse von der messenden Personen und von deren Umgebung wird wiederum in drei Aspekten der Objektivität unterschieden, die sich mit unterschiedlichen Aspekten personen- und kontext(un)abhängiger Ergebnisaufstellung beschäftigen; ausnahmslos mit dem Ziel, im Beurteilungs-

vorgang subjektive Einflüsse zu verringern (Ingenkamp & Lissmann, 2005; Sacher, 2009). Übergreifend gesichert werden soll im Sinne der Testtheorie, dass unterschiedliche Testleiter Tests möglichst identisch *durchführen*, *auswerten* und *interpretieren*, um dem Maßstab der Objektivität gerecht zu werden (Bühner, 2011).

Als *durchführungsobjektiv* wird ein Vorgang bezeichnet, bei dem während einer Leistungserhebung alle Lernenden den gleichen Bedingungen ausgesetzt sind. Dies würde im Schulkontext bedeuten, dass alle Schüler in einer Prüfungssituation die gleiche Aufgabenstellung, Bearbeitungszeit, Hilfestellung und Hilfsmittel erhalten, um dann – aufgrund einer vergleichbaren Ausgangslage – individuelle Noten zu bekommen (Ingenkamp & Lissmann, 2005; Sacher, 2009). Die ideale Umsetzung einer objektiv durchgeführten Prüfungssituation ist in Anbetracht der individuellen Schülervoraussetzungen vermutlich nur annäherungsweise sicherzustellen. So weisen empirische Ergebnisse bereits lange Zeit auf die individuellen Unterschiede in den Eingangsvoraussetzungen der Schüler hin, sodass fraglich ist, ob bei möglichst identischen Durchführungsbedingungen von gleichen sozialen wie kognitiven Voraussetzungen ausgegangen werden kann, die wiederum zu subjektiv unterschiedlich erlebten Testbedingungen führen (Ditton & Krüsken, 2006b; Heller, 1984). Die Durchführungsobjektivität schriftlicher wie mündlicher Testsituationen – auf welche sich der hier betrachtete Aspekt der Objektivität vermutlich eher bezieht – ist dagegen bisher weniger untersucht worden (Heller, 1984; Hochweber, 2010; Ingenkamp & Lissmann, 2005). Auf Grundlage von älteren Befunden zur unterschiedlichen Bewertung schriftlicher Arbeiten weisen Ingenkamp und Lissmann (2005) darauf hin, dass Diktate vermutlich niemals objektiv identisch durchgeführt werden können, unterscheiden sich die Lehrkräfte doch immer in der Art und Weise vorgetragener Diktatinhalt. Offen bleibt in diesem Zusammenhang ebenfalls, welche Rolle gängige, mündliche Nachprüfungen für eine objektive Prüfungssituation spielen, konstatieren Schulrecht und Empirie doch eine durchaus legitimierte individuelle Abweichung je nach Erhebungssituation (Füssel, 2010; Ingenkamp & Lissmann, 2005). Dabei werden mündliche Prüfungen häufig als individuelle

Überprüfung des Lernerfolgs eingesetzt, die vor allem bei Wiederholungsfragen zur Beginn einer Unterrichtsstunde von unterschiedlichem Schwierigkeitsgrad und damit von unterschiedlichen Durchführungsbedingungen zeugen (Ingenkamp & Lissmann, 2005).

Die *Auswertungsobjektivität* besagt, dass jeder Untersucher für dasselbe gezeigte Verhalten zum gleichen Messergebnis kommen muss (Gniewosz, 2015; Ingenkamp & Lissmann, 2005). Übertragen auf den Schulkontext würde das bedeuten, dass die Auswertungsschemata verschiedener Lehrkräfte bei derselben Schülerleistung zum gleichen Ergebnis kommen müssten. Dieser Baustein umfasst also letztendlich die Frage, welches Schülerverhalten auf welche Art und Weise zu bewerten ist, müsste diese Tatsache doch lehrerübergreifend stichhaltig festgelegt werden, um diesen Anspruch an Objektivität erfüllen zu können (Ingenkamp & Lissmann, 2005; Sacher, 2009). Die Auswertungsobjektivität schriftlicher Arbeiten wurde bereits zahlreich und vornehmlich in den 1970er Jahren untersucht (Ingenkamp & Lissmann, 2005). Eine der am häufigsten zitierten Studien sind die Befunde von Starch und Elliot (1976), welche bereits damals auf Beurteilungsdifferenzen in den Fächern Englisch, Geschichte und Mathematik hinwiesen. Sie kamen zu dem Ergebnis, dass die Urteile differenter Lehrkräfte über dieselbe Arbeit erheblich variieren und diese Ergebnisse auch über ein Fach hinweg nachzuweisen sind. Letzterer Befund erscheint dabei besonders interessant, ist doch zu vermuten, dass Bewertungen von Mathematikleistungen aufgrund ihrer klar vorgegebenen Lösung weniger auf subjektive Interpretationen angewiesen sind. Die Autoren bilanzierten daraufhin, dass „[d]ie Variabilität der Urteile [...] somit keine Funktion des Themas, sondern eine Funktion des Prüfers und der Prüfungsmethode“ ist (Starch & Elliot, 1976, S.88). Jene fachübergreifende Befundlage zu lehrkraftspezifischen Auswertungsschemata schriftlicher Arbeiten ist auch für die Auswertung mündlicher Schülerleistungen nachgewiesen worden. Die ebenfalls ältere und häufig zitierte Studie von Hartog und Rhodes (1976) verweist ebenfalls auf individuelle Abweichungen, wonach derselbe Prüfling bei der mündlichen Abfrage eines allgemein formulierten Themenbereichs je nach Lehrkraft unterschiedlich bewertet wurde.

Während aktuelle Abhandlungen zur Objektivität von Noten vor allem die in Anbetracht des Schulrechts wenig verwunderlichen Beurteilungsdivergenzen zwischen verschiedenen Lehrkräften für mündliche und schriftliche Leistungsabfragen thematisieren (Birkel, 2009; Birkel & Birkel, 2002), sind Ergebnisse zur Auswertungsdifferenz von Mathematikprüfungen – wie bei Starch und Elliot (1976) nachgewiesen – auch in jüngerer Zeit noch überraschend. Obwohl Ingenkamp und Lissmann (2005) auf die mangelnde Übertragbarkeit ältere Befunde für die aktuelle Schulpraxis hinweisen, zeigen auch aktuelle Befunde aus internationalen Schulleistungsuntersuchungen für den mathematischen Bereich, dass 15-jährige bei gleicher Mathematikleistung unterschiedliche Noten erhielten (Baumert, Trautwein & Artelt, 2003). Sacher (2009) weist auf vier Gründe für Abweichungen in der Auswertungsobjektivität hin: Differente Auswertungen derselben Prüfungsleistung seien auf die Anzahl der gefundenen Fehler, die festgelegte Maximalpunktzahl, den Notenschlüssel oder die Zuteilung von Schülerleistungen zu einer entsprechenden Punktzahl zurückzuführen. Zusätzlich dazu ist die individuell abweichende Milde und Strenge in der Benotung ebenfalls als Grund für die mangelnde Auswertungsobjektivität zu betrachten, die auch schon früh im Zusammenhang mit Fächerunterschieden diskutiert wurde (Ingenkamp, 1995a; Weiss, 1976c). Bereits Weiss (1976c) diskutierte die unterschiedlichen Bewertungsmaßstäbe in unterschiedlichen Fächern und merkte an, dass in Fächern mit schriftlichen Prüfungsleistungen (Deutsch, Mathematik, Fremdsprachen) deutlich strenger zensiert werde. Dieses Faktum scheint insbesondere bei der ungewichteten Verrechnung einzelner Fachnoten zu einem Durchschnitt relevant, wird dabei doch davon ausgegangen, dass Noten aus unterschiedlichen Fächern gleichbedeutend sind (Jachmann, 2003). Gerade in der Diskussion um fachspezifische Benotungsunterschiede weisen bereits Fingerhut und Langfeldt (1974) darauf hin, dass je Schulfach eher von fachtypischen Verteilungsformen ausgegangen werden muss und Schulnoten je nach Fach und Klasse nicht unbedingt normal verteilt sind. Demnach würden Normalverteilungsabsichten für eine kleine Stichprobe dazu führen, dass „die Note eines Schülers eher von den Leistungen seiner Mitschüler als von seiner eigenen Leistung bestimmt“ (Fingerhut & Langfeldt, 1974, S.257) werde.

An die Befundlage zur mangelnden Auswertungsobjektivität von Prüfungsleistungen schließt sich die zu sichernde *Interpretationsobjektivität* an, welche wohl die am schwierigsten umzusetzende Objektivitätsart im Schulalltag ist. Gewährleistet werden muss in diesem Sinne, dass jeder Auswerter möglichst zur gleichen Beurteilung bzw. Interpretation eines Testergebnisses kommt (Bühner, 2011). Dabei muss für Ingenkamp (1976a) – als einem der Hauptdiskutanten über die Objektivität von Noten – gerade für die Unabhängigkeit der Interpretation vom Beobachter dringend zwischen Messung und Bewertung des Schülerverhaltens unterschieden werden. Demnach besteht gerade im Schulkontext die Gefahr, bei der Wertung gleicher Testergebnisse individuelle Schülervoraussetzungen zu berücksichtigen und damit die objektive Leistungsmessung zu überlagern. Da im Schulkontext eine Ausschaltung intersubjektiver Interpretationen nur schwer möglich ist, sollte das Messergebnis zunächst unabhängig von einer Bewertung festgestellt werden (Ingenkamp, 1976a). Redundant dazu weisen auch spätere Arbeiten von Ingenkamp und Lissmann (2005) darauf hin, dass der pädagogische Ermessensspielraum bei Leistungsbewertungen die Beachtung individueller Voraussetzungen und Situationsabhängigkeiten zulässt, weswegen eine wertfreie Leistungsmessung ein hohes Gut sei, um unkontrollierten Einflussfaktoren entgegen zu wirken. Offen bleibt allerdings, ob ein solcher Messvorgang in der Schule jemals losgelöst von der bewertenden Lehrkraft erfolgen kann. Während empirische Studien zur Objektivität von Schulnoten einerseits Lehrkraftfaktoren und deren theoretisch wie empirisch konstatierte Einflüsse auf Noten prüfen (Brügelmann, 2006; Lüders, 2001; Schrader & Helmke, 2014), weisen andererseits zahlreiche Befunde der Schulforschung auf klasseninterne Bezugssysteme hin, die ebenfalls als Ursache mangelnder Objektivität von Schulnoten diskutiert werden (Maaz, Baeriswyl et al., 2011).

Übergreifendes Ergebnis aller aufgezeigten Analysen ist wohl die Tatsache, dass Lehrkräfte in ihrer Klasse keinen absoluten Bewertungsmaßstab anlegen, sondern die Prüfungsleistungen vor dem Hintergrund des klasseninternen Leistungsniveaus und individuell gewählten Bewertungsmaßstabs einstufen (Hochweber, 2010;

Ingenkamp & Lissmann, 2005). So verweisen bereits Ingenkamps (1976b) Analysen auf einen klasseninternen Bezugsrahmen, wonach in unterschiedlichen Klassen für unabhängig erfasste Leistungswerte unterschiedliche Noten vergeben werden. Allerdings wurde die leistungsbezogene Rangordnung innerhalb der Klasse von einer Lehrkraft recht ähnlich zur parallel erhobenen Testleistung eingeschätzt (Ingenkamp, 1976b). Aktuellere Ergebnisse von Thiel und Valtin (2002) weisen auf ähnliche Tendenzen hin, wonach innerhalb einer Klasse die Korrelation zwischen Leistung und Noten deutlich höher ausfällt als zwischen den Klassen. Ebenso fanden sich in der letzten Zeit auch gehäuft Belege für Effekte des klassenspezifischen Leistungsniveaus auf die Notengebung, wobei in leistungsstärkeren Klassen tendenziell strenger bewertet wird als in leistungsschwächeren Klassen (Hochweber, 2010; Lintorf, 2012; Trautwein & Baeriswyl, 2007). Die klassen- bzw. lehrkraftspezifische Relevanz für die Objektivität von Noten diskutiert bereits Weiss (1976b) vor dem Hintergrund äußerer und interner Bewertungsmaßstäbe, welche letztlich den Schulleistungsbegriff prägen. Während ein entsprechendes Außenkriterium eine von außen festgelegte Klassifizierung von Schulleistung sichern soll, bezieht sich das Innenkriterium auf den Bewertungsmaßstab des Lehrers, damit er „imstande ist, seinen zwar subjektiven Leistungsmaßstab objektiv, d.h. unbeeinflusst von leistungsfremden Motiven, anzuwenden.“ (Weiss, 1976b, S.112). Jene leistungsfremden Einflussfaktoren sind Bestandteil zahlreicher Analysen der Bildungsforschung (z.B. Ditton, 2007b, 2010; Prenzel et al., 2013). Auch wenn bereits Weiss (1976b) darauf hinweist, dass ein objektives Lehrerurteil nicht gleichzeitig ein *richtiges* Urteil bedingt, werden leistungsfremde Einflüsse auf Noten häufig vor dem Hintergrund benachteiligender Wirkungen von Schule für spezifische Gruppen diskutiert. Objektivitätsverletzungen werden beispielsweise in Abhängigkeit vom Schülergeschlecht und dem sozialen wie ethischen Hintergrund des Schülers diskutiert (Lintorf, 2012; Maaz, Baeriswyl et al., 2011; Tiedemann & Billmann-Mahecha, 2007a). Allgemeiner Tenor der Abhandlungen ist, dass sich je nach Gruppenzugehörigkeit intersubjektive Abweichungen auf Noten ergeben. Einflüsse, die durch zusätzlich eingesetzte Testhebungen aufgedeckt werden und dann wiederum häufig Grundlage

kritischer Abhandlungen zur mangelnden Objektivität von Schulnoten sind.

Aktuelle Befunde zu Geschlechtereffekten in der Notengebung weisen auf Geschlechtereffekte in der Benotungspraxis hin, wonach Mädchen in Deutsch bessere Noten erhalten als ihre Mitschüler und Jungen dagegen in Mathematik bessere Leistungen erzielen (Blossfeld et al., 2009; Kuhl & Hannover, 2012). Studien zu sozialschichtspezifischen Einflüssen auf die Notengebung deuten darauf hin, dass sich Lehrkräfte in der Beurteilung schriftlicher Arbeiten von der sozialen Herkunft der Schüler beeinflussen lassen und insgesamt die erhobenen Testleistungen in deutlichem Zusammenhang mit der sozialen Herkunft der Schüler stehen (Ditton & Krüsken, 2009; Holtappels & Heerdegen, 2005; Stahl, 2007; Weiss, 1976b; Wiese, 1982). Neuere Befunde zum expliziten Zusammenhang zwischen sozialer Herkunft und Notengebung sind allerdings rar, wobei der Zusammenhang von sozialen Hintergrundmerkmalen und Schulleistung häufig bei der Übertrittsempfehlung am Ende der Grundschulzeit untersucht wird (Ditton & Krüsken, 2006a; Maaz & Nagy, 2010). Die Ergebnisse von Maaz und Nagy (2010) weisen beispielsweise auf Effekte der sozialen Herkunft auf den Übertritt in die Sekundarstufe I hin, allerdings konnte die Frage nach der Bewertungsgrundlage der Lehrkräfte – als möglicher Mediator – in der Studie nicht beantwortet werden. Befunde zu Effekten des Migrationshintergrunds auf die Notengebung sind dagegen widersprüchlich (Lintorf, 2012). Gomolla und Radtke (2009) konnten mithilfe qualitativ erhobener Daten eine Benachteiligung von Kindern mit Migrationshintergrund bei der Vergabe von Übertrittsempfehlungen feststellen, während van Ewijk (2011) in seiner Experimentalstudie zur Aufsatzbenotung von Lehrkräften keine direkten Effekte feststellte. Ebenso konnten Ditton, Krüsken und Schauenberg (2005) lediglich einen schwachen direkten Effekt des Migrationshintergrunds feststellen.

Zusammenfassend zeigen die exemplarischen Befunde zur Objektivität der Notengebung, dass die differenzierten Objektivitätsaspekte in der Schulpraxis unzureichend erfüllt werden können. Überträgt man

jenen Gütemaßstab also auf die schulische Bewertungspraxis, treten Fragen zum Einfluss von Lehrkraft, Klasse und Gruppenzugehörigkeit auf. Die Objektivität in der Leistungsbeurteilung wäre gemäß den messtheoretischen Gütekriterien dann gewährleistet, wenn Schüler mit gleichen Schulleistungen unabhängig von der bewertenden Lehrkraft, ihrer zugehörigen Klasse und ihren individuellen Voraussetzungen für die gleiche Leistung gleich benotet werden. Betrachtet man die Objektivität in der Notengebung als das notwendige Kriterium für eine gerechte Verteilung, weist schon Jachmann (2003) auf die desillusionierende Wirkung der Forschungsbefunde zur mangelnden Objektivität von Noten hin. Dennoch ist die Sicherung der Objektivität die Voraussetzung für die Gültigkeit und Zuverlässigkeit einer (Leistungs-)Messung; erst wenn die Objektivität einer Leistungsbeurteilung gewährleistet ist, kann gemäß der testtheoretischen Grundlagen über Fragen der Zuverlässigkeit und Gültigkeit entschieden werden (Ingenkamp & Lissmann, 2005; Jachmann, 2003).

### 3.2 Reliabilität von Schulnoten

Die Reliabilität einer Messung gibt den Grad der Sicherheit oder Genauigkeit an, mit dem ein Test ein Merkmal misst, unabhängig davon, was der Test zu messen vorgibt (Bühner, 2011; Ingenkamp & Lissmann, 2005). Übertragen auf den Schulkontext bedeutet dies, dass nach der wahren Ausprägung des gezeigten Merkmals gefragt werden muss, um haltbare Aussagen über die Stabilität oder Zufälligkeit des gemessenen Merkmals treffen zu können (Sacher, 2009). Die Frage der Zuverlässigkeit betrifft im Schulkontext vor allem die Frage, ob man auf Grundlage einer einmalig gemessenen Leistung zuverlässige Aussagen über die Stabilität des Leistungsverhaltens eines Schülers treffen kann. Gemäß diesem Maßstab wäre ein Test dann umso zuverlässiger, wenn eine Lehrkraft dieselbe Klassenarbeit mit zeitlichem Abstand identisch beurteilen würde und eine Messung von stabil anzunehmenden Merkmalen vorliegt (für Ingenkamp (1976a) sind z.B. Rechtschreibleistung oder Interesse des Schülers stabilere Merkmale als die psychische Verfassung eines Schülers) (Ingenkamp & Lissmann, 2005).

Ausgehend von der Annahme, dass jeder Messung ein wahrer und ein zufallsbedingter Wert zugrunde liegt, überprüft man die Zuverlässigkeit des Messwerts – in der Testpsychologie und bei der Konstruktion von Schulleistungstests – mithilfe der Wiederholungs(Retest)-, Halbierungs(Split-Half)- und Paralleltestmethode. Erstere überprüft die Zuverlässigkeit der Messung mittels der Korrelation zwischen zwei zeitlich versetzten Testerhebungen, die den gleichen Inhalt vorgeben und von derselben Person bewertet werden. Die zweite Methode teilt ein Test in zwei inhaltlich vergleichbare Hälften und wertet getrennt aus. Letztere Methode entwickelt zwei strukturähnliche Tests und leitet aus Korrelationen beider Ergebnisse deren Erhebungszuverlässigkeit ab (Ingenkamp & Lissmann, 2005; Sacher, 2009). Bezogen auf den Schulkontext sind jene Reliabilitätsanforderungen auf eine Note zu übertragen. Dabei kommt bereits für Tent, Fingerhut und Langfeldt (1976) der Reliabilität von Schulnoten eine hohe praktische Bedeutung zu, gehe man doch im Schulalltag von einer gewissen Kontinuität von schulischen Leistungen aus. Ausgehend von der Annahme, dass die Reliabilität von Noten einerseits „von der Konstanz des Merkmals `Schulleistung` und der Konsistenz in der Beurteilung andererseits“ (Tent, Fingerhut & Langfeldt, 1976, S. 22) bestimmt wird, wurde in empirischen Studien bislang vor allem die zeitliche Konstanz von Leistungsbewertungen (Retestrelia-bilität) und die Bedeutung ähnlicher Erhebungsinstrumente für Noten (Paralleltestmethode) untersucht. Die Zuverlässigkeit in der Notengebung wäre dann gewährleistet, wenn zeitlich differente Leistungsmessungen desselben Schülermerkmals mit derselben Prüfungsform zum selben Leistungsergebnis kommen würden. Befunde aus der Schulpraxis zur zeitlichen Stabilität und Zuverlässigkeit eines Bewertungsvorgangs liegen erneut vor allem aus den 1970er Jahren vor. Bezüglich der Parallel-Retest-Reliabilität untersuchte Aschersleben (1971) vier Diktate und Mathematikarbeiten aus insgesamt 47 Klassen auf deren Zusammenhang, bilanzierend mit dem Ergebnis von Interkorrelation fast immer unter  $r=.60$ , wobei die Werte für die Mathematik insgesamt niedriger ausfielen (Aschersleben, 1971). Bezüglich der Zuverlässigkeit wiederholter Benotung derselben Prüfungsarbeiten in Geschichte und Geographie durch dieselben Lehrkräfte wies schon Eells (1976) auf eine unbefriedigende

Retest-Korrelation hin. Nur zehn von 61 teilnehmenden Lehrkräften kamen bei einem zweiten Korrekturdurchgang zu denselben Ergebnissen, die Antworten variierten in den Reliabilitätskoeffizienten von  $r=.25$  bis  $r=.51$  (Eells, 1976). Ähnlich dazu kam auch Dicker (1976) in seinen Analysen zur wiederholten Beurteilung von Mathematikarbeiten zu dem Ergebnis, dass sich für ein und dieselbe Arbeit signifikante Unterschiede zwischen den Beurteilungen ergaben ( $r=.46$ ) und Lehrkräfte bei einer wiederholten Korrektur strenger bewerteten. Bezüglich der zeitlichen Reihenfolge der Korrektur von unterschiedlichen Prüfungsarbeiten wies zudem Baurmann (1976) auf signifikante Effekte der Reihenfolge einer Bewertung von Aufsätzen hin. Zusätzlich dazu ergaben sich Effekte des Leistungsniveaus, indem schlechtere Arbeiten noch schlechter beurteilt wurden, wenn sie nach einer guten Arbeit korrigiert wurden (Ingenkamp, 1995b).

Während die exemplarischen Befunde aus der Schulpraxis auf eine gering ausgeprägte Reliabilität von Schulnoten hinweisen, kritisieren bereits Fingerhut und Langfeldt (1974) die mangelnde Übertragbarkeit der testtheoretischen Reliabilitätskonzepte auf die Schulpraxis. Demnach sind jene Konzepte zwar ein guter Maßstab, um auf instrumentelle Mängel von Noten hinzuweisen, sie sollten allerdings nicht dazu führen, Noten an die vorgestellten Konzepte anzupassen (Fingerhut & Langfeldt, 1974). Ebenso diskutiert Birkel (1984b) bereits die Frage, ob eine hohe Korrelation zwischen zwei zeitlich hintereinander liegenden Klassenarbeiten überhaupt wünschenswert sei, würde dies doch unter pädagogischen Gesichtspunkten bedeuten, dass ein schlechter Schüler auch nach einiger Zeit als schlechter Schüler beurteilt wird.

### 3.3 Validität von Schulnoten

Die Validität eines Messinstruments gibt Auskunft darüber, inwiefern das Testinstrument tatsächlich das misst, was es zu messen vorgibt (Langfeldt, 1984). Messwerte sind demnach valide, wenn sie fehlerfreie Rückschlüsse auf den Ausprägungsgrad des gemessenen Merk-

mals zulassen, sie also den inhaltlichen Aspekt des Messvorgangs realitätsnah festhalten (Lienert & Raatz, 1998).

Bezogen auf die Validität von Schulnoten wird danach gefragt, was Schulnoten konkret messen, welcher Sachverhalt also zu einer entsprechenden Ziffernbetitelung führt: „Schulnoten sind Maßzahlen für die vom Lehrer vorgenommene Einschätzung der schulischen Leistungen, wie sie in Klassenarbeiten, Prüfungen oder während des Unterrichtes wahrgenommen werden.“ (Fingerhut & Langfeldt, 1974, S.260). Fragen der Validität beschäftigen sich im Schulkontext folglich mit der Frage, welchen Inhalt Noten widerspiegeln (sollen), wie gut sie dies tun und wie weitreichend die Kategorisierung repliziert werden kann. Beispielsweise sind in einer Aufsatzbeurteilung die Gliederung und Formulierung eigener Gedanken zu berücksichtigen und nicht – wenn nicht explizit angemerkt – die äußere Form und Rechtschreibeleistung (Ingenkamp, 1976a). Die Validitätsprüfung ist bereits für Ingenkamp (1976a) in der pädagogischen Diagnostik eine entscheidende Größe, da ohne eine solche Prüfung auf Grundlage von Messwerten getroffene pädagogische Maßnahmen wenig zielführend seien:

Wenn also Aussagen über die intellektuelle Kapazität in Wirklichkeit Aussagen über die Anpassung an die Vorstellungen des Lehrers, über gutes Betragen und Fleiß sind, dann können sie keine Entscheidungshilfen für Maßnahmen sein, für die die intellektuelle Kapazität ein wichtiger Faktor ist. (Ingenkamp, 1976a, S.25).

Aussagen über eine alternative, korrektere Feststellung jener intellektuellen Kapazität werden in diesem Zusammenhang allerdings nicht getätigt. Ähnlich zu den vorangegangenen Gütemaßstäben werden trotz Diskussion über die Untrennbarkeit der einzelnen Konzeptaspekte auch für die Validität verschiedene Bereiche unterschieden, die aufgrund ihrer Vielfalt im Folgenden nicht alle ausgeführt werden. Sie finden sich aber in ihren Grundzügen in den folgenden drei wichtigsten Aspekten wieder. Unterschieden wird im Sinne der klassischen Testtheorie häufig zwischen inhaltlicher Validität, Konstruktvalidität

und Kriteriumsvalidität (Bühner, 2011; Lienert & Raatz, 1998; Lintorf, 2012).

Von *Inhaltsvalidität* wird gesprochen, wenn ein Test das zu messende inhaltliche Konstrukt präzise erfasst. Bezogen auf den Schulkontext beschreibt beispielsweise der Lehrplan den Inhalt, der vermittelt und durch Noten überprüft werden soll, wobei unspezifisch formulierte Lernziele vermutlich auch kontroversere Überprüfungsformen fördern (Ingenkamp & Lissmann, 2005). Ebenso wird unter diesem Aspekt danach gefragt, ob tatsächlich jene Merkmale abgefragt werden, die im Unterricht vermittelt wurden oder ob zur Beantwortung der außerschulische Kontext des Schülers vonnöten war (Sacher, 2009). Durch die aufgeführten Validitätsbeispiele im Schulkontext wird deutlich, dass eine inhaltliche Validitätsprüfung von Noten augenscheinlich schwierig scheint, solange kein allgemein anerkanntes Kriterium für das Schulleistungskonstrukt vorliegt, sowie der Notengebung mannigfache Funktionen zugeschrieben werden (Fingerhut & Langfeldt, 1974). Auch Ingenkamp und Lissmann (2005) sehen in den Interpretations- und Auswahlmöglichkeiten des Lehrplans begründet, dass sich empirische Analysen bislang selten mit der inhaltlichen Validität schriftlicher Arbeiten beschäftigen. Bezüglich der inhaltlichen Validität besteht von struktureller Seite aus aktuell die länderübergreifende Verpflichtung seitens der KMK auf einheitliche Bildungsstandards, die sich auf fachspezifische Inhaltsstandards sowie übergreifende Leistungsstandards beziehen; überprüft im Grundschulbereich neben der Internationalen Grundschulstudie IGLU mittels flächendeckender Vergleichsarbeiten (Köller, 2009).

Unter dem zweiten Validitätsaspekt, der *Konstruktvalidität*, werden häufig alle Validitätsarten zusammengefasst. Demnach beschreibt dieser Aspekt der Validität im eigentlichen Sinne die Übereinstimmung der gemessenen Eigenschaften mit dem zu erfassenden theoretischen Konstrukt (Bühner, 2011; Ziegenspeck & Lehmann, 1999). Bezogen auf den Schulkontext würde dies vermutlich bedeuten, dass der Prüfung theoretische Modelle zugrunde liegen müssten, die Vorstellungen darüber beinhalten, was als Leistung im vorliegenden Fach

oder Bereich betitelt wird und nach welchen Niveaustufen man die gezeigten Merkmale unterscheiden bzw. abfragen kann (Sacher, 2009). Exemplarisch würde eine geringe Konstruktvalidität vorliegen, wenn ein Mathematiktest die Rechtschreibeleistung des Schülers messen oder bei einer Aufsatzbewertung die äußere Form Einfluss auf den Messwert nehmen würde, obwohl dies nicht explizit Vorgabe der Prüfungsleistung war (Ingenkamp, 1976a).

Bei der *kriterienbezogenen Validität* schließlich wird häufig zwischen einer Prognosevalidität und einer Übereinstimmungsgültigkeit unterschieden, die zusammen genommen auch mehrfach als empirische Gültigkeitsmaße bezeichnet werden (Ingenkamp & Lissmann, 2005; Lienert & Raatz, 1998). Bezogen auf den Schulkontext wird im Sinne einer prognostischen Validitätsprüfung danach gefragt, ob Schulnoten jene Merkmale widerspiegeln, die für den weiteren Bildungsverlauf bedeutsam sind, den Schulerfolg also prognostisch valide vorhersagen. Fragen der Übereinstimmungsgültigkeit beziehen sich dagegen auf den Zusammenhang zwischen zwei Kriterien, wobei eine hohe Korrelation für eine hohe Kriteriumsvalidität spricht (Bühner, 2011). Gefragt wird hierbei beispielsweise nach der Vergleichbarkeit von mündlichen und schriftlichen Prüfungsformen oder nach der Korrelation zwischen Noten und zusätzlich erhobener Testleistung, auf deren Grundlage dann Aussagen zur Gültigkeit von Schulnoten getroffen werden (Sacher, 2009; Tent et al., 1976; Ziegenspeck & Lehmann, 1999). Obwohl bereits Fingerhut und Langfeldt (1974) die pädagogische Intention der Vorhersage-Validität von Noten kritisch diskutieren, sind sowohl die empirische Überprüfung prognostischer Validität von Schulnoten und die theoretische Herleitung prognostischer Lehrkraftinformationen als auch die Überprüfung der Zusammenhänge zwischen Schulleistungstests und Schulnoten in der empirischen Bildungsforschung aktuell wie historisch weit verbreitet (z.B. Baeriswyl, Trautwein, Wandeler & Lüdtke, 2010; Becker & Birkelbach, 2013; Krüsken, 2007a; Müller, 2013; Sauer & Gamsjäger, 1996; Sommer, 1983). Auffällig ist dabei, dass aktuelle Forschungsbefunde bezüglich der Validität von Schulnoten lediglich einzelne Teilaspekte prüfen; neben der Übereinstimmung von Testleistung und Schulnoten

wird die prognostische Validität von Schulnoten häufig bei der Übertrittsempfehlung am Ende der Grundschulzeit fokussiert (Baeriswyl et al., 2010; Böhnelt, 1996; Lintorf, 2012; Sauer & Gamsjäger, 1996; Sommer, 1983; Stubbe, Bos & Euen, 2012). Auf empirische Kernbefunde zu den beiden genannten Validitätsaspekten wird im Folgenden eingegangen.

Bezüglich der *Übereinstimmung von Testleistung und Schulnoten* wird also davon ausgegangen, dass dadurch einerseits ein Maß für die Güte lehrkraftspezifischer Urteils- und Diagnoseleistungen gegeben ist und andererseits subjektive bzw. kontextbezogene Einflüsse auf die Noten sichtbar gemacht werden können (Lintorf, 2012; Schrader, 1997). Jene zusätzlich zu Noten durchgeführte Testverfahren sind in der empirischen Bildungsforschung weit verbreitet und sollen Auskunft darüber geben, inwiefern Schulnoten oder Grundschulempfehlung am Ende der Primarstufe mit zusätzlich erfassten Testleistungen übereinstimmen (z.B. Baumert, Artelt et al., 2003; Böhnelt, 1996; Bos, Lankes, Prenzel, Schwippert, Valtin & Walther, 2003; Ditton, 2007b; Ingenkamp, 1976a; Stubbe et al., 2012). Als ein Teilbereich der internationalen Grundschuluntersuchung wurden die in IGLU 2011 erhobenen Testdaten sowohl mit den Übertrittsempfehlungen der Lehrkräfte als auch mit den Noten in Deutsch und Mathematik auf ihre Übereinstimmung geprüft (Stubbe et al., 2012). Während insbesondere die Deutschnote ein guter Indikator für die Übertrittsempfehlung war (97 % Prozent der Schüler mit einer sehr guten Deutschnote erhielten eine Gymnasialempfehlung), ergaben sich in Abhängigkeit vom sozialen Hintergrund für identische Testleistungen teils unterschiedliche Empfehlungen (Stubbe et al., 2012). Für die Daten aus IGLU 2006 liegen die Korrelationen zwischen Deutschnote und Lesekompetenztest in der Grundschule im Jahr 2006 bei  $r = -.57$  (Arnold, Bos, Richert & Stubbe, 2007). Dagegen zeigen Daten aus PISA 2006 für die anschließende Sekundarstufe geringere, signifikante Korrelationen zwischen naturwissenschaftlichen Kompetenztests und Noten in Biologie, Physik und Chemie von  $r = -.34$  bis  $r = -.36$  (Schütte, Frenzel, Aseburg & Pekrun, 2007). Für den Grundschulbereich fanden Sauer und Gamsjäger (1996) mittlere Effekte für den Zusammenhang zwi-

schen fachspezifischen Testleistungen und Noten. Die Analysen von Krüsken (2007a) ergaben höhere Zusammenhänge zwischen der Durchschnittsnote aus Deutsch und Mathematik und den entsprechenden Testleistungen im Längsschnitt (für die 3. Klasse  $r=-.66$  und 4. Klasse  $r=-.72$ ). Bei einem explizit klassenübergreifenden Übereinstimmungsvergleich zwischen Testleistung und vergebener Schulnote ergaben die Analysen von Ingenkamp (1968) in sechsten Grundschulklassen, dass die Übereinstimmung eines nach Lehrplan abgestimmten Leistungstest in Mathematik mit den Noten je nach Klasse variierte. So bekam ein Schüler mit einem Leistungswert von 30,33 Punkten in einer Klasse die Note 2, während in einer anderen Klasse für dieselbe Note eine Testpunktzahl von 50,85 notwendig war (Ingenkamp, 1968). Bilanzierend merkt Sacher (2009) an, dass eine Validitätsprüfung zwischen Testleistung und Schulnoten damit nur klassenintern bzw. je Lehrkraft sinnvoll sei. Analog zu den oben erwähnten Einflussfaktoren auf eine objektive Notengebung sind auch im Bereich der Validität von Schulnoten neben Klasseneffekten schicht- und fachspezifische, regionale sowie schulformabhängige Unterschiede in der Gültigkeit von Schulnoten gefunden worden, die im Folgenden aufgrund möglicher Überschneidungen nicht weiter ausgeführt werden (Sacher, 2009). Ebenso wurde an dieser Stelle auf eine Differenzierung zwischen unterschiedlichen Testprädiktoren verzichtet, da dies im Sinne der psychometrischen Testdiagnostik ein weiteres, umfangreiches Themengebiet umfasst und für den Fokus der vorliegenden Arbeit zu weitführend ist.

Auch wenn der Zusammenhang zwischen Testleistung und Schulnoten studienübergreifend recht stabil ist, kann man aufgrund der fehlenden vollständigen Übereinstimmung davon ausgehen, dass es für Schulnoten neben entsprechenden Testleistungen noch weitere prädiktive Faktoren gibt, weswegen zahlreiche Studien auch explizit den *Validitätskoeffizient von mündlichen und schriftlichen Prüfungen* untersuchen (Lintorf, 2012). Gemäß der Literatur ist es für die Übereinstimmung von Noten und Testleistung ebenfalls von Bedeutung, wie und aus welchen Prüfungsformen sich eine Schulnote zusammensetzt. Zusätzlich zu den oben angeführten Ergebnissen aus der Aufsatz-

forschung finden sich für schriftliche Prüfungsleistungen Effekte der äußeren Form, Aufsatzlänge, grammatikalischer und orthografischer Fehler, Reihenfolge der Korrektur und widersprüchliche Ergebnisse zum Einfluss von Vorabinformationen (Baurmann, 1976; Birkel, 2005; Birkel & Birkel, 2002; Osnes, 1976). Während Baurmann (1976) für Vorinformationen über Schüler keine Urteilsverzerrungen feststellen konnte, ergaben die Analysen mündlicher Zensuren von Birkel (1978) Effekte von vorher präsenten Schülerinformationen, wobei im Sinne eines Informationseffekts Schüler mit schlechten Vorleistungen signifikant schlechter bewertet wurden als jene, deren Vorleistungen als gut bezeichnet wurden. Ähnliche Befunde zeichnen sich in weiteren Studien zur Übereinstimmung mündlicher Prüfungsformen ab (Lintorf, 2012). Dabei besagt der Validitätsanspruch an mündliche Prüfungen – analog zu schriftlichen –, dass nur Informationen über die abgefragte Leistung Gegenstand eines Leistungsurteils sein sollten und individuelle Merkmale von Schüler und Prüfer keinen Einfluss nehmen (Birkel, 1984a). Betitelt als „die am wenigsten erforschte Form der Leistungskontrolle“ (Ingenkamp, 1995b, S.191) ergeben sich für die mündliche Prüfungsform – erneut auf Grundlage älterer Studien – Effekte für die prüfende Person, Informationen über die Vorleistung und den vorangegangenen Prüfungen, vornehmlich überprüft mit Übereinstimmungswerten bei schriftlichen oder mündlichen Prüfungen (Birkel, 1978, 1984a).

Die Analyse der *prognostischen Validität* von Schulnoten und der zusätzlich erhobenen Testwerte – als weiterer Aspekt der Kriteriumsvalidität – ist vermutlich einer der am häufigsten untersuchten Teilaspekte in der Diskussion um die messtheoretische Güte von Schulnoten. Während im Schulkontext häufig der prognostische Wert von Übertrittsempfehlungen untersucht wird, stützen sich die im Bildungsverlauf voran geschrittenen Studien beispielsweise auf die Prognosekraft von Abiturnoten für den Studienerfolg (Trapmann, Hell, Weigand & Schuler, 2007). So fand die auf Grundlage internationaler Studien entstandene Metaanalyse von Trapmann et al. (2007) für Durchschnittsnoten aus Deutsch und Mathematik am Ende der Schulzeit höhere Prognosewerte für den Studienerfolg als für entsprechende

Einzelnoten, wobei sich besonders in Deutsch höhere Validitäten fanden als für einzelne, studienfachnahe Schulnoten. Untersuchungen zur Prognosefähigkeit von Leistungsbeurteilungen während der Schulzeit beziehen sich wie erwähnt überwiegend auf die unterschiedlich verbindliche Übertrittsempfehlung am Ende der Grundschulzeit, vereinzelt analysieren allerdings auch aktuelle Schulleistungsstudien die Diskrepanzen zwischen gemessenen Kompetenzen, Schulnoten und prognostischen Lehrerurteilen (Becker & Birkelbach, 2013; Schütte et al., 2007). Ausgangslage ist die Frage, wie gut der Schulerfolg von Schülern vorhergesagt werden kann, wobei die besagten Prädiktoren je Studie zwischen Übertrittsempfehlung, Durchschnittsnote oder Einzelnoten schwanken und auch das Kriterium des späteren Lernerfolgs je Studie different ist (Ingenkamp, 1993; Ingenkamp & Lissmann, 2005). Böhnelt (1996) fand in ihrem österreichisch-ungarischen Kooperationsprojekt zur Prognostizierbarkeit von Schulerfolg anhand von Durchschnittsnoten der vierten Klasse eine Korrelation von  $r=.78$  mit jenen aus der achten Klasse der Sekundarstufe. Allerdings zeigen differenzierte Analysen weiter, dass für ca. 40% der Stichprobe keine hinreichende Erklärung der weiteren Schullaufbahn durch die Primarstufennote vorliegt (Böhnelt, 1996). Sauer und Gamsjäger (1996) fanden dagegen im selben Jahr für die gleichen Jahrgangsstufen in Österreich einen niedrigeren Zusammenhang zwischen Noten der vierten und achten Jahrgangsstufe. Während durch Grundschulnoten der Eingang in die Sekundarstufe I noch relativ gut vorhersagbar war (37,9% der Varianz in den Schulleistungen in der ersten Gymnasialstufe wurde durch die Grundschulnoten aufgeklärt), konnte die Varianz in den Schulleistungen der achten Klasse nur noch zu 30% aufgeklärt werden (für den Hauptschulzweig waren es sogar nur 6%). Zu beachten ist in dieser Studie allerdings, dass die Prädiktorwahl in der Grundschule für die prognostischen Urteile im weiteren Schulverlauf schwanken. Zudem stellen die Autoren fest, dass die Vorhersagekraft für Leistungstests inkl. Intelligenzvariablen insgesamt schwächer ist als für Grundschulnoten (Sauer & Gamsjäger, 1996). Ebenso wird darauf hingewiesen, dass im Sinne einer differenziellen Validität vor allem für durchschnittlich bis schlecht bewertete Schüler kaum prognostisch valide Vorhersagen getroffen werden können (Heller, 1997; Sauer &

Gamsjäger, 1996). Für einen grundschulinternen Prognosewert ermittelte die Scholastik-Studie stabile Zusammenhänge zwischen Zweit- und Viertklassnoten, wobei explizit erwähnt wird, dass ein Lehrerwechsel zwischen den Klassenstufen scheinbar keinen großen Einfluss nehme (Helmke, 1997). Bei einer aktuelleren, grundschulübergreifenden Längsschnittanalyse fand Müller (2013) keinen signifikanten Zusammenhang zwischen der Durchschnittsnote der zweiten Klassenstufe mit der Durchschnittsnote in der sechsten Klassenstufe. Allerdings zeichnete sich für die von der Lehrkraft eingeschätzten kognitiven Schülermerkmale (Begabung, sprachliche Fähigkeiten, Vorwissen, kft) ein signifikanter Zusammenhang ab ( $\beta = -.256$ ) (Müller, 2013).

Bei all den Analysen zum Prognosewert von Schulnoten und zusätzlich erhobenen Leistungstests weist bereits Ingenkamp (1993) auf die Relevanz der Prädiktorenauswahl für die Vorhersagegültigkeit hin. Demnach müsse man sich damit abfinden, dass mit einer Kombination mehrerer Prädiktoren nicht zwangsläufig ein erhöhter Prognosewert für den Sekundarschulerfolg zu finden sei (Ingenkamp, 1993). Ähnlich dazu fragten schon Fingerhut und Langfeldt (1974) nach dem Nutzen einer besseren prognostischen Validität von Schulnoten, beinhalte das doch eine Konstanz des nicht zwangsläufig stabilen Konstrukts Schulleistung. Eine Stabilität könnte dann wiederum „dem Anspruch der Schule auf Förderung der Schüler entgegenwirken, da eine Verbesserung der Schulleistungen, vor allem bei den schwächeren Schülern, zu einem Absinken der Vorhersage-Validität führen müsste.“ (Fingerhut & Langfeldt, 1974, S.262). In diesem Sinne diskutiert auch Birkelbach (2011) den Nutzen einer validen Lehrerprognose, könnte diese Einschätzung doch auch immer mit einer *Self-Fulfilling Prophecy* einhergehen, in dem Sinne, dass eine Situationsdefinition aufgrund von leistungsnahen aber auch leistungsfremden Merkmalen die Entscheidung der Schüler bzw. Eltern in eine bestimmte Richtung lenken könnte.

Bilanzierend kommt der Gültigkeit von Beurteilungen bei ihrer Vorhersagefunktion nichtsdestotrotz gerade an Bildungsübergängen eine besondere Bedeutung zu. Noten haben also nicht nur eine punktuelle Leistung festzustellen, sondern müssen in Anbetracht der Selektions-

bzw. Differenzierungsfunktion eine Einteilung für weiterführende Schulformen treffen und beziehen sich dabei – qua Gesetz – auf Beobachtungen über längere Zeiträume als ein punktueller Test (Böhnel, 1996; Schrader, 1997). Systematische Abweichungen scheinen sich dabei für die ungleiche Ausschöpfung des Wertebereichs, der Referenzgruppe und den individuellen Einflussfaktoren zu ergeben, auf die sich wiederum Analysen zu Beurteilungsfehlern schulischer Notengebung beziehen (Maier, 2015). In diesem Sinne kann eine Beobachtung bzw. bilanzierende Bewertung über einen längeren Zeitraum auch immer unsystematische Urteilsfehler umfassen. Betrachtet man nochmals analyseübergreifend den prognostischen Wert und die Übereinstimmungsvalidität von Test und Schulnoten, fordert bereits Ingenkamp (1976a) eine deutlich höhere Übereinstimmung zwischen Leistungstests und Schulnoten als bisher gefunden werden konnten, begründet vor allem dadurch, dass sich aus Noten weitreichende Folgen für Bildungslaufbahnen ergeben können. Und auch wenn die aufgezeigten Evidenzen zur Validität von Schulnoten teilweise älteren Datums sind, weisen neuere Befunde auf ähnliche Erkenntnisse hin.

Inwiefern eine solche Übereinstimmung zwischen Testleistung und Schulnoten erreichbar ist und wie sinnvoll eine Überprüfung von Noten anhand psychometrischer Testverfahren scheint, wird im letzten Unterkapitel zur messtheoretischen Güte von Schulnoten auf Grundlage einiger Argumente einführend diskutiert.

### 3.4 Zur Beziehung von messtheoretischen Gütekriterien und Notengebung in der Schule

In Anbetracht der exemplarischen Befunde aus den vorangegangenen Unterkapiteln zur Übertragbarkeit messtheoretischer Gütekriterien auf die Notengebung kann zusammenfassend festgehalten werden, dass die traditionelle Leistungsbewertung durch Schulnoten den messtheoretischen Gütekriterien nur ansatzweise entsprechen kann. Nach den Befunden zur eingeschränkten Objektivität in schulischen Bewertungsprozessen ist es wenig verwunderlich, dass Aspekte der Zuverlässigkeit

sigkeit und Güte einer Bewertung nur noch teilweise erfüllt werden können. Schulnoten können demnach nicht im eigentlichen Sinne als Leistungsmessung bezeichnet werden, da sie als Teil eines sozialen Prozesses vermutlich immer Schwierigkeiten haben werden, die psychometrischen Gütekriterien vollständig zu erfüllen (Jachmann, 2003). Was aber bedeutet das nun für die schulische Bewertungspraxis? Was sind Vor- und Nachteile der eingesetzten Leistungstests, welche Aufgaben kann entgegengesetzt dazu die traditionelle Bewertungsform durch Noten einnehmen und welchen Mehrwert bzw. Sinn hat der häufig getätigte Vergleich zwischen diesen zwei Erhebungsverfahren?

Ausgehend von dem Faktum, dass die Forderung nach der Einführung von Messstandards vornehmlich im Zuge der Bildungsungleichheitsdebatte diskutiert wird, bieten Schulleistungstests nicht nur die Möglichkeit, ganze Bildungssysteme, Schulformen, Klassen, Lehrkräfte und Schüler miteinander zu vergleichen, sondern können auch subjektiv und/oder strukturell bedingte Abweichungen einzelner Bewertungsprozesse sichtbar machen (Dietrich & Fricke, 2013; Prenzel et al., 2013). Erkenntnisse, die dann wiederum häufig auf ihre grundrechtliche Legitimität hin geprüft und diskutiert werden. Die Frage nach der grundrechtlichen Legitimität von Einflussfaktoren auf die Leistungsbewertung beschäftigte bereits Ingenkamp (1976a). Selbst für ihn bedeutet ein gerechtes Lehrerurteil über einen Schüler nicht gleich ein objektives Urteil. Ein gerechtes Urteil müsse bei einer Verhaltensbewertung zwar immer Ausgangslage, Selbst- und Fremdbemühung, Lernbedingungen und Leistungspotential abwägen, allerdings weist er explizit auf die Trennung der Informationen (Messung und Bewertung) hin, die für ein gerechtes Urteil unbedingt unabhängig voneinander vorliegen müssten. Ein Faktum, das für ihn in der Notengebung nicht erfüllt wird, da diese keine explizite Trennung zwischen subjektiven Urteilen und unabhängigen Informationsfeststellungen ermöglichen würde (Ingenkamp, 1976a).

Im Gegensatz zu der subjektiv konnotierten Notengebung werden Schulleistungstests als jene objektiven Verfahren der Lernstandsdiagnostik bezeichnet, die normorientiert oder kriteriumsorientiert sein

können<sup>13</sup> (Ingenkamp & Lissmann, 2005). Eine der Grundideen der Anwendung formeller Schulleistungstests<sup>14</sup> ist es, neben einer subjektiv konnotierten Beurteilung mündlicher und schriftlicher Leistungen durch Lehrkräfte ein objektives Messverfahren zu konstituieren, auf dessen Grundlage eine möglichst treffende Prognose zukünftiger Schulerfolge erfolgen soll (Ingenkamp, 1995b; Langfeldt, 1984). Ausgehend von der Tatsache, dass die Beurteilung von Schülern eine schwierige und fehleranfällige Aufgabe für den Lehrer darstellt, weisen auch Schrader und Helmke (2014) darauf hin, dass objektive, gültige und zuverlässige Zusatzinformationen über den Leistungsstand eines Schülers für die Lehrkraft eine wertvolle Hilfestellung sein können. Normorientierte Testverfahren würden demnach Rückmeldung zur Leistung eines Schülers im Vergleich zur Alters- oder Jahrgangsnorm geben sowie lehrzielorientierte Tests Aufschluss über die zu erreichenden Lehrziele geben (Schrader & Helmke, 2014). Positiv ist bei vergleichenden Schulleistungsmessungen vor allem die damit verbundene Möglichkeit, durch den Vergleich zwischen Schülern, Klassen, Schulformen und Schulsystemen eine wichtige Grundlage für bildungspolitische wie pädagogische Entscheidungen zu erhalten und damit Stärken und Schwächen in schulischen Vorgängen aufzuzeigen (Ingenkamp, 1995b; Weinert, 2014). Internationale Schulleistungsvergleiche wie PISA oder IGLU stellen demnach wertvolles Orientierungswissen dar, sowie landesweite Orientierungsstudien Informationen über Schulen geben, die wiederum schulische und unterrichtliche Qualitätsentwicklung ermöglichen (Weinert, 2014).

Bezüglich des Mehrwerts von Schulleistungstests für die Notengebung gibt es bisher allerdings nur vereinzelte Erkenntnisse. Strietholt und Bos (2010) weisen darauf hin, dass der Einsatz von standardisierten, administrierten Schulleistungstests im Sinne der eingeführten Vergleichsarbeiten in der Schule eher eine Ausnahme darstellt und

---

13 Eine ausführliche Explikation der verschiedenen Testarten ist bei Ingenkamp und Lissmann (2005) ab Seite 155 zu finden.

14 Bei aller Diskussion um die Leistungserhebung in der Schule muss selbstverständlich immer zwischen experimentellen Studien zur Notengebung und der large-scale Forschung unterschieden werden.

weiterhin die diagnostische Kompetenz und Einschätzung der Lehrer eine bedeutsame Rolle einnimmt. Bezüglich ihrer These zum unterstützenden Einsatz von standardisierten Leistungstests fanden sie in ihrer Untersuchung keine Effekte. Aufgetretene Varianzen zwischen Schülerleistung im Lesen und der Deutschnote konnten demnach nicht durch die Frage erklärt werden, ob Lehrkräfte bei der Notenvergabe auf die Ergebnisse standardisierter Testleistungen Bezug nehmen oder nicht (Strietholt & Bos, 2010). Zudem verweisen zentrale Implikationen der Empirie aufgrund der mangelnden Übereinstimmung zwischen Noten und Testleistung immer wieder auf eine sozialstratifizierte Notenvergabe, woraufhin nicht selten das meritokratische Grundprinzip der Schule hinterfragt wird (z.B. Bos, Voss et al., 2004; Radtke, 2004). Jene der Schulleistungsdiagnostik zugesprochene Kontrollfunktion mittels Übereinstimmung zwischen Testleistung und Noten erscheint allerdings nur dann sinnvoll, wenn die Unterschiede zwischen Noten und Testerhebungen ausreichend beleuchtet werden.

Die kontroversen Diskussionen um den Einsatz von Schulleistungsmessungen in der Schule umfassen auch kritische Stimmen, die sich nach Weinert (2014) generell gegen den Einsatz von Leistungsbewertung bzw. Leistungsfeststellung richten. Unterschieden werden muss deswegen im Folgenden zwischen den kritischen Stimmen, die sich allgemein auf die Anwendung von Schulleistungstests in der Schule beziehen<sup>15</sup> und solchen, die sich kritisch gegenüber dem Einsatz von Schulleistungstests zur Überprüfung der Notenqualität äußern (Ingenkamp, 1989, 1995a; Lintorf, 2012). Ein Hauptargument der zweiten Ansicht bezieht sich auf die bereits erwähnte Frage, ob ein an wissenschaftlichen Kriterien orientierter Leistungstest schulinternen Curricula jemals vollständig entsprechen kann. Demnach erscheint die Frage nach der Relevanz der Testinhalte für Aussagen zur Qualität von Noten maßgeblich (Heller, 1974). Nach Brügelmann (2015) kann nicht davon ausgegangen werden, dass eine Testleistung den wahren Wert einer Schülerleistung darstellt, wodurch es unsinnig sei, „die

---

15 Eine informative Zusammenfassung jenes Kritikstranges ist bei Ingenkamp (1989) zu finden.

Qualität des einen Verfahrens allein durch den Grad der Übereinstimmung mit den Ergebnissen des anderen zu bestimmen“ (Brügelmann, 2015, S.119). Bezüglich der Testinhalte wird zudem nach der Verallgemeinerbarkeit von Ergebnissen einer einmaligen Testuntersuchung gefragt, die ebenfalls von individuellen Einschränkungen beeinflusst werden könnten (Schrader & Helmke, 2014). Im Gegensatz dazu gelten Lehrerurteile in Form von Noten meist als die „vollständigeren und repräsentativeren Informationen“ (Schrader & Helmke, 2014, S.51) über den Leistungsstand des Schülers.

Die Frage nach der Objektivität von Schulleistungstests wird demnach in Vergangenheit und Gegenwart durchaus kritisch diskutiert. Bereits für Wiese (1982) erscheint die Objektivierung von Ausleseverfahren nur dann sinnvoll, wenn die Empfehlung der Lehrkraft durch deren Mittelschichtgebundenheit sozial stratifiziert ist. Wenn man dagegen davon ausgeht, dass Lehrkräfte mit ihren Empfehlungen die Erfolgchancen eines Schülers in Abhängigkeit von dessen sozialen Voraussetzungen abschätzen, sei eine Objektivierung nicht sinnvoll (Wiese, 1982). Ebenso übt bereits Langfeldt (1984) Kritik an dem Einsatz formeller Schulleistungstests. Er gibt dabei zu bedenken, dass ein – suggeriert gerechteres – objektives Messverfahren individuelle Voraussetzungen von Schülern bei der Lernstandserhebung vollkommen außer Acht lasse. Bezüglich Reliabilitäts- und Validitätsansprüchen argumentiert er zudem, dass diese die schulischen Verhältnisse in dem Sinn stabilisieren, als dass prognostisch und zeitlich übereinstimmende Bewertungen „die Chancengleichheit und das Recht aller Schüler auf optimale Förderung“ (Langfeldt, 1984, S.96) eher behindern als fördern. Obgleich diese Aussage bereits Jahre zurückliegt, fokussiert sie doch einen Punkt, der auch in der aktuellen Diskussion um die messtheoretische Güte beachtenswert ist, hat sich doch an den Kriterien bis heute strukturell nichts geändert. Während Bürgermeister (2014) im Zuge der Einführung von Bildungsstandards und neuen Bewertungsformen auf ein sich verändertes Verständnis der Leistungsbeurteilung hinweist, sind es nach wie vor die Noten, welche Schüler letztendlich für weiterführende Bildungsgänge qualifizieren. Fakt ist, dass

Schulnoten qua Gesetz mehr Aspekte umfassen (sollen) als in einem einmaligen Messzeitpunkt erfasst werden kann.

Welche Möglichkeiten und Folgen ergeben sich nun also aus der Übertragung messtheoretischer Gütekriterien auf die Noten? Neben der bereits beleuchteten Korrekturfunktion von zusätzlich durchgeführten Leistungstests spricht Sacher (2009) gar von einer schulischen Dilemma-Situation. Demnach würde eine Objektivierung einer Leistungsmessung – in diesem Fall der Notengebung – gleichzeitig immer auch eine Schematisierung und Vereinheitlichung bedeuten:

Das Bemühen um die Sicherung von Objektivität geht somit ersichtlich auf Kosten der Validität: Das, was einigermaßen objektiv erhoben und beurteilt werden kann, repräsentiert die Schülerleistung nur noch sehr ausschnitthaft und unvollständig. (Sacher, 2009, S.48).

Ausgehend von einer pädagogischen Sichtweise auf Erziehung und Bildung als fortwährender Prozess wird die Gefahr der Einführung von Bildungsstandards und klassen- bzw. schulübergreifenden Vergleichsarbeiten diskutiert, bei welchen die subjektive Einschätzung der Lehrkräfte zwar kontrolliert, aber auch ihres pädagogischen Charakters beraubt werden könnte (Sacher, 2009). Die Sicherung jener Objektivität beinhaltet – gerade im Schulkontext – vermutlich immer eine gewisse Einschränkung der Leistungsmöglichkeiten und den Zwang zur Vereinheitlichung von Darbietung und Interpretationsschemata, um den Gütemaßstab leichter umzusetzen (Ingenkamp & Lissmann, 2005). Dennoch bietet dies gleichzeitig die Chance, sozial stratifizierte Urteilsprozesse zu kontrollieren. Allerdings wird darauf hingewiesen, dass nicht alle gezeigten Verhaltensweisen gleichermaßen gut für eine Messung zugänglich sein können. Während sich die sogenannten konvergenten Lernprozesse durch klar definierte Ziele besser für einen Messvorgang eignen, zeichnen sich divergente Lernprozesse durch viele mögliche Lösungsmöglichkeiten aus und erscheinen so – ohne eine explizite Definition über zu messende Eigenschaften – für einen Messvorgang wenig zielführend (Ingenkamp & Lissmann, 2005; Weiss, 1988). Demnach orientieren sich auch Schulleistungstests an

relational festgelegten Normen, zu welchen bereits Ingenkamp (1989) keine Alternative sah. Im Zuge der Objektivitätsdiskussion sah er ebenfalls eine Gegenüberstellung von Lehrerurteil in Form von Noten und Testwerten als unangemessen. Zusätzliche Leistungsmessungen sollten Lehrkräften vielmehr als unterstützende Information dienen, um subjektive Beobachtungen ggf. zu korrigieren (Ingenkamp, 1989).

Wie bereits angemerkt, weisen empirische Evidenzen allerdings darauf hin, dass Schulnoten vermutlich immer mehr umfassen, als mithilfe von Testleistungen erfasst werden kann. Lintorf (2012) vermutet sogar, dass Noten und Testleistungen differente Kompetenzaspekte erfassen und diskutiert in ihrer Arbeit die allgemeine Übertragbarkeit von psychometrischen Güteverfahren auf Schulnoten, indem sie einer Rekonzeptualisierung von Güte nachgeht, welche im Folgenden aufgrund ihrer Differenziertheit nicht weiter ausgeführt wird. Festgehalten werden soll an dieser Stelle allerdings der Hinweis, nach dem allein aus rein methodischen Gründen die Validierung von Noten mittels Testverfahren immer potenziell asymmetrisch verläuft<sup>16</sup>. Eine Asymmetrie zwischen Prädiktor und Kriterium, die wiederum eine vollkommene Übereinstimmung im Sinne des Symmetrieprinzips verletzt. Demnach muss für einen perfekten Zusammenhang zwischen zwei Messungen das Verhältnis zwischen Kriterium und Prädiktor symmetrisch verlaufen (Schmidt-Atzert & Amelang, 2012). Gemäß den in [Kapitel 2.5](#) dargestellten schulrechtlichen Grundlagen ist allerdings zu vermuten, dass Testleistungen den Leistungsstand eines Kindes nur ausschnitthaft repräsentieren können, wonach letztendlich die Frage offen bleibt, was in jener – vermutlich asymmetrischen Beziehung – mit Schulnoten gemessen wird.

Die pädagogischen Kontroversen um die unterschiedlichen Erfassungsmöglichkeiten schulischer Leistung umfassen bis heute viele Argumente, die hinreichend für oder auch gegen die jeweils andere Perspektive sprechen können. Weinert (2014) warnt allerdings davor,

---

<sup>16</sup> Asymmetrie liegt nach Lintorf (2012) dann vor, wenn sich Prädiktor und Kriterium in ihrer Breite und Differenzierung unterscheiden, ein eng gefasster Prädiktor also beispielsweise ein breit angelegtes Kriterium erfassen soll und umgekehrt.

dadurch „die tatsächlichen Ziele, de[n] wahrscheinliche[n] Nutzen und die möglichen Gefahren bestimmter Leistungsanforderungen und Leistungsmessungen“ (S.18) aus dem Blick zu verlieren. Zum Ende des Kapitels bleibt die Frage offen, wie man zukünftig die Qualität und Bestandteile von Schulnoten erfassen kann. Letztlich scheint hinter jeder empirischen Analyse die eingangs thematisierte Frage zu stehen, welche Leistungsaspekte bei der Leistungsbewertung durch Noten Berücksichtigung finden (sollten) und wie gut diese erfasst werden. Welche Komponenten von Schulnoten in Theorie und Empirie bisher fokussiert wurden, wird im nächsten Kapiteln erläutert.

## 4 Ausgewählte Bestandteile von Schulnoten – theoretische Basis und empirische Evidenzen

Ausgehend von den verschiedenen Sichtweisen auf den Begriff Schulleistung und den Chancen und Grenzen ihrer Überprüfung mittels messtheoretischer Gütekriterien wird im folgenden Kapitel ein Überblick über mögliche Determinanten schulischer Leistung gegeben. Nach einer Einführung in die theoretische Modellierung von Schulleistung, wird in [Kapitel 4.2](#) die Bedeutung individueller Merkmale des Schülers für die Zusammensetzung einer Note beleuchtet. Es folgt eine Ausführung zum Zusammenhang zwischen familiären Merkmalen und Schulnoten, um dann in [Kapitel 4.4](#) auf die Bedeutung schulischer Rahmenbedingungen für die Notenzusammensetzung einzugehen. Obgleich diese Darstellung keinesfalls ein vollständiges Bild der Bestandteile von Noten liefern kann, wird eine Auswahl jener kindlichen, familiären und schulischen Bedingungsfaktoren dargestellt, die in Vergangenheit und Gegenwart häufig als Einflussfaktoren für Schulleistungskomponenten diskutiert werden.

### 4.1 Ausgewählte Bedingungsmodelle von Schulleistung

Wie in [Kapitel 2](#) dieser Arbeit bereits deutlich wurde, ist für die begriffliche Einordnung von Schulleistung bzw. Schulnoten maßgeblich, aus welchen Komponenten sich die schulische Leistung zusammensetzt und zu welchen Teilen diese in Bewertungen berücksichtigt werden. Neben schulrechtlichen Bestimmungen, die mehr oder minder (fach-)spezifische, präzise Anforderungen an die Schulleistungskomponenten stellen, gibt es in Theorie und Empirie zahlreiche Modelle, welche mögliche Einflüsse bzw. Bedingungsfaktoren schulischer Leistung in schematischen Einteilungen systematisieren (z.B. Heller, 1995; Helmke & Schrader, 2010; Helmke & Weinert, 1997; Krapp, 1973, 1976; Sauer & Gattringer, 1985; Tent et al., 1976). Gemeinsam ist

allen Modellen, dass sie verschiedene Determinantenbereiche benennen und die multikausale Bedingungsstruktur von Schulleistung sichtbar machen; neben Faktoren der Schülerpersönlichkeit werden meist familiäre Faktoren und schulische Einflussbereiche betrachtet. Allen gemein ist die Ansicht, dass den im jeweiligen Modell einfließenden Wirkdeterminanten übergeordnete Strukturen vorausgehen: Kulturelle, gesellschaftliche, historische sowie wirtschaftliche Rahmenbedingungen, welche wiederum die einzelnen Determinanten des Modells beeinflussen und so indirekt immer Gegenstand empirischer Überprüfungen sind. Unterschiede zwischen den Bedingungsmodellen ergeben sich u.a. aus den einbezogenen Wirkmechanismen und deren Erhebungszeitraum, der Kausalität der Einflüsse und Rückkopplungsprozesse, der Differenzierung in zeitstabile und zeitvariable Leistungsbedingungen sowie dem unterschiedlichen Komplexitätsgrad der Darstellung (z.B. Helmke & Schrader, 2010; Krapp, 1973, 1976; Sauer & Gamsjäger, 1996; Tent et al., 1976). An dieser Stelle nicht ausgeführt, aber dennoch erwähnt werden sollen zudem drei Besonderheiten der Arbeiten von Krapp (1973) bzw. (1976). In seinen Modellen zu Determinanten der Schulleistung weist er zum einen auf die Relevanz der Differenzierung zwischen Messung und Bewertung von Leistung hin. Zum anderen unterscheidet er Prozesse innerhalb der Schülerpersönlichkeit in habituelle, zeitstabile Leistungsvoraussetzungen (z.B. Intelligenz) und zeitlich variable Leistungsbedingungen des Schülers (z.B. Motivation), wobei Leistung erst durch die Wechselbeziehung zwischen zeitstabilen und aktuellen Leistungsbedingungen erklärt werden könne. Daraus schlussfolgert er, sich der begrenzten Aussagekraft von Querschnittanalysen bewusst zu sein und bestmöglich zwischen einer prozessualen und strukturellen Forschungsperspektive zu differenzieren, um stabile Aussagen über das Leistungsverhalten zu erhalten (Krapp, 1976).

Aufgrund der Vielzahl an Überblicksarbeiten zur multikausalen Bedingungsstruktur wird im Folgenden auf zwei für die Arbeit wesentliche Modelle eingegangen. Das erste Modell von Helmke und Weinert (1997) bezieht sich auf der Makro-Ebene auf die einzelnen Komponenten der drei häufigsten Determinationsbereiche. Das zweite

Modell von Heller (1995) fokussiert die Zusammenhänge zwischen individuellen Prädiktoren von Schulleistung, wobei insbesondere auf Moderationseffekte zwischen den Prädiktoren eingegangen wird.

#### 4.1.1 Makro-Modell der Bedingungsfaktoren schulischer Leistungen nach Helmke und Weinert (1997)

Die schematische Einteilung zur allgemeinen und bereichsspezifischen Modellkonzeption der Bedingungen schulischer Leistung von Helmke und Weinert (1997) dient der vorliegenden Arbeit als übergeordnetes Makro-Modell, welches die multikausale Bedingungsstruktur schulischer Leistungen aufzeigt. Das übergeordnete Modell fokussiert auf der Makro-Ebene die verschiedenen Perspektiven der Forschung, die zur Erklärung und Vorhersage von Schulleistung in ihren Zusammenhängen berücksichtigt werden sollten (vgl. [Abbildung 1](#)).

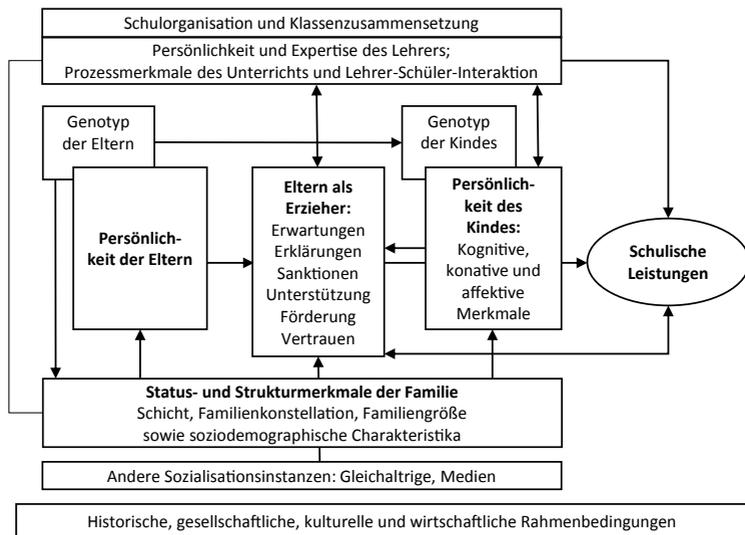


Abbildung 1: Makro-Modell zu Determinanten schulischer Leistung nach Helmke und Weinert (1997, S. 86)

Im Wesentlichen handelt es sich neben den bereits erwähnten kulturellen, gesellschaftlichen und wirtschaftlichen Rahmenbedingungen auch

in diesem Modell um drei unterschiedliche Perspektiven: Die individuellen Schülermerkmale, die familiären Hintergrundmerkmale sowie die schulischen Rahmenbedingungen. Zunächst erscheint diese Aufteilung jener von Krapp (1976) zu entsprechen, allerdings differenzieren die Autoren dann nochmals in spezifische Wirkmechanismen der drei Gegenstandsbereiche. So werden die individuellen Schülermerkmale in kognitive, konative und affektive Determinanten differenziert. Die familiären Merkmale werden in Status- und Strukturmerkmale und Prozessmerkmale unterschieden. Der Einfluss der Prozessmerkmale auf die Schulleistung erfolgt durch vier Funktionen elterlichen Verhaltens: Stimulation, Instruktion, Motivation und Imitation. Als *stimulierende* Effekte werden jegliche, unmittelbare Formen anregender, materieller Umwelt seit der frühen Kindheit bezeichnet, die *Instruktion* bezieht sich auf eine unmittelbare, kognitive Förderung wie z.B. Hausaufgabenunterstützung oder Fördereinrichtungen. Von der *motivationalen* Beeinflussung werden vornehmlich indirekte Effekte für die Schulleistung erwartet, dergestalt, dass beispielsweise die elterlichen Aspirationen die individuellen Schülermerkmale beeinflussen, die sich dann wiederum in der Schulleistung des Kindes widerspiegeln. Von der *Imitation* elterlichen Verhaltens bzgl. sprachlicher, motivationaler Lern- und Arbeitsstrategien wird in dem Übersichtsartikel von Helmke und Weinert (1997) ebenfalls von Effekten auf kindliche Leistungsmerkmale berichtet.

Ohne auf jede einzelne Komponente der schematischen Darstellung einzugehen, werden in der vorliegenden Arbeit in [Kapitel 4.2](#), [4.3](#) und [4.4](#) einzelne Determinanten des Modells erläutert. Während das Modell von Helmke und Weinert (1997) die multikausale Bedingungsstruktur schulischer Leistung eindrucksvoll darstellt, macht es aufgrund seiner übergeordneten Makrostruktur keine Aussagen über die Zusammenhänge zwischen den einzelnen individuellen Schülermerkmalen. Diese Lücke schließt das nachfolgende Modell von Heller (1995), welches die Bedingungsstruktur auf Schülerseite fokussiert.

### 4.1.2 Multikausales Bedingungsmodell schulischer Leistung nach Heller (1995)

Aufgrund der Vielzahl an inhaltlichen Überschneidungspunkten zwischen den Bedingungsmodellen sowie mannigfachen Operationalisierungsvarianten wird abschließend auf das Modell von Heller (1995) eingegangen, das u.a. innerhalb der individuellen Schülermerkmale zwischen kognitiven und nicht-kognitiven Prädiktoren differenziert und diese in einen Wirkzusammenhang zueinander stellt (vgl. [Abbildung 2](#)).

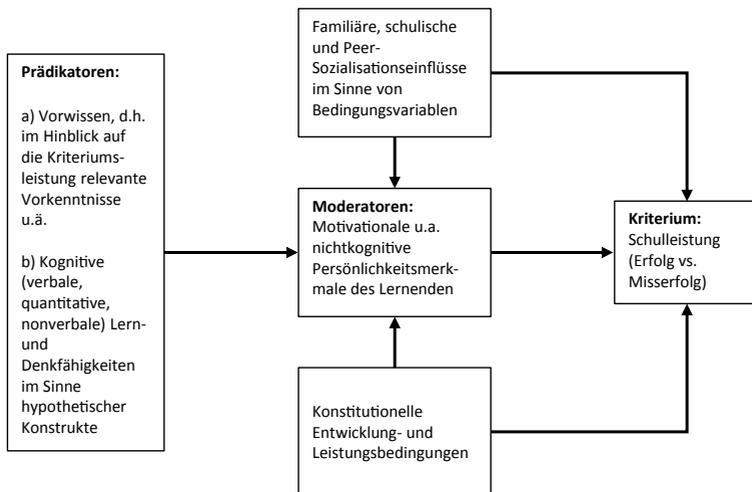


Abbildung 2: Allg. Bedingungsmodell der Schulleistung nach Heller (1995, S. 984)

Bemerkenswert ist an diesem Bedingungsmodell einmal, dass explizit zwischen *Prädiktor- und Kriteriumszusammenhängen* und *Moderatoreffekten* von Schulleistung unterschieden wird. Unter Prädiktoren von Schulleistung werden jene Variablen gefasst, die einen kausalen oder zumindest korrelativen Zusammenhang zur Schulleistung aufweisen. Als Moderatoren (in diesem Modell grafisch nicht zu verwechseln mit Mediationseffekten) werden jene Persönlichkeitsmerkmale gezählt, durch die der Zusammenhang zwischen kognitiven Merkmalen und Schulleistung moderiert wird. Während in diesem Modell also

beispielsweise Vorwissensstrukturen und kognitive Grundfähigkeiten des Kindes prädiktiven Charakter vorweisen, sind die nichtkognitiven Merkmale als Moderatoren aufgeführt, die wiederum von familiären, schulischen und konstitutionellen Leistungsbedingungen beeinflusst werden und zusammen genommen Effekte auf die Schulleistung des Kindes moderieren (Heller, 1995). Aufgrund der multikausalen Bedingungsstruktur des Kriteriums Schulleistung weist Heller (1995) darauf hin, *unterschiedliche diagnostische Informationsquellen* für Prädiktor- und Moderatorvariablen in Analysen miteinzubeziehen. Zudem verweist er auf eine zunächst eingeschränkte Gültigkeit von Schulerfolgsprognosen, die je nach Einbezug unterschiedlicher Prädiktor- und Moderatoreffekte, den beobachteten Zeitpunkten<sup>17</sup> (kurz- vs. langfristig) sowie deren Operationalisierung (Schulerfolg vs. Schulnote) variieren (Heller, 1995).

Nach Heller (1997) ergeben sich u.a. zwei wesentliche Fragestellungen:

1. Die Rolle kognitiver und nicht-kognitiver Schulleistungsindikatoren (Prädiktor-Kriteriumszusammenhänge) und
2. die Kovariationen bzw. Wechselwirkungen zwischen den einzelnen Indikatoren (Moderatoreffekte).

Unter Bezug auf den Leistungsstand des Kindes werden nach diesen Modellen unterschiedliche Effekte erwartet: Während eine gültige Vorhersage bei herausragenden Leistungen allein durch kognitive Variablen prognostiziert werden kann, bedarf es bei der Mehrzahl der Schüler einen Einbezug weiterer Moderatorvariablen (nicht kognitiv, konstitutionell, sozial). Folglich stellen kognitive Prädiktorvariablen „zwar eine notwendige, aber in vielen Fällen keine hinreichende Voraussetzung für treffsichere Schulleistungsprognosen“ (Heller, 1995, S.983) dar; gute Werte in den Moderatorvariablen führen zu engeren Prädiktor-Kriteriumszusammenhängen. Eine annähernd gültige Erklärung von Schulleistung bedarf also des Einbezugs sowohl kognitiver

<sup>17</sup> Heller (1995) empfiehlt im Schulkontext vornehmlich kurz- oder gegebenenfalls mittelfristige Schulleistungsprognosen, da neben individuellen Entwicklungsprozessen auch Effekte von pädagogischen Maßnahmen bzw. ihrer Erfolge zu erwarten sind.

als auch nicht-kognitiver Persönlichkeitsmerkmale, wobei letztere als Vermittlungsmechanismen zwischen kognitiven Prädiktoren und Leistungskriterien angenommen werden<sup>18</sup>.

Zusätzlich zu individuellen Prädiktoren schulischer Leistung verweist das Modell auf die Bedeutsamkeit sozialer Bedingungsvariablen. Heller (1995) subsumiert in seinem Modell sämtliche familiären, schulischen und peerbezogenen Sozialisierungseinflüsse unter einem Punkt, die sich sowohl direkt als auch vermittelt über die nicht-kognitiven Schülermerkmale auf die Schulleistung auswirken. Warum der indirekte Einfluss lediglich auf die nicht-kognitiven Persönlichkeitsmerkmale wirkt, wird nicht erläutert. Demgegenüber sei an dieser Stelle beispielsweise auf das Modell von Sauer und Gamsjäger (1996) verwiesen, welches entgegen dazu von mediiierenden Effekten zwischen familiären Hintergrundmerkmalen, kognitiven und nicht-kognitiven Schülermerkmalen und Schulleistung ausgeht.

Obwohl das vorgestellte Modell im Gegensatz zu den beiden vorangegangenen Bedingungsmodellen konkrete Aussagen über das Verhältnis zwischen kognitiven und nicht-kognitiven Bedingungs-faktoren von Schulleistung zulässt, ist es ebenfalls auf der Makroebene angesiedelt und kann so zu unterschiedlichen Operationalisierungsvarianten führen. Betrachtet man die zwei vorgestellten Modelle von Helmke und Weinert (1997) und Heller (1995) bezüglich ihrer Differenziertheit, ist es nicht verwunderlich, dass sich in der empirischen Umsetzung teils differente Untersuchungsergebnisse in Abhängigkeit von der allgemeinen oder bereichsspezifischen Modellzusammensetzung ergeben. Auch die geringe Vergleichbarkeit zwischen Untersuchungsstichproben, Indikatoren oder methodischen Analyseverfahren lassen eine allgemeingültige Bedingungsstruktur von Schulleistung nahezu unmöglich erscheinen (Heller, 1997). Während die vorgestellten

18 Heller (1995) weist in seiner Abhandlung zudem darauf hin, dass aufgrund von Messproblemen im Kindes- und Jugendalter immer nur von einer bedingten Vorhersagegültigkeit ausgegangen werden kann. Auch wenn bei den hier nachfolgend aufgeführten Schülermerkmalen von relativ stabilen Leistungsmerkmalen ausgegangen wird, muss bei Schülermerkmalen gleichzeitig immer die Variabilität bzw. Stabilität von Prognosemerkmalen bedacht werden (Heller, 1995).

Modelle vornehmlich auf übergeordneter Ebene die Bedingungsstruktur schulischer Leistungen fokussieren, wird im Folgenden auf diejenigen Bestandteile von Schulleistung näher eingegangen, welche in zahlreichen empirischen Analysen aufgegriffen werden und die für die nachfolgende, eigene empirische Untersuchung relevant sind.

Die folgende Ausführung beinhaltet sowohl theoretische Erklärungsansätze als auch empirische Ergebnisse aus Untersuchungen, die Aussagen zu Komponenten von Schulleistung zulassen, wobei kein Anspruch auf Vollständigkeit besteht. Nach Betrachtung individueller Bedingungsfaktoren auf Seiten des Schülers, folgt eine Analyse familiärer Einflussfaktoren, um abschließend auf die Bedeutung schulischer Rahmenbedingungen für die Notenzusammensetzung einzugehen. Die aufgeführten Bedingungsfaktoren fokussieren bewusst unterschiedliche theoretische und empirische Zugänge, um einen Einblick in die aktuellen wie historisch diskutierten Bedingungsfaktoren schulischer Leistung zu geben<sup>19</sup>. Im Fokus stehen vornehmlich jene Befunde, die sich mit dem Zusammenhang von Schulleistungskomponenten und Leistungsbewertungen befassen, häufig unterteilt in proximale (kognitive, motivationale und metakognitive Fähigkeiten, familiäre Faktoren) und distale Variablen (schulische und soziodemographische Merkmale) (Heller, 1997; Wang, Haertel & Walberg, 1993). Die daraus abgeleiteten Annahmen werden am Schluss des vierten Kapitels in einem eigenen Modell zusammengefasst und anschließend einer empirischen Prüfung unterzogen.

---

19 Die lerntheoretische Perspektive auf das Konstrukt Schulleistung ist in diesem Kontext bewusst nicht gewählt worden, da es bei der folgenden Untersuchung nicht um die Lernprozesse bei dem Schüler selbst geht, sondern bewusst die Fremdperspektive auf relevante Bedingungsfaktoren gewählt wurde. Weiterführende Übersichten über kognitive, sozialkognitive und konstruktivistische Lerntheorien sind bei Woolfolk (2008) und Helmke (2009) zu finden.

## 4.2 Individuelle Bedingungsfaktoren von Schulleistung

Als individuelle Bedingungsfaktoren von Schulleistung werden im Folgenden unter Bezug auf die in [Kapitel 4.1](#) eingeführten Bedingungsmodelle all jene individuellen Determinanten von Schulleistung betrachtet, welche in Theorie und Empirie häufig diskutiert werden. Zudem sind es Faktoren, die laut Schulrecht im Bewertungsvorgang der Lehrkräfte – zusätzlich zur Einhaltung messtheoretischer Gütekriterien – Beachtung finden können. Empirische Befunde stammen häufig aus quantitativen Korrelationsstudien, die meist im Zuge der Vorhersage schulischer Leistung entstanden sind und sich nicht selten auf Leistungsprozesse am Übertritt in die Sekundarstufe I beziehen. Die erwähnten Studien untersuchen meist Einflussfaktoren auf Schulleistung, Bildungs-/Schulerfolg oder Übertrittsempfehlungen und verweisen damit eher indirekt auf die Zusammensetzung von schulischer Leistung, ausgedrückt in Noten<sup>20</sup>. Je nach (impliziter) Definition von Schulerfolg unterscheidet sich in Studien also die abhängige Variable nach Leistungstests, Übertrittsempfehlung oder Noten; meist in Form von Durchschnittsnoten dargestellt. Da Noten für den weiteren Bildungsverlauf im Schulsystem und für die nachfolgende Untersuchung maßgeblich sind, konzentriert sich die folgende Forschungsperspektive vornehmlich auf Prädiktoren von Schulnoten und deren moderierende Effekte. Zusätzlich dazu wird immer beleuchtet, aus welcher Perspektive die erfassten Prädiktoren bzw. Determinanten von Noten erhoben wurden. So könnte man beispielsweise annehmen, dass sich Lehrkraftangaben zu individuellen Schülereigenschaften in ihrer Passung von Einschätzungen der Schüler selbst unterscheiden. Faktoren,

---

20 Bezüglich der Ursachenforschung schulischer Leistungsindikatoren weist bereits Gaedike (1974) darauf hin, dass in der pädagogischen Psychologie meist nur Aussagen über Ursachen-Wirkungs-Zusammenhänge getroffen werden können. Korrelative Zusammenhänge geben demnach – bis heute – keine Auskunft darüber, ob einzelne Indikatoren enthaltene Faktoren von Leistung sind oder diese gar verursachen. Beispielsweise könnte Vorwissen im Fach Deutsch ein Teil der Deutschnote sein, gleichzeitig diese Deutschnote aber auch bedingen; Intelligenz kann als Determinante wirken, aber auch durch schulisches Lernen beeinflusst werden usw. (Gaedike (1974); Helmke und Weinert (1997)). Aufgrund der mangelnden Überprüfbarkeit dieser Tatsache werden im Folgenden die beiden Begriffe Determinante und Faktor synonym verwendet.

die im Kind verankert sind, werden je nach theoretischer Modellierung und empirischer Operationalisierung beispielsweise in kognitive und nicht-kognitive, motivationale, volitional/konative, affektive, konstitutive oder sozial-emotionale Merkmale differenziert.

Internationale und nationale empirische Studien zu einzelnen Determinanten der Schulleistung existieren vielfach. Hauptsächlich sind es Überblicksarbeiten älteren Datums, die eine detaillierte Zusammenfassung von Schulleistungsdeterminanten – aus psychologischer und soziologischer Perspektive – liefern (z.B. Helmke & Weinert, 1997; Krapp, 1973; Lavin, 1965; Naylor, 1972; Wang et al., 1993). Als übergreifendes Ergebnis erscheint die Tatsache, dass die Zusammenhänge unter den Prädiktoren selbst sowie deren Prädiktionskraft auf Schulleistungskriterien je nach Stichprobe und Operationalisierung variieren (Helmke & Weinert, 1997).

Die individuellen Indikatoren von Schulleistung werden im deutschsprachigen Raum wiederum meist auf Grundlage empirisch ermittelter Beziehungen in lernrelevante, schulnahe, leistungsnahe oder leistungsferne Merkmale kategorisiert<sup>21</sup>, die aufgrund ihrer vielfältigen Interpretationsspielräume im Folgenden nicht weiter verfolgt werden<sup>22</sup>. Nachfolgend expliziert werden daher zunächst theoretische Bezüge und Befunde zu sieben individuellen Determinanten der Schulleistung; differenziert in kognitive und nicht-kognitive Schülermerkmale<sup>23</sup>. Die Betrachtung kognitiver Komponenten umfasst die

---

21 Obwohl die nachfolgend explizierten Determinanten nicht in dieses, wertende Kategoriensystem eingeordnet werden, wird angemerkt, dass jene Differenzierungen vermutlich implizit Aussagen darüber zulassen, was in der jeweiligen Gesellschaft sowie deren Bildungsinstitutionen als Leistung definiert wird.

22 Hinterfragt werden kann an dieser Stelle, was jene Betitelung in lern- oder leistungsrelevante Schülermerkmale zur Folge haben könnte. Jene Betitelungen könnten im Sinne der impliziten Persönlichkeitstheorie nach Hofer (1986) bei der Beurteilung von Schülereigenschaften als hinreichend genauer Indikator für eine Gymnasialeignung angesehen werden. Unschärfe, weit interpretierbare Begrifflichkeiten, die für Leistungsbeurteilungen als Argumentationsgrundlage dienen, ohne genau explizieren zu müssen, welches Schülerverhalten genau darunter verstanden wird.

23 Anzumerken ist an dieser Stelle, dass die erfolgte Zuordnung zu kognitiven bzw. nicht-kognitiven Bedingungsfaktoren vornehmlich aus empirischen Untersuchungen stammt. Überschneidungspunkte bzw. differente Einteilungen werden bewusst nicht betrachtet, auch besteht kein Anspruch auf Vollständigkeit.

individuellen Testleistungen, Intelligenz, Vorwissen, Begabungskomponenten und sprachliche Fähigkeiten. Das schulische Arbeits- und Sozialverhalten wird als nicht-kognitiver Prädiktor aufgeführt. Wie zu Beginn von [Kapitel 4](#) bereits angeführt, wird in modelltheoretischen Annahmen davon ausgegangen, dass der Zusammenhang zwischen den kognitiven Persönlichkeitsmerkmalen und dem Schulleistungskriterium von nicht-kognitiven Merkmalen moderiert wird, wobei dieser je nach Leistungsstand des Kindes unterschiedlich stark ausgeprägt sein kann. Für die Notenzusammensetzung erscheint hier nicht nur relevant, ob Schüler z.B. über eine hohe Intelligenz, sprachliche Fähigkeiten sowie über ein ausgeprägtes Vorwissen verfügen. Relevant scheint auch, ob sich diese Eigenschaften wiederum in einem entsprechenden Arbeitsverhalten bzw. motivationalen und sozial angepassten Verhaltensweisen zeigen, die dann wiederum die Schulnote beeinflussen. Empirische Evidenzen der Meta-Analyse von Wang et al. (1993) verweisen für den nordamerikanischen Raum auf ähnliche Tendenzen. Demnach kommen vornehmlich kognitiven, aber auch metakognitiven Kompetenzen und motivationalen Faktoren die größte Bedeutung bei der Erklärung von Schulleistung zu.

Unter konstitutiven Bedingungen schulischer Leistung wird neben Einflüssen des Alters häufig auch die Relevanz des Schülergeschlechts betrachtet, das in der vorliegenden Untersuchung als Kontrollvariable aufgenommen wird. Ebenso anzumerken ist an dieser Stelle, dass im Folgenden zwar ganz allgemein von Persönlichkeitseigenschaften des Kindes gesprochen wird, es aber nicht Ziel der Arbeit ist, sich mit den wissenschaftlichen Erkenntnissen der menschlichen Persönlichkeit auseinander zu setzen, nach welchen man sicherlich viel differenzierter zwischen statischen und variablen Eigenschaften trennen müsste (Gaedike, 1974).

#### 4.2.1 Kognitive Prädiktoren von Schulleistung

Die folgende Ausführung bezieht sich auf fünf ausgewählte Schülermerkmale, die häufig den kognitiven Fähigkeiten eines Kindes zugeschrieben werden. Nach einer begrifflichen Einführung werden für jeden Indikator empirische Befunde vorgestellt, die den Zusammenhang zwischen kindlichen Merkmalen und Schulleistung fokussieren, erhoben im Primarbereich und am Übertritt in die Sekundarstufe I.

#### Testleistung

Wie in [Kapitel 3.3](#) der vorliegenden Arbeit bereits thematisiert, deklarieren nationale wie internationale Befunde zu Determinanten von Noten einstimmig den stärksten Zusammenhang zwischen Schulnoten und der zusätzlich dazu erhobenen, fachspezifischen *Testleistung* (z.B. Ditton, 2007b; Heller, 1997; Lintorf, 2012; Weinert & Helmke, 1997). Auch wenn in einschlägiger Literatur der Zusammenhang zwischen dem Prädiktor Testleistung und Noten nicht immer gleich starke Effekte erzielt, ist der Zusammenhang durchgängig stabil (Hochweber, 2010; Krüsken, 2007a; Schütte et al., 2007). Dennoch wird aufgrund mangelnder vollständiger Übereinstimmung häufig diskutiert, welche Faktoren in Schulnoten – zusätzlich zur Testleistung – einfließen. Bisher veröffentlichte Ergebnisse der Koala-S Studie bestätigen ebenfalls den hohen Zusammenhang zwischen Testleistungen und Noten, deuten aber auch darauf hin, dass zusätzlich zu Testleistungen noch andere Eigenschaften des Schülers mit in dessen Benotung einfließen (Krüsken, 2007a). Diskutiert wird daraufhin neben Vermutungen über diagnostische Fehleinschätzungen von Lehrkräften nicht selten, dass Schulnoten qua Gesetz immer eine größere Anzahl an Schülereigenschaften erfassen, als dass mit einzelnen Testerhebungen möglich ist (z.B. Anders et al., 2010; Gräsel, Krolak-Schwerdt, Nölle & Hörtermann, 2010). Auch wird die zusätzliche Bedeutung der allgemeinen Intelligenz von Schülern für die Notenzusammensetzung diskutiert.

Zusammenfassend liegt die durchschnittliche Korrelation zwischen fachspezifischen Testleistungen und Schulnoten im Bereich von  $r=-.34$  bis  $r=-.72$  (Krüsken, 2007a; Schütte et al., 2007), womit weiterführend die Frage offen bleibt, welche Zusammenhänge die objektiv

erhobenen Testleistungen mit den von der Lehrkraft eingeschätzten Merkmalen aufweist, die wiederum ebenfalls die Notenzusammensetzung beeinflussen. Im Sinne einer diagnostisch validen Einschätzung müsste der Zusammenhang zwischen Einschätzungen, Testleistungen und Schulnoten relativ hoch sein. Welche Bedeutung die allgemeine Intelligenz des Kindes für die Zusammensetzung einer Note aufweist, wird im Folgenden betrachtet.

### **Intelligenz**

Die Betrachtung des Zusammenhangs zwischen Intelligenz und Schulnoten ist ein seit langer Zeit diskutiertes Thema (z.B. Gaedike, 1974; Heller, 1973). Einigkeit scheint bis heute darin zu bestehen, dass die Stärke des Zusammenhangs zwischen der Intelligenz eines Schülers und seiner Schulleistung je nach Art des Intelligenztests und Schulleistungskriteriums durchaus variieren kann (Gaedike, 1974; Helmke & Schrader, 2010; Wang et al., 1993). Während in der einschlägigen Literatur Intelligenz häufig mit Begabung gleichgesetzt wird, wird innerhalb der Intelligenzforschung nochmals zwischen allgemeiner Intelligenz und bereichsspezifischen Unterarten (z.B. sozialer, emotionaler und praktischer Intelligenz) differenziert, die dann wiederum in Anteile fluider und kristalliner Intelligenzfaktoren unterschieden werden können (Gut, Reimann & Grob, 2012; Heller, 1973; Hoyer, 2012; Rost, 2009). Bekannte Definitionsversuche beziehen sich nach Heller (1973) immer auf die Leistungsdisposition eines Individuums, die wiederum andere Merkmale beeinflussen kann. So konstatiert beispielsweise bereits Groffmann (1964) Intelligenz als

die Fähigkeit des Individuums, anschaulich oder abstrakt in sprachlichen, numerischen und raum-zeitlichen Beziehungen zu denken; sie ermöglicht erfolgreiche Bewältigung vieler komplexer und mit Hilfe jeweils besonderer Fähigkeitsgruppen auch ganz spezifischer Situationen und Aufgaben. (Groffmann, 1964, S.190).

Während die spezifischen Differenzierungen von Intelligenz im Folgenden nicht weiter expliziert werden, beziehen sich schulnahe Intelligenztests immer auf die Erfassung allgemeiner, kognitiver

Leistungsfähigkeit. Die Art des verwendeten kognitiven Fähigkeitstests, die jeweilig zugrundeliegende Stichprobe und das erfasste Leistungskriterium sind vermutlich Gründe, weswegen Ergebnisse zum Zusammenhang zwischen Intelligenz und Schulleistung erhebliche Unterschiede aufweisen (Gaedike, 1974; Heller, 1973, 1997; Helmke & Weinert, 1997). Der Intelligenztest für den Grundschulbereich KFT 1-3 bzw. CFT 20 – wie er auch in der nachfolgenden Untersuchung verwendet wird – gilt allerdings als relativ stabil (Heller & Geisler, 1983; Perleth & Sen, 2010; Weiß, 1998). So fanden Weinert und Helmke (1997) in ihrer vierjährigen Längsschnittstudie im Grundschulbereich für den Zusammenhang zwischen Testleistung in Mathematik und Intelligenzleistung in den Klassenstufen zwei bis vier Werte von  $r=.51$  bis  $r=.47$  (Helmke, 1997). Ebenso weisen Befunde der älteren Studie von Sauer und Gattringer (1985), in der Intelligenz mittels eines mehrfaktoriellen Tests erhoben wurde, auf Zusammenhänge zwischen Schulleistung (Faktor aus Zeugnisnote der 4.Volksschulklasse und Schulleistungstest) und Intelligenz von  $r=.74$  hin. Auch Sauer und Gamsjäger (1996) erfassen in ihrer Studie zur Vorhersage von Schulerfolg die kindliche Intelligenz als Prädiktor für Schulleistung. Dabei ergeben sich zwischen Intelligenz und Grundschulnote Ende der 4. Grundschulklasse Zusammenhänge von  $r=-.59$  im Fach Deutsch. Weiterführend weisen die Ergebnisse von Müller (2013) auf geringe Effekte zwischen Testintelligenz des Schülers (Kognitiver Fähigkeitstest Kft) und Schulleistungskriterien hin. Allerdings wirkt sich der Kft, vermittelt über den Leistungsindex der zweiten Klasse, signifikant auf die Viertklassnote aus ( $\beta=-.013$ ) (Müller, 2013). Differenziert in Zusammenhänge zwischen Mathematik- und Sachkundenoten der vierten Grundschulklasse, findet dagegen Lintorf (2012) für die Intelligenz eines Schülers – erhoben durch eine Skala aus dem KFT und in ihren Analysen als Begabung betitelt – nach der Testleistung die stärksten, signifikanten Zusammenhänge mit den Schulnoten ( $\beta=.167$  für Mathematik;  $\beta=.144$  für Sachkunde).

Nach diesen exemplarisch aufgeführten, teils widersprüchlichen Befunden weisen auch Helmke und Weinert (1997) in ihrem Forschungsüberblick darauf hin, dass ein differenzierter Blick auf

strukturelle Fähigkeitskomponenten von Intelligenz ein kompliziertes, gar unübersichtliches Forschungsfeld darstellt. Die übergreifende Frage zum Zusammenhang zwischen Intelligenz und Schulnoten ist allerdings – auch für die beiden Autoren – *wodurch* Zusammenhänge zwischen individuellen Schülermerkmalen und Schulleistung ursächlich beeinflusst werden (Helmke & Weinert, 1997; Sauer & Gamsjäger, 1996). Fragt man sich also in der Folge nach einem zugrundeliegenden Faktor, bleibt offen, welche direkten und indirekten Effekte die Intelligenzleistung auf die Grundschulnote eines Schülers ausübt. Vereinzelt Befunde weisen darauf hin, dass die Intelligenz als Vermittlungsfaktor eine bedeutsame Rolle einzunehmen scheint (Sauer & Gamsjäger, 1996; Sauer & Gattringer, 1985).

### **Vorwissen**

Zusätzlich zur Intelligenz weisen zahlreiche Forschungsbefunde wiederholt auf die Bedeutsamkeit des fachspezifischen Vorwissens als Prädiktor von Schulleistung hin (z.B. Ditton & Krüsken, 2006a, 2009; Heller, 1995; Helmke & Schrader, 2010; Müller, 2013). Unter Vorwissen werden dabei all jene Komponenten gefasst, „die Kenntnisse (Wissen, daß; deklaratives Wissen) und Fertigkeiten (Wissen, wie; prozedurales Wissen) einer Person in einem bestimmten Gegenstandsbereich“ umfassen (Renkl, 1996, S.175).

Während Ingenkamp (1995b) auf Grundlage älterer Studien auf die uneinheitliche Befundlage bezüglich des Zusammenhangs zwischen Schulnoten und Vorleistungen hinweist, heben vor allem neuere Befunde die Bedeutsamkeit der Differenzierung zwischen Intelligenz- und Vorwissenstrukturen für die Vorhersage von Schulnoten hervor. Demnach geht von bereichsspezifischem Vorwissen eine stärkere Prädiktionskraft für Schulleistung aus als von Intelligenzfaktoren (Schrader, 2009). Ohne näher auf mögliche Erklärungsansätze und Wirkweisen bezüglich des Zusammenhangs zwischen Vorwissen und Noten einzugehen, erscheint diese Tatsache nicht weiter verwunderlich, wenn man Lernen als konstruktiven Prozess betrachtet (Renkl, 1996). Ähnliche Ergebnisse ergaben auch die in der Scholastik-Studie gefundenen korrelativen Zusammenhänge zwischen Vorkenntnissen,

Intelligenz und Testleistungen. Vor allem die Bedeutung der bereichsspezifischen Vorwissenstrukturen wurde hervorgehoben, wonach sich nach Auspartialisieren der Vorkenntnisse der Zusammenhang zwischen Intelligenz und Testleistung deutlich reduzierte (Weinert & Helmke, 1997). Obwohl auch Schrader und Helmke (2008) in ihrem Übersichtsartikel auf die Bedeutsamkeit und Enge des Zusammenhangs zwischen Intelligenz und Vorwissen eingehen, weisen auch sie darauf hin, dass Wissensdefizite nicht allein durch vorhandene Intelligenz kompensiert werden können.

Ohne weiter auf die Grundlagen der Intelligenzforschung einzugehen, sind bei der vielfach nachgewiesenen, prädiktiven Kraft von Vorwissenstrukturen für Schulnoten – auf empirischer Ebene – zweierlei Punkte zu beachten. Zunächst erscheint es angesichts der Befunde zur Stabilität von Schulleistungsprädiktoren nicht überraschend, dass das Vorwissen eines Schülers für seine aktuellen Schulnoten bedeutsam ist, zumal die von Lehrkräften verfassten Schulnoten häufig einerseits als Prädiktoren und andererseits als Kriterium herangezogen werden (Heller, 1997). Diese Tatsache führt zu dem zweiten Punkt, dass bei empirischen Befunden zur prädiktiven Bedeutsamkeit von Vorwissen immer auch die entsprechenden Operationalisierungsvarianten von Vorwissen beachtet werden müssen. In Studien verwendete Vorwissenprädiktoren variieren z.B. je nach Stichprobe zwischen Angaben der Lehrkraft zum Vorwissen des Kindes zu Beginn der Schulzeit, vorausgegangenen Durchschnittsnoten und aus vorausgegangenen (fachspezifischen) Testleistungen des Schülers (z.B. Ditton & Krüsken, 2009; Dummert, Endlich, Schneider & Schwenck, 2014; Müller, 2013; Sauer & Gamsjäger, 1996; Wang et al., 1993; Weinert & Helmke, 1997).

Für den Grundschulbereich beispielsweise konnten Ditton und Krüsken (2006a) 51% der Varianz in den Viertklassnoten durch die Eingangstestleistung aus der zweiten Grundschulklasse erklären. Ebenso fanden die beiden Autoren (2009) mit Daten der Koala-S-Studie enge Zusammenhänge zwischen Testleistungen der zweiten Klassen und den Schülertestleistungen in der dritten und vierten Grundschul-

klasse (für den zusammengefassten Leistungsindex  $r=.77$ ,  $r=.80$ ). Ebenfalls mit Daten der Koala-S-Studie fand Müller (2013) bei der Vorhersage der Viertklassnote für die Durchschnittsnote der Hauptfächer der zweiten Klasse signifikant größere Effekte ( $\beta=.239$ ) als für den, deutlich geringeren Effekt des Leistungsindex der zweiten Klasse ( $\beta=-.097$ ). Obgleich jene Befunde auf unterschiedlichen Vorwissen-  
indikatoren beruhen, weisen sie doch auf eine gewisse Stabilität der Notengebung während der Grundschulzeit hin. Zusätzlich zu direkten Effekten der Vorkenntnisse auf zeitlich nachfolgende Leistungserhebungen ist bisher selten untersucht worden, welche Bedeutung die von der Lehrkraft zeitlich vorausgegangene Benotung – im Sinne eines schulisch erfassbaren Vorwissens – für die Notenzusammensetzung am Ende der Grundschulzeit spielt und welche direkten und vermittelnden Effekte über die kindliche Fähigkeitseinschätzung zu erwarten sind. Erste Ergebnisse der Studie von Gräsel, Krolak-Schwerdt, Nölle und Höstermann (2010) zur diagnostischen Urteilsbildung von Lehrkräften am Übergang in die Sekundarstufe I fanden bezüglich der Lehrkräfteeinschätzung zur Bedeutsamkeit von Vorkenntnissen für die Notenbildung heraus, dass die Entwicklung der Leistung über Zeit ein wichtiges Kriterium für die Übertrittsempfehlung darstellt. Obwohl daraus keine Erkenntnisse für die tatsächliche Notenbildung gewonnen werden können, deuten diese ersten Ergebnisse einmal mehr auf eine äußerst differenzierte Urteilsbildung von Lehrkräften am Ende der Grundschulzeit hin. Bezüglich der Bedeutung von Vorwissen-  
komponenten für die Zensurenvergabe der Lehrkräfte in Deutsch und Mathematik fand Stocké (2010) neben Hinweisen zu schichtspezifischen Benotungstendenzen Zusammenhänge zwischen dem Ausgangsniveau und der zeitlichen Entwicklung schulischer Kompetenzen. Lehrkräfte scheinen gemäß dieser Ergebnisse eine positive Leistungsentwicklung sowie eine hohe Anstrengungsbereitschaft mit besseren Noten zu belohnen (Stocké, 2010).

Zusammenfassend weisen die exemplarisch aufgeführten Befunde auf die Relevanz des kindlichen Vorwissens für die Notenzusammensetzung hin. Welche Bedeutung dabei das von der Lehrkraft eingeschätzte

Vorwissen für die Notenzusammensetzung in der vierten Jahrgangsstufe spielt, wurde bislang selten thematisiert.

### Sprachliche Fähigkeiten

Sprachliche Fähigkeiten werden nicht nur als ein wichtiger Teil der kognitiven Entwicklung im Grundschulalter bezeichnet, sie stellen auch in mündlicher und schriftlicher Form einen relevanten – dem kognitiven Bereich zugeordneten – Faktor für den schulischen Differenzierungsprozess dar und sind damit eine bedeutsame Determinante für die Schulleistung eines Kindes (Krapp, 1976; Schrader, 2009; Schrader, Helmke & Hosenfeld, 2008; Stahl, 2007; Stern, 2002). Ohne auf die Besonderheiten der sprachlichen Entwicklung und deren Erklärungsansätze einzugehen<sup>24</sup>, wird unter sprachlicher Kompetenz jene Fähigkeit verstanden, die sprachliches Handeln in Alltagssituationen ermöglicht (Chudaske, 2012). Differenziert werden muss hier allerdings immer in jene Komponenten, welche die in der Familie gesprochene Sprache betreffen von solchen, die das Kind selbst im schulischen Kontext zeigt. Auch wenn die Familiensprache vermutlich immer mit den sprachlichen Fähigkeiten des Kindes konfundiert ist, werden diese Komponenten und deren Einfluss auf Schulleistungen in empirischen Studien meist getrennt geprüft (z.B. Aulinger, 2009; Bos et al., 2003; Chudaske, 2012; Lehmann, Peek & Gänsfuß, 1997; Schneider, 2011; Tiedemann & Billmann-Mahecha, 2007b; Weinert & Helmke, 1997). Vornehmlich diskutiert wird der Zusammenhang zwischen Familiensprache und Schulleistung häufig vor dem Hintergrund migrations- und oder schichtspezifischer Sprachmuster, die sich auf die schulischen Leistungen des Kindes auswirken. Eine historisch bekannte und vielfach diskutierte Theorie zum schichtspezifischen Sprachgebrauch hebt den Zusammenhang zwischen Sozialschicht und sprachlichen Fähigkeiten hervor, wodurch sich – unabhängig vom Grad der Intelligenz – unterschiedlich versierte Sprachformen entwickeln. Folgen für die Schule würden sich dann dadurch ergeben, dass Kinder aus höheren sozialen Schichten durch elaborierte Sprach-

---

<sup>24</sup> Für einen ausführlichen Überblick zur Entwicklung sprachlicher Kompetenzen sei an dieser Stelle auf Chudaske (2012) verwiesen.

kenntnisse erfolgreicher Lernprozesse durchlaufen können als Kinder aus niedrigeren Sozialschichten, die über restringierte Sprachkenntnisse verfügen (Bernstein, 1972). Empirische Nachweise für diese These sind allerdings widersprüchlich (Oevermann, 1970; Roth & Sauer, 1981). Empirische Untersuchungen zum Zusammenhang zwischen Migrationshintergrund, sprachlichen Fähigkeiten und Schulleistung thematisieren wiederum die Wechselwirkung zwischen sozialer Herkunft und Migrationshintergrund, wobei häufig die soziale Herkunft als ausschlaggebendes Kriterium diskutiert wird (McElvany & Razakowski, 2013). Auf Seiten der Schulrechtsverordnungen heben Bestimmungen von Kultusministerkonferenz und Grundschulrecht ebenfalls die Bedeutung sprachlicher Fähigkeiten für den Schulerfolg hervor. So sieht beispielsweise das bayerische Schulrecht vor, bei der Bewertung schriftlicher Arbeiten die äußere Form, die Sprachrichtigkeit und schwere Ausdrucksmängel zu berücksichtigen (Bayerisches Staatsministerium für Unterricht und Kultus, 2008).

Betrachtet man den empirischen Forschungsrahmen zur Überprüfung sprachlicher Fähigkeit, differiert diese je nach theoretischer Konzeption und Fragestellung der Studie. Weitestgehende Einigkeit besteht in den nationalen wie internationalen Erklärungssträngen dahingehend, dass sprachliche Fähigkeiten – als Unterrichtssprache und bewertete Kompetenz – sowohl direkt als auch indirekt die schulische Leistung beeinflussen und damit einen zentralen Faktor für Schulleistungserfolge und integrative Prozesse darstellen (Auling, 2009; Bernstein, 1972; Bos et al., 2003; Chudaske, 2012; Esser, 2006; Oevermann, 1970; Stahl, 2007; Tiedemann & Billmann-Mahecha, 2007b). Während empirische Ergebnisse laut Krapp (1976) bereits damals je nach Zeitpunkt der Berücksichtigung der sprachlichen Fähigkeiten variierten (sprachliche Entwicklung im Längsschnitt vs. sprachliche Fähigkeiten zu einem Zeitpunkt), sind Effekte der von der Lehrkraft eingeschätzten sprachlichen Kompetenzen der Schüler auf die Schulnoten – unabhängig von Migrationsaspekten – bisher deutlich seltener untersucht worden. In Analysen von Ditton und Krüskens (2006a) können 16% der Varianz in der Durchschnittsnote der vierten Grundschulklasse durch die von der Lehrkraft eingeschätzten sprachlichen

Fähigkeiten des Schülers erklärt werden, wobei diese Einschätzung wiederum relativ hoch mit der Testleistung zusammenhängt. Bereits veröffentlichte Befunde der Koala-S-Studie weisen auf einen engen Zusammenhang zwischen der Beurteilung von Sprach- und Lesefähigkeiten mit der Übertrittsempfehlung und den Schulnoten hin, wobei die Übereinstimmung der Lehrereinschätzung mit zusätzlich durchgeführten Lese- und Rechtschreibetests ebenfalls im mittleren Bereich liegt (z.B. Zusammenhang zwischen Lehrerbeurteilung und Leseleistung 4. Klasse  $r=.49$ ,  $p<.01$ ) (Stahl, 2007).

Aus prognostischer Perspektive analysierten Gut, Reimann und Grob (2012) mithilfe einer längsschnittlich angelegten Stichprobe im Grundschulbereich ( $N=263$ ), inwiefern die aus Schülerperspektive erhobene Intelligenz, sprachliche, mathematische und sozial-emotionale Kompetenzen (erhoben mittels IDS; Intelligence and Development Scales) deren Schulnoten in Deutsch, Mathematik und Sachkunde drei Jahre später vorhersagen können. Die Ergebnisse hierarchischer Regressionen zeigen, dass nach Kontrolle von Alter, Schulform, Geschlecht, Migrationshintergrund und sozialem Hintergrund, neben Intelligenz und logisch-mathematischen Faktoren, auch die sprachlichen Kompetenzen die Schulnoten in Mathematik und Deutsch signifikant vorhersagen können ( $\beta=.21$ ;  $\beta=.23$ ). In zusätzlich durchgeführten Kommunalitätsanalysen (der IDS klärt für die Mathematiknote 21%, für die Deutschnote 18% Varianz auf) wurde für die sprachlichen Kompetenzen in den Mathematik- und Deutschnoten rund 17 % spezifische Varianz aufgeklärt (Gut et al., 2012).

Die exemplarisch aufgeführten Befunde zeigen einmal mehr die Bedeutsamkeit der sprachlichen Kompetenzen für die schulische Leistung eines Kindes auf. Während unabhängig von migrationspezifischen Fragestellungen vornehmlich die Relevanz von Lese- und Rechtschreibleistungen untersucht wurden, existieren bislang nur wenige Erkenntnisse zum Zusammenhang zwischen Lehrkräfteeinschätzungen zum Sprachstand der Schüler und deren Noten.

## Begabung

Die Diskussion um den Begabungsbegriff umfasst vielschichtige Aspekte, die in wissenschaftlichen und öffentlichen Diskursen vorgebracht werden. Vornehmlich im alltäglichen Sprachgebrauch scheint der Begabungsbegriff hinlänglich mit Bedeutung versehen. Nicht zuletzt wissenschaftliche Abhandlungen zeigen jedoch auf, welche teils widersprüchlichen und differenten Aspekte mit dem Begabungsbegriff verknüpft sind; auch werden die Determinanten von Begabung kontrovers diskutiert.

Die historische Betrachtung des Begabungsbegriffs reicht bis in das letzte Drittel des 19. Jahrhunderts zurück, in welchem der Begabungsbegriff vornehmlich in medizinischen, psychologischen und philosophischen Zusammenhängen diskutiert wurde (Baerwald, 1896; Hoyer, 2012; Stern, 1967). Neben Diskussionen um die Frage nach anlage- und umweltbedingten Ursachen von Begabung waren es im 20. Jahrhundert die Themen der Hochbegabung sowie der Diagnostik und Förderung individueller Begabungen, die vornehmlich in Psychologie und Pädagogik vor dem Hintergrund auftretender Bildungsungleichheiten bis heute thematisiert werden (Heller, 1973; Roth, 1970; Stamm, 2014; Tenorth, 2007). Der anfänglich statische Blick auf verschiedene Begabungsbereiche fokussierte im pädagogischen Bereich dann zunehmend den Zusammenhang zwischen Begabung und Leistung, wobei der gesellschaftliche und ökonomische Nutzen stärker betrachtet wurde (Behrensen & Solzbacher, 2016). Aktuelle Bezüge des 21. Jahrhunderts verortet Stamm (2014) vorwiegend in entwicklungspsychologischen und bildungssoziologischen Fragenstellungen, die sich unter anderem mit fröhpädagogischen Fördermaßnahmen, kognitiven Entwicklungsprozessen im Lebenslauf sowie gesellschaftlichen Einflüssen und Reproduktionsmechanismen von Begabung beschäftigen.

Der Begabungsbegriff umfasst eine Fülle an Bedeutungen, die teilweise synonym, ergänzend oder konträr eingesetzt werden<sup>25</sup>. Neben der mangelnden Trennschärfe ist unklar, in welchem Zusammenhang beispielsweise die Begriffe Fähigkeit, Talent und Eignung zum Begabungsbegriff stehen und welchem Begriff welche zeitliche Stabilität und Bedeutung für schulische Leistungen zugesprochen wird (Frey, 1973; Hany, 2012; Heller, 1976; Mühle, 1970; Stamm, 2014; Süllwold, 1976). Während Ballauff und Hettwer (1967) in ihrem Sammelband zu Begabungsförderung und Schule einen differenzierten, historischen Blick auf die Begabungsdiskussion werfen, verweisen auch neuere Abhandlungen auf diese Bedeutungsvielfalt mit dem Schluss, Begabung als ein theoretisches, nicht direkt beobachtbares Konstrukt zu begreifen, um – vornehmlich im pädagogisch-psychologischen Sinne – unterschiedliche Leistungsausprägungen zu erklären (Hany, 2012; Hoyer, 2012). Betrachtet man den Begriff unabhängig von einem Fachbereich, umfasst er immer eine bestimmte Begabungsart (allgemein vs. bereichsspezifisch), die dann wiederum in ihrer jeweiligen Ausprägungsstärke variiert (Bundesministerium für Bildung und Forschung, 2015; Süllwold, 1976). Ebenfalls von allgemeiner Natur, aber unter explizitem Einbezug von Umweltbedingungen heben Trautwein, Maaz und Baeriswyl (2014) in ihrem breit angelegten Begabungsbegriff das individuelle Leistungspotential hervor, welches sich je nach sozialen und kulturellen Lebensbedingungen individuell entwickelt.

Für die begriffliche Einordnung kann festgehalten werden, dass sich das Begabungskonzept zunächst als ein theoretisch zu rekonstruierendes Konstrukt auszeichnet, das sich dann – theoretisch aufgezeigt und oder empirisch gemessen – immer auf ein individuelles Merkmal bezieht, welches das Individuum zu einem bestimmen Verhalten befähigt und im Vergleich zur Bezugsgruppe qualitativ oder quantitativ überdurchschnittlich stark ausgeprägt ist (Behrens, 2000; Giesinger, 2014). Die Frage der Erhebungsperspektive, deren Bedeutung für den Ausprägungsgrad festgestellter Begabungen und letztlich deren

---

25 Engelmayer (1967) ging bereits davon aus, dass man den Begabungsbegriff erst durch eine Abgrenzung verwandter Begriffe definieren könne.

Einflussstärke auf gemessene Schulleistungen wird in diesem Zusammenhang allerdings selten angesprochen.

Hinter jedem begrifflichen Einordnungsversuch scheint – ähnlich wie beim Leistungsbegriff – letztlich die Frage nach der Funktion der Begriffsdifferenzierung zu stehen, die je nach Fachdisziplin unterschiedliche Aspekte fokussiert. Wollersheim (2014) sieht im aktuellen sowie historischen Kontext eine Verbindungslinie zum jeweilig vorherrschenden Bildungsbegriff. Demnach seien durch breiter werdende Zugänge zu Bildung Abgrenzungsversuche bestimmter sozialer Gruppen zu berücksichtigen, die sich mit Fragen der Inhalte von Begabung und mit der Suche nach einer wissenschaftlich angemessenen Begabungsdiagnostik beschäftigen (Wollersheim, 2014). Der damaligen Tendenz, Bildungsunterschiede als Begabungsunterschiede zu begreifen (Solga, 2009), steht die historische Forderung Roths (1967) nach einem pädagogischen Begabungsbegriff gegenüber, der die Förderbarkeit, Prozesshaftigkeit und Umweltbedeutung hervorhebt (Roth, 1967; Wollersheim, 2014). Der Fokus der philosophisch orientierten Perspektive auf die Begabungsforschung diskutiert ungleich verteilte Begabungen und deren Folgerungen für – (un)legitimierte – Ungleichheitsprozesse. So verweisen kritische Stimmen auf die dabei mitschwingende Legitimität sozialer Ungleichheitsprozesse und diskutieren – bei bestmöglicher Gleichheit aller Individuen – die selektiven Förderstrukturen bestimmter Gesellschaftsgruppen bei gleichzeitiger Legitimation unterschiedlicher – als natürlich betrachteter – Begabungsunterschiede (Giesinger, 2014).

Ohne auf die vielfältigen Facetten theoretischer Konzepte zur Begabungsforschung einzugehen, wird deutlich, dass der Drang nach der Greifbarkeit des Begriffs eine Fülle an Theoriebezügen und Operationalisierungsvarianten mit sich bringt. Allen (fachübergreifenden) Abhandlungen gemein ist der Leistungsbezug, der meist als Bezugspunkt von Begabung gilt. Der mehr oder weniger eindeutige Begabungsbegriff gilt nicht nur häufig als Erklärung unterschiedlicher Leistungspotentiale, Lernzuwächse und –ergebnisse, sondern wird auch als Erklärung für Leistungsverhalten herangezogen, das in Referenzgrup-

pen gar nicht erst gezeigt wird<sup>26</sup> (Mühle, 1970; Roth, 1970). Schon früh wurde allerdings auf die Problematik hingewiesen, „Leistungsmängel mit Begabungsmängeln gleichzusetzen“ (Mühle, 1970, S.69), die zu mangelnden Förderbemühung von Lehrkräften führen könnten (Mühle, 1970; Roth, 1967). So kann anhand der im Schulkontext eingesetzten Schulleistungstests noch nichts über die Ursachen der Leistungsunterschiede ausgesagt werden. Begründet man diese durch unterschiedliche Begabungen, muss in dessen Folge auch über die Ursachen von Begabung diskutiert werden (Süllwold, 1976; Susteck, 1995). Diese Ursachensuche individueller Begabungen wird schon lange Zeit vor dem Hintergrund der allseits bekannten Anlage-Umwelt Debatte diskutiert, die sich unter anderem mit der Frage nach Förderung, Entwicklung und Entstehung von Begabungen beschäftigt<sup>27</sup> (Süllwold, 1976).

Ausgangspunkt dieser Debatte sind historische Forschungskonzepte, die von einer Anlagebedingtheit ausgehen und Begabung als ein angeborenes, individuell differierendes Merkmal ansehen (Behrens, 2000; Giesinger, 2014). Entgegengesetzt dazu gehen Vertreter der Umwelt davon aus, dass die Umweltbedingungen das prägende Kriterium für individuelle Begabungsunterschiede sind. Anlagebedingte Prozesse wurden dabei in ihrer Wirksamkeit vollkommen zurückgewiesen<sup>28</sup> (Behrens, 2000). Die wohl am häufigsten zitierte Abhandlung von Stern (1967) verwies allerdings schon auf den sog. Konvergenzstandpunkt, nachdem das Zusammenwirken beider Bedingungsgruppen entscheidend sei. Dennoch galt lange Zeit die Meinung, dass Umwelt-

---

26 Das IQB (2005) differenziert beispielsweise in der Hochbegabungsdiskussion nochmals zwischen Kompetenzen und Performanz. Kompetenzen werden als als Leistungsvermögen, als eine vorhandene Grundbegabung bezeichnet, während sich die Performanz – aufgrund einer Begabung – wiederum in konkret messbarer Leistung zeige.

27 Der Begriff *Umwelt* umfasst in der Begabungsforschung meist einen etwas engeren Umweltbegriff. Im Zusammenhang mit der individuellen Schulleistung werden als Indikatoren meist Merkmale des sozialen Hintergrunds (Schulbildung, Beruf, Einkommen, Familienstruktur u.ä.) herangezogen. Dagegen vereint der Begriff *Anlage* komplexe, biologisch begründete Faktoren (Heller, 1976).

28 Für weiterführende Informationen zu theoretischen Erklärungsmodellen zum Zusammenhang zwischen Anlage-Umwelt Faktoren und menschlichen Entwicklungsprozessen sei auf Heller (1976) ab Seite 43 verwiesen.

faktoren nur für nachträgliche Verhaltensänderungen verantwortlich seien. Vornehmlich wissenschaftliche Abhandlungen berücksichtigten aber schon damals beide Faktoren gleichermaßen, obgleich die relative Bedeutung beider bis heute diskutiert wird (Bundesministerium für Bildung und Forschung, 2015; Giesinger, 2014; Heller, 1976). Dass die Diskussion um die Federführung von Umwelt- und Anlageeinflüssen aber letztlich niemals geklärt werden könne, fasst Susteck (1995) in seiner Abhandlung zum Zusammenhang zwischen Begabung und Schulleistung trefflich zusammen. Da milieuspezifische Faktoren schon von der frühesten Kindheit an vorhanden sind, sei es praktisch unsinnig, einen nach Ursachen differenzierenden Begabungsbegriff zu suchen (Susteck, 1995). Solga (2009) verweist in diesem Zusammenhang auf die Bedeutung sozial vereinbarter Kriterien zur Definition von Begabungen, die sich dann letztlich auch in der Leistungsdefinition wiederfinden. Gehe man davon aus, dass Bildung das Ergebnis von Begabung und Anstrengung ist, wäre das Ergebnis von Bildungszertifikaten letztlich die gezeigte Anstrengung des Schülers. Damit „wäre allerdings nicht Begabung, sondern Fleiß das eigentliche Platzierungskriterium“ (Solga, 2009, S.67), wodurch angeborene Begabungen im Bildungssystem wiederum nur zu Erfolg führen, wenn sie sich durch entsprechende Anstrengungen auszeichnen (Solga, 2009).

Roth (1967) hebt bereits damals die Frage nach Anlage oder Umwelt als problematische Fragestellung hervor, deren Antworten sowohl die theoretische Diskussion als auch den pädagogischen Alltag massiv beeinflussen. Würde man davon ausgehen, dass Begabungsunterschiede im hohen Maße von sozialen Faktoren abhängig sind und maßgeblich die Leistungen von Schülern beeinflussen, wären die als objektiv bezeichneten Leistungskriterien gleichzeitig auch soziale Reproduktionsmechanismen, dergestalt dass Kinder aus höheren Schichten mehr von jenen Eigenschaften aufweisen, deren Ursache spezifischen Begabungen zugeschrieben wird und in dessen Folge sie höhere Leistungen erzielen<sup>29</sup> (Hoyer, 2012; Stamm, 2007). Geht man davon aus,

---

29 Die Bedeutung sozialer Merkmale könnte hierbei auf alle individuellen Schülermerkmale übertragen werden, bei denen von einer zeitlichen Stabilität ausgegangen werden kann. Ihre Bedeutung für Begabungsunterschiede wird in diesem Sinne lediglich stellvertretend vertieft.

dass schichtspezifische Umweltfaktoren die (kulturellen) Lernbedingungen eines Kindes schon seit der frühesten Kindheit an prägen und sich in unterschiedlichem schulischen Lernverhalten zeigen, kann die Bedeutung der Umweltfaktoren für Begabungen kaum mehr geleugnet werden (Roth, 1970). Die Frage nach der Relevanz von Umwelt- und Anlagefaktoren für individuelle Begabungen umschließt also letztlich auch die Frage nach sonstigen Einflussfaktoren, die zusätzlich oder vermittelt über den Faktor Begabung für Schulleistungen bedeutsam werden könnten (Roth, 1970). Einflüsse, deren empirische Überprüfung bis heute nur vereinzelt existieren.

Bei der Betrachtung der empirischen Zugänge in der Begabungsforschung erscheint es vor dem Hintergrund der vielfältigen Diskussionspunkte nicht verwunderlich, dass auch in der Empirie vielfältige Herangehensweisen existieren. Auch die Forschung zu Begabungsindikatoren muss sich den Fragen des Gütemaßstabs der Messung, der einbezogenen Kennzeichen, des Vergleichsmaßstabs, des Erhebungszeitpunktes und der Signifikanzschwelle von (Hoch-)Begabungen stellen (Ziegler, 2009). Als Erhebungszeitpunkt von individuellen Begabungen im Schulkontext werden häufig Schnittstellen im Bildungssystem fokussiert; neben der Begabungsfeststellung am Übergang vom Elementar- in den Primarbereich erfolgt auch der anschließende Übergang in die Sekundarstufe vor dem Hintergrund individueller Begabungen (Susteck, 1995; Trautwein et al., 2014). Unklar ist dabei häufig, inwiefern eine solche Differenzierung auf Grundlage eines wenig greifbaren Begabungsbegriffs getroffen werden sollte (Trautwein et al., 2014), zumal die Debatte um Anlage und Umwelt unterschiedliche Schlussfolgerungen über die Förderbarkeit von Begabung konstatiert<sup>30</sup>.

---

30 In der Argumentation von Giesinger (2014) kann jene begabungsorientierte, dreigliedrige Struktur des Bildungswesens zur Reproduktion von sozialen Ungleichheiten beitragen. Würde man gemäß der Anlagebedingtheit davon ausgehen, dass Begabungen naturbedingt, zeitlich stabil und nahezu unveränderbar sind, wäre eine, an individuelle Begabung angepasste Bildungsgangdifferenzierung sinnvoll. Eine häufig nachgewiesene Kopplung von sozialen Hintergrundvariablen und höheren Bildungsgängen lässt den Autor allerdings nach der Kausalität jener Zuweisung fragen (Giesinger 2014).

Die methodische Umsetzung in empirischen Untersuchungen zum schulischen Begabungsbegriff unterscheidet sich häufig in ihrer Begriffsdefinition, was unvermeidlich zu abweichenden Indikatoren und Erhebungsperspektiven führt. Die Überprüfung bezieht sich vornehmlich auf korrelative- und faktoranalytische Auswertungsverfahren, wobei begabungsbezogene Variablen als Prädiktoren zu faktoranalytischen Konstrukten zusammengefasst werden, die ihrerseits als Moderatorvariablen fungieren und oder einen direkten Leistungsbezug vorweisen (Arnold, Kleß & Prescher, 2014).

Bereits die Ergebnisse der Untersuchung von Steinkamp (1976) in Hamburger Volksschulen von der 2. bis zur 6. Klasse zum Zusammenhang zwischen von der Lehrkraft als relevant erachteten Schülermerkmalen und der Übertrittsempfehlung für die Sekundarstufe verweisen auf die starke Bedeutung individueller Begabungen. Während sich 21% der Lehrkräfte in ihrer Empfehlung ausschließlich an der Schulleistung orientieren, gaben 57% der Befragten an, einem begabten Kind trotz negativem Arbeits- und Sozialverhalten eine höhere Schulempfehlung zu geben. Ebenfalls verweisen die Befunde auf einen Zusammenhang zwischen Begabungseinschätzung und sozialer Herkunft des Schülers (Steinkamp, 1976). Dagegen erfasst die ebenfalls frühe Studie von Höhn (1980) in der ersten Teilerhebung mittels mündlicher Befragung die individuelle Lehrkraftmeinung über die drei schlechtesten Schüler ihrer Klasse. Obwohl eine fehlende Explikation des Adjektivs *schlecht* nur indirekt erfolgt, wird diese Studie bis heute vielfach zitiert, wenn von kindlichen Eigenschaftsausprägungen gesprochen wird, die schlechte Schulleistungen auszeichnen. In der Beschreibung von 90 Schülern wurden 44 % der Fälle mit Begabungsmängeln ausgezeichnet, was wiederum nicht mit Faulheit oder mangelndem Fleiß in Zusammenhang gebracht wurde (Höhn, 1980). Obwohl in der Studie nicht ersichtlich wird, ob die Lehrkraftmeinungen zusätzlich durch Leistungstests validiert wurden und was genau Autorin und Lehrkräfte unter Begabungsmängeln verstehen, bleibt doch bemerkenswert, welche versierte Eigenschaften Lehrkräfte schlechten Schülern zusprechen.

Inwiefern Lehrkräfte die Begabung eines Schülers treffend einschätzen können, beschäftigt die Arbeitsgruppe von Heller, Reimann und Senfter (2005), welche Begabung mit Intelligenz gleichsetzt. Angelehnt an das Münchner Hochbegabungsmodell wurden die Lehrkräfte von fünf Testschulen gebeten, jeweils die intelligentesten 10 Prozent sowie die sozial kompetentesten und kreativsten Schüler der dritten und vierten Klassenstufen zu benennen. Die Überprüfungen des korrelativen Zusammenhangs zwischen Lehrkräfteeinschätzung und parallel durchgeführten Testerhebungen ergaben die höchste Übereinstimmung für die Intelligenzeinschätzung in der dritten Klasse ( $r=.43$ ). Die Einschätzung zur Kreativität und Sozialkompetenz fiel dagegen recht niedrig aus ( $r=.11$  bzw.  $r=.05$ ) (Heller et al., 2005).

Bei ihrer Untersuchung zur Differenziertheit von Lehrkräfteeinschätzungen am Übergang in die Sekundarstufe I kommen Anders, McElvany und Baumert (2010) zu dem Ergebnis, dass die von der Lehrkraft vorgenommene kindliche Begabungs- und Fähigkeitseinschätzung die größte Vorhersagekraft für die Halbjahresnoten der vierten Klasse in Deutsch und Mathematik hat. Im Vergleich der beiden Schulfächer konnten für die Halbjahresnoten in Mathematik noch etwas stärkere Effekte erzielt werden (Mathematiknote:  $\beta=-.75$ ; Deutschnote:  $\beta=-.66$ ). Die Ergebnisse basieren auf einem Modell aus ursprünglich zehn, als lernrelevant betrachteten Faktoren, die aufgrund ihrer hohen Interkorrelationen mittels explorativen Faktorenanalysen zu drei Faktoren zusammengefasst wurden (Begabung und Leistung; soziale Fähigkeiten und Sozialverhalten; Motivation und Lerntugenden). Der Faktor Begabung und Leistung wurde durch neun Merkmale gemessen, die sich sowohl auf allgemeine schulische Fähigkeiten als auch begabungsspezifische Unterbereiche beziehen (akademische Begabung, kognitive Fähigkeiten, Gedächtnis, Mitarbeit im Unterricht, Belastbarkeit, Selbstwertvertrauen in Leistungssituationen, mathematische Fähigkeiten, Lesekompetenz, selbstständiges Arbeiten/Lernen) (Anders et al., 2010). Als ebenfalls relevant für die Schulnoten Ende der vierten Grundschulstufe bezeichnete Stahl (2009) die Begabungseinschätzung der Lehrkraft (Skala aus mathematischer und sprachlicher Begabung, Intelligenz und Kreativität), die zusätzlich zur

individuellen Schülerleistung einen signifikanten Zusammenhang zur Deutschnote aufweist. Dass das Begabungs-konstrukt je nach Studie in seiner Differenziertheit und Perspektive variiert, zeigt weiter die Studie von Lintorf (2012) zur Prädiktionskraft individueller und kontextspezifischer Merkmale für Grundschulnoten, die Begabung mithilfe des kognitiven Grundfähigkeitstests erhebt. Bei den aus Schülerperspektive erhobenen Individualmerkmalen zeichnet sich nach der individuellen Testleistung des Schülers dessen Begabung als zweitstärkster Prädiktor für die Mathematiknote ab ( $\beta = .167$ ) (Lintorf, 2012).

Betrachtet man abschließend die exemplarisch aufgeführten Forschungsergebnisse wird deutlich, dass die Begabungseinschätzung durch die Lehrkraft durchaus eine wichtige Komponente für Bewertungen einnimmt. Nahezu nebulös bleibt allerdings bei fast allen Studien, welche Merkmale ein besonders begabtes Kind nun abschließend aufweisen muss, um eine gute Schulnote zu erhalten. Vermutet werden kann, dass insbesondere von der Lehrkräfteeinschätzung zur Begabung eine hohe Prädiktionskraft ausgeht, was konkret diese allerdings unter jener Begrifflichkeit versteht und mit welchen kindlichen Merkmalen dies zusammenhängt, bleibt weiterhin offen.

Abschließend kann festgehalten werden, dass sich die kognitiven Indikatoren aufgrund der geteilten Relevanz für die abhängige Variable Schulleistung bzw. Schulnote sicherlich auch gegenseitig beeinflussen und damit relativ hohe Korrelationen zwischen den einzelnen Indikatoren zu erwarten sind. Demnach ist davon auszugehen oder gar wünschenswert, dass eine hohe Schülertestleistung in Lesen und Mathematik gleichzeitig von der Lehrkraft positive Einschätzungen bezüglich Begabung und Sprache erhält. In Anlehnung an das multikausale Bedingungsmodell schulischer Leistung nach Heller (1995) ist allerdings davon auszugehen, dass kognitive Fähigkeiten allein eine *notwendige*, aber keine *hinreichende* Bedingung für schulische Leistungen darstellen, sondern auch nicht-kognitive bzw. motivationale und volitionale Faktoren relevant sind (Helmke & Weinert, 1997). Welche Rolle das häufig den nicht-kognitiven Einflussfaktoren zugeordnete Arbeits- und Sozialverhalten für die Zusammensetzung einer Note spielen, wird nachfolgend betrachtet.

#### 4.2.2 Nicht-kognitive Prädiktoren von Schulleistung

Wie zu Beginn des vierten Kapitels bereits angemerkt, können Determinanten der Leistungsbeurteilung neben jenen, dem kognitiven Bereich zugeschriebene Schülermerkmale, auch nicht-kognitive, affektive, soziale oder psychomotorisch-praktische Prädiktoren umfassen (Zumhasch, 2001). In Anlehnung an die in [Kapitel 2.5](#) eingeführten schulrechtlichen Grundlagen werden diese in der Grundschule meist unter den Begriffen Lern-, Arbeits- und Sozialverhalten zusammengefasst und zusätzlich zu fachspezifischen Leistungen bewertet. Da sich die nachfolgende empirische Untersuchung mit der Bedeutung jener Merkmale für die Notenzusammensetzung aus Lehrkraftsicht beschäftigt, wird abweichend von der häufig vorgebrachten Differenzierung in motivationale und sozial-emotionale Determinanten, vornehmlich die Relevanz jener nicht-kognitiver Merkmale betrachtet, welche unter schulrechtlich relevante Eigenschaften zusammengefasst werden<sup>31</sup>. So wird in der nachfolgenden Argumentation davon ausgegangen, dass sich die motivationalen Faktoren eines Schülers – aus Lehrkraftsicht – in dessen Arbeits- bzw. Lernverhalten zeigen, die sich wiederum in den Schulnoten des Kindes widerspiegeln können (Heller, 1984; Trautwein et al., 2014).

#### **Arbeits- und Sozialverhalten**

In der Zeit, in der Begabung und Intelligenz als genetisch festgelegte, stabile Merkmale erachtet wurden, lag der Fokus wissenschaftlicher Abhandlungen weniger auf der Bedeutung nicht-kognitiver Merkmale für schulische Leistungsprozesse (Krapp, 1976). Studien neueren Datums betrachten jedoch verstärkt die Bedeutung des Zusammenhangs zwischen Schulleistung und den als konativ und volitional bezeichneten Determinanten (Helmke & Weinert, 1997). Die Ziele der Beurteilung nicht-kognitiver Merkmale bezeichnete Ingenkamp (1976a) allerdings nur als vage und widersprüchlich, sodass eine exakte Fassung und Prüfung nur schwer möglich sei. Bilanzierend für die Forschungsbefunde der siebziger Jahre beschrieb Krapp (1976)

---

31 Welche Bedeutung die motivationalen Schülermerkmale für den Schulerfolg spielen, wurde u.a. von Kaufmann (2008) mit Daten der Koala-S-Studie untersucht.

die Untersuchungen zu jenem Merkmalsbereich als äußerst vielfältig, wobei sicherlich fast alle kindlichen Merkmale schon einmal auf Zusammenhänge zu Schulleistungskomponenten untersucht worden seien. Neben zahlreichen anderen Faktoren verwies er dabei auf die Bedeutung des Arbeits- und Sozialverhaltens, welches in Anlehnung an schulrechtliche Grundlagen bis heute relevant ist (Krapp, 1976).

Während bundeslandübergreifend durch das Schulrechthandbuch von Avenarius und Füssel (2010) die Bewertung von Mitarbeit und Sozialverhalten im Zeugnis legitimiert wird, halten auch die Grundschulordnungen von Bayern und Sachsen in individueller Ausprägung die Bewertung von Sozial-, Lern- und Arbeitsverhalten fest (vgl. Kap. 2.5). Unabhängig von der Übertrittsempfehlung in die Sekundarstufe I soll in der bayerischen Grundschule der verantwortliche Klassenlehrer nach Rücksprache mit den Lehrerkollegen, neben den Zeugnisnoten der Pflichtfächer, das Arbeits-, Lern- und Sozialverhalten des Schülers in einer vierstufigen Notenskala festhalten und erläutern (Bayerisches Staatsministerium für Unterricht und Kultus, 2008). Eine Abhandlung des Staatsinstituts für Schulqualität und Bildungsforschung München (2015) differenziert das recht weit gefasste Lern- und Arbeitsverhalten in die ebenfalls weitgefassten Oberbegriffe Interesse und Motivation, Lern- und Arbeitsweise sowie Konzentration und Ausdauer des Schülers. Das Sozialverhalten wird dagegen in soziale Verantwortung, Kooperation, Kommunikation sowie Konfliktverhalten differenziert (Staatsinstitut für Schulqualität und Bildungsforschung München, 2015). In Sachsen bezieht sich die Schulordnung dagegen ganz explizit auf die Bewertung von Mitarbeit, Fleiß, Ordnung und Betragen, welche wiederum in einer fünfstufigen Skala bewertet werden. Ohne nochmals ausführlich auf diese Charakterisierung einzugehen, umfasst in Sachsen beispielsweise die Ordnung eines Schülers dessen Sorgfalt, Zuverlässigkeit, Pünktlichkeit und Bereithalten von Lernmaterialien. Fleiß wird differenziert in Lernbereitschaft, Ausdauer, Regelmäßigkeit, Zielstrebigkeit sowie die Erfüllung von Aufgaben (Sächsische Staatskanzlei, 2013).

Welche Bedeutung diese im Schulrecht aufgeführten kindlichen Merkmale nun tatsächlich für deren Schulnoten vorweisen, wurde bereits für einige Merkmale empirisch untersucht; auffällig sind dabei vor allem die vielfältigen Varianten der verwendeten, dem nicht-kognitiven Bereich zugeordneten Indikatoren. Exemplarisch aufgeführt wird für die frühen Untersuchungen zum Zusammenhang jenes Merkmalsbereichs und Schulleistung die bereits oben aufgeführte Studie von Steinkamp (1976). In seiner Untersuchung zum Zusammenhang zwischen den von der Lehrkraft als relevant erachteten Schülermerkmalen und der Übertrittsempfehlung für die Sekundarstufe in Hamburger Volksschulen von der 2. bis zur 6. Klasse wies er, neben der Bedeutung individueller Begabungen, auf die Bedeutung der Arbeitshaltung und weiterer charakterlicher Eigenschaften für Übertrittsempfehlung hin. 55 % der Lehrkräfte (N=261) gaben in mündlichen Befragungen an, die unter der Arbeitshaltung gefassten Komponenten Fleiß, Ausdauer, Konzentration, Mitarbeit, Leistungswille, Interesse, Gewissenhaftigkeit und Ordnung zusätzlich zur Leistung heranzuziehen. 33 % der Befragten erachteten auch die charakterlichen Eigenschaften Ehrlichkeit, Aufrichtigkeit, Gehorsam, Höflichkeit, Wahrhaftigkeit und Disziplin als relevant für Übertrittsempfehlungen, was heute unter das Sozialverhalten des Schülers subsumiert werden könnte (Steinkamp, 1976).

Bezüglich der Effekte des Arbeitsverhaltens für unterschiedliche Leistungshöhen kamen die frühen Analysen von Kemmler (1967) in einer Lehrerbefragung Ende des dritten Schuljahres zu dem Schluss, dass die Leistungshöhe maßgeblich mit der Arbeitshaltung des Schülers zusammenlag; die gezeigte Arbeitsweise also ebenso wichtig sei wie die Begabungs- und Denkfähigkeiten der Schüler.

Mit Daten der IGLU Studie aus dem Jahr 2001 ergab die Studie von Stubbe und Bos (2008) zur Schullaufbahneempfehlung von Lehrkräften am Ende der Grundschulzeit in einem Pfadmodell, dass die von den Eltern eingeschätzte Anstrengungsbereitschaft des Kindes nach der Testleistung die Durchschnittsnote aus Deutsch und Mathematik am besten erklärt (-.29). Befunde von Stubbe, Bos und Euen (2012) für

Daten der IGLU Studie 2006 bestätigen diesen Befund. Hochweber (2010) weist in seiner Arbeit bezüglich der Anstrengungsbereitschaft darauf hin, dass hinter dieser Kategorisierung ebenso Merkmale wie Fleiß oder Motivation des Schülers stehen können. Die Mitzensierung jener Anstrengung sowie dieser naheliegenden Merkmale (z.B. Mitarbeit) muss dabei nicht bewusst bzw. aktiv erfolgen (Hochweber, 2010).

Mit Daten der TIMSS Studie von 2007 verweist auch Lintorf (2012) auf die unterschiedliche Zusammensetzung fachspezifischer Schulnoten. Während die stärkste Prädiktionskraft für die Mathematik- und Sachkundenote der vierten Grundschulklasse von der Fachleistung ausgeht ( $\beta=.515$  bzw.  $\beta=.374$ ), folgt für die Mathematiknote – neben der Begabung ( $\beta=.167$ ) – die aus Schülerperspektive erfasste Gewissenhaftigkeit ( $\beta=.094$ ) und Anstrengungsbereitschaft ( $\beta=.079$ ). Dagegen ist für die Sachkundenote die Gewissenhaftigkeit der zweitstärkste Prädiktor ( $\beta=.203$ ), wohingegen der Anstrengungsbereitschaft keine signifikante Bedeutung zukommt. Differenzierte Analysen lassen die Autorin dann aber zu dem Schluss kommen, dass sich die sogenannten Sekundärtugenden gegenseitig beeinflussen und damit eher als Ganzes in eine Note einfließen (Lintorf, 2012).

Steinmayr und Meißner (2013) zeigen in ihrer Replikationsstudie der Untersuchung von Helmke (1992), dass je nach Indikator schulischer Leistung kognitive und motivationale Prädiktoren in ihren Effekten variieren. In ihrer Untersuchungsstichprobe der 8. Klasse in Gymnasium und Realschule ( $N=463$ ) fanden sie Unterschiede im Zusammenhang zwischen den Kriterien standardisierter Mathematikleistungstests und Mathematiknoten mit der allgemeinen Intelligenz und dem selbstberichteten Fähigkeitsselbstkonzept (FSK) der Schüler. Während sich die Ergebnisse der Mehr-Gruppen-Modelle zwischen den Schulformen auf latenter Ebene hinsichtlich ihrer Modellparameter nicht unterscheiden, ergab die Prüfung statistisch signifikante Unterschiede zwischen den Varianzanteilen der Prädiktoren, wobei das Fähigkeitsselbstkonzept bei der Mathematiknote stärkere Varianzaufklärungen erzielte. Bilanzierend resümieren die Autorinnen, dass Effekte des FSKs vermutlich durch die Lehrkraftbeurteilung des Lern-

und Anstrengungsverhaltens in die Note einfließen könnten (Steinmayr & Meißner, 2013). Auch sie kommen zu dem Schluss, dass in der Notenbildung neben der Schulleistung vermutlich immer weitere Informationen und Eindrücke der Lehrkraft einfließen, „sodass Noten eine heterogene Variable darstellen und nicht nur die reine Leistung widerspiegeln.“ (Steinmayr & Meißner, 2013, S. 280).

Bezüglich der lehrkraftspezifischen Ansichten zum Inhalt einer Leistungsbewertung geben auch die Befunde von Schumacher (2002) aus einer qualitativen Befragung von  $N=514$  Lehrkräften Aufschluss. Lediglich 15% der Befragten gaben an, bei der Leistungsbewertung „eher“ und „in erster Linie“ die kognitiven Fähigkeiten eines Kindes zu bewerten. Neben der Berücksichtigung der individuellen Jahresleistung bei der Bewertungspraxis (76% stimmen mit „genau zutreffend“ zu), gaben 73% der Lehrkräfte an, gute Umgangsformen und ein positives Sozialverhalten der Schüler positiv zu berücksichtigen (Schumacher, 2002). In ihrer Pilotstudie zur Perspektive der Lehrkräfte auf übergangsrelevante diagnostische Informationen kamen Nölle et al. (2009) zu dem Ergebnis, dass die befragten Grundschullehrkräfte ( $N=52$ ) neben Noten der Hauptfächer (43 Nennungen) und der Leistungsentwicklung über die Zeit (42 Nennungen) auch Ängstlichkeit, Ausdauer und Fleiß berücksichtigen (29 Nennungen). Bezüglich der prädiktiven Kraft von Lehrkräfteeinschätzungen zu Schülermerkmalen ( $N=231$ ) für Schulnoten und Übergangsempfehlung ergaben die Befunde von Anders et al. (2010) ein ähnliches Bild. Für den Zusammenhang zwischen Halbjahresnote in Deutsch und Mathematik fand sich zunächst für den Faktor soziale Fähigkeiten/Sozialverhalten (Konfliktverhalten, Empathie, Selbstbeherrschung, Kontrolle von Impulsen, Teamfähigkeit, Sozialverhalten, Betragen in der Schule, positive Beeinflussung der Gefühle anderer) ein Zusammenhang von  $r=-.43$  bzw.  $r=-.34$ . Unter Kontrolle der zu zwei Faktoren zusammengefassten Begabung/Leistung (akademische Begabung, allg. kognitive Fähigkeiten, Gedächtnis, Mitarbeit im Unterricht, Belastbarkeit, Selbstwertvertrauen in Leistungssituationen, mathemat. Fähigkeiten, Lesekompetenz, selbstständiges Arbeiten/Lernen) sowie Motivation/Tugenden (Aufmerksamkeit im Unterricht, Interesse am Lernen, Lern-

motivation, Anstrengungsbereitschaft, Ausdauer, Gewissenhaftigkeit und Fleiß, Pünktlichkeit und Disziplin, Leistungsmotivation, Ehrgeiz) gingen höhere Beurteilungen der sozialen Fähigkeiten/Sozialverhalten mit schlechteren Noten einher, was die Autoren durch einen Supresoreffekt erklären (Anders et al., 2010).

Wie akkurat übergreifend die Lehrkräfteeinschätzung zu individuellen Schülermerkmalen sein kann, analysierte Spinath (2005) in ihrer Grundschulstichprobe der Klassen eins bis vier. Gegenüber gestellt wurden dazu die Intelligenzkenwerte und selbstberichtete Angaben zu Fähigkeitsselbstwahrnehmung, Lernmotivation und Schulängstlichkeit von N=723 Grundschulern mit den Einschätzungen von N=43 Lehrkräften zu den vier Schülermerkmalen. Mittels Differenzierung der akkuraten Einschätzung in Niveau-, Differenzierungs- und Rangkomponenten weisen die Befunde der Studie auf eine nur durchschnittlich akkurate Lehrkräfteeinschätzung hin, die gemäß der Autorin wiederum nicht auf eine allgemeine Fähigkeit zur akkuraten bzw. diagnostisch treffenden Personenbeurteilung verweise (Spinath, 2005). Mit einer amerikanischen Stichprobe von N=794 Lehrkräfte aus den Städten Ohio und New York untersuchten Bennett et al. (1993) den Zusammenhang zwischen Schulnoten, Testleistungen und lehrkräftspezifischen Beurteilungen unterschiedlicher Leistungsbereiche und gegebenen Noten für Anstrengung (effort) und Verhalten (conduct). Die Untersuchungsstichprobe der Schüler bezog sich auf drei unterschiedliche Jahrgangsstufen von der Vorschule bis zur zweiten Klassenstufe. Als übergreifendes Ergebnis weisen diese Befunde auf die Bedeutsamkeit der lehrkräftspezifischen Verhaltenseinschätzung für die Zeugnisnote hin. Nach Kontrolle der Testleistung und des Geschlechts bleibt die durchschnittliche Verhaltensnote für die Vorhersage der Noten relevant (Bennett et al., 1993).

Fasst man die empirischen Tendenzen der exemplarisch dargestellten Befunde zusammen, verweisen diese auf eine hohe Bedeutsamkeit nicht-kognitiver Schülermerkmale für Schulnoten, aus Schüler- und Lehrkraftperspektive. Bezüglich der Lehrerschaft kann man aufgrund der meist fehlenden Validierung von tatsächlich gezeigtem Verhalten

und daraus geschlossenen Bewertungen nur indirekt und aufgrund der Ergebnisse von Spinath (2005) vorsichtig darauf schließen, dass Lehrkräfte im Mittel jenes Schülerverhalten – ohne Beeinflussung durch Drittvariablen – treffend einschätzen können. Geht man davon aus, dass hohe Werte im Arbeits- und Sozialverhalten folglich immer zur besseren Leistungsbewertung führen, weist schon Schröder (1990) auf den möglichen Fehlschluss von Prämisse (hoher Fleiß führt zu guten Leistungen) zur Konklusion (die Leistung des Schülers ist gut, also weist er hohen Fleiß auf) hin, da sich diese nicht zwangsläufig gegenseitig bedingen.

Während auf Schulrechtsseite für den Grundschulbereich zwischen den Zeugnisnoten der Pflichtfächer und der Bewertung des Arbeits-, Lern- und Sozialverhaltens explizit unterschieden wird, bleibt weiterhin offen, inwiefern diese Lehrkräfteeinschätzung auch in die Zeugnisnoten der jeweiligen Pflichtfächer miteinfließt. Anzunehmen ist, dass der explizit zugesprochene Beurteilungs- und Bewertungsspielraum die Möglichkeit offenlässt, in den entsprechenden Fachnoten ebenfalls das Lern-, Arbeits- und Sozialverhalten einfließen zu lassen; für das Lern- und Arbeitsverhalten wird dies durch das Schulrecht legitimiert. Welche Rolle in diesem Zusammenhang Erwartungseffekte auf Seiten der Lehrkraft spielen, wird an dieser Stelle nicht weiter vertieft. Festzuhalten ist an dieser Stelle allerdings, dass Menschen anhand von nahezu identischen Informationen durchaus zu unterschiedlichen Schlussfolgerungen kommen können (z.B. Hochweber, 2010). Welche Bedeutung die gegenseitige Beeinflussung von kognitiven und nicht-kognitiven Schülermerkmalen im Sinne moderierender Effekte für die Notenzusammensetzung spielen, wurde bisher nur am Rande erwähnt. So diskutiert Solga (2009) die Annahme, ob Begabungs-, Intelligenz- oder Vorwissensunterschiede allein notwendigerweise zu Bildungserfolgen führen, oder für entsprechende Erfolge zusätzlich auch entsprechende, nicht-kognitive Merkmale wie die Anstrengung des Schülers vorliegen müssten.

### 4.2.3 Moderatoreffekte individueller Schülermerkmale

In Anlehnung an das in [Kapitel 4.1.2](#) eingeführte Modell von Heller (1995) und unter Bezug zahlreicher Forschungsbefunde wird in diesem Unterkapitel nochmals die Wechselwirkung zwischen kognitiven und nicht-kognitiven Schülermerkmalen fokussiert.

Wie die aufgeführten Forschungsbezüge deutlich machen, nehmen individuelle Determinanten auf Seiten des Kindes eine bedeutsame Rolle bei der Zusammensetzung von Schulnoten ein; hier exemplarisch fokussiert durch Testleistung, Intelligenz, Vorwissen, sprachliche Fähigkeiten, Arbeits- und Sozialverhalten. Alle aufgeführten Merkmale sind – auf theoretischer und empirischer Ebene – bedeutsame Prädiktoren für die Notenzusammensetzung. Dabei muss von einer multikausalen Bedingtheit von Schulleistung ausgegangen werden, bei welcher sich auch die aufgeführten Prädiktoren gegenseitig bedingen. Ein moderierender Effekt würde sich demnach einstellen, wenn der Zusammenhang zwischen kognitiven Schülermerkmalen und Schulleistung (Prädiktor-Kriteriums-Zusammenhang) durch nicht-kognitive Schülermerkmale systematisch variiert werden würde (Heller, 1997). So fanden bereits Heller, Rosemann und Steffens (1978) in ihrer Längsschnittstudie zur Schullaufbahnberatung u.a. moderierende Effekt für das von der Lehrkraft eingeschätzte Arbeitsverhalten, die Selbstständigkeit und Konzentration des Schülers, welche den Zusammenhang zwischen Intelligenz und Schulleistung beeinflussen (Heller, 1997). Dabei verweisen die Autoren auf den Aspekt der differentiellen Validität, bei dem ein Prädiktor je nach Personengruppe unterschiedliche Effekte vorweist, so dass Schüler mit positiver Intelligenzleistung und positiven Persönlichkeitsmerkmalen entsprechend höhere Leistungen erzielen als diejenigen mit weniger positiven Merkmalsausprägungen (Heller et al., 1978). Bezüglich des Einflusses von Intelligenztestleistungen auf die Grundschulnoten am Ende der vierten Klasse fanden Sauer und Gamsjäger (1996) unter Kontrolle des Lehrkrafturteils zum Arbeitsverhalten des Kindes und dessen Testleistung keine direkten Effekte der Intelligenz mehr. Obgleich in dieser Untersuchung damit ein mediiender Effekt vermutet werden kann,

scheint in diesen Analysen die Einschätzung der Lehrkraft mit den größten Effekt für die Grundschulnote zu haben.

Ebenso untersuchten Gut, Reimann und Grob (2012) mithilfe der längsschnittlich angelegten Stichprobe im Grundschulbereich (N=263), inwiefern die aus Schülerperspektive erhobene Intelligenz, sprachliche, mathematische und sozial-emotionale Kompetenzen (erhoben mittels IDS; Intelligence and Development Scales) deren Schulnoten in Deutsch, Mathematik und Sachkunde drei Jahre später vorhersagen können. Neben direkten Effekten individueller Schülermerkmale ergaben sich in den Analysen Moderatoreffekte der sozial-emotionalen Kompetenz auf den Zusammenhang zwischen Intelligenz, Sprache und Mathematiknote sowie auf den Zusammenhang zwischen Intelligenz und Deutschnote (Gut et al., 2012). Obwohl die Autoren selbst auf eine mangelnde Berücksichtigung weiterer, vielfach nachgewiesener außerschulischer Einflussfaktoren verweisen, kann aus den gefundenen Ergebnissen auf eine kompensierende Wirkung sozial-emotionaler Kompetenzen geschlossen werden. Hohe sozial-emotionale Kompetenz konnte also negative Werte für die Intelligenz und die sprachlichen Kompetenzen ausgleichen, wobei der Effekt für Schüler mit hoher Intelligenz und hohen sprachlichen Fähigkeiten weniger stark wurde (Gut et al., 2012).

Zusammenfassend deuten die aufgeführten Befunde sowie der Forschungsüberblick von Heller (1997) und dessen theoretische Modellierung (vgl. [Kap. 4.1.2](#)) darauf hin, dass der Zusammenhang zwischen kognitiven Schülermerkmalen und Schulleistung durch nicht-kognitive Merkmale kovariiert wird. Zugleich existieren in Anlehnung an die Ergebnisse von Gut et al. (2012) wenige Untersuchungen zu moderierenden Effekten zwischen weiteren aus Lehrkraftperspektive erfassten kognitiven und nicht-kognitiven Schülermerkmalen. Gemäß der Thesen von Heller (1997) sollten zur Vorhersage von Schulleistung nicht nur einzelne Prädiktoren und Moderatoren berücksichtigt, sondern auch deren Interaktion in Analysen miteinbezogen werden. Ob durch die Perspektive der Lehrkraft auf die Schülermerkmale von einer höheren Passung mit den von ihr gebildeten Schulnoten ausgegangen

werden kann, lässt sich an dieser Stelle nur vermuten und muss im weiteren Verlauf empirisch geprüft werden.

Die bisher gefundenen Ergebnisse machen zudem immer wieder deutlich, dass für die Entwicklung und Bewertung individueller Schülermerkmale die familiäre Umgebung eine bedeutsame Vermittlungsinstanz im Zusammenhang zwischen schulischer Leistung und kindlichen Merkmalen einnimmt (z.B. Gerleigner, 2013; Giesinger, 2014; Wiese, 1982). In Anlehnung an die bereits dargestellten Zusammenhänge zwischen individuellen Schülermerkmalen und Schulleistung folgt nun eine Analyse der etwas weiter gefassten individuellen Schülermerkmale: eine theoretische wie empirische Analyse der Zusammenhänge zwischen familiären Hintergrundmerkmalen und schulischer Leistung, wobei in [Kapitel 4.3.2](#) insbesondere auf medierende Effekte zwischen familiären Hintergrundindikatoren, Schülermerkmalen und Schulleistung eingegangen wird.

### 4.3 Familiäre Bedingungsfaktoren von Schulleistung

Folgt man den oben exemplarisch vorgestellten Modellen zu Bedingungsfaktoren schulischer Leistung (vgl. [Kap. 4.1](#)), kommt auch dem familiären Hintergrund eine maßgebliche Rolle bei schulischen Bildungsprozessen zu. Gegenstand des nachfolgenden Kapitels sind die vielfach theoretisch und empirisch untersuchten familiären Hintergrundmerkmale, die häufig als maßgebliche Einflussfaktoren schulischer Leistungen, Bildungsbeteiligungen und der daraus resultierenden Bildungserfolge diskutiert werden (Ditton, 2007a). Nach einer kurzen Einführung zum Zusammenhang zwischen Familie und Schulleistung folgt in [Punkt 4.3.1](#) ein theoretischer Überblick zu Erklärungsansätzen sozialschichtspezifischer Bildungserfolge. Das nachfolgende Kapitel beschäftigt sich dann aus empirischer Perspektive mit Effekten der sozialen Herkunft auf die Schulleistungen von Kindern, differenziert nach Struktur- und Prozessmerkmalen familiärer Herkunft.

In der Bildungsforschung richten sich Fragen der Reproduktion sozialer Disparitäten in Bildungsbeteiligung und schulischen Leistungen u.a. auf Prozesse in Familie und Schule, um die Genese sozialer Selektivität und deren Vermittlungsmechanismen aufzuspüren (Dietrich & Fricke, 2013; Ditton et al., 2005; Ditton & Maaz, 2011; Maaz, Baumert & Trautwein, 2011). Jene durch familiäre Herkunft vermittelten Effekte werden häufig in *direkte* und *indirekte Schichteffekte* auf Bildungsbeteiligung und schulische Leistungen unterschieden. Während Merkmale wie die soziale Schichtzugehörigkeit eines Kindes deren Leistung indirekt über Vermittlungsmechanismen (z.B. Unterstützung durch zusätzliches Lernmaterial, elterliche Hausaufgabenhilfe etc.) beeinflussen können, würde man z.B. von direkten Schichteffekten sprechen, wenn Lehrkräfte die soziale Herkunft des Kindes unabhängig von dessen Noten in der Übertrittsempfehlung berücksichtigen (Schrader & Helmke, 2008; Wiese, 1982). Direkte Schichteffekte auf die Leistungsbewertung könnten sich aber auch durch die elterliche Unterstützung bei Hausaufgaben ergeben, die dann wiederum von der Lehrkraft als günstige Bedingung für höhere, prognostisch valide Schullaufbahnen angesehen werden und unabhängig von der gezeigten Schulleistung die Lehrkraft in ihrer Benotung beeinflusst (Wiese, 1982). Im Sinne der bildungssoziologischen Reproduktionstheorien wird zudem hinterfragt, ob die schulinternen Vorstellungen eines „guten Schülers“ nicht eher mit Verhaltensmerkmalen von Schülern höherer Sozialschichten übereinstimmen, wodurch sich wiederum indirekte Schichteffekte auf die Schulnote des Kindes ergeben könnten (Bourdieu, 1982; Ditton, 2011; Wiese, 1982).

Empirische Befunde und theoretische Ansätze zu sozialschichtspezifischen Bildungsverläufen beziehen sich dabei häufig auf Bildungsentscheidungen am – für die weitere Bildungskarriere bedeutsamen – Übergang von der Grundschule in die Sekundarstufe I, der im Folgenden ebenfalls im Fokus steht (z.B. Ditton, 2007b; Maaz, Hausen, McElvany & Baumert, 2006; McElvany & Razakowski, 2013). Nachfolgend werden aus vornehmlich theoretischer Perspektive die Ansätze von Boudon (1974), Bourdieu (1982) und Coleman (1988) zu sozialschichtspezifischen Bildungsprozessen vertieft. Ein Bezug zu weiteren

schulischen Erklärungsansätzen folgt in [Kapitel 4.4](#) der vorliegenden Arbeit, wobei eine vollständige Ausführung der Erklärungen sozial ungleicher Bildungserfolge hier nicht erfüllt werden kann<sup>32</sup>.

#### 4.3.1 Ausgewählte Erklärungsansätze zu familiär bedingten Ungleichheiten im Schulsystem

Ausgangspunkt vieler Erklärungsansätze zu sozial ungleichen Bildungschancen ist die sogenannte Zirkelhypothese, welche generationsübergreifende Bildungsungleichheiten mithilfe eines zirkulären Prozesses beschreibt, die sich aufgrund unterschiedlicher Schulerfolge stetig reproduzieren (Ditton, 2011; Rolff, 1997). Demnach sind Angehörige höherer Sozialschichten besser mit den Regeln, Abläufen, Möglichkeiten und Anforderungen des Bildungssystems vertraut und können größere Ressourcen aufbringen, um in die für sie sozialschichtert haltenden Bildungsprozesse zu investieren. Eine höhere Ressourcenausstattung führt gemäß dieser Annahme dazu, bei der Wahl der Bildungslaufbahn höhere Risiken einzugehen, da diese – bei mangelndem Erfolg – durch außerschulische Unterstützungsmöglichkeiten gefördert werden können (Ditton, 2011). Die, im Rahmen humankapitalistischer Erklärungsansätze entstandenen Rational-Choice-Modelle gehen davon aus, dass Bildungsentscheidungen durch Kosten-Nutzen-Abwägungen möglicher Bildungserträge getroffen werden, wobei subjektive Kosten und erwartete Erträge gewichtet werden (Boudon, 1974; Breen & Goldthorpe, 1997; Ditton, 1992; Erikson & Jonsson, 1996; Esser, 1999). Soziale Bildungsungleichheiten werden demnach durch sozialstratifizierte Bewertungen der anfallenden Kosten und Erfolgserwartungen erklärt, die sich durch entsprechende Unterschiede in Bildungsentscheidungen und der daraus folgenden Bildungsteilnahme stetig reproduzieren (Maaz, Baumert et al., 2011). Als Ausgangspunkt vieler, später entwickelten Theorien zum rationalen Entscheidungsverhalten unterscheidet Boudon (1974) zwischen primären und sekundären

---

32 Eine umfassende Übersicht zu bildungssoziologischen Erklärungen von Ungleichheiten an Bildungsübergängen ist beispielsweise bei Kristen (1999) zu finden. Des Weiteren ist anzumerken, dass sich die vorliegenden Ansätze ausschließlich auf die Erklärung *sozialer* Ungleichheiten im Bildungssystem beziehen. Einen Überblick zu Erklärungsansätzen für die Benachteiligung von Migranten geben u.a. Ditton und Auling (2011).

Effekten der sozialen Ungleichheit. Primäre Effekte beziehen sich auf unterschiedliche Schulleistungen von Kindern, die durch deren soziale Herkunft bzw. die sozioökonomische Lage des Elternhauses zustande kommen und damit zur sozialen Ungleichheit der Bildungschancen beitragen. Sekundäre Effekte betreffen dagegen das sozialschichtabhängige, subjektive Entscheidungsverhalten der Eltern, unabhängig von der Schulleistung ihres Kindes (Becker & Lauterbach, 2016).

Soziale Disparitäten am Übertritt in die Sekundarstufe I können demnach durch zweierlei Effekte erklärt werden: Einerseits wird die Schulleistung des Kindes durch sozialschichtspezifische Bedingungen – Kapitalien im Sinne Bourdieus (1983) – beeinflusst, die wiederum die Bildungsentscheidung bedingen. Andererseits wirkt sich die soziale Herkunft direkt und unabhängig von der Schulleistung auf die Kosten-Nutzen-Abwägungen bei der Übertrittsentscheidung aus. Während bei sekundären Herkunftseffekten unterschiedliche Entscheidungskomponenten betrachtet werden, beziehen sich die primären Effekte der sozialen Herkunft nach Boudons (1974) „*simple theoretical scheme*“ vornehmlich auf die schulischen Leistungen eines Kindes („*school achievement*“). Angenommen wird, dass die in der Schule erworbenen Leistungen neben anlagebedingten Einflussfaktoren von monetären und vornehmlich nichtmonetären, familialen Ressourcen beeinflusst werden. Diese herkunftsspezifischen Sozialisationsbedingungen (sozialschichtspezifische Sprachkultur, Lern- und Bildungsmotivation, habitualisierte Lerngewohnheiten, kulturelle Anregungen, Vorwissen etc.) werden wiederum als maßgebliche Vermittlungsmechanismen für schulische Leistungen diskutiert und wurden bereits zahlreichen empirischen Überprüfungen unterzogen (z.B. Datcher-Loury, 1989; Ditton, 2007b; Gresch, Baumert & Maaz, 2010; Maaz, Baumert et al., 2011; Maaz et al., 2006; Maaz & Nagy, 2010). Während in Boudons (1974) Abhandlung die sozialstratifizierten Schulleistungen durch verbale Tests („*Achievement in a Verbal Test*“) und schulische Leistungen („*school achievement*“) erhoben wurden, unterscheiden sich die empirischen Abhandlungen zur Überprüfung jener Annahmen in der Wahl des Indikators Schulleistung (Maaz & Nagy, 2010). Ohne auf die spezifischen Unterschiede der Abhand-

lungen einzugehen, wird selten thematisiert, welche Leistungsfaktoren konkret als sozialschichtspezifisch determiniert betrachtet werden; welche ganz konkreten Leistungsvariablen also durch familiäre Vermittlungsmechanismen beeinflusst und als unterschiedlich wahrgenommen werden. Eine unterschiedliche Wahrnehmung schulischer Leistung am Übertritt in die Sekundarstufe I betrifft dabei die Einschätzung der kindlichen Kompetenzen durch die Lehrkraft. So weisen Befunde von Maaz und Nagy (2010) zur Übertrittsempfehlung darauf hin, dass unabhängig von schulischen Leistungen direkte Effekte der sozialen Herkunft auf die Lehrkräftempfehlung bestehen bleiben, was wiederum als sekundärer Effekt der Leistungsbewertung<sup>33</sup> bezeichnet wird. Gleichzeitig weisen McElvany und Razakowski (2013) in ihrer Abhandlung auf die Notwendigkeit weiterer Untersuchungen zur differenziellen Notengebung hin, da diese auch die – im Sinne motivationaler Merkmale – Mitarbeit oder Anstrengung beeinflussen könnten, was wiederum häufig nicht kontrolliert wird.

Boudons Grundannahmen wurden in neueren Modellen zur rationalen Wahl aufgenommen und weiterentwickelt (Maaz, Baumert & Trautwein, 2010). Kritisiert wird unter anderem die Vernachlässigung der Sozialisationswirkung der Schule selbst sowie das ausschließlich als rational angenommene Entscheidungsverhalten (z.B. Maaz et al., 2010; Schauenberg, 2007). Neben Beschreibungen sozialungleicher Bildungsprozesse zeichnen sich die Modelle jüngerer Zeit durch Erklärungsversuche von Ursache-Wirkungs-Zusammenhängen zwischen sozialer Herkunft und Bildungsungleichheiten aus (Becker & Lauterbach, 2016; Erikson & Jonsson, 1996; Esser, 1990). Demnach werden nach Becker und Lauterbach (2016) vornehmlich die sozialen Vermittlungsmechanismen zwischen Ursache und Wirkung fokussiert, um soziale Reproduktionsmechanismen zu identifizieren. Als Beispiel führen die Autoren die zur Gruppe der Rational-Choice-Theorien

---

33 Maaz und Nagy (2010) entwickeln in Anlehnung an Boudon (1974) die Identifikation und Operationalisierung von primären und sekundären Herkunftseffekten weiter. Sie differenzieren dabei u.a. zwischen der chronologischen Abfolge von Leistungsbeurteilungen (Noten, Schullaufbahneempfehlung), wobei objektive Schulleistungen und Leistungsbeurteilungen unterschieden werden müssen.

zugeordneten werterwartungstheoretischen Modelle individueller Bildungsverläufe an, die sich auf jene sozialen Vermittlungsmechanismen fokussieren (Becker & Lauterbach, 2016). Ohne weiter auf Einzelheiten dieser Ansätze einzugehen, lassen sich die sozialstratifizierten Leistungsunterschiede laut Maaz, Baumert und Trautwein (2011) vornehmlich durch die unterschiedliche Ausstattung in ökonomischem, kulturellem und sozialem Kapital und deren Wirkmechanismen erklären, die auf eltern- und lehrkraftspezifisches Entscheidungsverhalten Einfluss nehmen können.

Der zugrundeliegende Erklärungsansatz zur Reproduktion von sozialer Ungleichheit stammt aus den Aufsätzen Bourdieus (z.B. Bourdieu, 1976, 1982, 1983; Bourdieu & Passeron, 1971), der entgegen gesetzt zu den Rational-Choice-Modellen nicht von rational getroffenen Bildungsentscheidung ausgeht. Stattdessen analysieren dessen Abhandlungen die verdeckten, sozial reproduzierenden Gesellschaftsstrukturen, die wiederum sozialschichtabhängige Schulleistung erklären (Gerleigner, 2013). Ohne weiter auf die umfassenden Facetten der bourdieuschen Reproduktionstheorie einzugehen, ist für die vorliegende Untersuchung von Bedeutung, durch welche Vermittlungsmechanismen hier sozial ungleiche Schulleistungen erklärt werden. Bourdieus Erklärung beruht hierzu auf der Annahme klassenspezifischer Kulturformen, die für die Aufrechterhaltung sozial ungleicher Gesellschaftsverhältnisse verantwortlich sind (Hradil, 2005). Thematisiert wird auch die Rolle des Schulsystems selbst, das sich als *Reproduktionsagent* durch die Vergabe von Bildungstiteln – im positiven Sinne – auszeichnet, aber durch die daraus folgende Statuszuweisung auch stigmatisierende, hierarchisch gestufte Klassen erzeugen kann (Bourdieu, 1982, 1983). So definieren Bourdieu und Passeron (1971) soziale Ungleichheit in der Schule als schichtspezifischen Mechanismus, der durch die institutionelle Werteordnung, durch vorgeschriebene Regeln und Sprachcodes in der Schule gefördert wird und bestimmte soziale Gruppen in ihren Handlungen begünstigt (Bourdieu & Passeron, 1971). Ausgehend von der Annahme, dass man gesellschaftliche Strukturen und Funktionen erst durch „den Begriff des Kapitals in allen seinen Erscheinungsformen“ (Bourdieu, 1983, S. 184) verstehen

könne, differenziert Bourdieu (1983) – entgegengesetzt zur rein wirtschaftlichen Perspektive – zwischen ökonomischem, kulturellem und sozialem Kapital. Während das ökonomische Kapital jeglichen materialen Wohlstand umfasst, bezieht sich das kulturelle Kapital auf drei Kapitalformen: das inkorporierte, objektivierte und institutionalisierte kulturelle Kapital.

Das *inkorporierte kulturelle Kapital* bezieht sich auf körpergebundene, kulturelle Bildungsprozesse, die vom Individuum selbst verinnerlicht werden. Jene Kapitalform ist maßgeblich von der Zeit ihres Aneignungsprozesses abhängig und kann nicht kurzfristig erworben und weitergegeben, wohl aber durch die ökonomische Lage der Familie beeinflusst werden (finanzieller Erwerbsdruck vs. Befreiung von ökonomischen Verpflichtungen). Die Inkorporierung kultureller Fähigkeiten kann dabei ohne Erziehungsmaßnahmen erfolgen und beeinflusst typische Sprechweisen und die Persönlichkeit selbst (Bourdieu, 1983). Das *objektivierte kulturelle Kapital* umfasst dagegen sämtliche kulturellen Güter von wirtschaftlichem Wert (z.B. Kunstwerke, Bücher, Instrumente). Während die objektivierten Kulturgüter ähnlich wie ökonomisches Kapital übertragbar sind, benötigt das Individuum selbst kulturelle Fähigkeiten – jenes inkorporierte kulturelle Kapital – „die den Genuß eines Gemäldes oder den Gebrauch einer Maschine erst ermöglichen.“ (Bourdieu, 1983, S. 188). Schließlich wird das *institutionalisierte kulturelle Kapital* als diejenige Kapitalform bezeichnet, die im erweiterten Sinne der nachfolgenden Untersuchung als abhängige Variable dient. Jene dritte kulturelle Kapitalform bezieht sich auf gesellschaftlich legitimierte, schulische oder akademische Bildungstitel, die dem Individuum seine „kulturellen Eigenschaften“ in einen rechtlich anerkannten Wert übertragen. Diese, durch das Unterrichtssystem bestätigte institutionelle Bewertung bzw. verliehene Anerkennung wird wiederum als Verbindungsglied zwischen ökonomischem und kulturellem Kapital bezeichnet: Einerseits ist für den Erwerb eines Titels ein bestimmter Geldwert erforderlich und andererseits bedingt dieser Titel auf dem Arbeitsmarkt vermutlich mehr ökonomisches Kapital. Jene institutionalisierte Kulturkapitalform führt gemäß der Annahmen dann wiederum dazu, gesellschaftliche Anerkennung zu

erlangen, die – bedenkt man die sozial ungleichen Bildungserfolge – sozialschichtspezifische Gesellschaftstrukturen bedingen könnten (Bourdieu, 1983).

Gemäß dieser theoretischen Annahmen sind jene kulturellen Kapitalformen die entscheidende Einflussgröße für den Bildungserfolg des Kindes und werden – zumindest implizit – von Generation zu Generation weitergegeben. Nach Bourdieu (1983) werden durch diese theoretischen Thesen Ungleichheiten in schulischen Leistungen zwischen sozialen Klassen erst greifbar, da diese – so die Annahme – nicht allein auf natürliche Fähigkeiten oder ökonomische Bildungsinvestitionen zurückzuführen sind. Demnach müsse man berücksichtigen, dass Begabungen und Fähigkeiten „das Produkt einer Investition von Zeit und kulturellem Kapital“ seien (Bourdieu, 1983, S.186) und schulische Erträge maßgeblich vom kulturellen Kapital einer Familie abhängen sowie die verschiedenen Kapitalformen direkt miteinander zusammenhängen und sich gegenseitig bedingen (Bourdieu, 1983).

Wenn man davon ausgeht, dass das institutionalisierte kulturelle Kapital gesellschaftliche Anerkennung beeinflusst, benennt Bourdieu (1983) als dritte Kapitalart die *soziale Kapitalform*. Als Sozialkapital wird „die Gesamtheit der aktuellen und potentiellen Ressourcen, die mit dem Besitz eines dauerhaften Netzes von mehr oder wenig institutionalisierten Beziehungen gegenseitigen Kennens oder Anerkennens verbunden sind“ (Bourdieu, 1983, S. 190) bezeichnet. Diese Ressourcen beruhen auf einer bestimmten Gruppenzugehörigkeit, die Zugang zu (institutionellen) Beziehungen und Netzwerken gewährt und so beispielsweise die ökonomische Kapitalaneignung ermöglicht (McElvany & Razakowski, 2013). Das Ausmaß des Sozialkapitals hängt wiederum von der Breite des Beziehungsnetzwerkes ab sowie von den zur Verfügung stehenden, kulturellen und ökonomischen Kapitalien der sozialen Interaktionspartner. Anders als das ökonomische Kapital ist das Sozialkapital bzw. das Beziehungsnetz ein Produkt individueller oder gemeinschaftlicher Investitionsbeziehungen, die (un)bewusst zu Schaffung und Erhalt von sozialen Beziehungen getätigt werden (Bourdieu, 1983). Jene rein zum Selbstzweck

gebildeten sozialen Kapitalformen wurden in Colemans Theorie um eine rationale Komponente erweitert (Gerleigner, 2013). Während er einerseits Bourdieus Annahmen der Nutzung sozialer Kapitalformen zur individuellen Zielverwirklichung sowie ökonomischen Konvertierbarkeit bestärkt, verweist er andererseits auf die wechselseitigen Erwartungen und sozialen Kontrollen sowie auf diejenigen Ressourcen, die insbesondere die kognitive und soziale Kindesentwicklung fördern (Baumert, Watermann & Schümer, 2003; Coleman, 1988; Gerleigner & Auling, 2017). Auch in der Sozialkapitaltheorie nach Coleman (1988) werden Formen von Sozialkapital unterschieden (obligations and expectations, information channels, social norms). Entgegengesetzt zu Bourdieus Annahmen wird soziales Kapital nicht als Ausdruck finanzieller Ressourcen betrachtet, sondern es wird von einer negativen Beziehung zwischen diesen ausgegangen. Bezüglich dem Verhältnis zwischen kulturellem und sozialem Kapital wird außerdem angenommen, dass sich das kulturelle Kapital einer Familie erst durch die Qualität und Intensität innerfamiliärer Beziehungsstrukturen auf die intellektuelle Entwicklung des Kindes auswirken kann; Kinder also erst durch die elterliche Zuwendung von deren kulturellen Kapitalien profitieren können (Coleman, 1996). Zusätzlich dazu hebt Coleman (1988) auch die Intensität der elterlichen Kontakte zur Schule ihrer Kinder hervor, durch welche die Eltern untereinander engere Verbindungen schließen, die wiederum unterstützendes Verhalten für die Kinder und effektive Schul- bzw. Bildungsnormen hervorbringen könnten. Demnach wird bei einer stärkeren Nähe zwischen Schule und Elternhaus von einem positiveren Bildungsklima ausgegangen, das sich auf die Intensität der Förderung innerhalb der Familie auswirkt, und so auch die Motivation und das Schulumengagement des Schülers beeinflusst. In diesem Sinne wird von einer vermittelnden Wirkung ausgegangen: Das schulbezogene Sozialkapital der Familie wirkt über deren Förder- und Erwartungsverhalten auf die schulbezogenen Eigenschaften des Kindes, die sich dann wiederum auf dessen Leistung auswirken. Der Zusammenhang zwischen Sozialkapital und Schulleistung sollte also vollständig über jene Indikatoren vermittelt werden (Stocké, 2010).

Bedingt durch das soziale, kulturelle und ökonomische Kapital ist in diesem Zusammenhang auch das bourdieusche Konzept des *Habitus* von zentraler Bedeutung. Der Habitus einer Person wird als „Erzeugungsprinzip objektiv klassifizierbarer Formen von Praxis und Klassifikationssystem“ (Bourdieu, 1982, S.277) bezeichnet, durch die sich Individuen ihre soziale Welt repräsentieren und sich in ihr zurechtfinden. Bourdieu (1983) setzt jenes Habitus Konzept mit dem inkorporierten kulturellen Eigentum einer Person gleich: „aus „Haben“ ist „Sein“ geworden.“ (Bourdieu, 1983, S. 187). Bedeutsam für die vorliegende Abhandlung ist dabei die Tatsache, dass jene Wahrnehmungs-, Handlungs- und Bewertungsschemata das gesamte Verhalten einer Person beeinflussen. Da diese Wahrnehmungen aber wiederum eine Teilung in soziale Lagen verinnerlichen, tragen sie gleichzeitig zu einem „System von Differenzen von unterschiedlichen Positionen“ (Bourdieu, 1982, S. 279) bei, wodurch sich soziale Identitäten voneinander abgrenzen sowie entsprechend unterschiedliche Lebensstile verfolgen, welche wiederum als systematische Produkte des Habitus bezeichnet werden (Bourdieu, 1982). Übertragen auf den Kontext Schule könnte dies zur Folge haben, dass sich Familien und Lehrkräfte unterschiedlicher Schichtzugehörigkeit in ihren Wahrnehmungen systematisch voneinander unterscheiden und je nach sozialer Umgebung unterschiedliche Verhaltensweisen aufzeigen, die sich letztlich auch in der Leistungsbeurteilung eines Schülers und schlussendlich in dessen Bildungserfolg widerspiegeln könnten.

Zusammenfassend festgehalten werden muss nach der Darstellung ausgewählter Erklärungsansätze zur Entstehung sozialer Ungleichheiten nochmals, dass diese nicht allein durch Bedingungen im familiären Kontext erklärt werden können. Dennoch geht man davon aus, dass im Bildungssystem jenen, über die primären Effekte vermittelnden Bildungsungleichheiten die größte Bedeutung zukommt (Ditton, 2011). Während die bisherige Abhandlung hauptsächlich aus der theoretischen Perspektive argumentiert hat, wird nachfolgend ein mehrheitlich empirischer Fokus gewählt, welcher sich auf die theoretische Basis von Boudon (1974), Bourdieu (1983) und Coleman (1988) bezieht

und vornehmlich die individuellen Vermittlungsmechanismen zwischen sozialer Herkunft und Schulleistung fokussiert.

#### 4.3.2 Zum Zusammenhang zwischen familiären Struktur- und Prozessmerkmalen und individuellen Schulleistungskomponenten

In Anlehnung an die oben eingeführten familialen Ansätze zur Erklärung sozialer Ungleichheiten im Schulsystem, ergeben sich für die Empirie zahlreich zu prüfende Einflüsse auf die schulischen Leistungen eines Kindes, die durch die familiäre Ressourcenausstattung bedingt werden. Diese außerschulischen Effekte werden unter Bezug auf das oben eingeführte Modell von Helmke und Weinert (1997) häufig danach unterschieden, ob sie den strukturellen, als zeitstabil angenommenen Einflussfaktoren zugeordnet werden oder ob sie sich auf prozesshafte, zeitvariable Merkmale beziehen. Strukturelle Merkmale umfassen in der Empirie u.a. Faktoren des elterlichen Bildungshintergrunds, Berufsstands und andere strukturelle Schichtungsmerkmale sowie Faktoren der Familiengröße,-konstellation und Geschwisterreihung. Die Differenzierung nach Prozessmerkmalen der Familie umfasst jegliche schichtspezifisch ausgeformten (kulturellen) Praktiken, wie z.B. die elterlichen Bildungswünsche, deren Schulpräsenz bzw. schulisches Kommunikationsverhalten sowie die Unterstützungsleistung der Schulkinder (Gerleigner, 2013; Krapp, 1976; Richert, 2012; Stocké, 2010).

Auch Baumert, Watermann und Schümer (2003) unterscheiden in ihrer Studie in Struktur- und Prozessmerkmale familialer Herkunft, wobei in ihrem theoretischen Modell die Strukturmerkmale den Prozessmerkmalen vorangestellt werden. Die strukturelle Ausstattung der Familie, die sowohl vermittelt über die kulturellen und sozialen Handlungen der Familie als auch direkt auf den Kompetenzerwerb wirkt. Die Differenzierung beruht dabei auf der Einteilung nach ökonomischen, kulturellen und sozialen Kapitalien im Sinne Bourdieus (1983). Die Autoren gehen davon aus, dass zum Verständnis generationsübergreifender sozialer Disparitäten auch andere Aspekte familiärer Lebensverhältnisse betrachtet werden müssen, die im Sinne der

Wert-Erwartungstheorie vermittelnde Effekte beschreiben können (Baumert, Watermann et al., 2003). Jene Einteilung dient zahlreichen Studien als Ausgangspunkt. Während beispielsweise Effekte des sozioökonomischen Status – als struktureller Aspekt – bereits lange Zeit als Erklärung von Schulleistungen betrachtet werden, wird auf Seiten der Prozessmerkmale häufig das kulturelle und soziale Kapital der Familie analysiert. In der Bildungssoziologie werden dabei häufig die sozialen Kapitalien der Familie fokussiert, worunter – ähnlich wie beim Leistungsbegriff – inhaltlich weitgefasste Faktoren subsumiert werden (Stocké, 2010). Stocké (2010) fasst den bildungssoziologischen Forschungsstrang zur Sozialkapitaltheorie in zwei Gruppen zusammen: während einmal Prozesse in der Familie wie z.B. der Beziehungintensität zwischen Eltern und Kind, der Häufigkeit schulischer Unterstützung sowie Gespräche über schulische Angelegenheiten fokussiert werden, beschäftigt sich der andere Forschungszeitraum mit den verschiedenen Facetten des Kontaktes zwischen Elternhaus und Schule. Einzelne Aspekte dieser beiden Determinantenbereiche werden im Folgenden näher betrachtet. Grundlage der nachfolgenden Abhandlung ist die Frage nach den Wirkmechanismen der Reproduktion von Bildungsungleichheiten, auf Struktur- und Prozessebene familiärer Merkmale. Fokussiert werden für beide Bereiche aus empirischer Perspektive zunächst die direkten Effekte familiärer Herkunft auf die Schulleistung des Kindes. Anschließend folgen zentrale Kernbefunde zu indirekten familiären Effekten, die vermittelt über die individuellen Schülermerkmale auf die Leistung des Kindes festgestellt wurden.

### **Strukturmerkmale**

Zentraler Bezugspunkt der sozialwissenschaftlichen Bildungsungleichheitsforschung zum Zusammenhang zwischen familiären Strukturmerkmalen und Bildungsprozessen ist der sozioökonomische Status, der meist über das Einkommen, den Beruf und das Bildungsniveau der Eltern definiert wird (Ditton & Maaz, 2015). Die Kategorisierung unterschiedlicher sozialer Positionen bezieht sich auf verschiedene Modelle gesellschaftlicher Strukturen, abgeleitet aus Klassen- und Schichtungstheorien. Empirische Untersuchungen beziehen sich meist auf kontinuierliche Maße und Einteilungen jener Schicht- und

Klassenmerkmale; häufig verwendet werden der ISEI (Ganzeboom, de Graaf & Treiman, 1992) und das EGP-Klassenschema (Erikson, Goldthorpe & Portocarero, 1979). Der ISEI (International Socio-Economic Index of Occupational Status) basiert auf Angaben zu Beruf, Einkommen und Bildungsniveau und kann als kontinuierlicher Maßstab Werte zwischen 16 (z.B. Reinigungskräfte) und 90 (z.B. Richter) annehmen, wobei häufig der höchste Berufsabschluss im Haushalt als familiäres Hintergrundmaß herangezogen wird (Ditton & Maaz, 2015; Ganzeboom et al., 1992). Das kategoriale EGP-Klassenschema bezieht sich auf die berufliche Tätigkeit, Beschäftigungsstatus und berufliche Stellung, wobei zwischen Arbeitgebern, Arbeitnehmern und Selbständigen (ohne Mitarbeiter) differenziert wird. Bei der Einteilung berücksichtigt werden die Art der Tätigkeit, die Stellung im Beruf und die Weisungsbefugnis, wobei von ursprünglich elf Klassen meist sechs Klassen unterschieden werden (Ditton & Maaz, 2015). Als weiterer Indikator sozialer Positionen wird ebenfalls häufig der Bildungsstatus der Familie herangezogen, wobei der höchste erworbene Schulabschluss in der Familie operationalisiert wird (Ditton & Krüsken, 2009).

Ausgehend von der Annahme, dass schichtspezifische Einflüsse auf die Notengebung durch unterschiedliche Mechanismen erfolgen können, untersuchte die frühe Studie von Wiese (1982) Effekte der Eltern- und Lehrkräfteeinstellung auf die Übertrittsempfehlung am Ende der Grundschulzeit. Neben direkten Schichteffekten auf die elterliche Schulformpräferenz und der stärkeren Notenorientierung der Lehrkräfte (.634) verweisen die Ergebnisse der Pfadanalyse auf einen indirekten Schichteffekt auf die Übertrittsempfehlung am Ende der Grundschulzeit, der durch den Zusammenhang zwischen sozialer Herkunft und Notengebung (.272) bedingt wird (Wiese, 1982).

Die internationalen Schulleistungsstudien PISA und IGLU untersuchen ebenfalls sozialschichtspezifische Einflüsse auf die schulische Leistungsentwicklung, wobei primär nicht der Einfluss auf Schulnoten, sondern auf die (über)fachlichen Kompetenzen der Schüler getestet wird. Watermann und Baumert (2006) analysierten auf Basis der

PISA Daten von 2000 den Zusammenhang zwischen sozialer Herkunft und schulischen Kompetenzen. In Anlehnung an das Modell von Baumert, Watermann und Schümer (2003) analysierten sie auf Basis eines psychologisch-institutionellen Mediationsmodells die bourdieuschen Reproduktionsmechanismen auf den schulischen Kompetenzerwerb. Neben direkten Effekten der einbezogenen Strukturmerkmale (sozioökonomischer Status, Bildungsniveau und Migrationsstatus der Familie) wurden diese überwiegend durch die eingezogenen, kulturellen Prozessmerkmale auf den Kompetenzerwerb vermittelt. Die Bedeutung familiärer Hintergrundmerkmale auf schulische Noten wurden nicht in die Analysen miteinbezogen. Dagegen weisen die Ergebnisse der IGLU-Studie für den Grundschulbereich auf einen weniger engen Zusammenhang zwischen sozialer Herkunft und Schulleistung hin. Im Grundschulbereich wurden dennoch für das Jahr 2001 Differenzen in der Übertrittsempfehlung nach sozialem Status festgestellt, wobei Grundschüler aus oberen Schichten unter Kontrolle ihrer kognitiven Grundfähigkeiten und Lesekompetenz eine 2,63-fach größere Chance auf eine Gymnasialempfehlung haben (Bos, Lankes et al., 2004).

Mit Daten der vorliegenden KOALA-S-Studie konnte auch Krüsken (2007a) unter Kontrolle individueller Schülermerkmale einen direkten, signifikanten Effekt des beruflichen Status der Eltern auf die Viertklassnote feststellen ( $\beta=.06$ ). Ebenfalls auf Basis der KOALA-Daten fand Müller (2013) einen direkten, signifikanten Effekt der Strukturmerkmale der Familie (ISEI und Bildungsstatus) auf die Durchschnittsnote der 4. Grundschulklasse. Welche direkten Effekte sich unter Kontrolle weiterer individueller kognitiver und nicht-kognitiver Schülermerkmale auf die Deutsch- und Mathematiknote der vierten Grundschulklasse ergeben, wurde in diesem Zusammenhang nicht getestet. Während die aufgeführten Befunde insgesamt auf geringe, aber durchaus bestehende direkte Effekte der familiären Strukturmerkmale auf Leistungsvariablen verweisen, wird auch auf den vermittelnden Effekt über die individuellen Schülermerkmale hingewiesen. Auch Maaz, Baeriswyl und Trautwein (2011) konstatieren anhand der ELEMENT-Daten für die vierte bis sechste Grundschulstufe bzw. für die TIMSS-Übergangsstudie unter Kontrolle der individuellen Testleistungen direkte, schwach ausgeprägte, aber signifikante Effekte des

sozioökonomischen Status auf die Durchschnittsnote am Ende der Grundschulzeit. Die Autoren benennen diese Zusammenhänge allerdings als „Bruttoeffekte“, dürfe man doch nicht davon ausgehen, dass Testleistung ein valides Prüfkriterium für Schulnoten ist. In einer weiteren Teilstudie untersuchen die Autoren den vermittelnden Effekt motivationaler Schülermerkmale und nehmen die selbstberichtete Anstrengungsbereitschaft und Gewissenhaftigkeit bei schulischen Aufgaben in ihre Analysen auf. Auf Basis von Daten zum Übergang in die Sekundarstufe I prüften die Autoren, ob für die familiären Herkunftseffekte unter Einbezug der individuellen Schülermerkmale eine partielle oder vollständige Mediation konstatiert werden kann. Die Ergebnisse ergeben allerdings nur eine geringfügige Änderung des direkten Effekts auf die Grundschulnote, wodurch die sozialschichtspezifische Benotung nicht durch die einbezogenen motivationalen Merkmale erklärt werden konnte (Maaz, Baeriswyl et al., 2011). Weitere Kernbefunde zu Mediationsprozessen zwischen familiären Strukturmerkmalen und individuellen Schülermerkmalen sowie deren Einfluss auf Schulnoten wurden bislang nur für einzelne kindliche Merkmale geprüft, wobei häufig selbstberichtete Schülermerkmale einbezogen werden.

Entgegengesetzt zu den Befunden von Maaz, Baeriswyl et al. (2011) ergab die Studie von Roth und Sauer (1981) unter Einbezug einer längsschnittlich erhobenen Stichprobe unterschiedlichen Alters (Kindergarten, Grundschulbeginn- und Ende), dass der Effekt der elterlichen Schulbildung auf die Schulleistung hauptsächlich über die kognitiven Leistungen des Schülers vermittelt wird. Ob der Einfluss der sprachlichen Fähigkeiten des Kindes den Effekt zwischen familiären Hintergrundmerkmalen und Schulleistung vermittelt, analysierte die Studie von Helmke und Reich (2001) auf Grundlage der MARKUS Daten aus der 8. Sekundarstufe unterschiedlicher Schularten. Schulartübergreifend ergeben sich unter Einbezug der elterlichen Bildungsnähe ( $r=.31$ ) und Deutsch als Primärsprache ( $r=.15$ ) positive Einflüsse. Während der sprachliche Hintergrund der Schüler allein rund zwei Prozent der Leistungsunterschiede innerhalb eines Bildungsgangs erklärt, reduziert sich der Effekt nochmals durch die Hinzunahme der familiären Bildungsnähe (beide Merkmale erklären 10,44 % der Leis-

tungsunterschiede) (Helmke & Reich, 2001). Neben der Bedeutung sprachlicher Fähigkeiten analysiert die Studie von Stahl (2009) unter anderem den Zusammenhang zwischen Lehrkräfteeinschätzungen zur kindlichen Begabung und Schulnoten sowie deren Abhängigkeit von familiären Hintergrundmerkmalen. Mittels Regressionsanalysen zur Vorhersage der Deutsch- und Mathematiknote am Ende der vierten Grundschulklasse weisen die Ergebnisse auf einen vermittelnden Effekt zwischen sozialer Herkunft und den einbezogenen individuellen Schülermerkmalen hin. Unter Kontrolle der Testleistung wird der darüberhinausgehende Effekt der sozialen Herkunft auf die Noten signifikant über die Einschätzung der Lehrkraft zur Begabung ( $\beta = -.56$ ) und Anstrengung ( $\beta = -.10$ ) vermittelt. Die Varianzaufklärung für die Deutschnote ist hier etwas größer als für die Mathematiknote (Stahl, 2009).

Auch Trautwein, Maaz und Baeriswyl (2014) beschäftigen sich in ihrem Übersichtsartikel mit der Rolle der Begabung am Übertritt in die Sekundarstufe I. Ausgehend von der gesellschaftlichen Annahme, dass die Schulartzuweisung nach der Grundschulzeit eine Identifikation bzw. Differenzierung bestimmter Begabungen beinhaltet, geben die Autoren zu bedenken, dass dabei Schüler mit vergleichbaren Begabungen in Abhängigkeit von sozialen Hintergrundmerkmalen und Geschlecht unterschiedliche Chancen erhalten; eine Differenzierung nach Begabung also durchaus an ihre Grenzen stößt. Auch die Arbeiten von Stubbe und Bos (2008) bzw. Stubbe, Bos und Euen (2012) analysieren auf Basis der IGLU-Daten den Zusammenhang zwischen Schullaufbahneempfehlungen- bzw. Entscheidungen von Lehrkräften und Eltern. Mittels Pfadanalysen wurde u.a. der Zusammenhang zwischen Testleistungen, individuellen Schülermerkmalen (kognitive Grundfähigkeiten, Geschlecht, Anstrengungsbereitschaft, Prüfungsangst), familiären Faktoren (ISEI, Migrationshintergrund, Anzahl Bücher) zu den Deutsch- und Mathematiknoten analysiert. Für beide Modelle zur Erklärung der Deutsch- und Mathematiknote weisen die Testleistungen den stärksten Zusammenhang zu den Noten auf. Unter Kontrolle dieser individuellen Leistungsergebnisse und der Anstrengungsbereitschaft des Schülers, blieben signifikante Effekte

der familiären Hintergrundmerkmale bestehen (Stubbe & Bos, 2008; Stubbe et al., 2012).

Neben zahlreichen anderen Autoren halten dennoch auch Helmke und Schrader (2010) in ihrer Abhandlung zu Determinanten der Schulleistung fest, dass von Struktur- und Statusvariablen (Berufstätigkeit der Eltern, Familienkonstellationen, Schichtzugehörigkeit etc.) lediglich ein indirekter Erklärungswert für individuell unterschiedliche Schulleistungen ausgeht. Neben den gerade fokussierten individuellen Schülermerkmalen werden als weitere Vermittlungsvariablen Merkmale familiärer Lernumwelt diskutiert. Fokussiert werden neben kulturellen Hintergrundmerkmalen im Sinne Bourdieus (1983) elterliche Erwartungen, Einstellungen und Verhaltensweisen, materielle und bildungsbezogenen Ressourcen oder Interaktionen, welche Schulleistungen beeinflussen (Helmke & Schrader, 2010; Schrader & Helmke, 2008). Jene, ausgewählten familiären Prozessmerkmale gilt es im Folgenden zu betrachten.

### Prozessmerkmale

In Anlehnung an das Makro-Modell von Helmke und Weinert (1997) und die theoretischen Annahmen von Bourdieu und Passeron (1971) und Coleman (1988) zum Zusammenhang zwischen familiären Hintergrundmerkmalen und Schulleistung wird im Folgenden auf die vermittelnde Instanz der prozesshaften, familiären Interaktionsformen eingegangen. Während Studien zu Prozessmerkmalen familiärer Herkunft häufig den Einfluss kultureller Kapitalien auf Schulleistungskomponenten fokussieren<sup>34</sup>, stehen nachfolgend insbesondere jene Merkmale im Blickpunkt, die einen direkten Bezug zur Schule vorweisen und insbesondere für die Leistungsbewertung der Lehrkraft relevant sein könnten. Ähnlich wie die Darstellung zentraler Kernbefunde zu den strukturellen Hintergrundmerkmalen, werden hier direkte und indirekte Effekte ausgewählter Prozessmerkmale auf die

---

<sup>34</sup> Baumert, Watermann und Schümer (2003) heben die kulturelle Praxis der Familie als den entscheidenden Faktor für den Kompetenzerwerb des Kindes hervor. Demnach wirkt das Bildungsniveau der Eltern vor allem durch die kulturelle Praxis in der Familie auf den Kompetenzerwerb des Schülers.

Schulleistung fokussiert. Im Fokus stehen empirische Evidenzen zu theoretisch deklarierten Mechanismen zwischen Struktur- und Prozessmerkmalen, individuellen Schülermerkmalen und Schulleistung (vgl. Kap. 4.1 bzw. Kap. 4.3.1). Während das theoretische Modell von Helmke und Weinert (1997) davon ausgeht, dass die elterlichen Verhaltensweisen auf unterschiedlicher Art das Lernverhalten beeinflussen (Motivation, Instruktion, Stimulation und Modelllernen), wurden in Studien u.a. die Bedeutung der elterlichen Bildungsaspiration, deren Unterstützungsverhalten sowie die Bedeutung des Kontakts zwischen Elternhaus und Schule fokussiert.

Im Sinne sozialkapitalistischer Annahmen würde man davon ausgehen, dass der Zusammenhang zwischen familiärem Sozialkapital und Schulleistung über die elterliche Bildungsaspiration, deren Förderverhalten sowie das individuelle Schülerverhalten vermittelt wird (Stocké, 2010). Ohne auf die theoretisch deklarierten Zusammenhänge zwischen elterlichen Bildungsaspirationen und Bildungserfolgskomponenten sowie deren Entstehung einzugehen, ergab die bereits erwähnte, frühe Untersuchung von Sauer und Gattringer (1985) bedeutsame Effekte der elterlichen Bildungsaspiration auf die Intelligenz und die Schulleistung (Faktor aus Zeugnisnote der 4. Volksschulklasse und Schulleistungstest) des Kindes. Während die Sozialschichtzugehörigkeit einerseits im starken Maß die Bildungsaspiration der Eltern beeinflusst, ergab die pfadanalytische Auswertung direkte Effekte der elterlichen Bildungsaspiration auf die Schulleistung. Auch wurden Effekte der Bildungsaspiration auf die Intelligenz und die Leistungsmotivation des Schülers gefunden (Sauer & Gattringer, 1985). Während die theoretische Erklärung für den Zusammenhang zwischen Bildungsaspiration und individuellen Schülermerkmalen in der Studie nahezu ausbleibt, zeigen die Befunde, dass nicht nur Mediationseffekte zwischen dem sozialen Status der Familie und den elterlichen Bildungsaspirationen zu bestehen scheinen, sondern diese auch über die individuellen Schülermerkmale vermittelt werden können. Auch Stocké (2010) prüfte in seiner Studie den Zusammenhang zwischen schulbezogenem Sozialkapital der Familie und dem Schulerfolg der Kinder. Mit Daten des Mannheimer Bildungspanels für die dritte und vierte Grundschul-

stufe lagen, neben Informationen zum Sozialstatus der Familie, Angaben von Lehrkräften zur Häufigkeit der Elternabendbesuche und dem Leistungsverhalten der Kinder und Angaben der Eltern zur Leistungsmotivation ihres Kindes sowie zur Häufigkeit der Elternabendbesuch vor. Die Häufigkeit der Elternabendbesuche wurden mit den Angaben der Lehrkräfte gemittelt und als Indikator für das schulbezogene Sozialkapital der Eltern verwendet. Zudem lagen standardisierte Testergebnisse der Schüler sowie deren Halbjahresnoten in Deutsch und Mathematik in der vierten Klasse vor. Der Einfluss des schulbezogenen Sozialkapitals auf die Notenvergabe in Deutsch und Mathematik erfolgte dabei in mehreren Schritten, ebenfalls wurde die zeitliche Entwicklung der Indikatoren berücksichtigt, die im Folgenden nicht thematisiert wird. Zentrale Befunde ergaben sich erstens durch Effekte der elterlichen Bildung auf die Notenvergabe, wobei die mütterliche Bildung stärker Einfluss nahm. Auch für die Deutschnote wurden insgesamt stärkere Effekte gefunden. Im Sinne der theoretischen Annahme zum Zusammenhang zwischen Struktur- und Prozessmerkmalen verkleinerte sich der direkte Effekt des elterlichen Sozialstatus, wenn der Schulkontakt der Eltern einbezogen wurde, der ebenfalls signifikante Effekte auf die Viertklassnoten ergab. Unter Kontrolle der individuellen Testleistung verringerte sich der direkte Effekt des schulbezogenen Sozialkapitals, blieb aber weiterhin für beide Fächer signifikant bestehen. Unter Berücksichtigung der schülerbezogenen Leistungsmotivation, der Unterrichtskonzentration und der Hausaufgabensorgfalt (die dritte Komponente nahm keinen zusätzlichen Einfluss auf die Benotung) wurde der direkte Effekt des schulbezogenen Sozialkapitals für das Fach Mathematik nicht mehr signifikant, für die Deutschnote blieb er weiterhin signifikant bestehen. Auch unter zusätzlicher Kontrolle der elterlichen Bildungsaspiration, die sich direkt auf die Leistungsbeurteilung in beiden Fächern auswirkt, blieb der direkte Effekt des schulbezogenen Sozialkapitals auf die Deutschnote nahezu unverändert. Neben dem markanten Ergebnis, dass der Elternabendbesuch unter Kontrolle leistungsbezogener Merkmale die Deutschnote in der vierten Klasse beeinflusst, unterscheidet sich dieser nur unmerklich zwischen den sozialen Gruppen. Sozialschichtspezifisch stärker beeinflusst waren dagegen die Testleistung, die Bildungsaspiration und das

schulbezogene Leistungsverhalten der Kinder (Stocké, 2010). Ob eine von der Lehrkraft eingeschätzte Bildungsaspiration eine stärkere Passung mit der Notenvergabe im Grundschulbereich vorweist, wurde dagegen bislang selten untersucht.

Welche Bedeutung kommt nun aber dem elterlichen Unterstützungsverhalten als weitere – den elterlichen Prozessmerkmalen zugeordnete – Einflussdeterminante der Schulleistung zu? Gemäß dem Modell von Helmke und Weinert (1997) ist diese Determinante den Strukturmerkmalen der Familie nachgelagert und beeinflusst sowohl direkt als auch indirekt über die kindlichen Merkmale die schulischen Leistungen. Im Sinne der Reproduktionstheoretischen Annahmen würde man davon ausgehen, dass die häusliche Unterstützung je nach sozialem Hintergrund der Schüler variiert und diese wiederum die Erfolgchancen der Kinder über eine bessere Benotung beeinflusst (Geißler, 2006). Dabei kann sich das elterliche Unterstützungsverhalten auf unterschiedliche Art und Weise äußern und Einfluss nehmen. Studien fokussieren dabei u.a. das elterliche Unterstützungsverhalten bei schulischen Übungen und Hausaufgaben und zeigen je nach einbezogener Unterstützungshäufigkeit, Art und erhobenen Zeitpunkten unterschiedliche Ergebnistendenzen (Helmke, Schrader & Hosenfeld, 2004). Helmke, Schrader und Hosenfeld (2004) heben hervor, dass auch gegenläufige Effekte plausibel sind: Einerseits könnten Eltern erst dann Unterstützung anbieten, wenn ihr Kind schlechte schulische Leistungen zeigt (negative Korrelationen). Andererseits könnte elterliches Unterstützungsverhalten auch eine lernförderliche Wirkung haben (positive Korrelationen). Welche Rolle die kindlichen Merkmale für das elterliche Unterstützungsverhalten unabhängig vom sozialen Status zeigen, untersuchten die Autoren auf Basis des Projekts SALVES, aus dem Daten von Schüler-, Eltern- und Lehrkraftfragebögen aus fünften und sechsten Klassen unterschiedlicher Schularten vorliegen. Analysiert wurde u.a. inwiefern das elterliche Unterstützungsverhalten (Kontrolle, Übung, Lernmanagement und Erklären) mit den schulischen Leistungen und individuellen Merkmalen des Kindes zusammenhängt. Die Ergebnisse weisen u.a. auf einen negativen Zusammenhang zwischen schulischer Leistung und

Unterstützungshäufigkeit hin; je besser die Schulleistung und je höher die Schulform, desto weniger Unterstützungsverhalten wurde gezeigt. Bezüglich des Zusammenhangs zwischen kindlichen Persönlichkeitsmerkmalen und elterlichem Unterstützungsverhalten zeigte sich, dass Eltern auch hier ihre Unterstützung eher kompensierend einsetzen; es wurden negative Zusammenhänge mit dem Fähigkeitsselbstkonzept und positive Zusammenhänge mit Konzentrationsstörungen festgestellt (Helmke et al., 2004)

Während die Studie von Helmke et al. (2004) aus Eltern- bzw. Schülerperspektive die Relevanz elterlicher Unterstützung für schulische Leistungskomponenten fokussiert, betrachtet die Analyse von Nölle, Hörstermann, Krolak-Schwerdt und Gräsel (2009) aus Lehrkraftperspektive relevante Leistungskomponenten. Mittels halbstandardisierter Interviews wurden Grundschullehrkräfte nach übergangsrelevanten Eigenschaften gefragt. Die Häufigkeits- bzw. Clusteranalysen zeigten, dass Lehrkräfte neben Leistungs-, Arbeits- und Sozialverhalten auch das elterliche Unterstützungsverhalten als relevant erachteten (Nölle et al., 2009). Inwiefern die Einschätzung elterlicher Unterstützung durch die Lehrkraft als Mediator für soziale Unterschiede in der Leistungsbewertung fungiert, untersuchte Stahl (2009) auf Basis der ersten Erhebung von Koala-S. Während der elterliche Sozial- und Bildungsstatus signifikant mit dem elterlichen Unterstützungsverhalten korrelierte, ergaben sich unterschiedliche Ergebnisse für die Vorhersage der Deutsch- und Mathematiknote. Unter Kontrolle individueller Leistungskomponenten (Kft, Testleistung) sowie dem Sozial- und Bildungsstatus der Familie, ergaben sich signifikante Ergebnisse für den Zusammenhang zwischen der Einschätzung elterlicher Unterstützung und der Deutschnote. Für die Mathematiknote ergaben sich keine Effekte (Stahl, 2009). Dagegen fanden Ditton und Krüsken (2006a) für die Übertrittsempfehlung am Ende der Grundschulzeit keine signifikanten Effekte des durch die Lehrkraft eingeschätzten elterlichen Unterstützungsverhaltens, wohingegen die Bildungsaspiration der Eltern signifikante Effekte ergab. Obwohl in diesem Fall nicht die Note des Schülers als abhängige Variable gewählt wurde, fokussierte diese Studie die Perspektive der Lehrkraft auf relevante Bildungsprozesse im

Grundschulbereich. Auch wenn Helmke und Schrader (2010) darauf hinweisen, dass die kausale Interpretation von Querschnittdaten zum Zusammenhang zwischen familiären Prozessmerkmalen und Schulleistung wenig über die Wirksamkeit kausaler Beziehungen aussagt, kann studienübergreifend festgehalten werden, dass jene Merkmale durchaus Effekte auf die Schulleistung ergeben, die den Zusammenhang zwischen familialen Strukturmerkmalen mediiieren.

Die dargestellten Befunde zu den ausgewählten Aspekten familiärer Prozessmerkmale ergeben einerseits direkte Effekte auf die schulischen Leistungen des Kindes, weisen aber auch auf vermittelnde Effekte zwischen familiären Struktur- und Prozessmerkmalen, individuellen Schülermerkmalen und Schulleistung hin. Die Befunde zu den strukturellen Einflussdeterminanten ergeben ebenfalls sowohl direkte, als auch indirekte, über die Schülermerkmale vermittelte Effekte auf die Schulnoten. Für die familialen Prozessmerkmale sind insbesondere die Bildungsaspirationen der Eltern, aber auch deren Kontakt zur Schule sowie deren Unterstützungsverhalten für die Schulnotenvergabe relevant. Zudem zeigen die Ergebnisse unterschiedliche Effekte für die Deutsch- und Mathematiknotenvergabe. Bislang wenig umfassend analysiert wurde dagegen die Bedeutung des von der Lehrkraft eingeschätzten elterlichen Verhaltens für die – ebenfalls von der Lehrkraft erfassten – Schulnoten im Grundschulbereich. Im Sinne der Coleman'schen These müssten die sozialen Herkunftseffekte vollständig durch die sozialen Kapitalien der Familie erklärt werden. Ist dies nicht der Fall und bleibt unter Kontrolle der Schülereigenschaften bestehen, verweist Stocké (2010) auf die Rolle weiterer schulischer bzw. lehrkraftspezifischer Determinanten, die im Folgenden einführend vorgestellt werden. Ohne weiter auf die Bedeutsamkeit der gefundenen Zusammenhänge für gesellschaftliche Ungleichheitsprozesse einzugehen, bleibt die Frage, inwiefern die gefundenen Zusammenhänge zwischen familiären Merkmalen und Schulleistungskomponenten schulrechtlich legitimiert werden bzw. welche Rolle das Schulsystem selbst in diesem Zusammenhang einnimmt und zukünftig einnehmen könnte.

## 4.4 Schulische Bedingungsfaktoren von Schulleistung

Da die theoretischen Annahmen zur multikausalen Bedingtheit von Schulleistung nicht nur Faktoren auf Seiten des Kindes und dessen Familie betrachten, kommt diese Arbeit nicht umhin, ausschnitthaft schulische Bedingungsfaktoren von Leistung zu fokussieren. Jene, als distale Faktoren bezeichneten Merkmale, können das Kind unabhängig und zusätzlich von seinen individuellen Hintergrundmerkmalen in seinen Leistungen beeinflussen und damit zu Bildungsungleichheiten beitragen. Während die vorausgegangene Betrachtung individueller und familiärer Bedingungsfaktoren immer wieder die Relevanz der Lehrkraftperspektive auf die Komponenten von Schulleistung fokussierte, beschäftigt sich die nachfolgende Ausführung explizit mit der Frage, welche Bedeutung die Institution Schule selbst und die darin tätigen Akteure für die Bewertung schulischer Leistung einnehmen. Fokussiert werden dafür, neben Faktoren der Lehrkraft, die Bedeutung der Klassenzusammensetzung; auch wird betrachtet, inwiefern in unterschiedlichen Schulfächern differente Leistungskomponenten gewichtet werden. Faktoren des Unterrichts sowie der (länderspezifischen) Schulentwicklung werden dabei nicht aufgegriffen.

### **Faktoren und Bewertungsmechanismen auf Seiten der Lehrkraft**

Der Beschäftigung mit Faktoren auf Seiten der Lehrkraft liegt im Folgenden die Frage zu Grunde, welche Faktoren die Beurteilungsprozesse einer Lehrkraft beeinflussen und durch welche diese geleitet werden können. Wie in [Kapitel 2](#) der vorliegenden Arbeit deutlich wurde, beinhaltet der schulische Bewertungsprozess mannigfache Aufgaben und Entscheidungssituationen, in welchen als Leistung anerkannte Verhaltensergebnisse von der Lehrkraft bewertet werden. Dabei sind u.a. zwei Entscheidungen zu treffen<sup>35</sup>: der gewählte Vergleichsmaßstab bei der Einordnung individueller Leistungen sowie der Einbezug von Informationen bzw. deren Gewichtung in der Notengebung.

---

35 Inwiefern die Lehrkraft diese Entscheidung selbst fällen muss oder im Austausch mit dem Lehrerkollegium fällt und ob Vorgaben der Schulleitung vorliegen, kann in schulrechtlichen Vorgaben nicht festgemacht werden und erfolgt vermutlich individuell unterschiedlich.

Häufig wird bei dem Vergleich zwischen Notengebung und messtheoretischen Gütekriterien die Lehrkraft als Ursache der Abweichung zwischen Testleistung und Noten angesehen. Betrachtet man die gegebene Schulnote aber als ein komplexeres, aus vielen Einzelinformationen zusammengesetztes Konstrukt, sind Fragen der Einflussnahme weniger leicht aufzudecken. Fest steht, dass die subjektive Erfassung von Kompetenzen immer der Wahrnehmung und Bewertung Seitens der Lehrkraft unterliegt, womit sich die Notengebung durch einen sozial strukturierten Wahrnehmungs- bzw. Bewertungsprozess auszeichnet. Jenem Prozess schließt sich dann eine Förderung und Ausbildung der festgestellten Kompetenzen an, an welchen Lehrkräfte und Eltern maßgeblich beteiligt sind. Diskutiert wird dabei, ob jene Prozesse jemals herkunftsunabhängig sein können, da eine Förderung und Bewertung immer von einer Vorleistung ausgeht, die bekanntlich sozial beeinflusst wird (Solga, 2009). Während vornehmlich in den 1970er Jahren die Rolle der Lehrkraft im Benotungsprozess diskutiert wurde, sind es in der aktuellen Forschung meist qualitative Ansätze, die sich explizit mit Bewertungsprozessen einer Lehrkraft beschäftigen sowie darüber hinaus häufig der prognostische Wert der Bewertung fokussiert wird (Birkelbach, 2011; Höhn, 1980; Ingenkamp, 1976a; Kalthoff, 1996; Pohlmann-Rother, 2010; Wagner, 1980).

Bei Betrachtung des Lehrberufs lassen sich nach Terhart (2011) in der deutschen Erziehungswissenschaft derzeit drei Ansätze zum professionellen Lehrerhandeln unterscheiden. Während der strukturtheoretische Ansatz die strukturellen, teils widersprüchlichen Aufgaben an den Lehrberuf fokussiert, betrachtet der berufsbiografische Bestimmungsansatz die berufsbiografischen Entwicklungsprozesse im Lehrberuf. Der dritte, kompetenztheoretische Ansatz beschäftigt sich explizit mit den individuellen Voraussetzungen der Lehrkraft, die sich bspw. durch Überzeugungen, Fakten, Einstellungen und Handlungsrouninen auf die Lernerfolge von Schülern auswirken und im Sinne der Lehrerprofessionalität förderbar sind (Terhart, 2011).

Theoretische Ansätze zur Erklärung und Effekten von Lehrerhandeln zeichnen sich im deutschen und angelsächsischen Raum durch eine

Vielzahl an theoretischen Konzepten aus, deren Abgrenzung voneinander nicht immer trennscharf ist (Hofer, 1986; Höhn, 1980; Pajares, 1992; Rosenthal & Jacobson, 1971). Eine übergeordnete Gemeinsamkeit ist wohl der Bezug zu subjektiven Theorien, welche sozial strukturierte Handlungsweisen fokussieren und sich mit Fragen nach dem *Warum* von Lehrerhandlungen beschäftigen (Sander, 1997). Während sich das Konzept des *Teachers Beliefs* auf zeitlich stabile Vorstellungen bspw. zur Gestaltung der Lernumgebung, Entscheidungen und Kommunikationsformen bezieht, wird bei Bewertungssituationen häufig der Bezug zu impliziten Persönlichkeitstheorien hergestellt (Sander, 1997; Stahl, 2009). Studien zu sozialen Urteilsprozessen beschäftigen sich u.a. mit der Frage, wie sich persönliche Einstellungen und Erfahrungen auf die Wahrnehmung und Beurteilung anderer Personen auswirken können. Jene subjektiven, naiven, alltagspsychologischen Annahmen werden unter den impliziten Persönlichkeitstheorien zusammengefasst und vereinfachen soziale Entscheidungsprozesse (Hofer, 1986; Sander, 1997). Vereinfachung erfolgt z.B. durch Stereotypisierung. Demzufolge werden übermäßig viele Informationen durch subjektive Wahrnehmungen vereinfacht, um kritische oder schnell zu meistern Situationen zu bewältigen, können gleichzeitig aber auch zu Verzerrungen der Realität führen, die Vorurteile und Fehleinschätzungen beinhalten (Sander, 1997). Stereotypisierte Erwartungshaltungen führen dazu, von einzelnen Verhaltensmerkmalen aus unterschiedlichen Bereichen (z.B. Arbeits-, Sozial- und Leistungsverhalten) auf die Gesamtpersönlichkeit des Schülers zu schließen, wodurch guten bzw. schlechten Schülern entsprechende Attribute zugeschrieben werden (Wagner, 1980). Als eine der bekanntesten deutschen Studien zu sozialen Beurteilungsprozessen analysiert Höhn (1980) implizite Persönlichkeitstheorien von Lehrkräften und deren Stereotypen von guten bzw. schlechten Schülern. Eine stereotype Zuschreibung erfolgte dann, wenn auf Grundlage von wenigen, direkten Interaktionsprozessen auf spezifische Eigenschaften eines Individuums geschlossen wurde, die auf verschiedene Situationen übertragen werden sowie affektive und moralische Färbungen beinhalten (Hurrelmann, 1975). Mittels freier Beschreibungen von Lehrkräften zu Eigenschaften von guten bzw. schlechten Schülern kommt die Studie zu dem Ergebnis, dass

der Begriff des schlechten Schülers in der Praxis ein Konglomerat aus ungünstigen Charaktereigenschaften beinhaltet, die sich auf Wissens- und Begabungsmängel beziehen und eine schlechte soziale Anpassung beinhalten (Höhn, 1980).

Aspekte der Attribution von gezeigtem Schülerverhalten befassen sich wiederum mit der kausalen Erklärung von Verhaltensweisen. Unterschieden wird zwischen zeitlich stabilen und variablen sowie internalen und externalen Verhaltensverankerungen, die das Verhalten der Lehrkraft steuern. Demnach wird das schlechte Leistungsergebnis eines Schülers, der sonst als begabt gilt, von der Lehrkraft als zufälliges Ereignis deklariert und durch den Mangel an Fleiß oder Anstrengung erklärt. Der Erfolg eines wenig begabten Kindes wird dagegen externalen, variablen Ursachen wie Glück zugeschrieben (Dann, 2008). Effekte dieser Verhaltenszuschreibung sowie Folgen, die im Sinne einer Self-Fulfilling Prophecy aus impliziten und expliziten Bewertungen entstehen, werden in diesem Zusammenhang nicht weiter ausgeführt<sup>36</sup>. Ebenso werden theoretische Argumentationen zu Fehlerquellen in der Bewertungspraxis sowie Aspekte der institutionellen Diskriminierung nicht vertieft, da diese nicht im Fokus jener Abhandlung stehen.

Studien zur Urteilsgenauigkeit von Lehrkräften beschäftigen sich hauptsächlich mit der diagnostischen Kompetenz bei Bewertungsprozessen (z.B. Gräsel et al., 2010; Spinath, 2005). Ein übergreifendes Ergebnis dieses Forschungsstrangs ist die Tendenz, dass Lehrkräfte die Schüler ihrer Klasse relativ gut in eine Rangreihenfolge bringen können, Einschätzungen des individuellen Leistungsniveaus sowie Vergleiche mit anderen Klassen aber von den zusätzlich erfassten Leistungsmaßen abweichen (Strietholt & Bos, 2010). Studien zur Diskrepanz der Bewertung von gleichen Leistungen durch unterschiedliche Lehrkräfte wurden bereits im Kapitel zur messtheoretischen Güte von

---

<sup>36</sup> Welche Relevanz Lehrerurteile im Lebensverlauf einnehmen können, diskutiert Birkelbach (2011) in seinem Artikel unter der Frage, ob Leistungsurteile valide Prognosen- oder im Sinne der Self-Fulfilling Prophecy eine selbsterfüllende Prophezeiung darstellen.

Schulnoten fokussiert (vgl. [Kap. 3.1](#) – [Kap. 3.3](#)) und werden an dieser Stelle nicht betrachtet. Weiterführend offen bleibt bei all diesen Erkenntnissen zum Zusammenhang zwischen Lehrkraft und Schulnoten die Frage, welche Komponenten unterschiedliche Lehrkräfte in einer Schulnote bewerten und ob sich in diesen Komponenten intersubjektive Gemeinsamkeiten finden lassen. So ist davon auszugehen, dass kindliche Merkmale durch die Wahrnehmung der Lehrkraft gleichzeitig immer eine subjektive Färbung beinhalten, wodurch sich Diskrepanzen in den erfassten Perspektiven ergeben können.

Geht man weiterführend davon aus, dass die Institution Schule selbst durch gesellschaftliche Erwartungen und Wertvorstellungen geprägt wird, unterliegt auch die Lehrkraft solchen Normvorstellungen. Demnach wird der Schule eine Nähe zur Mittelschicht zugesprochen, die Schüler aus dieser Gesellschaftssicht begünstigen. Demzufolge entspricht der Sozialcharakter von Schülern aus bildungsnäheren Schichten eher den schulischen Anforderungen, durch den sie schulinternen Verhaltenserwartungen besser entsprechen können und bessere Leistungen erzielen (Ditton, 2011; Rolff, 1997). Effekte der sozialen Herkunft werden in diesem Zusammenhang dahingehend diskutiert, ob es sich dabei um eine gezielte Diskriminierung des Lehrpersonals handelt oder ob jener Mechanismus – im Sinne einer institutionellen Diskriminierung – in schulischen Strukturen angelegt ist (Ditton, 2011; Gomolla & Radtke, 2009). Neben Fragen nach Fehlerquellen und systemgeprägten Determinanten bei der Benotung einer Schülerleistung wird als weiterer Einflussfaktor häufig der klasseninterne Bezugsrahmen diskutiert. Angemerkt wird, dass in verschiedenen Klassen nicht nur Abhängigkeiten von der Lehrkraft eine Rolle spielen, sondern auch der Klassenkontext selbst Einfluss auf die individuelle Leistung nimmt. Dies gilt es im Folgenden zu betrachten.

### **Faktoren auf Seiten der Klasse: Referenzgruppeneffekte**

Effekte des Kontextes auf die Bildungsbeteiligung, den Kompetenzerwerb und deren Relevanz für bestehende Bildungsungleichheit sind bereits lange Zeit Gegenstand der empirischen Sozialforschung (Ditton, 2013b; Dumont, Neumann, Maaz & Trautwein, 2013;

Ingenkamp, 1976a). Kontexteffekte beziehen sich auf die Frage, inwiefern Individuen durch spezifische, sie umgebende Kontexte beeinflusst werden, die über die individuellen Merkmale hinauswirken. Sie können sich auf spezifische Merkmale und Ebenen der Umgebung beziehen (global, analytisch, strukturell) und u.a. zu Interaktionseffekten zwischen individuellen und kontextuellen Einflüssen führen (Ditton, 2013b). Während sich Kontexte sowohl auf familiäre, nachbarschaftliche und regionale Effekte beziehen können, die im weiteren nicht weiter vertieft werden, umfassen sie auch institutionelle Rahmenkontexte, wie unterschiedliche Schulformen und Lernumwelten. Eine weitere Spezifizierung im Schulkontext bezieht sich auf Effekte, die sich aus der Zusammensetzung eines Kontextes für die Leistungsentwicklung eines Schülers ergeben und als Kompositionseffekte bezeichnet werden.

Kompositionsmerkmale werden in der Forschung vornehmlich nach drei Arten unterschieden: nach der sozialen, ethnischen und leistungsbezogenen Gruppenzusammensetzung (Ditton, 2013b; Dumont et al., 2013). Angenommen wird, dass sich durch die unterschiedlichen Lernangebote, Schülerzusammensetzungen und Förderkulturen differenzielle Lern- und Entwicklungsmilieus ergeben, die sich in der leistungsbezogenen Selbst- und Fremdbewertung von Schülern widerspiegeln (Ditton, 2013b). Bei der Prüfung von Kompositionseffekten wird in der Empirie häufig von Referenzgruppeneffekten oder Big-Fish-Little-Pond Effekten gesprochen sowie überwiegend die leistungsbezogene Klassenzusammensetzung in ihrer Wirkweise überprüft wird (Baeriswyl, Wandeler & Trautwein, 2011; Dumont et al., 2013; Marsh, 2005). Empirisch unterschiedlich gehandhabt werden dabei Fragen der Interpretation, methodischen Herangehensweise und Erklärung von postulierten Zusammenhängen und umfassen einige Anforderungen bei deren Umsetzung (Definition und Abgrenzung der Kontexte, Erklärung der Art und Wirkweise von erwarteten Effekten etc.) (Ditton, 2013b). In der Empirie beziehen sich Kompositionseffekte im Schulkontext meist auf aggregierte Individualmerkmale, die auf Klassen- und oder Schulebene – zusätzlich zur individuellen Voraussetzung – Einfluss auf die Leistung des Schülers nehmen und

bestmöglich längsschnittlich erhoben werden. Demnach würde man von einem Kompositionseffekt sprechen, wenn zwei Schüler mit gleichen Leistungen in unterschiedlichen Schulklassen unter Kontrolle weiterer Individualmerkmale unterschiedliche Lernfortschritte und Bewertungen erzielen (Dumont et al., 2013). Neben Effekten der Leistungszusammensetzung werden häufig auch Fragen der sozialen Klassenzusammensetzung diskutiert, die im Folgenden zusätzlich zur Leistungszusammensetzung fokussiert werden.

Bezüglich der Effekte aggregierter Merkmale einer Klasse weist die Befundlage also auf die Relevanz leistungsbezogener Gruppenmerkmale hin. Häufig hervorgehoben wird dabei die Relevanz des aggregierten, kognitiven Fähigkeitsniveaus, das über die Individualebene hinaus Einfluss auf Testleistungen und Fähigkeitsselbstkonzepte der Schüler nimmt (Marsh, 2005; Tiedemann & Billmann-Mahecha, 2004, 2007a). Zusätzlich dazu wurden Effekte eines hohen Klassenanteils von bildungsfernen sozialen Gruppen auf die Individualleistungen festgestellt, hier ist die Befundlage jedoch weniger konsistent und erst unter Bezug auf die jeweilig kontrollierten Faktoren zu interpretieren (Ditton & Krüsken, 2006b; Schümer, 2004). Während diese exemplarischen Befunde den Effekt von Kontextmerkmalen auf die individuelle Leistung eines Kindes fokussieren, müssen sie wiederum von Ergebnissen differenziert werden, die Effekte der Klasse auf die Benotungspraxis der Lehrkraft analysieren. So verweist bereits Ingenkamp (1976b) auf Effekte des klasseninternen Bezugssystems, das sich auf die Benotung von Schülern auswirken kann. Nach seinen Ergebnissen werden leistungsgleiche Schüler in unterschiedlich leistungsstarken Klassen different bewertet; die Benotung in der leistungsstärkeren Klasse also schlechter ausfällt als in der leistungsschwächeren Bezugsgruppe und Schulnoten somit über eine Klasse hinweg nicht vergleichbar sind (Ingenkamp, 1976b).

Auch aktuell wurden ähnliche Referenzgruppeneffekte im Klassenkontext gefunden. So deutet auch die Studie von Trautwein und Baeiswyl (2007) auf Referenzgruppeneffekte am Übertritt in die Sekundarstufe I hin. Unter Kontrolle individueller Leistungen fanden sich

u.a. negative Effekte der mittleren Klassenleistung auf die Einschätzung der Lehrkraft zu Schulnoten und kognitiven Fähigkeiten der Schüler. Bei gleicher individueller Testleistung wurden Schüler in leistungsstärkeren Klassen also weniger positiv benotet. Zu ähnlichen Effekten kommt auch die Studie von Trautwein, Lüdtke, Marsh, Köller und Baumert (2006), die für den Sekundarschulbereich ebenfalls auf einen Zusammenhang zwischen Schulnoten und mittlerem Testleistungsniveau der Klasse verweisen. Im Grundschulbereich analysierten Treutlein, Roos und Schöler (2008) für die dritte Jahrgangsstufe den Zusammenhang zwischen dem Leistungsniveau der Klasse und den Lese-, Rechtschreibe- und Deutschnoten der Kinder. Mithilfe von random-intercept-Modellen kamen sie für diese drei Noten zu unterschiedlichen Ergebnissen. Während die Rechtschreibnote bei gleicher Leistung in rechtschreibeschwächeren Klassen um .36 besser ausfiel, ergaben sich für die Leseleistung keine Einflüsse der aggregierten Klassenleistung. Für die Deutschnote ergeben sich ähnliche Effekte. Sie wird negativ von der durchschnittlichen Rechtschreibleistung beeinflusst, während sich die Lesefertigkeit der Klasse nicht auf die Deutschnote auswirkt; insgesamt sind die Effekte der Studie allerdings relativ schwach (Treutlein et al., 2008). Ergebnisse bezüglich Effekten der sozialen Komposition auf die Benotung sind im Grundschulbereich rar. Bisher liegen nur ausgewählte Ergebnisse für die Sekundarstufe vor. So fanden Maaz, Neumann, Trautwein, Wendt, Lehmann und Baumert (2008) für Sechstklässler zunächst schwache, negative Effekte der mittleren sozioökonomischen Klassenzusammensetzung. Schüler aus ähnlichen sozialen Verhältnissen wurden in sozial günstigeren Klassenzusammensetzungen unter Kontrolle von Individualmerkmalen weniger gut bewertet. Der Effekt der mittleren sozioökonomischen Klassenzusammensetzung war allerdings stark mit der leistungsbezogenen Klassenzusammensetzung konfundiert und ging unter Kontrolle dieser zurück. Ähnliches ergab sich für die zusätzlich analysierte Lernkompetenzeinschätzung der Lehrkräfte (Maaz et al., 2008).

Die exemplarisch aufgeführten Befunde zu Effekten der Klassenzusammensetzung auf die individuelle Leistung eines Schülers lassen

Fragen bezüglich der Relevanz für die Benotungspraxis in der Grundschule offen. Wenn man gemäß den Befunden davon ausgeht, dass sich Merkmale der Klasse auf die Benotung auswirken, kommt gerade im gegliederten Schulsystem der Schulnote am Ende der Grundschulzeit aufgrund ihrer Bedeutung für die weitere Bildungslaufbahn eine bedeutsame Rolle zu. Differenzen in der Benotung in verschiedenen Klassen bei gleichen individuellen Testleistungen deuten darauf hin, dass es für die Bestandteile einer Note zusätzlich zur individuellen und familiären Lernvoraussetzung darauf ankommt, welcher Schulklasse und Lehrkraft der Schüler zugeordnet wird. Diese Referenzgruppeneffekte bzgl. der Leistung und sozialen Klassenzusammensetzung gilt es für den Grundschulbereich zu prüfen.

### **Fachspezifische Einflussfaktoren**

Fragen nach fachspezifischen Differenzen in der Benotungspraxis ergeben, dass es durchaus Unterschiede sowohl in der Gewichtung von Leistung innerhalb eines Fachs als auch zwischen den verschiedenen Fächern gibt. Die strukturelle Verteilung in Haupt- und Nebenfächer sowie die Aufteilung in Wochenstunden und Zensierungsstrenge obliegen der geschichtlichen Entwicklung der Bildungskonzeption und werden im Folgenden nicht weiter ausgeführt (Ziegenspeck & Lehmann, 1999). Eine unterschiedliche Bewertungsstrenge wurden vor allem zwischen Haltungsfächern (Arbeits-, Sozialverhalten) und Hauptfächern festgestellt, wobei in höheren Schulstufen und in Fächern mit hohem schriftlichen, quantifizierbaren Arbeitsaufwand (Deutsch, Mathematik, Fremdsprachen) generell strenger bewertet wird (Ingenkamp, 1995b). Für Sechstklässler fand Lintorf (2012) tendenziell Unterschiede zwischen Komponenten der Mathematik- und Sachkundenote. Während für die Mathematiknote der stärkste Einzelprädiktor die kognitiven Grundfähigkeiten des Schülers waren, ergab sich für die Sachkundenote die Gewissenhaftigkeit als stärkster Einzelprädiktor sowie insgesamt den Sekundärtugenden in der Sachkundenote eine vergleichsweise größere Bedeutung zukam. Empirische Forschungsbefunde, die explizit Unterschiede in den Benotungskomponenten zwischen Hauptfächern im Grundschulbereich analysieren, sind bislang selten. Wenn als aufzuklärende Leistungskomponente

nicht Durchschnittsnoten herangezogen werden, wird selten diskutiert, ob und weswegen von unterschiedlichen Notenkomponenten ausgegangen wird. Für eine Berliner Grundschulstichprobe analysierten Maaz, Baeriswyl und Trautwein (2011) den Einfluss individueller Schülermerkmale für die Notenvergabe am Ende der Grundschulzeit. Für den Vergleich zwischen Deutsch- und Mathematiknote klärt die kognitive Fähigkeit mehr Varianz in der Mathematiknote auf als im Fach Deutsch. Dagegen klärte das Geschlecht deutlich mehr Varianz in der Deutschnote auf als in Mathematik.

Im Sinne der messtheoretischen Ansprüche an die Notenzusammensetzung wäre das Konstrukt einer Note valide (vgl. [Kap. 3.3](#)), wenn die Mathematiknote am besten durch die Mathematikleistung des Schülers vorhergesagt werden könnte und dagegen dessen Rechtschreibung weniger relevant wäre. Von einer niedrigen Konstruktvalidität würde man bspw. sprechen, wenn bei der Deutschnote die äußere Form und der Fleiß des Schülers gemessen werden würde (Ingenkamp, 1976a). Bezüglich der in [Kapitel 3.1](#) angesprochenen Auswertungsobjektivität sei ebenso nochmals auf den übergreifenden Befund verwiesen, dass die Zusammensetzung einer Schulnote je nach Fach variieren kann. Exemplarisch angeführt seien an dieser Stelle nochmals die frühen Befunde von Starch und Elliot (1976), die anfänglich davon ausgingen, dass die Bewertung sprachbezogener Fächer mehr durch subjektive Faktoren beeinflusst wird; ihre Ergebnisse dann allerdings auf eine ähnliche Variabilität in Mathematik- und Sprachnoten hinwiesen (Starch & Elliot, 1976). Da diese und aktuelle Befunde vornehmlich für den Sekundarstufenbereich vorliegen, überprüft die vorliegende Untersuchung die Relevanz fachspezifischer Notenkomponenten für zwei Hauptfächer im Grundschulbereich. Da die Befunde zu Deutsch- und Mathematiknoten häufig eher implizit Unterschiede in der Zusammensetzung aufzeigen, ist diese Fragestellung eher von explorativer Natur. Fokussiert wird die Frage, ob auch für den Grundschulbereich (keine) Unterschiede in der Notenzusammensetzung zwischen dem eher subjektiv konnotierten Fach Deutsch und dem eher objektiv wirkenden Fach Mathematik gefunden werden.

## 4.5 Eigenes Gesamtmodell zu Komponenten von Schulleistung

Im Folgenden werden die bisher vorgestellten theoretischen Zusammenhänge und Forschungsbefunde zu Bedingungsfaktoren von Schulnoten in einem eigenen theoretischen Gesamtmodell zusammenfassend festgehalten. Expliziert wird, von welchen Zusammenhängen in diesem theoretischen Gesamtmodell ausgegangen wird, um anschließend die darin postulierten Zusammenhänge einer empirischen Überprüfung zu unterziehen.

Während die theoretischen Modelle zu Determinanten von Schulleistung häufig einen expliziten Bezug zu Schulnoten vermeiden und stattdessen von Schulleistung (Tests vs. Noten), Bildungserfolg oder Kompetenzerwerb sprechen, wurde in den vorausgegangenen Kapiteln der Fokus auf Noten in der Grundschule gelegt und betrachtet, welche Einflüsse sich auf diese ergeben können.

Hauptaugenmerk lag erstens auf der Frage, zu welchen Anteilen individuelle Schülermerkmale in die Notenzusammensetzung einfließen. Zusätzlich dazu galt es, aus bildungssoziologisch geprägter Forschungsperspektive den Zusammenhang zwischen familiären Hintergrundmerkmalen, individuellen Schülermerkmalen und Schulnoten zu betrachten.

Bei der Vielfalt an Einflussmerkmalen auf Seiten des Schülers wurde in der vorliegenden Abhandlung nach kognitiven und nicht-kognitiven Prädiktoren unterschieden, die kausale Zusammenhänge zu Schulnoten vorweisen und auch im Gesamtmodell so modelliert werden. Zweitens wurden im Sinne Hellers (1997) die moderierenden Wechselwirkungen zwischen kognitiven und nicht-kognitiven Schülermerkmalen im Modell aufgenommen. Bilanzierend gilt in den bisherigen Forschungsbefunden die Testleistung des Schülers als stärkste Einflusskomponente der Schulnote. Offen bleibt dabei häufig die Frage, welche Zusammenhänge die objektiv erhobenen Testleistungen mit den von der Lehrkraft eingeschätzten Merkmalen aufweisen. Im

Sinne einer diagnostisch validen Einschätzung wird von einem positiven Zusammenhang zwischen Einschätzungen, Testleistungen und Schulnoten ausgegangen. Gleiches gilt für die mittels Test erhobene Intelligenz bzw. kognitive Grundfähigkeit der Schüler.

Bezüglich des kindlichen Vorwissens wird von einem direkten Effekt auf die zeitlich nachfolgenden Leistungserhebungen ausgegangen. Fokussiert wird im Gesamtmodell der direkte Effekt der von der Lehrkraft zeitlich vorausgegangenen Einschätzung auf die Notenzusammensetzung am Ende der Grundschulzeit.

Bezüglich sprachlicher Einflussdeterminanten wird zudem von einem direkten Effekt der von der Lehrkraft eingeschätzten sprachlichen Kompetenzen auf die Schulnoten ausgegangen. Während unabhängig von migrationsspezifischen Fragestellungen vornehmlich die Lese- und Rechtschreibleistung im Zusammenhang mit Schulnoten untersucht wurde, existieren bislang nur wenige Erkenntnisse zum Zusammenhang zwischen Lehrkräfteeinschätzungen zum Sprachstand der Schüler und deren Noten.

Auch ist der Zusammenhang zwischen Begabungskomponenten und Schulnoten weniger eindeutig. Häufig mit Intelligenz gleichgesetzt bleibt weiterhin offen, in welchem Zusammenhang die Begabungseinschätzung der Lehrkraft mit sozialen Hintergrundmerkmalen der Familie und anderen individuellen Schülermerkmalen steht. Im Modell wird davon ausgegangen, dass die Begabungseinschätzung einen direkten Effekt auf die Notenvergabe aufweist.

Im Sinne der schulrechtlichen Argumentation aus [Kapitel 2.5](#) werden im vorliegenden Gesamtmodell als nicht-kognitive Einflussdeterminanten das Arbeits- und Sozialverhalten des Kindes fokussiert. Zunächst werden in den nachfolgenden Analysen die direkten Effekte jener Merkmale auf die Schulnoten geprüft, wobei gemäß schulrechtlicher Grundlagen davon auszugehen ist, dass das Sozialverhalten keine Relevanz für die Notenzusammensetzung hat. Zudem wird im Sinne Solgas (2009) davon ausgegangen, dass kognitive Merkmale allein

nicht zum Bildungserfolg führen, sondern zusätzlich eine entsprechende Anstrengung bzw. Fleiß des Schülers vorliegen muss. Folglich wird unter Bezug auf das Modell von Heller (1995) angenommen, dass der Zusammenhang zwischen kognitiven Schülermerkmalen und Schulleistung durch nicht-kognitive Merkmale kovariiert wird. Höhere Werte in den beiden Moderatorvariablen Arbeits- und Sozialverhalten also zu engeren Prädiktor-Kriteriums-Zusammenhängen führen. Ebenso wird davon ausgegangen, dass sich zwischen den erfassten, individuellen Schülermerkmalen Wechselwirkungen ergeben bzw. die kognitiven Merkmale die nicht-kognitiven Merkmale beeinflussen. Zudem wird das Geschlecht des Schülers als Kontrollvariable in den Analysen berücksichtigt.

Im Sinne der Modelle von Helmke und Weinert (1997) und Heller (1995) und unter Bezug auf exemplarische Kernbefunde wird angenommen, dass sich Effekte der familiären Struktur- und Prozessmerkmale auf die Notenzusammensetzung ergeben, die Notenvergabe durch die Lehrkraft also nicht unabhängig von dem sozialen Hintergrund der Schüler erfolgt.

Im Gesamtmodell wird postuliert, dass der Effekt des sozioökonomischen und bildungsbezogenen Status der Familie nicht nur direkt, sondern auch über Vermittlungsmechanismen zum Tragen kommt. Als ein Vermittlungsindikator werden ausgewählte Prozessmerkmale der Familie fokussiert, die sich sowohl direkt als auch vermittelt über die individuellen Schülermerkmale auf die Schulnote auswirken. Während das Modell von Baumert, Watermann und Schümer (2003) als Prozessmerkmale die kulturelle Praxis der Familie analysiert, fokussiert das vorliegende Modell im Sinne Colemans (1988) das schulbezogene Sozialkapital der Familie. Angenommen wird, dass sich das strukturelle Kapital der Familie über das elterliche Förder- und Erwartungsverhalten auf die schulbezogenen Eigenschaften des Kindes und dessen Leistung auswirkt. Konkret wird auf Seiten der Prozessmerkmale die Bedeutung der Lehrkräfteeinschätzung zur elterlichen Bildungsaspiration, zum elterlichen Unterstützungsverhalten sowie zum Kontakt zwischen Elternhaus und Schule fokussiert.

Als zweiter Vermittlungsmechanismus zwischen familiären Hintergrundmerkmalen und Schulnoten wird überprüft, inwiefern die individuellen Schülermerkmale sozialschichtspezifisch determiniert sind. Angenommen wird, dass der direkte Einfluss familiärer Merkmale unter Einbezug der Schülermerkmale nahezu verschwindet, die Einschätzung durch die Lehrkraft also von den familiären Hintergrundmerkmalen beeinflusst wird.

Als dritter Determinantenbereich wird unter Bezug auf das Modell von Helmke und Weinert (1997) geprüft, ob sich unter Kontrolle der auf Individualebene erfassten Schülermerkmale zusätzliche Referenzgruppeneffekte der leistungsbezogenen und sozialen Klassenzusammensetzung auf die Schulnotenvergabe ergeben.

Nahezu explorativer Natur ist außerdem für den Grundschulbereich die Frage der fachspezifischen Notenzusammensetzung, hierzu werden die zwei, für den Grundschulübertritt relevanten Schulnoten in Mathematik und Deutsch auf eventuelle Unterschiede geprüft. Insgesamt wird mit dem nachfolgenden Gesamtmodell – unter Bezug auf Heller (1995) – der multikausalen Bedingungsstruktur dadurch Sorge getragen, dass unterschiedliche diagnostische Informationsquellen in die Analysen miteinbezogen werden. Damit ergibt sich das nachfolgende Gesamtmodell, die eingezeichneten Pfeile entsprechen den postulierten Zusammenhängen. Bevor auf die methodische Umsetzung und Operationalisierung der Konstrukte eingegangen wird, werden im folgenden Kapitel die Ziele und Fragestellungen mitsamt der abgeleiteten Hypothesen vorgestellt.

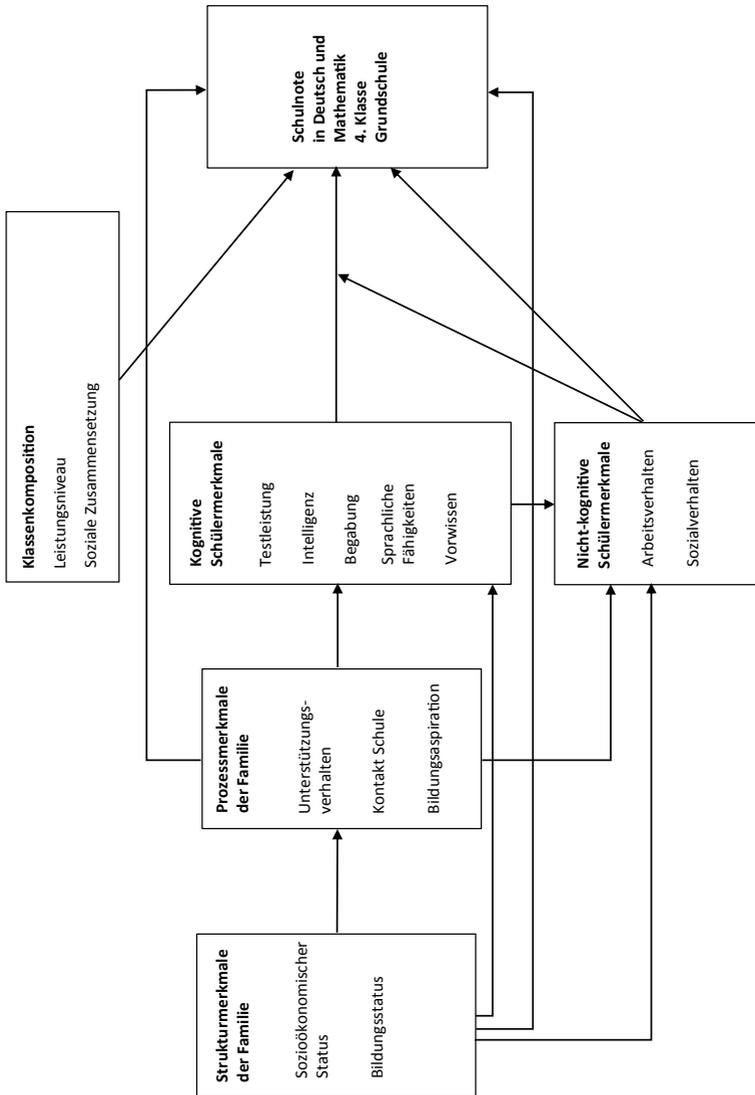


Abbildung 3: Eigenes theoretisches Gesamtmodell zu Komponenten von Schulnoten

Anmerkungen: Einbezogene Kontrollvariablen, wie das Geschlecht des Schülers, werden nicht eigens in der theoretischen Modellierung betrachtet, finden aber in den empirischen Analysen Berücksichtigung.



## 5 Ziele und Fragestellungen

Die theoretische Basis der vorliegenden Arbeit zeigt, dass sich eine Vielzahl von theoretischen Ansätzen und empirischen Untersuchungen mit den Bedingungen ungleicher Schulleistung beschäftigen, häufig untersucht am Übertritt in die Sekundarstufe I. Eine umfassende Erklärung zur Existenz von Bildungsungleichheiten im Schulsystem und jenen zugrundeliegenden schulischen Vermittlungsmechanismen scheint jedoch weiterhin ein nicht vollständig erschlossenes Forschungsfeld. Die wenig eindeutige Vorhersagekraft schulischer Leistungen am Ende der vierten Klasse und die Tatsache, dass unzählige Faktoren – zusätzlich zu und unabhängig von den familiären Einflussfaktoren – auf die Leistungsbeurteilung Einfluss nehmen, wirft tiefgreifende Fragen zur Beurteilungspraxis in der Grundschule auf.

Während im mehrgliedrigen Schulsystem die Übertrittsempfehlungen je nach Land in ihrer Verbindlichkeit variieren, gelten Schulnoten in den verbindlichen Empfehlungen als ein Hauptentscheidungskriterium für den Übertritt in die Sekundarstufe I. Dabei wird bezüglich der Schulnoten häufig diskutiert, welchen Gütekriterien diese entsprechen können bzw. sollten. Obgleich der Einsatz messtheoretischer Gütekriterien bei der Überprüfung schulischer Bewertungsleistungen als Regulativ für die Qualität des Unterrichts und der Lehrkraftaktivitäten gelten kann, sollte bei der Kritik an der Validität, Reliabilität und Objektivität der Notengebung beachtet werden, dass standardisierte Schulleistungstests die vom Schulrecht vorgegebenen Beurteilungsräume nicht ausreichend erfassen können (Trautwein et al., 2014). Um zu beurteilen, welche Qualität schulische Bewertungsprozesse aufweisen, muss zunächst Klarheit darüber herrschen, welche Bestandteile Noten in der Grundschule aufweisen und welche Schülermerkmale – individueller und familiärer Art – von der Lehrkraft also tatsächlich in fachspezifischen Noten berücksichtigt werden. Obwohl die globalen Tendenzen zur Notenzusammensetzung hinlänglich bekannt sind und meist indirekt erschlossen werden können, fehlt bislang eine systematische Aufarbeitung vom schulischen Leistungsbegriff bis hin zu den

bewerteten Leistungen im Grundschulbereich. Während die Arbeit von Müller (2013) ebenfalls mit Daten der Koala-S-Studie den prognostischen Wert der Lehrkräfteeinschätzungen aus der zweiten Klasse auf die Durchschnittsnote aus Mathematik, Deutsch und Heimat- und Sachkunde am Ende der vierten bzw. sechsten Klasse untersuchte, macht es sich die vorliegende Arbeit zum Ziel, schulrechtliche und psychometrische Kriterien auf ihre Stimmigkeit mit empirischen Korrelaten der Deutsch- und Mathematiknoten Ende der vierten Klasse zu prüfen und (außer)schulische Einflussfaktoren sichtbar zu machen. Dabei sollen die gefundenen Ergebnisse keinerlei Aussagen zur Qualität der schulischen Notengebung machen. In dem Bewusstsein, dass das vorgestellte Gesamtmodell sicherlich kein vollständiges und kausales Bedingungsmodell ist und nicht alle leistungsrelevanten Merkmale erfasst, sollen vielmehr weitere Informationen darüber gewonnen werden, was aus der Perspektive der Lehrkraft unter Schulleistung – ausgedrückt in Schulnoten – verstanden werden kann. Dabei ist es nicht das Ziel, zeitlich konstante Bedingungsfaktoren festzumachen, sondern einen Einblick in die Bewertungspraxis zu geben. Zunächst steht als Schwerpunkt des empirischen Teils die Frage, durch welche individuellen Prädiktoren Schulnoten in Mathematik und Deutsch in der vierten Grundschulklasse vorhergesagt werden können. Anschließend gilt es die Abhängigkeit der Prädiktoren von familiären Hintergrundmerkmalen zu prüfen, um in einem dritten Schritt die Abhängigkeit der Notenzusammensetzung von der leistungs- und sozial bedingten Klassenzusammensetzung zu analysieren. Effekte des Bundeslandes sowie der Migration werden in diesem Zusammenhang nicht untersucht. Gemäß der beiden schulrechtlichen Länderverordnungen wird nicht von systematischen Unterschieden ausgegangen sowie Effekte der Migration meist auf sozialschichtspezifische und individuelle Schülermerkmale zurückgeführt werden.

Mithilfe der Koala-S-Daten ist es möglich, jene Perspektive der Lehrkraft auf die Notenzusammensetzung im Längsschnitt zu explizieren und die hierarchische Struktur der Daten zu berücksichtigen. Obwohl Angaben zu individuellen Schülermerkmalen durch den Schüler selbst, aus Beobachtungen oder durch die Eltern den Vorteil bieten, eine indi-

viduelle Einschätzung der Merkmale und alltagsgetreue Beobachtungen zu erhalten, ist die Erfassung der Lehrkraftmeinung letztlich jene Perspektive, aus welcher die Zeugnisnote errechnet wird. Jene Zeugnisnoten, die – zumindest in Bundesländern mit verpflichtender Übertrittsempfehlung – häufig als richtungsweisend für die Schulkarriere eines Kindes gelten und vielfach auf ihre soziale Selektivität geprüft werden.

Die bisherigen Erkenntnisse und übergreifenden Ziele der Arbeit zu Bedingungsfaktoren schulischer Noten werden im Folgenden in drei Fragestellungen und entsprechenden Hypothesen festgehalten, die Abfolge orientiert sich dabei an den in Kapitel vier aufgeführten Faktoren. Ebenfalls ist anzumerken, dass sich die vorgestellten Hypothesen, wenn nicht explizit anders vermerkt, auf in der vierten Grundschulklasse erhobene Indikatoren beziehen.

## 5.1 Individuelle Schülermerkmale und Schulnoten

Fragestellung I: Welche Rolle spielen individuelle Merkmale des Schülers bei der Zusammensetzung der Deutsch- und Mathematik Note der 4. Klasse und wie werden diese gewichtet?

**Hypothese 1: Die von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale haben einen stärkeren Einfluss auf die Viertklassnote in Deutsch und Mathematik als die individuelle Test- und Intelligenzleistung des Schülers.**

H1.1: Bei gleichen Test- und Intelligenzleistungen haben Schüler mit einer positiveren Einschätzung in ihren kognitiven Fähigkeiten (Begabung, sprachliche Fähigkeiten, Vorwissen) eine bessere Note in Deutsch und Mathematik als Schüler, die von der Lehrkraft weniger positiv eingeschätzt werden.

H1.2: Die kognitiven Schülermerkmale (Testleistung, Intelligenz, Begabung, sprachliche Fähigkeiten, Vorwissen) sind stärkere Prädikto-

ren der fachspezifischen Viertklassnote als die nicht-kognitiven Schülermerkmale (Arbeits- und Sozialverhalten).

H1.3: Unter Kontrolle der kognitiven Merkmale hat auch das von der Lehrkraft eingeschätzte Arbeitsverhalten einen positiven Effekt auf die Benotung. Dabei fließt das Arbeitsverhalten stärker in die fachspezifische Viertklassnote ein als das Sozialverhalten des Kindes.

H1.4: Das von der Lehrkraft eingeschätzte Vorwissen des Schülers aus der zweiten Jahrgangsstufe ist ein wesentlicher Prädiktor für die fachspezifische Note der vierten Klasse.

**Hypothese 2: Der Zusammenhang zwischen den eingeschätzten kognitiven Schülermerkmalen und Schulnoten wird durch nicht-kognitive Merkmale moderiert.**

## 5.2 Familiäre Hintergrundmerkmale und Schulnoten

Fragestellung II: Welche Rolle spielen familiäre Hintergrundmerkmale des Schülers bei der Zusammensetzung der Deutsch- und Mathematik Note der 4. Klasse und wie werden diese gewichtet?

**Hypothese 3: Strukturelle Merkmale der Familie (Bildungs- und sozialer Status) haben einen signifikanten Einfluss auf die Viertklassnoten in Deutsch und Mathematik.**

H3.1: Schüler mit höheren Werten in den Strukturmerkmalen erhalten eine positivere kognitive und nicht-kognitive Merkmaleinschätzung als Schüler mit niedrigeren Werten in den Strukturmerkmalen.

H3.2: Der Einfluss struktureller Hintergrundmerkmale auf die fachspezifische Schulnote wird über die individuellen kognitiven und nicht-kognitiven Merkmaleinschätzungen des Kindes vermittelt.

**Hypothese 4: Schulbezogene Prozessmerkmale der Familie (Lehrkräfteeinschätzung bzgl. des elterlichen Unterstützungsverhaltens, Schulkontakts, Bildungsaspiration) haben einen signifikanten Einfluss auf die Vierklassnoten in Deutsch und Mathematik.**

H4.1: Die schulbezogenen Prozessmerkmale der Familie vermitteln den Effekt zwischen familiären Strukturmerkmalen und Schulnoten: Je höher die Strukturmerkmale einer Familie, desto höher werden die Prozessmerkmale der Familie eingeschätzt und desto niedriger ist die fachspezifische Schulnote.

H4.2: Die schulbezogenen Prozessmerkmale beeinflussen direkt und vermittelt über die Einschätzung der kognitiven und nicht-kognitiven Schülermerkmale die fachspezifische Viertklassnote eines Schülers.

### 5.3 Schulische Merkmale und Schulnoten

Fragestellung III: Welche Rolle spielen klassen- und fachspezifische Rahmenbedingungen für die Zusammensetzung der Deutsch- und Mathematik Note der 4. Klasse und wie werden diese gewichtet?

**Hypothese 5: Merkmale der Schulklasse wirken sich auf die Viertklassnoten in Deutsch und Mathematik aus.**

H5.1: Es besteht ein Zusammenhang zwischen dem durchschnittlichen Leistungsniveau der Klasse und der Schulnote eines Schülers. Schüler erhalten in einer leistungsstärkeren Klasse schlechtere Noten als in einer leistungsschwächeren Klasse.

H5.2: Es besteht ein Zusammenhang zwischen der durchschnittlichen sozialen Klassenzusammensetzung und der Schulnote des Schülers. Schüler erhalten in einer sozialschwächeren Klasse bessere Noten.

**Hypothese 6: Es bestehen Unterschiede in der Zusammensetzung der Viertklassnote je nach Fach.**



## 6 Datengrundlage

In diesem Kapitel wird die Datengrundlage der vorliegenden Arbeit vorgestellt. Ziel des Kapitels ist ein Überblick über die verwendete Studie, deren Durchführung, Stichprobe und die daraus resultierenden Indikatoren. Die Auswahl der Darstellung beruht dabei auf der Relevanz für die vorliegende Untersuchung, weiterführende Informationen zu Grundlagen der Studie finden sich bei Ditton (2007b).

### 6.1 Die Koala-S-Studie

Die Datenbasis dieser Untersuchung ist die von der Deutschen Forschungsgesellschaft geförderte Studie „**Kompetenzaufbau und Laufbahnen im Schulsystem**“ (Koala-S). Übergreifendes Ziel des Projekts unter Leitung von Prof. Dr. Hartmut Ditton war eine längsschnittlich angelegte Analyse bestehender Disparitäten der Bildungsbeteiligung und des Bildungserfolgs, erhoben durch Fragebögen an Lehrkräfte, Eltern und Schüler aus Grundschulen in Bayern und Sachsen. Erfasst wurden an zwei aufeinanderfolgenden Messzeitpunkten kognitive, nicht-kognitive und konstitutionelle Schülermerkmale, familiäre Hintergrundmerkmale sowie institutionelle und regionale Bedingungsfaktoren. Die Studie Koala-S umfasst zwei Erhebungswellen (2003-2007) mit Nacherhebungen zwischen den Jahren 2009 bis 2011 (vgl. [Tabelle 3](#)). Zahlreiche Studien konnten bereits auf dieser Datengrundlage Mechanismen sozialer Disparitäten im Schulsystem nachweisen (z.B. Ditton & Krüsken, 2009; Gerleigner, 2013; Müller, 2013; Wohlkinger, 2014).

Die vorliegende Untersuchung bezieht sich ausschließlich auf Daten der zweiten Erhebungswelle (2005-2007), durchgeführt jeweils zum Ende des Schuljahres in den Grundschulstufen zwei bis vier in Bayern und Sachsen, weswegen im Folgenden nur die für diesen Zeitraum relevanten Informationen berichtet werden.

Tabelle 3: Übersicht Erhebungszeitpunkte von Koala-S

Erhebungswelle	Jahrgangsstufe	Jahr	Bundesland
Erste Erhebungswelle	Jahrgangsstufe 3	2003	Bayern
	Jahrgangsstufe 4	2004	Bayern
Zweite Erhebungswelle	Jahrgangsstufe 2	2005	Bayern/Sachsen
	Jahrgangsstufe 3	2006	Bayern/Sachsen
	Jahrgangsstufe 4	2007	Bayern/Sachsen
Nacherhebung zur zweiten Erhebungswelle	Jahrgangsstufe 7	2009	Bayern/Sachsen
	Jahrgangsstufe 9	2011	Bayern/Sachsen

## 6.2 Durchführung der Untersuchung

Die Befragung der Teilnehmer und die Datenaufbereitung erfolgte durch das Data Processing Center der (DPC) der International Association for the Evaluation of Educational Achievement (IEA). Die wissenschaftliche Konzeption des Projekts sowie die Datenauswertung oblagen der Projektgruppe um Prof. Dr. Hartmut Ditton. Die Befragung erfolgte am Ende des Schuljahres an zwei aufeinanderfolgenden Tagen in einem festgelegten Zeitraum von zwei Wochen. Befragt wurden Schüler, Eltern und Lehrkräfte; die Befragung der Schüler wurde außerdem noch durch eine schulkontextunabhängige, testleitergeführte Leistungstestung erweitert. Diese umfasste jeweils einen standardisierten Leistungstest in Lesen, Rechtschreibung und Mathematik, zusätzlich dazu wurden die kognitiven Grundfähigkeiten des Schülers getestet (Heller & Geisler, 1983). Die Schülerbefragung (SFB) erfolgte jeweils im Klassenkontext. Eine Befragung der Eltern (EFB) fand ebenfalls durch einen Fragebogen statt, welcher außerhalb des Schulkontexts von den Eltern bearbeitet wurde. Parallel dazu wurden die Lehrkräfte der beteiligten Klassen gebeten, für jeden befragten Schüler einen Schülerbeurteilungsbogen (SBB) auszufüllen, eine zweite Befragung der Lehrkräfte erfolgte zu Aspekten der Lehrperson selbst und deren Einstellungen bezüglich sozialer Leistungsbezüge und schulkontextbezogener Informationen (LFB). Die Teilnahme der Befragung war für alle Untersuchungsteilnehmer freiwillig und für

jeden einzelnen Probanden vollkommen anonym. Eine individuelle Codezuweisung ermöglicht im Nachhinein eine Zuweisung der Schüler zu weiteren Untersuchungsmerkmalen.

### 6.3 Stichprobe

Die Datenbasis der vorliegenden Studie bilden Angaben einer geschichteten Zufallsstichprobe aus öffentlichen Grundschulen in Bayern und Sachsen. Als Stratifizierungsmerkmale wurden der Schultyp (Grundschule vs. Grund- und Teilhauptschule), die Schulgröße (ein- bis zweizügige Schulen vs. mehrzügige Schulen) sowie der Urbanisierungsgrad der regionalen Umgebung (Teil eines Ballungsraums vs. ländliche Region) herangezogen. Innerhalb dieser Auswahl wurden die Schulen zufällig ausgewählt, um daran anschließend innerhalb dieser Schule jeweils eine Klasse zufällig in die Koala-S Stichprobe aufzunehmen. Ausgewählt wurden dann diejenigen Fälle, für die eine Teilnahmegenehmigung der Eltern vorlag und deren Schulklassen zu allen drei Erhebungszeitpunkten an der Untersuchung teilnahmen. Nach anfänglich 84 ausgewählten Schulklassen nahmen somit insgesamt 77 Schulklassen (N= 1453) an allen drei Erhebungszeitpunkten der Befragung teil (Bayern: N= 871 aus 42 Klassen; Sachsen: N= 582 aus 35 Klassen). Da aufgrund der schulrechtlichen Länderregelungen (vgl. [Kap. 2.5](#)) von keinem systematischen Unterschied in der Zusammensetzung der Schulnoten nach Land ausgegangen wird, umfasst die Untersuchungsstichprobe beide Bundesländer ohne Differenzierung. Ausgeschlossen wurden aus der anfänglichen Gesamtstichprobe (N= 1453) für die nachfolgenden Analysen diejenigen Fälle, die in allen Untersuchungsvariablen fehlende Werte vorweisen (Untersuchungsstichprobe N= 1447).

Die nachfolgende Tabelle gibt einen Überblick über die Zusammensetzung der Untersuchungsstichprobe der Schüler, differenziert nach Geschlecht, Migration und sozialen Hintergrundmerkmalen.

Tabelle 4: Demographische Merkmale der Schülerstichprobe

<b>Schülerstichprobe</b>		
Geschlecht	Mädchen	660 (49,3%)
	Jungen	679 (50,7%)
Migrationshintergrund	Kein Elternteil im Ausland geboren	1242 (88,3%)
	Ein Elternteil im Ausland geboren	88 (6,3%)
	Beide Eltern im Ausland geboren	77 (5,5%)
Bildungsstatus <sup>a</sup>	Nicht höher als Hauptschulabschluss	251 (18,0%)
	Mittlerer Schulabschluss	643 (46,0%)
	Allgemeine Hoch-/Fachhochschulreife	503 (36,0%)
Sozialer Status (ISEI) <sup>b</sup>	Niedrig	391 (28,4%)
	Mittel	553 (40,2%)
	Hoch	431 (31,3%)

Anmerkungen: Die Teilnehmerzahl der Schülerstichprobe umfasst N= 1447. Die Werte und Prozentangaben der einzelnen demografischen Merkmale können aufgrund fehlender Werte in der Summe von der Gesamtstichprobenzahl abweichen.

a Der Bildungsstatus bezieht sich auf den höchsten Bildungsstatus im Haushalt.

b Der Soziale Status der Familie bezieht sich auf die ISEI-Kategorisierung: Niedrig = ISEI ≤ 33; Mittel = 33 > ISEI ≤ 51; Hoch = ISEI > 51.

Aus der beschriebenen Gesamtstichprobe werden in der folgenden Untersuchung Angaben der Lehrkraft zu individuellen Schülermerkmalen aus der zweiten und vierten Jahrgangsstufe, die Testleistungen der Schüler sowie deren familiäre Hintergrundmerkmale verwendet. Eine gesonderte Betrachtung von Schülern mit Migrationshintergrund wird nicht vorgenommen, da sich in vorausgehenden Analysen keine signifikanten Unterschiede zwischen den Teilnehmern ergaben sowie diese Unterscheidung für die vorliegende Fragestellung nicht von Relevanz ist.

## 6.4 Instrumente

Während in der Längsschnittstudie Koala-S durch die verschiedenen Erhebungsperspektiven und Messzeitpunkte eine Vielzahl an Informationen gewonnen wurde, werden im Folgenden nur jene Variablen aus dem ersten und dritten Messzeitpunkt (zweite und vierten Jahrgangsstufe) aufgeführt, die für anschließende Analysen herangezogen werden. Von Interesse sind Angaben, durch die Informationen zu Notenkomponenten der Schüler gewonnen werden können: Neben Angaben zu Schulnoten und Leistungstests werden individuelle Schülermerkmale und familiäre Hintergrundmerkmale expliziert. Angaben zur Erhebungsperspektive werden entsprechend kenntlich gemacht (LK=Lehrkräfteeinschätzung). Alle Eltern (EFB)- und Schülereinschätzungen durch die Lehrkraft (SBB) wurden mittels Fragebögen erhoben.

### 6.4.1 Schulnoten und Leistungstests

Während die Schulnoten der 4. Grundschulklasse in Deutsch und Mathematik die abhängigen Variablen der nachfolgenden Untersuchung darstellen, werden zusätzlich erhobene Testleistungen in Lesen, Rechtschreibung und Mathematik als kognitive Determinanten herangezogen.

#### Schulnoten

Die Deutsch- und Mathematiknoten der Schüler stammen aus den Halbjahres- bzw. Übertrittszeugnissen der vierten Klasse und bilden in dieser Untersuchung die abhängigen Variablen. Die ursprüngliche Polung in den Noten wird beibehalten, so dass höhere Werte in der abhängigen Variable schlechtere Noten implizieren. Obgleich gemäß der theoretischen Grundlage aus [Kapitel 2.3](#) Noten ordinal skaliert sind, werden sie in den folgenden Analysen – wie in der Bildungsforschung meist üblich – auf Intervallskalenniveau gemessen, vergleichende Analysen mit einer ordinal skalierten abhängigen Variable kamen zu den gleichen Ergebnistendenzen. [Tabelle 5](#) gibt einen Überblick zu der vorliegenden Verteilung der Deutsch- und Mathematiknoten in der vierten Jahrgangsstufe über alle Klassen hinweg.

Sowohl für die Deutsch- als auch für die Mathematiknote befinden sich die Durchschnittsnoten noch in einem guten Leistungsbereich, wobei die Noten in Deutsch geringfügig besser sind.

Tabelle 5: Zeugnissnoten in Deutsch und Mathematik 4. Klasse Grundschule

<b>Zeugnissnoten Deutsch, Mathematik 4.Klasse</b>	<b>N</b>	<b>M</b>	<b>SD</b>
Schulnote Deutsch	1219	2.46	.838
Schulnote Mathematik	1219	2.51	.890

Anmerkungen: Die Items wurden auf einer sechsstufigen Skala erfasst (sehr gut – ungenügend). Niedrigere Werte entsprechen einer besseren Schulnote.

## Testleistungen

Parallel zu der Schüler-, Eltern- und Lehrkraftbefragung wurden zu allen drei Messzeitpunkten Leistungstests in Mathematik, Rechtschreibung und im Leseverständnis eingesetzt; entwickelt und wissenschaftlich begleitet vom Institut für Schulpädagogik und Bildungsforschung bzw. Didaktikern der Ludwig-Maximilians-Universität München. Alle Testinhalte orientierten sich an lehrplanspezifischen Themen. Im mathematischen Bereich erfolgte die Testung in den Disziplinen Geometrie, Arithmetik und Zahlenverständnis; ebenfalls getestet wurden das Größen- und Messwissen sowie das sachbezogene Mathematikverständnis. Zur Erfassung der Rechtschreibeleistung wurden den Schülern standardisierte Wortdiktate von unterschiedlicher Länge vorgelegt. Einige Wörter wurden wiederholt eingesetzt, während andere jahrgangsspezifisch und in ihrem Schwierigkeitsgrad variierten. Die Testung des Leseverständnisses wurde an allen drei Messzeitpunkten durch sechs differente und pilotierungserprobte Lesetexte (Geschichte und altersgerechte Sachtexte) erfasst. Im Fokus der Testung stand das sinnverstehende, text- und wissensbasierte Lesen (Krüsken, 2007b). Aus den drei Leistungstests wurde für die vierte Klasse ein Leistungsindex gebildet, der z-skaliert in die Untersuchung aufgenommen wird (N= 1240). Höhere Werte entsprechen einem besseren Testergebnis. Einen Überblick über die Testinstrumente und deren

Reliabilitätskennwerte bietet die nachfolgende [Tabelle 6](#) (vgl. Ditton & Krüsken, 2009, S. 39).

Tabelle 6: Übersicht Testleistungen 4. Klasse

Testbereiche 4. Klasse	Itemanzahl	Reliabilität (Cronbachs Alpha)
Mathematik	15	$\alpha=.71$
Leseverständnis	26	$\alpha=.65$
Rechtschreiben	64	$\alpha=.90$

#### 6.4.2 Individuelle Schülermerkmale

Zusätzlich zu den erfassten Testleistungen und Zeugnisnoten wurden die Lehrkräfte zum dritten Messzeitpunkt Ende der vierten Klasse gebeten, die Schüler hinsichtlich ihrer Fähigkeiten einzuschätzen. Die Erhebung erfolgte durch Fragebögen und umfasst u.a. Angaben zu sprachlichen Fähigkeiten, fachspezifischer Begabung und sonstigen Merkmalen des Kindes, die auf Grundlage der theoretischen Annahmen für die Notenzusammensetzung relevant sein könnten. Zu betonen ist, dass es sich mit Ausnahme der Intelligenz (erfasst durch den kft) bei den nachfolgenden Schülermerkmalen um *Lehrkräfteeinschätzungen* handelt. Mit Ausnahme des Vorwissens Anfang der zweiten Klasse geht eine niedrigere Bewertung eines Schülermerkmals analog zur Notenskalierung mit einer positiveren Einschätzung der Lehrkraft einher und kann keine Aussagen über die tatsächliche Passung zu dem „wahren“ Verhalten der eingeschätzten Schüler treffen.

#### Intelligenz

Die Intelligenz des Schülers wurde in Anlehnung an die Befunde aus [Kapitel 4.2.1](#) durch die *kognitiven Grundfähigkeiten* des Schülers erfasst. Diese wurden mittels des kognitiven Fähigkeitstests für den Grundschulbereich – kurz KFT 1-3 bzw. CFT 20 – von Heller und Geisler (1983) und Weiß (1998) erhoben. Mit dem KFT 1-3 wurden in der dritten Klasse die intellektuellen Fähigkeiten und die Leistungs- und Lernvoraussetzungen der Schüler ermittelt. Die Testung in der vierten Klasse umfasste noch die schweren Items des KFT 1-3, ergänzt

durch eine Kurzversion des CFT 20 (N= 1210). Ergebnisse des Tests der vierten Klasse werden z-skaliert als Intelligenzwert des Schülers in den Analysen berücksichtigt.

### Vorwissen

Zusätzlich zu Angaben aus der vierten Klasse wird in den folgenden Analysen auch die vorausgegangene Vorwissenseinschätzung der Viertklässler durch die Lehrkraft berücksichtigt. Grund dafür ist einmal der strukturelle Rahmen, nach dem sich die Lehrkräfte einer Schule über die vorausgegangenen Leistungen der Schüler informieren können (vgl. Kap. 2.5). Zum anderen wird theoriegeleitet davon ausgegangen, dass die Noten eines Viertklässlers nicht unabhängig von dessen Vorwissen betrachtet werden können. Als Vorwissensindikator wird in der nachfolgenden Analyse die Lehrkräfteeinschätzung aus der zweiten Klasse zu Fähigkeiten des Kindes zu Schuleintritt herangezogen: Gefragt wurde nach den Fähigkeiten in Lesen, Schreiben und Rechnen. Die nachfolgende Tabelle gibt einen Überblick zu den vier Items der Skala. Die interne Konsistenz beträgt  $\alpha=.733$ .

Tabelle 7: Skala Vorwissen 2. Klasse (LK)

<b>Vorwissen (LK)</b> M <sub>Skala</sub> =2.47; SD <sub>Skala</sub> =.688; N <sub>Skala</sub> =1088; Cronbachs Alpha: $\alpha=.733$ ; N=936	<b>M</b>	<b>SD</b>	<b>Korr. Item- Skala- Korrela- tion</b>	<b>Cronbachs Alpha, wenn Item weg- gelassen</b>
Kind konnte beim Schuleintritt...				
...schon bis 10 zählen.	3.16	.876	.608	.627
...schon lesen.	1.39	.730	.410	.735
...leichte Rechenaufgaben lösen.	2.26	.988	.619	.617
...den eigenen Namen schreiben.	3.04	.939	.486	.701

Anmerkungen: Die Items wurden auf einer vierstufigen Skala erfasst (gar nicht – sehr gut). Höhere Werte entsprechen einem höher eingeschätzten Vorwissen.

### Sprachliche Fähigkeiten

Die Lese-, Sprach- und Ausdrucksfähigkeit des Kindes wurde von der Lehrkraft mittels SBB zum Ende der vierten Klasse ( $t_3$ ) erhoben. Abgefragt wurde mit einer sechsstufigen Ratingskala (sehr gut=1 bis ungenügend=6) die Lehrkräfteeinschätzung zur Lesefähigkeit des Kindes und dessen Sprach- und Ausdrucksfähigkeit, entwickelt im Rahmen des Koala-S-Projekts. [Tabelle 8](#) gibt einen Überblick zu den zwei Items der Skala und den deskriptiven Statistiken. Die interne Konsistenz beträgt  $\alpha=.876$ .

Tabelle 8: Skala Sprachliche Fähigkeiten 4. Klasse (LK)

<b>Sprachliche Fähigkeiten (LK)</b> M <sub>Skala</sub> =2.37; SD <sub>Skala</sub> =.944; N <sub>Skala</sub> =1230; Cronbachs Alpha: $\alpha=.876$ ; N=1230	<b>M</b>	<b>SD</b>	<b>Korr. Item-Skala-Korrelation</b>
Lesen	2.32	.977	.779
Sprach- und Ausdrucksfähigkeit	2.41	1.024	.779

Anmerkungen: Die Items wurden auf einer sechsstufigen Skala erfasst (sehr gut – ungenügend). Niedrigere Werte entsprechen höher eingeschätzten sprachlichen Fähigkeiten.

### Begabung

Zusätzlich zu den sprachlichen Fähigkeitseinschätzungen wurden die Lehrkräfte zum Ende der vierten Klasse um ihre Einschätzung zur Begabung des Schülers in den Fächern Deutsch und Mathematik gebeten. Die Bewertung erfolgte mittels einer vornehmlich für Koala-S entwickelten fünfstufigen Ratingskala (stark über Durchschnitt=1 bis stark unter Durchschnitt=5) und bezieht sich ausdrücklich auf die generelle Begabung des Kindes in Deutsch oder Mathematik; eingeschätzt durch dessen Klassenlehrkraft. Die nachfolgende Tabelle zeigt die deskriptiven Statistiken der beiden Begabungseinschätzungen. Aufgrund der vorliegenden Fragestellung werden die beiden Einschätzungen bewusst fachspezifisch getrennt und nicht zu einer Skala zusammengefasst und fließen als Prädiktoren mit in die Analysen ein.

Tabelle 9: Begabung in Deutsch und Mathematik 4. Klasse (LK)

Items Begabung	N	M	SD
Begabung Mathematik	1231	2.67	.967
Begabung Deutsch	1231	2.66	.944

Anmerkungen: Die Items wurden auf einer fünfstufigen Skala erfasst (stark über Durchschnitt – stark unter Durchschnitt). Niedrigere Werte entsprechen höher eingeschätzten Begabungen.

## Arbeitsverhalten

Vor dem Hintergrund der anfänglich vorgestellten, theoretischen und empirischen Bezüge wurden jene Merkmaleinschätzungen für das Arbeitsverhalten ausgewählt, die den in Schule und Lehrplänen benannten Begriff bestmöglich repräsentieren. Die Merkmaleinschätzung zum Schülerverhalten erfolgte ebenfalls durch die Lehrkraft am Ende der vierten Klasse; abgefragt durch eine fachunspezifische, fünfstufige Ratingskala (sehr gut=1 bis mangelhaft=5). Höhere Werte deuten – analog zu den Schulnoten – auf eine schlechtere Bewertung durch die Lehrkraft hin. [Tabelle 10](#) gibt einen Überblick zu den fünf Items der Skala und deren deskriptive Statistiken. Die vorgestellte Skala wird in den folgenden Analysen als Prädiktor in das Modell aufgenommen. Die interne Konsistenz beträgt  $\alpha=.911$ .

Tabelle 10: Skala Arbeitsverhalten 4. Klasse (LK)

Arbeitsverhalten (LK)	M	SD	Korr. Item-Skala-Korrelation	Cronbachs Alpha, wenn Item weggelassen
$M_{Skala}=2.31; SD_{Skala}=.845;$ $N_{Skala}=1231; Cronbachs$ $Alpha: \alpha=.911; N=1231$				
Ehrgeiz	2.34	.979	.834	.878
Ordnung	2.35	1.026	.692	.908
Fleiß	2.28	.991	.852	.874
Mitarbeit	2.30	.936	.670	.911
Ausdauer	2.29	.991	.826	.879

Anmerkungen: Die Items wurden auf einer fünfstufigen Skala erfasst (sehr gut – mangelhaft). Niedrigere Werte entsprechen einem höher eingeschätzten Arbeitsverhalten.

## Sozialverhalten

Ähnlich zum Arbeitsverhalten baut auch die Skala zum Sozialverhalten der Schüler auf den anfänglich vorgestellten Erkenntnissen und den schulrechtlichen Grundlagen auf. Die fachunspezifische Merkmalseinschätzung stammt aus dem SBB der vierten Jahrgangsstufe; mithilfe von fünfstufigen Ratingskalen (sehr gut=1 bis mangelhaft=5) wurden die Lehrkräfte um ihre fachunspezifische Merkmalseinschätzung gebeten. In [Tabelle 11](#) sind die drei Items der Skala und deren deskriptive Statistiken aufgeführt. Auch diese Skala wird in den nachfolgenden Analysen als Prädiktor in das Modell aufgenommen. Die interne Konsistenz beträgt  $\alpha = .883$ .

Tabelle 11: Skala Sozialverhalten 4. Klasse (LK)

Sozialverhalten (LK) M <sub>Skala</sub> = 1.89; SD <sub>Skala</sub> = .722; N <sub>Skala</sub> = 1231; Cronbachs Alpha: $\alpha = .883$ ; N = 1231	M	SD	Korr. Item- Skala-Korre- lation	Cronbachs Alpha, wenn Item weggelassen
Ehrlichkeit	1.83	.808	.784	.821
Höflichkeit	1.77	.749	.782	.827
Sozialverhalten	2.08	.848	.753	.851

Anmerkungen: Die Items wurden auf einer fünfstufigen Skala erfasst (sehr gut – mangelhaft). Niedrigere Werte entsprechen einem höher eingeschätzten Sozialverhalten.

## Geschlecht

Als letztes individuelles Schülermerkmale wird das Geschlecht des Schülers in den nachfolgenden Analysen als Kontrollvariable berücksichtigt. Die Variable Geschlecht fließt Dummy-kodiert (weiblich=0, männlich=1) in die Analysen ein. Das Geschlechterverhältnis in der vorliegenden Stichprobe ist nahezu ausgewogen (Mädchen: 49,3 %; Jungen: 50,7%).

### 6.4.3 Familiäre Hintergrundmerkmale

Zusätzlich zu den erfassten individuellen Schülermerkmalen liegen von den Eltern der befragten Schüler strukturelle Hintergrundinformationen vor. Die Erhebung erfolgte durch Fragebögen und umfasst u.a. Angaben zum familiären Sozialstatus und dem im Haushalt vorliegenden höchsten Bildungsgrad. Zusätzlich dazu wurden die Lehrkräfte gebeten, den Schulkontakt der Eltern, deren schulisches Unterstützungsverhalten und die Wichtigkeit eines hohen Schulabschlusses der Eltern mittels Fragebogen einzuschätzen. In Anlehnung an [Kapitel 4.3](#) erfolgt die Darstellung der verwendeten Instrumente differenziert in Struktur- und Prozessmerkmale der Familie.

#### 6.4.3.1 Strukturmerkmale

Die Strukturmerkmale familiärer Herkunft basieren in dieser Untersuchung auf Elternangaben zum familiären Sozialstatus und den im Haushalt vorliegenden Bildungsgraden; erhoben mittels Fragebögen im Rahmen des EFB zu allen drei Messzeitpunkten.

Der *Bildungsstatus der Familie* beruht auf dem höchsten Bildungsabschluss im Haushalt und wird in dieser Untersuchung in drei Kategorien zusammengefasst verwendet: niedrig = nicht höher als Hauptschule; mittel = mittlerer Abschluss bis Klasse 10; hoch = Allgemeine/fachgebundene Hochschulreife.

Der *sozioökonomische Status der Familie* wurde in der vorliegenden Untersuchung mithilfe der ISEI-Kategorisierung erhoben (International Socio-Economic Index of Occupational Status nach Ganzeboom, de Graaf & Treiman, 1992). Er umfasst Angaben zur beruflichen Stellung, Einkommen und die damit einhergehende Statusposition durch Zahlenwerte von 16 (z.B. landwirtschaftliche Hilfskräfte) bis 90 (z.B. Richter). Grundlage der in Koala-S erhobenen sozioökonomischen Hintergrundmerkmale sind Angaben der Schülerfamilie zu Beruf und Einkommen, die in dieser Arbeit in zusammengefasster Form in drei Kategorien verwendet werden. Angaben der ISEI-Werte sowie des höchsten Bildungsstatus der Untersuchungsstichprobe sind [Tabelle 12](#) zu entnehmen. Angaben mit höheren Werten zeichnen Familien

aus höheren ISEI-Statusgruppen bzw. mit höheren Bildungsabschlüssen aus. Die strukturellen Hintergrundmerkmale der Familie werden in den folgenden Analysen als Prädiktor in das Modell aufgenommen.

Tabelle 12: Übersicht familiäre Strukturmerkmale der Familie

<b>Strukturmerkmale der Familie: Bildungsstatus und Sozialer Status</b>		
Bildungsstatus <sup>a</sup>	Nicht höher als Hauptschulabschluss	251 (18,0%)
	Mittlerer Schulabschluss	643 (46,0%)
	Allgemeine Hoch-/Fachhochschulreife	503 (36,0%)
Sozialer Status (ISEI) <sup>b</sup>	Niedrig	391 (28,4%)
	Mittel	553 (40,2%)
	Hoch	431 (31,3%)

Anmerkungen: Die Teilnehmerzahl der Schülerstichprobe umfasst N= 1447. Die Werte und Prozentangaben der einzelnen Strukturmerkmale können aufgrund fehlender Werte in der Summe von der Gesamtstichprobenzahl abweichen.

- a Der Bildungsstatus bezieht sich auf den höchsten Bildungsabschluss im Haushalt.  
 b Der Soziale Status der Familie bezieht sich auf die ISEI-Kategorisierung: Niedrig = ISEI ≤ 33; Mittel = 33 > ISEI ≤ 51; Hoch = ISEI > 51.

#### 6.4.3.2 Prozessmerkmale

Die Prozessmerkmale familiärer Herkunft basieren in der vorliegenden Untersuchung auf Lehrkräfteeinschätzungen, erhoben zum dritten Messzeitpunkt in der 4. Klasse. Unterschieden wird zwischen der Lehrkräfteeinschätzung zum elterlichen Unterstützungsverhalten, der Qualität und Häufigkeit des Elternkontakts mit der Schule und der Einschätzung der Lehrkraft zur Bildungsaspiration der Eltern ihrer Schüler.

#### Elterliches Unterstützungsverhalten

Zusätzlich zur Einschätzung von individuellen Schülermerkmalen wurden die Lehrkräfte gebeten, eine Einschätzung hinsichtlich des elterlichen Unterstützungsverhaltens ihrer Schüler vorzunehmen. Die Bewertung erfolgte mittels einer vornehmlich für Koala-S entwickelten vierstufigen Ratingskala (stimmt nicht=1 bis stimmt genau=4), in

der die Lehrkraft angeben sollte, ob die Unterstützung des Elternhauses vorhanden ist. Die nachfolgende Tabelle zeigt die deskriptiven Statistiken zum elterlichen Unterstützungsverhalten. Die Einschätzung zum Unterstützungsverhalten der Familie wird in den folgenden Analysen als Prädiktor in das Modell aufgenommen.

Tabelle 13: Übersicht Unterstützung Elternhaus 4. Klasse (LK)

Item Unterstützung Elternhaus	N	M	SD
Die Unterstützung vom Elternhaus des Kindes ist vorhanden.	1207	3.28	.799

Anmerkungen: Das Item wurde auf einer vierstufigen Skala erfasst (stimmt nicht – stimmt genau). Niedrigere Werte entsprechen einem niedriger eingeschätzten Unterstützungsverhalten.

### Schulkontakt

Die Angaben zum Schulkontakt basieren auf Lehrkräfteeinschätzungen in der vierten Klasse, bei der die Lehrkräfte mittels Fragebogen um ihre Einschätzung zur Beziehung bzw. Kontakt zwischen Elternhaus und Lehrkraft gebeten wurden. Abgefragt wurde einmal mit einer eigens für Koala-S entwickelten, vierstufigen Ratingskala (stimmt nicht=1 bis 4=stimmt genau), inwiefern die Lehrkräfte den Elternkontakt als sehr gut empfinden und inwieweit sie das Elternhaus bzw. die Verhältnisse im Elternhaus kennen. Außerdem wurde mit einer fünfstufigen Ratingskala nach der Häufigkeit des Kontakts mit den Eltern gefragt. Die Auswahl der Items erfolgte aufgrund der theoretischen Annahmen. Auch die Skala zum Schulkontakt wird in den folgenden Analysen als Prädiktor in das Modell aufgenommen. [Tabelle 14](#) gibt einen Überblick zu den vier Items der Skala und den deskriptiven Statistiken. Die interne Konsistenz beträgt  $\alpha = .839$ .

Tabelle 14: Skala Schulkontakt 4. Klasse (LK)

<b>Schulkontakt (LK)</b> M <sub>Skala</sub> = 2.93; SD <sub>Skala</sub> = .681; N <sub>Skala</sub> = 1231; Cronbachs Alpha: $\alpha$ = .839; N = 1206	<b>M</b>	<b>SD</b>	<b>Korr. Item- Skala-Korre- lation</b>	<b>Cronbachs Alpha, wenn Item weggelassen</b>
Meine Beziehung zu den Eltern des Kindes ist sehr gut	3.05	.797	.663	.805
Ich kenne die Verhält- nisse im Elternhaus des Kindes genau	2.59	.868	.694	.791
Ich kenne die Eltern sehr gut	2.73	.889	.810	.735
Häufigkeit Kontakt mit den Eltern im Verlauf des letzten Schuljahrs	3.36	.760	.547	.850

Anmerkungen: Die Items 1-3 wurden auf einer vierstufigen Skala erfasst (stimmt nicht – stimmt genau). Niedrigere Werte entsprechen einem niedriger eingeschätzten Elternkontakt. Die Häufigkeit des Elternkontakts wurde auf einer fünfstufigen Skala erfasst (nie – sehr häufig). Niedrigere Werte entsprechen einer geringeren Kontakthäufigkeit.

## Bildungsaspiration

Zusätzlich zum Schulkontakt des Elternhauses wurden die Lehrkräfte in der vierten Klasse gefragt, wie wichtig den Eltern des Kindes ein hoher Schulabschluss ist. Dies wurde mit einer sechsstufigen Ratingskala (1=völlig unwichtig bis 6=ganz besonders wichtig) abgefragt und fließt als Prädiktor in die folgenden Analysen ein. Die nachfolgende Tabelle zeigt die deskriptiven Statistiken zur elterliche Bildungsaspiration.

Tabelle 15: Bildungsaspiration Elternhaus 4. Klasse (LK)

<b>Item Bildungsaspiration</b>	<b>N</b>	<b>M</b>	<b>SD</b>
Was meinen Sie: Wie wichtig ist den Eltern des Kindes, dass das Kind einmal einen möglichst hohen Schulabschluss erreichen wird?	1231	4.72	.995

Anmerkungen: Das Item wurde auf einer sechsstufigen Skala erfasst (völlig unwichtig – ganz besonders wichtig). Niedrigere Werte entsprechen einer niedriger eingeschätzten Wichtigkeit.

#### 6.4.4 Übersicht zu den Analysevariablen

Einen Überblick über die in dieser Untersuchung verwendeten Instrumente gibt die nachfolgende [Tabelle 16](#). Alle aufgeführten Bereiche entstammen der Studie Koala-S; hier differenziert nach Erhebungsbereich, eingeschlossenen Variablen bzw. Skalen, der Operationalisierung und Skalierung und dem verwendeten Zeitpunkt der Befragung (MZP). Einschätzungen der Lehrkraft werden entsprechend gekennzeichnet (LK). Auf eine getrennte Darstellung von Merkmalen auf Individual- und Klassenebene wird verzichtet, da auf Klassenebene lediglich aggregierte Individualmerkmale analysiert werden.

Tabelle 16: Übersicht zu den Analysevariablen

Bereich	Variablen	Operationalisierung	Skalierung	MZP
Schulnoten (LK)	Zeugnisnote Mathematik	1 Item	Notenstufen 1 bis 6	4.Klasse
	Zeugnisnote Deutsch	1 Item	Notenstufen 1 bis 6	4.Klasse
Leistungs-tests	Intelligenz (kognitive Grundfähigkeiten)	KFT	z-skaliert	4.Klasse
	Leistungsindex	Ergebnisse der Testleistungen in Lesen, Rechtschreibung und Mathematik	z-skaliert	4.Klasse
Kognitive Schülermerkmale (LK)	Begabung Mathematik	1 Item	Fünfstufige Skala	4.Klasse
	Begabung Deutsch	1 Item	Fünfstufige Skala	4.Klasse
	Sprachliche Fähigkeiten	Skala aus 2 Items	Mittelwert aus 2 sechsstufigen Items	4.Klasse
	Vorwissen	Skala aus 4 Items	Mittelwert aus 4 vierstufigen Items	2.Klasse
Nicht-kognitive Schülermerkmale (LK)	Arbeitsverhalten	Skala aus 5 Items	Mittelwert aus 5 fünfstufigen Items	4.Klasse
	Sozialverhalten	Skala aus 3 Items	Mittelwert aus 3 fünfstufigen Items	4.Klasse
Familiäre Strukturmerkmale	Sozialer Status (ISEI)	ISEI Wert des Vaters (bei fehlend. Wert der Mutter)	3-stufig (niedrig-mittel-hoch)	2.–4. Klasse
	Bildungsstatus	Höchster Schulabschluss im Haushalt	3-stufig (niedrig-mittel-hoch)	2.–4. Klasse
Familiäre Prozessmerkmale (LK)	Schulkontakt	Skala aus 4 Items	Mittelwert aus 4 vierstufigen/ fünfstufigen Items	4.Klasse
	Unterstützungsverhalten	1 Item	Vierstufige Skala	4.Klasse
	Bildungsaspiration	1 Item	Sechsstufige Skala	4.Klasse
Geschlecht		1 Item	Mädchen=0 Jungen=1	2.–4. Klasse

## 6.5 Datenaufbereitung und statistische Auswertung

Dieses Kapitel erläutert das für die Untersuchung gewählte methodische Vorgehen. Dabei geht es im Folgenden vornehmlich darum, die gewählten Analyseverfahren und Richtlinien zur Nachvollziehbarkeit der Ergebnisse darzulegen.

Die Datenaufbereitung der vorliegenden Koala-S-Studie erfolgte durch das IEA-DPC in Hamburg und durch die Projektgruppe an der Ludwig-Maximilians-Universität München. Die Ergebnisanalysen der vorliegenden Arbeit wurden mithilfe der Statistiksoftware IBM SPSS Statistics 23 für deskriptive und regressionsanalytische Auswertungen sowie mit MPLUS 7 für komplexere (Mehrebenen-)Strukturgleichungsmodellierungen (Muthén & Muthén, 2012) durchgeführt. Zur Überprüfung der genannten Fragestellungen wird in einem ersten Schritt auf deskriptiver Basis die korrelative Zusammenhangsstruktur zwischen den verwendeten Prädiktoren analysiert; die Darstellung orientiert sich dabei an der theoretischen Untergliederung der Arbeit. Als Maß für den Zusammenhang werden Korrelationskoeffizienten berichtet, wobei Werte von .00 bis .20 sehr schwache Korrelationen; .20 bis .40 schwache Korrelationen; .40 bis .60 mittlere Korrelationen; .60 bis .80 starke Korrelationen und .80 bis 1.00 als sehr starke bzw. perfekte Korrelationen interpretiert werden (Brosius, 2013). Für eine vertiefte Analyse der Hypothesen werden nachfolgend schrittweise Regressionsanalysen durchgeführt, um u.a. Moderator-effekte zu prüfen. Die Aufnahme der eingeführten Prädiktoren erfolgt theoriegeleitet nach deren vermuteten Regressionsgewichten. Die Modellgüte der Regressionsmodelle wird mithilfe des korrigierten  $R^2$  bewertet. Zur inferenzstatistischen Signifikanzprüfung wird das Alpha-Niveau auf 5 Prozent (.05) festgesetzt. Bedeutsame Zusammenhänge werden dann mit Strukturgleichungsmodellen berechnet. Die Güte der Modellanpassung wird vornehmlich mit dem Root Mean Square Error of Approximation (RMSEA), Standardized-Root-Mean-Square-Residual (SRMR) und dem Comparative-Fit-Index (CFI) beurteilt. Der RMSEA berücksichtigt Stichprobengröße und Freiheitsgrade, Werte  $\leq 0.05$  entsprechen einem guten Modell. Zur Gesamtbewer-

Die Analyse der Residuen zeigt ein SRMR-Koeffizient von 0, was einen sehr guten Modell-Fit anzeigt, aber auch Werte unter  $\leq .08$  gelten noch als akzeptabel. Als inkrementeller Fit-Index gibt der CFI die Passung des postulierten Modells mit dem Unabhängigkeitsmodell an, höhere Werte ( $\geq .95$ ) geben eine bessere Modellpassung an (Christ & Schlüter, 2010; Geiser, 2011; Hu & Bentler, 1999). Fehlende Werte wurden mithilfe des Full-Information-Maximum-Likelihood-Verfahrens (FIML) in MPLUS geschätzt (Muthén & Muthén, 2012; Reinecke, 2005).



## 7 Ergebnisse

Im folgenden Abschnitt werden die Ergebnisse der Analysen vorgestellt, welche die in [Kapitel 5](#) vorgestellten Fragestellungen zu beantworten suchen. Gefragt wird erstens, welche Bedeutung individuelle kognitive und nicht-kognitive Schülermerkmale für die Notenzusammensetzung in Deutsch und Mathematik am Ende der Grundschulzeit haben, um dann deren Zusammenhänge untereinander und mit familiären sowie klassenbezogenen Hintergrundmerkmalen zu klären. In einem ersten Teil werden dazu zunächst die deskriptiven Zusammenhänge innerhalb und zwischen den Prädiktorvariablen vorgestellt. In einem zweiten Teil werden dann Effekte der individuellen kognitiven und nicht-kognitiven Schülermerkmale auf die Deutsch- und Mathematiknote sowie deren Interaktionseffekte geprüft. In [Kapitel 7.3](#) bzw. [7.4](#) werden dann die gesamten Zusammenhänge des theoretisch postulierten Modells (vgl. [Kap. 4.5](#)) geprüft; fokussiert werden die Zusammenhänge zwischen familiären Struktur- und Prozessmerkmalen, individuellen Schülermerkmalen und Viertklassnoten. Eine Trennung nach Fachnoten wird in den Modell- und Zwischenüberschriften explizit kenntlich gemacht.

### 7.1 Deskriptive Befunde: Individuelle und familiäre Bedingungsfaktoren von Schulnoten

Nachfolgend werden die deskriptiven Merkmale sowie die Zusammenhänge zwischen kognitiven und nicht-kognitiven Schülermerkmalen und den Viertklassnoten in Mathematik und Deutsch aufgezeigt. Ebenfalls auf deskriptiver Basis folgt eine Analyse des Zusammenhangs zwischen familiären Hintergrundmerkmalen und Schulnoten sowie zwischen allen genannten Prädiktorvariablen. Nachfolgend wird zunächst die Verteilung der Fachnoten in Deutsch und Mathematik dargestellt.

### 7.1.1 Schulnoten in Deutsch und Mathematik am Ende der Grundschulzeit

Während [Kapitel 6.4.1](#) die Mittelwerte der Schulnoten in Deutsch und Mathematik aufzeigt (Deutsch:  $M=2.46$ ,  $SD=.838$ ; Mathematik:  $M=2.51$ ,  $SD=.890$ ), geben die nachfolgenden Abbildungen einen Überblick zu der Verteilung der Zeugnisnoten in der vierten Grundschulstufe.

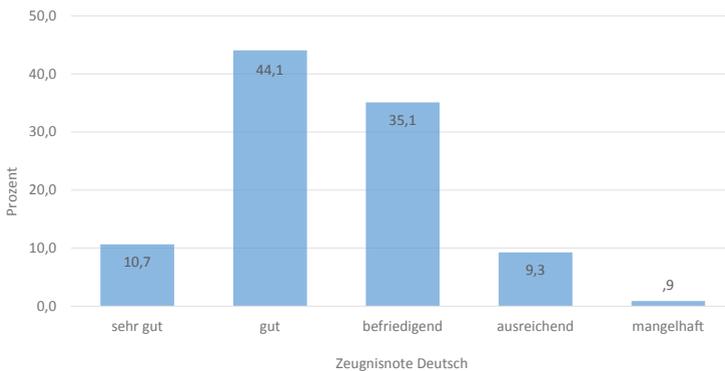


Abbildung 4: Verteilung Zeugnisnoten in Deutsch in Klasse 4

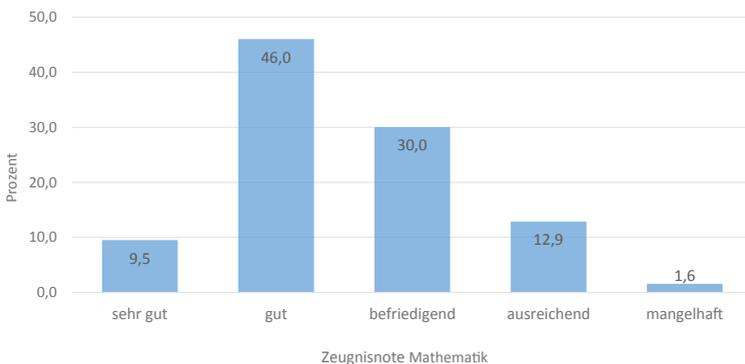


Abbildung 5: Verteilung Zeugnisnoten in Mathematik in Klasse 4

Die Verteilung der Notenskala von Note 1-6 ergab für die vorliegende Stichprobe ( $N=1219$ ), dass kein Schüler der vierten Klasse in Deutsch und Mathematik mit einer ungenügenden Note 6 bewertet wurde.

Ebenfalls ersichtlich wird, dass die meisten Schüler im guten bis befriedigenden Bewertungsrahmen liegen: 44,1 Prozent der Schüler erhalten eine gute Zeugnisnote in Deutsch, im Fach Mathematik trifft das für 46 Prozent der Fälle zu. Gleichzeitig erhalten etwas mehr Schüler in Deutsch eine sehr gute Note als im Fach Mathematik.

### 7.1.2 Kognitive und nicht-kognitive Schülermerkmale und Schulnoten

In einem nächsten Schritt geht es um die deskriptiven Merkmale der kognitiven und nicht-kognitiven Schülereigenschaften, die mit Ausnahme des Vorwissens aus Klassenstufe zwei alle aus Angaben der vierten Grundschulklasse stammen.

#### Kognitive Schülermerkmale

Die kognitiven Schülermerkmale der Untersuchung bestehen sowohl aus objektiv erhobenen Leistungstests (Allg. Leistungsindex aus Rechtschreibung, Leseverständnis, Mathematik und kognitiver Grundfähigkeitstest) als auch aus Lehrkräfteeinschätzungen zu Merkmalen des Kindes (Begabung in Mathematik/Deutsch, sprachliche Fähigkeiten, Vorwissen). Während höhere Werte in den objektiv erhobenen Testleistungen und dem eingeschätzten Vorwissen bessere Leistungen kennzeichnen, sind die anderen eingeschätzten kognitiven Merkmale gemäß einer Notenskala zu lesen; niedrigere Werte deuten bessere Einschätzungen an.

Tabelle 17: Kognitive Schülermerkmale

Kognitive Schülermerkmale	M	SD	N
Testleistung	1.17	.675	1240
Intelligenz (kft)	.99	1.065	1210
Begabung Deutsch	2.66	.944	1231
Begabung Mathematik	2.67	.967	1231
Sprachliche Fähigkeiten	2.37	.944	1230
Vorwissen	2.47	.688	1088

Anmerkungen: Intelligenz (kft), Testleistungen: z-standardisierten Koeffizienten, höhere Werte entsprechen besseren Ergebnissen; Sprachliche Fähigkeiten, fachspezifische Begabungen: Niedrigere Werte entsprechen höheren Einschätzungen; Vorwissen: Höhere Werte entsprechen höheren Einschätzungen.

Die Übersicht zu den deskriptiven Eigenschaften der eingeschätzten Schülermerkmale zeigt, dass diese durchschnittlich ähnlich gut bewertet werden, die Begabung in Deutsch und Mathematik allerdings einen tendenziell höheren Mittelwert zeigt, diese also schlechter bewertet wurden.

### Nicht-kognitive Schülermerkmale

Die nicht-kognitiven Schülermerkmale der Untersuchung bestehen aus Lehrkräfteeinschätzungen zu Merkmalen des Kindes. Die Skalen der nicht-kognitiven Merkmale sind gemäß einer Notenskala zu lesen; niedrigere Werte geben bessere Einschätzungen wieder.

Tabelle 18: Nicht-kognitive Schülermerkmale

Kognitive Schülermerkmale	M	SD	N
Arbeitsverhalten	2.31	.845	1231
Sozialverhalten	1.89	.722	1231

Anmerkungen: Arbeitsverhalten, Sozialverhalten: Niedrigere Werte entsprechen höheren Einschätzungen.

Während das eingeschätzte Arbeitsverhalten äquivalent zu den kognitiven Merkmaleinschätzungen im guten Bereich liegt, zeigt das mittlere Sozialverhalten eine niedrigere (also bessere) Einschätzung an. Die Schüler der Stichprobe zeichnen sich also alle durch ein relativ gutes Sozialverhalten aus.

### Interkorrelationen kognitiver und nicht-kognitiver Schülermerkmale und Schulnoten in Deutsch und Mathematik

Um zu überprüfen, inwiefern die kognitiven und nicht-kognitiven Schülermerkmale untereinander und mit den Viertklassnoten in Deutsch und Mathematik zusammenhängen, wurden Korrelationen zwischen diesen berechnet. [Tabelle 19](#) zeigt die Ergebnisse der Produkt-Moment-Korrelationen nach Pearson<sup>37</sup>.

<sup>37</sup> Die nicht dargestellten Korrelationsanalysen nach Spearman-Rho kamen zu vergleichbaren Ergebnissen.

Tabelle 19: Interkorrelationen kognitiver, nicht-kognitiver Schülermerkmale und Schulnoten in Deutsch und Mathematik

	1	2	3	4	5	6	7	8	9	10
1.Note Deutsch	1	.662**	-.694**	-.401**	.797**	.600**	.806**	-.466**	.677**	.359**
2.Note Mathematik	.662**	1	-.639**	-.456**	.623**	.794**	.621**	-.477**	.571**	.264**
3.Testleistung	-.694**	-.639**	1	.502**	-.704**	-.633**	-.698**	.496**	-.523**	-.274**
4.Intelligenz (kft)	-.401**	-.456**	.502**	1	-.392**	-.465**	-.382**	.320**	-.298**	-.135**
5.Begabung Deutsch	.797**	.623**	-.704**	-.392**	1	.689**	.824**	-.496**	.627**	.323**
6.Begabung Mathematik	.600**	.794**	-.633**	-.465**	.689**	1	.593**	-.473**	.503**	.211**
7.Sprachliche Fähigkeiten	.806**	.621**	-.698**	-.382**	.824**	.593**	1	-.459**	.663**	.390**
8.Vorwissen	-.466**	-.477**	.496**	.320**	-.496**	-.473**	-.459**	1	-.362**	-.225**
9.Arbeitsverhalten	.677**	.571**	-.523**	-.298**	.627**	.503**	.663**	-.362**	1	.582**
10.Sozialverhalten	.359**	.264**	-.274**	-.135**	.323**	.211**	.390**	-.225**	.582**	1

Anmerkungen: N=915-1240; \* p<.05; \*\* p<.01

Betrachtet man zunächst die Zusammenhänge zwischen den eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen fällt u.a. auf, dass die eingeschätzten Merkmale durchweg hochsignifikant mittel bis stark ausgeprägt miteinander zusammenhängen. Das von der Lehrkraft eingeschätzte Sozialverhalten ist dasjenige nicht-kognitive Merkmal, das am schwächsten mit den anderen eingeschätzten kognitiven Schülermerkmalen zusammenhängt. Dagegen korreliert das zweite nicht-kognitive Merkmal Arbeitsverhalten mit Ausnahme der Vorwissenseinschätzung ( $r=-.362$ ,  $p<.01$ ) durchweg auf mittlerem, positiven Niveau mit den anderen eingeschätzten Merkmalen. Insgesamt zeichnen sich für das Vorwissen aus der zweiten Klasse mit Ausnahme der nicht-kognitiven Merkmale mittlere, negative Zusammenhänge mit den eingeschätzten kognitiven Merkmalen aus der vierten

Klasse ab. Der stärkste positive Zusammenhang ergibt sich zwischen der Begabungseinschätzung in Deutsch und der sprachlichen Fähigkeitseinschätzung ( $r=.824$ ,  $p<.01$ ).

Bezüglich der nicht eingeschätzten kognitiven Merkmale ergeben sich für die Testleistung mit Ausnahme des schwachen negativen Effekts mit dem Sozialverhalten ( $r=-.274$ ,  $p<.01$ ) durchweg mittlere bis hohe Korrelationen mit den anderen eingeschätzten Schülermerkmalen. Die Intelligenzleistung (kft) korreliert dagegen lediglich auf schwachem bis mittlerem Niveau mit den weiteren kognitiven bzw. nicht-kognitiven Schülermerkmalen. Der stärkste Zusammenhang ergab sich zwischen Testleistung und der Begabungseinschätzung in Deutsch ( $r=-.704$ ,  $p<.01$ ).

Betrachtet man die Zusammenhänge der individuellen Merkmale der Schüler mit deren Noten in Mathematik und Deutsch fällt auf, dass die Deutschnote am stärksten positiv mit der Merkmalseinschätzung der Lehrkraft zu den sprachlichen Fähigkeiten ( $r=.806$ ,  $p<.01$ ) und der Begabungseinschätzung in Deutsch ( $r=.797$ ,  $p<.01$ ) zusammenhängt. Der Zusammenhang zwischen Deutschnote und Testleistung ist stark negativ ( $r=-.694$ ,  $p<.01$ ). Der Zusammenhang zum eingeschätzten Vorwissen besteht auf mittlerem, negativen Niveau ( $r=-.466$ ,  $p<.01$ ) sowie das Arbeitsverhalten stark positiv mit der Deutschnote korreliert ( $r=.677$ ,  $p<.01$ ). Die Intelligenz (kft) ( $r=-.401$ ,  $p<.01$ ) und das Sozialverhalten ( $r=.359$ ,  $p<.01$ ) weisen die schwächsten Zusammenhänge zur Deutschnote auf, welche jedoch auch hochsignifikant sind.

Bezüglich der Mathematiknote ergibt sich ein ähnliches Bild, allerdings weist die mathematische Begabungseinschätzung den stärksten Zusammenhang zur Mathematiknote auf ( $r=.794$ ,  $p<.01$ ), die sprachlichen Zusammenhänge sind ebenfalls noch hochsignifikant stark positiv ausgeprägt ( $r=.621$ ,  $p<.01$ ). Der Zusammenhang zur Testleistung ist stark negativ ( $r=-.639$ ,  $p<.01$ ). Der Zusammenhang zum eingeschätzten Vorwissen besteht auf mittlerem, negativen Niveau ( $r=-.477$ ,  $p<.01$ ) sowie das Arbeitsverhalten stark positiv mit der Mathematiknote korreliert ( $r=.571$ ,  $p<.01$ ). Die Intelligenz (kft) ( $r=-.456$ ,  $p<.01$ ) und das

Sozialverhalten ( $r=.264, p<.01$ ) weisen die schwächsten, jedoch hochsignifikanten Zusammenhänge zur Mathematiknote auf.

Alle Interkorrelationen gemeinsam betrachtet, zeichnen sich für das eingeschätzte Sozialverhalten, das eingeschätzte Vorwissen und die Testintelligenz ( $kft$ ) die geringsten Zusammenhänge mit den anderen individuellen Schülermerkmalen und deren Noten ab.

### 7.1.3 Familiäre Struktur- und Prozessmerkmale und Schulnoten

Die familiären Struktur- und Prozessmerkmale stammen in der vorliegenden Untersuchung aus Elternangaben zum sozioökonomischen- und höchsten bildungsbezogenen Status der Familie und Lehrkräfteeinschätzungen zu schulbezogenen Prozessmerkmalen. Höhere Werte kennzeichnen einen höheren Sozialstatus und eine höhere (bessere) Einschätzung der Lehrkraft.

Tabelle 20: Struktur- und Prozessmerkmale der Familie

<b>Familiäre Hintergrundmerkmale</b>	<b>M</b>	<b>SD</b>	<b>N</b>
Bildungsstatus	2.18	.712	1397
Sozialer Status (ISEI)	2.03	.773	1375
Schulkontakt	2.93	.681	1231
Unterstützungsverhalten	3.28	.799	1207
Bildungsaspiration	4.72	.995	1231

Anmerkungen: Bildungsstatus, sozialer Status (ISEI), Schulkontakt, Unterstützungsverhalten, Bildungsaspiration: Höhere Werte entsprechen höheren (eingeschätzten) Herkunftsmerkmalen.

Die deskriptiven Kennwerte der familiären Struktur- und Prozessmerkmale zeigen, dass sich die Familien der ausgewählten Schüler im Durchschnitt durch mehrheitlich mittlere Ausprägungen auszeichnen, lediglich das Unterstützungsverhalten und die Bildungsaspiration zeichnen sich durch etwas höhere mittlere Einschätzungen aus.

## Interkorrelationen familiärer Struktur- und Prozessmerkmale und Schulnoten in Deutsch und Mathematik

Um zu überprüfen, inwiefern die familiären Struktur- und Prozessmerkmale untereinander und mit den Schulnoten in Deutsch und Mathematik in der vierten Klasse Grundschule zusammenhängen, wurden Korrelationen zwischen diesen berechnet. Um zu überprüfen, inwiefern die familiären Struktur- und Prozessmerkmale untereinander und mit den Schulnoten in Deutsch und Mathematik in der vierten Klasse Grundschule zusammenhängen, wurden Korrelationen zwischen diesen berechnet. [Tabelle 21](#) gibt einen Überblick zu den Produkt-Moment-Korrelationen nach Pearson. [Tabelle 21](#) gibt einen Überblick zu den Produkt-Moment-Korrelationen nach Pearson<sup>38</sup>.

Tabelle 21: Interkorrelationen familiärer Struktur- und Prozessmerkmale und Schulnoten in Deutsch und Mathematik

	1	2	3	4	5	6	7
1.Note Deutsch	1	.662**	-.336**	-.213**	-.207**	-.377**	-.503**
2.Note Mathematik	.662**	1	-.301**	-.192**	-.161**	-.352**	-.448**
3.Bildungsstatus	-.336**	-.301**	1	.437**	.143**	.217**	.379**
4.Sozialer Status (ISEI)	-.213**	-.192**	.437**	1	.147**	.192**	.277**
5.Schulkontakt	-.207**	-.161**	.143**	.147**	1	.567**	.307**
6.Unterstützungsverhalten	-.377**	-.352**	.217**	.192**	.567**	1	.499**
7.Bildungsaspiration	-.503**	-.448**	.379**	.277**	.307**	.499**	1

Anmerkungen: N=1160-1397; \* p<.05; \*\* p<.01

Die Korrelationsanalysen zwischen den beiden Strukturmerkmalen geben zunächst an, dass der Bildungsstatus und der soziale Status der Familie auf mittlerem, hochsignifikanten Niveau positiv zusammenhängen ( $r=.437$ ,  $p<.01$ ). Zwischen den Prozessmerkmalen ergeben sich schwache- bis mittel ausgeprägte Zusammenhänge. Am stärksten positiv korreliert das von der Lehrkraft eingeschätzte Unterstützungsverhalten mit dem ebenfalls durch die Lehrkraft eingeschätzten

38 Die nicht dargestellten Korrelationsanalysen nach Spearman-Rho kamen zu vergleichbaren Ergebnissen.

Schulkontakt ( $r=.567, p<.01$ ) und der elterlichen Bildungsaspiration ( $r=.499, p<.01$ ). Dagegen ergibt sich zwischen dem Schulkontakt und der Bildungsaspiration ein lediglich schwacher, positiver Zusammenhang ( $r=.307, p<.01$ ).

Bezüglich der Beziehung zwischen den Struktur- und Prozessmerkmalen der Familie ergeben sich insgesamt schwache, positive Zusammenhänge, wobei insgesamt der Zusammenhang zwischen den familiären Prozessmerkmalen und dem Bildungsstatus etwas stärker ausfällt als für den sozialen Status der Familie. Während der Schulkontakt am schwächsten positiv mit den beiden Strukturmerkmalen zusammenhängt, korreliert der Bildungsstatus der Familie am stärksten positiv mit der eingeschätzten elterlichen Bildungsaspiration ( $r=.379, p<.01$ ). Dieser Zusammenhang ist analog zum Unterstützungsverhalten ( $r=.217, p<.01$ ) allerdings auch schwach ausgeprägt.

Betrachtet man die Beziehungen der familiären Struktur- und Prozessmerkmale mit den Grundschulnoten der vierten Klasse in Deutsch und Mathematik, ergeben sich schwache bis mittlere negative Zusammenhänge. Bezüglich der Strukturmerkmale weist der Bildungsstatus im Vergleich zum sozialen Status der Familie einen stärkeren negativen Zusammenhang zu den Noten in Deutsch ( $r=-.336, p<.01$ ) und Mathematik ( $r=-.301, p<.01$ ) auf. Bezüglich der familiären Prozessmerkmale korreliert die Deutschnote am stärksten negativ mit der eingeschätzten Bildungsaspiration der Eltern ( $r=-.503, p<.01$ ), am schwächsten hängt die Deutschnote mit dem elterlichen Schulkontakt zusammen ( $r=-.207, p<.01$ ). Der Zusammenhang zum elterlichen Unterstützungsverhalten besteht für die Deutschnote ebenfalls auf schwachem negativen Niveau ( $r=-.377, p<.01$ ). Für die Mathematiknote ergibt sich ein ähnliches Bild: am stärksten negativ korreliert die von der Lehrkraft eingeschätzte Bildungsaspiration mit der Mathematiknote ( $r=-.448, p<.01$ ), der Zusammenhang zum Schulkontakt ist schwach negativ ausgeprägt ( $r=-.161, p<.01$ ). Der Zusammenhang zum elterlichen Unterstützungsverhalten ist für die Mathematiknote ebenfalls schwach negativ ausgeprägt ( $r=-.352, p<.01$ ).

### 7.1.4 Beziehung zwischen individuellen und familiären Bedingungsfaktoren

Nachdem die familiären Struktur- und Prozessmerkmale untereinander und im Zusammenhang zu den Schulnoten in Deutsch und Mathematik dargestellt wurden, gilt es in einem nächsten Schritt, die Beziehung zwischen den familiären Merkmalen und den individuellen Schülermerkmalen am Ende der vierten Klasse zu analysieren. Dazu werden zunächst die Ergebnisse der Interkorrelationen vorgestellt, um anschließend auf deskriptiver Basis zu prüfen, inwiefern die Merkmalseinschätzungen in Abhängigkeit von ausgewählten familiären Hintergrundmerkmalen variieren.

#### Interkorrelationen individueller und familiärer Bedingungsfaktoren

Um die Beziehung zwischen den individuellen und familiären Bedingungsfaktoren der vorliegenden Untersuchung zu klären, werden zunächst in [Tabelle 22](#) die Ergebnisse der Produkt-Moment-Korrelationen nach Pearson vorgestellt<sup>39</sup>.

Tabelle 22: Interkorrelationen individueller und familiärer Bedingungsfaktoren

	Bildungs- status	Sozialer Status (ISEI)	Schul- kontakt	Unterstützungs- verhalten	Bildungs- aspiration
Testleistung	.302**	.222**	.087**	.277**	.392**
Intelligenz (kft)	.229**	.118**	.043	.141**	.202**
Begabung Deutsch	-.310**	-.236**	-.175**	-.345**	-.446**
Begabung Mathematik	-.273**	-.224**	-.146**	-.305**	-.389**
Sprachliche Fähigkeiten	-.350**	-.235**	-.230**	-.385**	-.509**
Vorwissen	.196**	.102**	.051	.147**	.210**
Arbeitsverhalten	-.243**	-.167**	-.294**	-.533**	-.525**
Sozialverhalten	-.127**	-.102**	-.245**	-.381**	-.233**

Anmerkungen: N=904-1397; \* p<.05; \*\* p<.01

39 Die nicht dargestellten Korrelationsanalysen nach Spearman-Rho kamen zu vergleichbaren Ergebnissen.

Die Zusammenhänge zwischen den familiären Struktur- und Prozessmerkmalen und den kognitiven und nicht-kognitiven Schülermerkmalen ergeben, dass die objektiv erhobenen Testleistungen des Kindes (Leistungsindex, kft) insgesamt schwach, aber hochsignifikant positiv mit den Struktur- und Prozessmerkmalen der Familie zusammenhängen.

Bezüglich der eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale ergeben sich unterschiedliche Ergebnisse für die Struktur- und schulbezogenen Prozessmerkmale der Familie. Während die eingeschätzten Individualmerkmale insgesamt schwach, aber hochsignifikant mit den familiären Strukturmerkmalen korrelieren, ergeben sich für die ebenfalls von der Lehrkraft eingeschätzten Prozessmerkmale stärkere Zusammenhänge. Insgesamt hängt die Bildungsaspiration der Eltern am stärksten und hochsignifikant negativ mit der Begabungseinschätzung in Deutsch ( $r = -.446$ ,  $p < .01$ ), der sprachlichen Fähigkeitseinschätzung ( $r = -.509$ ,  $p < .01$ ) und dem Arbeitsverhalten des Schülers ( $r = -.525$ ,  $p < .01$ ) zusammen. Ebenfalls auf mittlerem Korrelationsniveau besteht eine hochsignifikante negative Beziehung zwischen dem Arbeitsverhalten und dem Unterstützungsverhalten des Elternhauses ( $r = -.533$ ,  $p < .01$ ). Den insgesamt schwächsten Zusammenhang zu den individuellen Schülermerkmalen ergibt der durch die Lehrkraft eingeschätzte Schulkontakt. Für den Zusammenhang zwischen eingeschätztem Schulkontakt, der Intelligenzleistung (kft) und dem eingeschätzten Vorwissen der zweiten Klasse ergeben sich nur sehr schwache und nicht signifikante Zusammenhänge.

Obwohl die Interkorrelationen zwischen den Struktur- und Prozessmerkmalen und den individuellen Schülermerkmalen insgesamt eher mittel bis schwach ausgeprägt sind, ergaben sich bilanzierend zwischen den eingeschätzten Indikatoren von Schulleistung höhere Zusammenhänge als zwischen den Testvariablen der Schüler.

In einem nächsten Schritt wurde auf deskriptiver Basis geprüft, inwiefern die Merkmalseinschätzung der kognitiven und nicht-kognitiven Schülermerkmale in Abhängigkeit von den familiären Hintergrund-

merkmalen variieren. Da die vorgestellten Interkorrelationen (vgl. [Tabelle 22](#)) der Struktur- und Prozessmerkmale mit den individuellen Schülermerkmalen für den Bildungsstatus und die eingeschätzte Bildungsaspiration die höchsten Korrelationen ergaben, werden diese exemplarisch dargestellt.

[Abbildung 6](#) und [Abbildung 7](#) geben einen Überblick über die vorliegenden Verteilungen zwischen dem Bildungsstatus der Familie (dreistufig), den von der Lehrkraft eingeschätzten elterlichen Bildungsaspirationen (sechstufig) und den eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen. Für die Abbildungen gilt zu beachten, dass mit Ausnahme der Vorwissenseinschätzung aus der zweiten Klassenstufe alle Merkmalseinschätzungen gemäß einer Notenskala zu interpretieren sind, höhere Werte also eine schlechtere Einschätzung implizieren. Das Vorwissen des Schülers zu Beginn der zweiten Klasse ist gegenläufig kodiert, höhere Werte geben also eine bessere Einschätzung an.

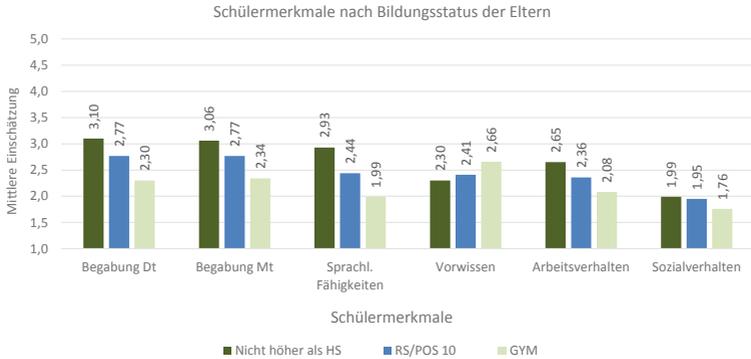


Abbildung 6: Lehrkräfteeinschätzung Schülermerkmale 4. Klasse nach Bildungsstatus der Eltern

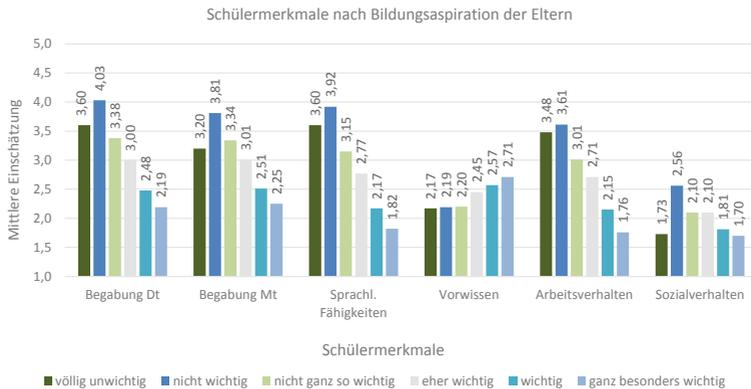


Abbildung 7: Lehrkräfteeinschätzung Schülermerkmale 4. Klasse nach Bildungsaspiration der Eltern

Wie in [Abbildung 6](#) ersichtlich, haben Schüler, deren Eltern über einen Gymnasialabschluss verfügen, in ihren kognitiven und nicht-kognitiven Schülermerkmalen im Schnitt niedrigere, also bessere Einschätzungen als Kinder, deren Eltern maximal über einen Hauptschulabschluss oder einen Realschulabschluss verfügen. Die stärksten Unterschiede ergeben sich in der sprachlichen Fähigkeitseinschätzung, gefolgt

von den fachspezifischen Begabungseinschätzungen. Die geringsten Abweichungen ergeben sich für das Sozialverhalten.

**Abbildung 7**, welche die deskriptive Verteilung der von der Lehrkraft eingeschätzten Schülermerkmale nach der Beurteilung der elterlichen Bildungsaspiration darstellt, zeigt ähnliche Tendenzen. Schüler, deren Eltern von der Lehrkraft mit einer hohen Wichtigkeit eines hohen Schulabschlusses (*ganz besonders wichtig*) eingeschätzt werden, haben in diesen Daten bessere Merkmalseinschätzungen. Die Schüler der Eltern, bei welchen dies als *nicht wichtig* wahrgenommen wurde, zeichnen sich durch die schlechtesten Merkmalsausprägungen aus.

Obwohl durch die Betrachtung der deskriptiven Verteilungen keinerlei Aussage über signifikante Mittelwertsunterschiede zwischen den Gruppen gemacht werden kann, zeigen die Abbildungen durchaus eine gewisse Tendenz. Durchschnittlich betrachtet verfügen Schüler, deren Eltern über einen hohen Bildungsabschluss sowie über eine hohe eingeschätzte Bildungsaspiration verfügen über bessere Merkmalseinschätzungen als ihre Mitschüler. Inwiefern diese Unterschiede kausal zu interpretieren sind, gilt es in den nachfolgenden Analysen zu klären.

Bilanzierend ergeben die deskriptiven Ergebnisse signifikante Zusammenhänge zwischen den individuellen und familiären Merkmalen einerseits und andererseits mit den Noten der vierten Jahrgangsstufe in Deutsch und Mathematik. Um zu überprüfen, welche individuellen und familiären Indikatoren von Schulleistung wie stark auf die Notenzusammensetzung am Ende der Grundschulzeit wirken, werden nachfolgend die Ergebnisse der Regressionsanalysen vorgestellt.

## 7.2 Effekte der individuellen Schülermerkmale auf die Deutsch- und Mathematiknote der 4. Klasse

Während die deskriptiven Ergebnisse bereits erste Aussagen zum Zusammenhang zwischen den Prädiktoren und den Deutsch- und Mathematiknoten der 4. Grundschulklasse zulassen, werden im Nachfolgenden zunächst auf regressionsanalytischer Basis kausale Effekte der kognitiven und nicht-kognitiven Schülermerkmale für die Notenzusammensetzung vorgestellt.

Gefragt wird, welche Relevanz testbasierte und lehrkraftspezifische Einschätzungen von Schülermerkmalen für die Notenzusammensetzung haben und inwiefern sich Unterschiede zwischen den Fächern Deutsch und Mathematik ergeben. Die abhängigen Variablen werden dabei im Folgenden – wie in der Forschungsgemeinschaft weit verbreitet – als intervallskalierte Variablen behandelt. Hier nicht dargestellte Berechnungen mit ordinalen abhängigen Variablen kamen allerdings zu vergleichbaren Ergebnistendenzen. Der Aufbau der hierarchischen Regressionsanalysen erfolgt gemäß der Forschungsbasis nach der Stärke der vermuteten Prädiktionskraft; differenziert nach individuellen Testleistungen und eingeschätzten Verhaltensmerkmalen. Gemäß den zugrundeliegenden Hypothesen werden zwei unterschiedliche Berechnungen vorgestellt: Für die Beantwortung von Hypothese 1 umfasst die Regressionsanalyse ausschließlich kognitive und nicht-kognitive Schülermerkmale. In einer weiteren Regression werden zur Beantwortung von Hypothese 2 zusätzlich zu den kognitiven und nicht-kognitiven Schülermerkmalen noch die Interaktionseffekte zwischen den eingeschätzten Schülermerkmalen eingeführt. Die Darstellung erfolgt dabei differenziert nach Deutsch- und Mathematiknote.

### 7.2.1 Effekte kognitiver und nicht-kognitiver Schülermerkmale auf die Deutschnote

Um zu prüfen, inwiefern die von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale einen stärkeren Einfluss auf die Zeugnisnote in Deutsch haben als die individuellen Testleistun-

gen des Schülers und inwiefern kognitive und nicht-kognitive Merkmale in ihrem Einfluss auf die Deutschnote variieren, zeigt [Tabelle 23](#) das Ergebnis der hierarchischen Regression.

Tabelle 23: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale

Variable	B	SE B	$\beta$	R <sup>2</sup> (korr.)
<b>Modell 1</b>				.480**
Testleistung	-.807	.036	-.650**	
Intelligenz (kft)	-.048	.022	-.064*	
Geschlecht	.108	.041	.065**	
<b>Modell 2</b>				.710**
Testleistung	-.195	.036	-.157**	
Intelligenz (kft)	-.018	.016	-.024	
Geschlecht	.082	.031	.049**	
Begabung Deutsch (LK)	.277	.031	.309**	
Sprachliche Fähigkeiten (LK)	.359	.030	.401**	
Vorwissen (LK)	-.061	.027	-.050*	
<b>Modell 3</b>				.729**
Testleistung	-.181	.035	-.146**	
Intelligenz (kft)	-.016	.016	-.021	
Geschlecht	.050	.031	.030	
Begabung Deutsch (LK)	.240	.030	.268**	
Sprachliche Fähigkeiten (LK)	.285	.031	.318**	
Vorwissen (LK)	-.056	.026	-.046*	
Arbeitsverhalten (LK)	.210	.027	.215**	
Sozialverhalten (LK)	-.050	.026	-.043	

Anmerkungen: N=870; \* p<.05; \*\* p<.01

Zur Beantwortung von Hypothese 1 wurden schrittweise die Prädiktoren der Deutschnote in das Regressionsmodell aufgenommen. Modell 1 umfasst die kognitiven Testergebnisse des Kindes (Testleistung, kft), Modell 2 nimmt zusätzlich die eingeschätzten kognitiven Merkmale auf, Modell 3 analysiert zusätzlich die nicht-kogni-

tiven Merkmale, das Geschlecht des Kindes wird in allen Schritten kontrolliert<sup>40</sup>.

Aus [Tabelle 23](#), Modell 3 wird ersichtlich, dass in der hierarchischen Regressionsanalyse die acht Prädiktoren insgesamt 72,9 Prozent der Varianz in der Deutschnote der vierten Klasse erklären.

Insgesamt erklärt das erste Modell 48,0 Prozent der Varianz in der Deutschnote der vierten Klasse (ANOVA:  $F(3,866)=268.74, p<.01$ ). In diesem Modell 1, das zunächst ausschließlich die Testleistung, Intelligenz (kft) und das Geschlecht des Schülers berücksichtigt, ist die Testleistung des Schülers die stärkste Einflusskomponente ( $\beta = -.650, p<.01$ ), die Intelligenz (kft) und das Geschlecht nehmen schwach, aber hochsignifikant Einfluss. In Modell 2, das zusätzlich die Lehrkräfteeinschätzung zu den kognitiven Schülermerkmalen aufnimmt, steigt die Varianzaufklärung auf 71 Prozent an (ANOVA:  $F(6,863)=355.41, p<.01$ ). Das Beta-Gewicht der Testleistung sinkt durch die Hinzunahme der eingeschätzten kognitiven Schülermerkmale stark ab, bleibt aber weiterhin hochsignifikant (von  $\beta = -.650$  auf  $\beta = -.157, p<.01$ ). Die Intelligenz des Schülers nimmt dagegen keinen signifikanten Einfluss mehr, der Effekt des Geschlechts bleibt dagegen hochsignifikant. Von den eingeschätzten kognitiven Schülermerkmalen ergibt sich für die sprachliche Fähigkeitseinschätzung der stärkste positive Effekt auf die Deutschnote ( $\beta = .401, p<.01$ ), ebenso nehmen die anderen kognitiven Merkmale alle signifikant Einfluss.

Im Gesamtmodell (vgl. [Tabelle 23](#), Modell 3), das zusätzlich die nicht-kognitiven Schülermerkmale berücksichtigt, steigt die Varianzaufklärung dann nochmals um knapp 2 Prozent auf 72,9 Prozent an (ANOVA:  $F(8,861)=292.73, p<.01$ ). Während sich für die Testleistung des Kindes weiterhin hochsignifikante, aber insgesamt schwächere negative Effekte auf die Deutschnote ergeben ( $\beta = -.146, p<.01$ ), bleibt die Intelligenz weiterhin nicht signifikant, gleiches gilt für das

---

40 Die Variable Geschlecht wurde Dummy-kodiert (weiblich=0, männlich=1) in die nachfolgenden Regressionsanalysen aufgenommen.

kontrollierte Geschlecht. Bezüglich der kognitiven Merkmalseinschätzungen ist der stärkste Prädiktor für die Deutschnote der vierten Klasse die sprachliche Fähigkeitseinschätzung ( $\beta = .318, p < .01$ ), gefolgt von der Begabungseinschätzung in Deutsch ( $\beta = .268, p < .01$ ). Schüler mit gleicher Testleistung, die über eine schlechtere sprachliche Fähigkeitseinschätzung verfügen, erhalten somit eine um .318 höhere, also schlechtere Deutschnote. Für das Vorwissen des Schülers aus der zweiten Klasse ergibt sich ebenfalls ein schwacher, aber signifikant negativer Effekt ( $\beta = -.046, p < .05$ ), was aufgrund der Polung als schwach positiver Effekt zu deuten ist. Bezüglich der nicht-kognitiven Merkmalseinschätzungen ergeben sich erwartungsgemäß stärkere Effekte für das Arbeitsverhalten als für das Sozialverhalten des Schülers ( $\beta = .215, p < .01$  vs.  $\beta = -.043, n.s.$ ). Auffällig ist, dass unter Berücksichtigung der nicht-kognitiven Schülermerkmale der Effekt des Geschlechts nicht signifikant wird, der Einfluss also durch das Arbeitsverhalten des Kindes vermittelt zu werden scheint (hier nicht dargestellte Analysen unter ausschließlicher Berücksichtigung des Arbeitsverhalten ergaben ebenfalls vermittelnde Effekte). Zusätzlich zum Schülergeschlecht ergeben sich im Gesamtmodell zur Erklärung der Deutschnote der vierten Klasse für das Sozialverhalten des Schülers nicht signifikante, schwache Effekte, weswegen die beiden Prädiktoren (Geschlecht, Sozialverhalten) in den Strukturgleichungsmodellierungen zur Erklärung der Deutschnote keine Berücksichtigung mehr finden.

Mittels einer weiteren hierarchischen Regressionsanalyse wird der zweiten Hypothese – der Frage nach zusätzlichen Moderatoreffekten zwischen den eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen auf die Schulnoten – nachgegangen.

Unter Bezug auf die bereits vorgestellten Ergebnisse wird für die nicht-kognitiven Schülermerkmale allein das Arbeitsverhalten für die Interaktionen berücksichtigt, da aufgrund der mangelnden Effekte des Sozialverhaltens auf die Deutschnote nicht von einer zusätzlichen Beeinflussung ausgegangen wird. Die Auswahl der Interaktionen ausschließlich zwischen den *eingeschätzten* kognitiven Schülermerkmalen

und dem Arbeitsverhalten liegt außerdem darin begründet, dass in der vorliegenden Arbeit u.a. nach der Perspektive der Lehrkraft gefragt wird; inwiefern die Wahrnehmung der Lehrkraft zum Arbeitsverhalten des Kindes also den Einfluss der kognitiven Merkmalseinschätzung auf die Schulnoten variiert. Gefragt wird an dieser Stelle konkret, ob die Relevanz der sprachlichen Fähigkeiten, Begabung und des Vorwissens für die Deutschnote in Abhängigkeit von der Einschätzung des Arbeitsverhaltens variiert.

Modelle 1 bis 3 der nachfolgenden Tabelle umfassen dieselben Prädiktorschritte wie das bereits vorgestellte Basismodell zu den individuellen Schülermerkmalen (vgl. [Tabelle 23](#)). Das anschließend eingeführte vierte Modell berücksichtigt zusätzlich dazu die Interaktionsterme zwischen den eingeschätzten kognitiven Schülermerkmalen und dem Arbeitsverhalten des Schülers. [Tabelle 24](#) zeigt die Ergebnisse der hierarchischen Regressionsanalyse zur Überprüfung von Moderatoreffekten auf die Deutschnote.

Tabelle 24: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale unter Berücksichtigung von Interaktionseffekten

Variable	B	SE B	β	R <sup>2</sup> (korr.)
<b>Modell 1</b>				.480**
Testleistung	-.807	.036	-.650**	
Intelligenz (kft)	-.048	.022	-.064*	
Geschlecht	.108	.041	.065**	
<b>Modell 2</b>				.710**
Testleistung	-.195	.036	-.157**	
Intelligenz (kft)	-.018	.016	-.024	
Geschlecht	.082	.031	.049**	
Begabung Deutsch (LK)	.277	.031	.309**	
Sprachliche Fähigkeiten (LK)	.359	.030	.401**	
Vorwissen (LK)	-.061	.027	-.050*	
<b>Modell 3</b>				.729**
Testleistung	-.181	.035	-.146**	
Intelligenz (kft)	-.016	.016	-.021	
Geschlecht	.050	.031	.030	
Begabung Deutsch (LK)	.240	.030	.268**	
Sprachliche Fähigkeiten (LK)	.285	.031	.318**	
Vorwissen (LK)	-.056	.026	-.046*	
Arbeitsverhalten (LK)	.210	.027	.215**	
Sozialverhalten (LK)	-.050	.026	-.043	
<b>Modell 4</b>				.731**
Testleistung	-.183	.035	-.147**	
Intelligenz (kft)	-.016	.016	-.021	
Geschlecht	.051	.031	.031	
Begabung Deutsch (LK)	.244	.030	.272**	
Sprachliche Fähigkeiten (LK)	.294	.031	.329**	
Vorwissen (LK)	-.051	.026	-.042	
Arbeitsverhalten (LK)	.214	.027	.218**	
Sozialverhalten (LK)	-.051	.026	-.044*	
Arbeitsverhalten*Begabung Deutsch	.045	.027	.060	
Arbeitsverhalten*sprachli- che Fähigkeiten	-.067	.026	-.094**	
Arbeitsverhalten*Vorwissen	-.033	.017	-.039	

Modelle 1 bis 3 in [Tabelle 24](#) ergeben dieselben Ergebnisse wie im bereits vorgestellten Basismodell ausschließlich zu den individuellen Schülermerkmalen für die Deutschnote (vgl. [Tabelle 23](#)). Unter Berücksichtigung der Interaktionseffekte ergibt sich im vierten Modell im Vergleich zu Modell 3 eine unmerklich höhere Varianzaufklärung von insgesamt 73,1 Prozent (ANOVA:  $F(11,858)=215.55, p<.01$ ).

Zusätzlich dazu ergeben sich für das vierte Modell – unter Berücksichtigung der Interaktionen – teilweise veränderte direkte Effekte.

Bezüglich der einzelnen kognitiven Schülermerkmale ergeben sich für die Testleistung und die Intelligenz (kft) keine bedeutsamen Unterschiede zu den vorangegangenen Modellen. Ebenso nimmt das Geschlecht des Schülers weiterhin keinen signifikanten Einfluss. Allerdings wird unter Berücksichtigung der Interaktionen der direkte Effekt der sprachlichen Fähigkeiten im Vergleich zu Modell 3 etwas stärker (Modell 3:  $\beta = .318, p<.01$  vs. Modell 4:  $\beta = .329, p<.01$ ). Gleiches gilt für den direkten Effekt der Begabungseinschätzung (Modell 3:  $\beta = .268, p<.01$  vs. Modell 4:  $\beta = .272, p<.01$ ). Zudem wird der ohnehin schwache Effekt des Vorwissens in Modell 4 nicht mehr signifikant (Modell 3:  $\beta = -.046, p<.05$  vs. Modell 4:  $\beta = -.042, n.s.$ ).

Auch die Effekte der nicht-kognitiven Schülermerkmale auf die Deutschnote werden unter Berücksichtigung der Interaktionen etwas stärker. Der hochsignifikant direkte Effekt des Arbeitsverhaltens auf die Deutschnote wird im Vergleich zu Modell 3 in Modell 4 geringfügig stärker (Modell 3:  $\beta = .215, p<.01$  vs. Modell 4:  $\beta = .218, p<.01$ ). Zudem wird der schwache direkte Effekt des Sozialverhaltens auf die Deutschnote signifikant ( $\beta = -.044, p<.05$ ).

Zusätzlich dazu zeigt die Prüfung der Interaktionsterme für die Interaktion zwischen Arbeitsverhalten und sprachlichen Fähigkeiten einen schwachen, hochsignifikant negativen Effekt auf die Deutschnote ( $\beta = -.094, p<.01$ ). Die anderen Interaktionsterme sind nicht signifikant.

Nach der Betrachtung der Prädiktionskraft individueller Schülermerkmale für die Deutschnote der vierten Klasse wird in einem nächsten Schritt deren Effekte auf die Mathematiknote geprüft.

### 7.2.2 Effekte kognitiver und nicht-kognitiver Schülermerkmale auf die Mathematiknote

Die Analyse der Effekte kognitiver und nicht-kognitiver Schülermerkmale auf die Mathematiknote der vierten Klasse verläuft analog zum Vorgehen in [Kapitel 7.2.1](#) zur Erklärung der Deutschnote anhand individueller Schülermerkmale. Um zu prüfen, inwiefern die von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale einen stärkeren Einfluss auf die Halbjahresnote in Mathematik haben als die individuellen Testleistungen des Schülers und inwiefern kognitive und nicht-kognitive Merkmale in ihrem Einfluss auf die Mathematiknote variieren, wird in [Tabelle 25](#) das Ergebnis der hierarchischen Regression für das Fach Mathematik vorgestellt.

Tabelle 25: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale

Variable	B	SE B	$\beta$	R <sup>2</sup> (korr.)
<b>Modell 1</b>				.450**
Testleistung	-.746	.039	-.565**	
Intelligenz (kft)	-.141	.024	-.176**	
Geschlecht	-.212	.045	-.120**	
<b>Modell 2</b>				.665**
Testleistung	-.182	.041	-.138**	
Intelligenz (kft)	-.053	.019	-.066**	
Geschlecht	-.023	.036	-.013	
Begabung Mathematik (LK)	.516	.026	.562**	
Sprachliche Fähigkeiten (LK)	.129	.027	.136**	
Vorwissen (LK)	-.064	.031	-.050*	
<b>Modell 3</b>				.677**
Testleistung	-.177	.041	-.134**	
Intelligenz (kft)	-.056	.018	-.069**	
Geschlecht	-.071	.037	-.040	
Begabung Mathematik (LK)	.481	.026	.524**	
Sprachliche Fähigkeiten (LK)	.050	.030	.052	
Vorwissen (LK)	-.062	.030	-.047*	
Arbeitsverhalten (LK)	.181	.032	.174**	
Sozialverhalten (LK)	-.025	.030	-.020	

Anmerkungen: N=870; \* p<.05; \*\* p<.01

Zur Beantwortung von Hypothese 1 wurden für das Fach Mathematik ebenfalls schrittweise die Prädiktoren der Mathematiknote in das Regressionsmodell aufgenommen. Modell 1 umfasst die kognitiven Testergebnisse des Kindes (Testleistung, kft), Modell 2 nimmt zusätzlich die eingeschätzten kognitiven Merkmale auf, Modell 3 analysiert zudem die nicht-kognitiven Schülermerkmale, das Geschlecht des Kindes wird in allen Schritten kontrolliert<sup>41</sup>.

41 Die Variable Geschlecht wurde Dummy-kodiert (weiblich=0, männlich=1) in die nachfolgenden Regressionsanalysen aufgenommen.

Aus der hierarchischen Regressionsanalyse in [Tabelle 25](#), Modell 3 wird ersichtlich, dass die acht Prädiktoren insgesamt 67,7 Prozent der Varianz in der Mathematiknote der vierten Klasse erklären und damit die Aufklärung etwas geringer ausfällt als für die Deutschnote der vierten Klasse.

Insgesamt erklärt das erste Modell 45,0 Prozent der Varianz in der Mathematiknote der vierten Klasse (ANOVA:  $F(3,866)=237.85$ ,  $p<.01$ ). In diesem ersten Modell, das zunächst ausschließlich die Testleistung, die Intelligenz (kft) und das Geschlecht des Schülers berücksichtigt, ist die Testleistung des Schülers die stärkste Einflusskomponente ( $\beta = -.565$ ,  $p<.01$ ). Konträr zur Erklärung der Deutschnote ergeben sich im ersten Modell insgesamt stärkere Effekte für die Intelligenzleistung ( $\beta = -.176$ ,  $p<.01$ ) und das Geschlecht ( $\beta = -.120$ ,  $p<.01$ ) des Schülers. In Modell 2, das zusätzlich die Lehrkräfteeinschätzung zu den kognitiven Schülermerkmalen umfasst, steigt die Varianzaufklärung auf 66,5 Prozent an (ANOVA:  $F(6,863)=288.23$ ,  $p<.01$ ). Analog zu den Ergebnissen für die Deutschnote sinkt das Beta-Gewicht der Testleistung durch die Hinzunahme der eingeschätzten kognitiven Schülermerkmale stark ab, bleibt aber weiterhin hochsignifikant (von  $\beta = -.565$  auf  $\beta = -.138$ ,  $p<.01$ ). Konträr zur Deutschnote ergeben sich für die Mathematiknote in Modell 2 weiterhin schwache hochsignifikante Effekte der Intelligenz (kft) auf die Mathematiknote. Zusätzlich wird das Geschlecht des Schülers nicht signifikant. Der hochsignifikante Effekt des Geschlechts aus Modell 1 scheint durch die kognitive Fähigkeitseinschätzung der Lehrkraft in Modell 2 vermittelt zu werden. Ebenfalls unterschiedlich zur Deutschnote nimmt von den kognitiven Merkmaleinschätzungen die fachspezifische Begabungseinschätzung durch die Lehrkraft mit Abstand den stärksten, hochsignifikanten Einfluss auf die Mathematiknote der vierten Klasse ( $\beta = .562$ ,  $p<.01$ ), aber auch die sprachlichen Fähigkeiten und das Vorwissen bleiben bedeutsam.

Im Gesamtmodell (vgl. [Tabelle 25](#), Modell 3), das zusätzlich die nicht-kognitiven Schülermerkmale berücksichtigt, steigt die Varianzaufklärung dann um knapp ein Prozent auf 67,7 Prozent an (ANOVA:  $F(8,861)=228.86$ ,  $p<.01$ ). Während sich für die Test-

leistung des Kindes weiterhin hochsignifikante, aber insgesamt schwächere negative Effekte auf die Mathematiknote ergeben ( $\beta = -.134, p < .01$ ), bleibt auch die Intelligenz weiterhin hochsignifikant, das kontrollierte Geschlecht bleibt dagegen unter zusätzlicher Kontrolle der nicht-kognitiven Schülermerkmale nicht signifikant. Erneut konträr zur Deutschnote ist für die Mathematiknote der vierten Klasse weiterhin die fachspezifische Begabungseinschätzung der Lehrkraft der stärkste Prädiktor ( $\beta = .524, p < .01$ ), Effekte der sprachlichen Fähigkeitseinschätzung sind dagegen schwach und nicht signifikant ( $\beta = .052, n.s.$ ). Schüler erhalten bei gleicher Testleistung, aber einer höheren, also schlechteren Begabungseinschätzung eine um .524 höhere, also schlechtere Mathematiknote. Für das Vorwissen des Schülers aus der zweiten Klasse ergibt sich ebenfalls ein schwacher, aber signifikanter Effekt ( $\beta = -.047, p < .05$ ). Bezüglich der nicht-kognitiven Merkmalseinschätzungen ergeben sich auch für die Mathematiknote erwartungsgemäß stärkere Effekte für das Arbeitsverhalten als für das Sozialverhalten des Schülers ( $\beta = .174, p < .01$  vs.  $\beta = -.020, n.s.$ ). Insgesamt sind die Effekte des Arbeitsverhaltens auf die Mathematiknote allerdings schwächer ausgeprägt als für die Deutschnote. Zusätzlich zu Effekten des Schülergeschlechts, welche in Modell 3 weiterhin nicht signifikant bleiben, ergeben sich im Gesamtmodell zur Erklärung der Mathematiknote der vierten Klasse also für die sprachlichen Fähigkeiten und das Sozialverhalten des Schülers nicht signifikante, schwache Effekte, weswegen diese drei Prädiktoren (Geschlecht, sprachliche Fähigkeiten, Sozialverhalten) in den Strukturgleichungsmodellierungen zur Erklärung der Mathematiknote keine Berücksichtigung mehr finden.

Zur Beantwortung der zweiten Hypothese – der Frage nach möglichen Moderatoreffekten zwischen den eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen auf die Schulnoten – werden in [Tabelle 26](#) auch für die Mathematiknote die Ergebnisse einer weiteren hierarchischen Regressionsanalyse vorgestellt.

Analog zum Vorgehen in [Kapitel 7.2.1](#), wird auch für die Mathematiknote für die nicht-kognitiven Schülermerkmale ausschließlich

das Arbeitsverhalten des Schülers für die Interaktionen berücksichtigt, da aufgrund der mangelnden Effekte des Sozialverhaltens für die Mathematiknote nicht von einer zusätzlichen Beeinflussung ausgegangen wird. Die Begründung der Auswahl der Interaktionen für die Mathematiknote erfolgte auf gleicher Argumentationsbasis wie für die Deutschnote. Analysiert wird also auch an dieser Stelle, ob die Effekte der sprachlichen Fähigkeiten, Begabung und des Vorwissens für die Mathematiknote in Abhängigkeit von der Einschätzung des Arbeitsverhaltens variieren.

Modelle 1 bis 3 der nachfolgenden Tabelle umfassen dieselben Prädiktorschritte wie das bereits dargestellte Basismodell für die Mathematiknote unter Berücksichtigung der individuellen Schülermerkmale (vgl. [Tabelle 25](#)). Das zusätzlich betrachtete Modell 4 nimmt außerdem die Interaktionsterme zwischen den eingeschätzten kognitiven Schülermerkmalen und dem Arbeitsverhalten mit in die Analysen auf. Die Ergebnisse der hierarchischen Regressionsanalyse zur Prüfung möglicher Moderatoreffekte auf die Mathematiknote zeigt [Tabelle 26](#).

Tabelle 26: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch kognitive und nicht-kognitive Schülermerkmale unter Berücksichtigung von Interaktionseffekten

Variable	B	SE B	$\beta$	R <sup>2</sup> (korr.)
<b>Modell 1</b>				.450**
Testleistung	-.746	.039	-.565**	
Intelligenz (kft)	-.141	.024	-.176**	
Geschlecht	-.212	.045	-.120**	
<b>Modell 2</b>				.665**
Testleistung	-.182	.041	-.138**	
Intelligenz (kft)	-.053	.019	-.066**	
Geschlecht	-.023	.036	-.013	
Begabung Mathematik (LK)	.516	.026	.562**	
Sprachliche Fähigkeiten (LK)	.129	.027	.136**	
Vorwissen (LK)	-.064	.031	-.050*	
<b>Modell 3</b>				.677**
Testleistung	-.177	.041	-.134**	
Intelligenz (kft)	-.056	.018	-.069**	
Geschlecht	-.071	.037	-.040	
Begabung Mathematik (LK)	.481	.026	.524**	
Sprachliche Fähigkeiten (LK)	.050	.030	.052	
Vorwissen (LK)	-.062	.030	-.047*	
Arbeitsverhalten (LK)	.181	.032	.174**	
Sozialverhalten (LK)	-.025	.030	-.020	
<b>Modell 4</b>				.680**
Testleistung	-.184	.041	-.140**	
Intelligenz (kft)	-.057	.018	-.071**	
Geschlecht	-.069	.037	-.039	
Begabung Mathematik (LK)	.478	.026	.521**	
Sprachliche Fähigkeiten (LK)	.053	.031	.056	
Vorwissen (LK)	-.065	.030	-.050*	
Arbeitsverhalten (LK)	.174	.032	.167**	
Sozialverhalten (LK)	-.023	.030	-.018	
Arbeitsverhalten*Begabung Mathematik	.060	.023	.066**	
Arbeitsverhalten*sprachliche Fähigkeiten	-.019	.020	-.025	
Arbeitsverhalten*Vorwissen	-.013	.020	-.015	

Anmerkungen: N=870; \* p&lt;.05; \*\* p&lt;.01

Modelle 1 bis 3 in [Tabelle 26](#) ergeben dieselben Effekte wie im bereits vorgestellten Basismodell ausschließlich zu den individuellen Schülermerkmalen (vgl. [Tabelle 25](#)). Unter Berücksichtigung der Interaktionseffekte ergibt sich im vierten Modell im Vergleich zu Modell 3 eine unmerklich höhere Varianzaufklärung von insgesamt 68,0 Prozent (ANOVA:  $F(11,858)=168.74, p<.01$ ).

Analog zur Präsentation der Interaktionseffekte für die Deutschnote ergeben sich in Modell 4 unter Berücksichtigung der Interaktionen teilweise veränderte direkte Effekte auf die Mathematiknote.

Bezüglich der einzelnen Prädiktoren zeigen sich für die kognitiven Merkmale Testleistung und Intelligenz (kft) etwas stärkere Effekte auf die Mathematiknote als im vorangegangenen Modell 3, das Geschlecht des Schülers bleibt dagegen weiterhin nicht signifikant. Bezüglich der eingeschätzten kognitiven Merkmale, wird der direkte Effekt der mathematischen Begabung unter Berücksichtigung der Interaktionen etwas schwächer (Modell 3:  $\beta = .524, p<.01$  vs. Modell 4:  $\beta = .521, p<.01$ ). Der ohnehin schwache Effekt der sprachlichen Fähigkeiten bleibt weiterhin nicht signifikant (Modell 3:  $\beta = .052, n.s.$  vs. Modell 4:  $\beta = .056, n.s.$ ). Der direkte Effekt des Vorwissens wird dagegen etwas stärker und bleibt signifikant (Modell 3:  $\beta = -.047, p<.05$  vs. Modell 4:  $\beta = -.050, p<.05$ ).

Bezüglich der nicht-kognitiven Schülermerkmale bleibt in Modell 4 der direkte Effekt des Sozialverhaltens nahezu unverändert und nicht signifikant (Modell 3:  $\beta = -.020, n.s.$  vs. Modell 4:  $\beta = -.018, n.s.$ ). Für das Arbeitsverhalten ergeben sich durch die Aufnahme der Interaktionsterme – äquivalent zur Deutschnote – dagegen etwas veränderte Effekte. Der direkte Effekt des Arbeitsverhaltens auf die Mathematiknotenzusammensetzung wird unter Berücksichtigung der Interaktionen im Modell 4 im Vergleich zu Modell 3 etwas schwächer, bleibt aber weiterhin hochsignifikant (Modell 3:  $\beta = .174, p<.01$  vs. Modell 4:  $\beta = .167, p<.01$ ).

Zusätzlich dazu zeigt die Prüfung der Interaktionsterme für die Interaktion zwischen Arbeitsverhalten und mathematischer Begabungseinschätzung einen schwachen, hochsignifikant positiven Effekt auf die

Mathematiknote ( $\beta = .066, p < .01$ ), dagegen sind die anderen Interaktionsterme nicht signifikant.

### 7.2.3 Zusammenfassung der bisherigen Ergebnisse

Im Vordergrund der vorangegangenen Analysen stand die Frage, welche Rolle individuelle kognitive und nicht-kognitive Schülermerkmale bei der Zusammensetzung der Deutsch- und Mathematiknote der vierten Klasse spielen und wie diese gewichtet werden. Dazu wurden zwei Teilfragen untersucht: Die erste Frage behandelte die Effekte der kognitiven und nicht-kognitiven Schülermerkmale für die fachspezifische Notenzusammensetzung. Die zweite Teilfrage berücksichtigte dann in einem zusätzlichen Regressionsmodell mögliche Interaktionseffekte zwischen den individuellen Schülermerkmalen.

Bezüglich der ersten Frage wurde fachübergreifend ersichtlich, dass die eingeschätzten kognitiven Schülermerkmale insgesamt den stärksten Erklärungswert für die Schulnoten der vierten Klasse einnehmen, die individuelle Testleistung also in den Gesamtmodellen weniger bedeutsam wird. Die deutliche Abnahme der Effektstärke der Testleistung durch Hinzunahme der Merkmaleinschätzungen spricht allerdings dafür, dass diese teilweise über die individuellen Testleistungen vermittelt werden und sich damit durch einen gemeinsamen Varianzanteil auszeichnen. Auch für die Vorwissenseinschätzung der Lehrkraft aus der zweiten Klasse ergeben sich schwache, aber signifikante Effekte für die Notenzusammensetzung in der vierten Klasse. Zudem ergeben sich ebenfalls fachübergreifend insgesamt stärkere Effekte für die kognitiven als für die nicht-kognitiven Schülermerkmale. Für die einbezogenen nicht-kognitiven Schülermerkmale wird das Arbeitsverhalten insgesamt für die Notenzusammensetzung bedeutsamer als das Sozialverhalten des Schülers.

Zudem wurden fachspezifische Unterschiede in der Notenzusammensetzung der Viertklassnoten in Deutsch und Mathematik gefunden.

Für die Deutschnote der vierten Klasse ergeben sich insgesamt die stärksten, hochsignifikanten Effekte für die sprachliche Fähigkeits-

einschätzung der Lehrkraft, etwas weniger stark, aber dennoch hochsignifikant findet auch die Begabungseinschätzung im Fach Deutsch Berücksichtigung. Nicht signifikante Ergebnisse zeigen sich im Fach Deutsch für die Prädiktoren Intelligenz (kft), Sozialverhalten und Geschlecht des Schülers, allerdings zeichnen sich für das Schülergeschlecht Vermittlungseffekte durch die nicht-kognitive Fähigkeitseinschätzung ab.

Dagegen ist bei der Erklärung der Mathematiknote die mathematische Begabungseinschätzung der Lehrkraft mit Abstand der einflussstärkste Prädiktor, die sprachlichen Fähigkeiten sind im Gegensatz zur Deutschnote für die Mathematiknotenzusammensetzung nahezu irrelevant. Parallel zur Deutschnote zeichnen sich für das Geschlecht und das Sozialverhalten des Schülers keine signifikanten Effekte ab, allerdings wird der Geschlechtereffekt für die Mathematiknote bereits durch die Hinzunahme der kognitiven Merkmalsbewertungen nicht signifikant. Insgesamt ist die Varianzaufklärung der Deutschnote durch die einbezogenen Prädiktoren etwas höher.

Bezüglich der zweiten Teilfragestellung, der Frage nach zusätzlichen Interaktionseffekten zwischen den eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen auf die Notenzusammensetzung, ergeben sich ebenfalls fachspezifische Unterschiede.

Für die Deutschnote zeigen sich unter Berücksichtigung der Interaktionen für die Testleistung und der Intelligenz (kft) im Vergleich zum Basismodell keine bedeutsamen Effektunterschiede. Allerdings werden die direkten Effekte der sprachlichen Fähigkeiten und der Begabungseinschätzung etwas stärker sowie der ohnehin schwache Effekt des Vorwissens unter Berücksichtigung der Interaktionen nicht mehr signifikant wird. Zusätzlich dazu wird der direkte Effekt des Arbeitsverhaltens unter Berücksichtigung der Interaktionen etwas stärker. Bezüglich der einbezogenen Interaktionsterme ergibt sich zudem für den Interaktionsterm zwischen Arbeitsverhalten und sprachlichen Fähigkeiten ein schwacher, hochsignifikant negativer Effekt. Diese Interaktion ist im Sinne eines leicht abschwächenden Effekts zu

deuten: Berücksichtigt man im Modell also neben den beiden Haupteffekten (direkte Pfade von Arbeitsverhalten, sprachliche Fähigkeiten) noch den zusätzlichen Effekt zwischen sprachlichen Fähigkeiten und Arbeitsverhalten, kann dieser Koeffizient im Sinne eines leicht kompensatorischen Effekts interpretiert werden. Dies bedeutet, dass ein Schüler, der in seinen sprachlichen Fähigkeiten und seinem Arbeitsverhalten sehr gut eingeschätzt wird, unter Berücksichtigung der Interaktion nicht eine noch bessere Deutschnote erhalten kann. Damit fällt für den ohnehin guten Schüler der kompensatorische Effekt der Interaktion etwas geringer aus als für einen Schüler, der schlecht eingeschätzt wird. Im Vergleich zwischen den eingeschätzten Koeffizienten bleiben die sprachlichen Fähigkeiten allerdings weiterhin am bedeutsamsten.

Für die Mathematiknote zeigt sich ein etwas anderes Bild. Unter Einbezug der Interaktionen ergeben sich für die kognitiven Merkmale Testleistung und Intelligenz (kft) etwas stärkere direkte Effekte auf die Mathematiknote. Bezüglich der eingeschätzten kognitiven Merkmale wird der direkte Effekt der Begabungseinschätzung etwas schwächer. Die ohnehin schwachen Effekte von Vorwissen und sprachlichen Fähigkeiten werden etwas stärker. Bezüglich der nicht-kognitiven Schülermerkmale wird der direkte Effekt des Arbeitsverhaltens unter Berücksichtigung der Interaktionen etwas schwächer, nimmt aber weiterhin hochsignifikant Einfluss, während das Sozialverhalten weiterhin schwach und nicht signifikant ist.

Der direkte Effekt des Arbeitsverhaltens wird unter Berücksichtigung der Interaktionen in seiner Bedeutung für die Notenzusammensetzung also etwas abgeschwächt, dagegen bleibt der direkte, starke Effekt der mathematischen Begabung nahezu unverändert. Bezüglich der Interaktionsterme nimmt die Interaktion zwischen der mathematischen Begabungseinschätzung und dem Arbeitsverhalten zusätzlich hochsignifikant positiven Einfluss auf die Mathematiknote. Der positive Interaktionsterm ist im Sinne eines verstärkenden Effekts zu deuten. Berücksichtigt man im Modell also neben den beiden Haupteffekten (direkte Pfade von Arbeitsverhalten, mathematische Begabung)

noch den zusätzlichen Effekt zwischen mathematischer Begabung und Arbeitsverhalten, ist die Begabung eines Schülers wesentlich bedeutsamer für die Mathematiknote als dessen Arbeitsverhalten.

Nachdem nun die erste Fragestellung der vorliegenden Arbeit analysiert wurde, wird nachfolgend die zusätzliche Relevanz familiärer Hintergrundmerkmale für die Notenzusammensetzung der Deutsch- und Mathematiknote der vierten Jahrgangsstufe beleuchtet.

### 7.3 Effekte der familiären und individuellen Hintergrundmerkmale auf die Deutsch- und Mathematiknote der 4. Klasse

Während das vorangegangene Kapitel bereits erste Aussagen über den kausalen Zusammenhang zwischen individuellen kognitiven und nicht-kognitiven Schülermerkmalen und den Schulnoten der vierten Jahrgangsstufe in Deutsch und Mathematik zulässt, wird im Folgenden nun die zusätzliche Relevanz familiärer Hintergrundmerkmale analysiert. Gemäß der theoretischen Basis der vorliegenden Arbeit wird hierbei zwischen Struktur- und Prozessmerkmalen familiärer Herkunft unterschieden. Während die Strukturmerkmale aus Elternangaben stammen und als relativ konstant angenommen werden, handelt es sich bei den einbezogenen familiären Prozessmerkmalen ausschließlich um Lehrkräfteeinschätzungen zu schulbezogenen Familienmerkmalen. Bewusst ist dabei, dass diese sicherlich zeitlich veränderlichen Strukturen unterliegen, dennoch ermöglichen sie einen ersten Einblick in, zu diesem Zeitpunkt, kausale Zusammenhangsstrukturen. Gefragt wird also, ob die familiären Hintergrundmerkmale eine eigenständige Bedeutung für die Notenzusammensetzung haben oder ob sie vollständig durch die Einschätzung individueller Schülermerkmale mediiert werden. Ebenso wird beleuchtet, inwiefern die Strukturmerkmale die familiären Prozessmerkmale beeinflussen und ob sich unterschiedliche Effekte auf Testleistungen, Merkmalseinschätzungen und Schulnoten ergeben. Während für die Beantwortung dieser zweiten Fragestellung eine erste regressionsanalytische Untersuchung schrittweise die kau-

salen Effekte für die Notenzusammensetzung in Deutsch und Mathematik betrachtet, folgt daran anschließend in [Kapitel 7.4](#) mithilfe von Strukturgleichungsmodellierungen eine Analyse der komplexeren Zusammenhangsstrukturen, erneut getrennt dargestellt nach der abhängigen Variable, der Viertklassnote in Deutsch und Mathematik.

### 7.3.1 Effekte familiärer und individueller Schülermerkmale auf die Deutschnote

Um zu prüfen, inwiefern die Zeugnisnoten in Deutsch in der vierten Jahrgangsstufe von strukturellen und prozessbezogenen familiären Hintergrundmerkmalen beeinflusst werden und inwiefern die Effekte durch die Hinzunahme der von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmalen sowie deren Testleistungen variieren, wird in [Tabelle 27](#) das Ergebnis der hierarchischen Regressionsanalyse vorgestellt.

Tabelle 27: Regressionsanalyse zur Vorhersage der Deutschnote in der vierten Klasse durch familiäre und individuelle Schülermerkmale

	Variable	B	SE B	$\beta$	R <sup>2</sup> (korr.)
<b>Modell 1</b>					.123**
	Bildungsstatus	-.330	.042	-.282**	
	Sozialer Status (ISEI)	-.098	.038	-.091*	
	Geschlecht	.226	.054	.136**	
<b>Modell 2</b>					.286**
	Bildungsstatus	-.161	.040	-.137**	
	Sozialer Status (ISEI)	-.035	.035	-.032	
	Geschlecht	.188	.049	.113**	
	Schulkontakt (LK)	-.001	.044	-.001	
	Unterstützungsverhalten (LK)	-.149	.045	-.133**	
	Bildungsaspiration (LK)	-.308	.031	-.362**	
<b>Modell 3</b>					.548**
	Bildungsstatus	-.061	.032	-.052	
	Sozialer Status (ISEI)	.021	.028	.019	
	Geschlecht	.107	.039	.064**	
	Schulkontakt (LK)	-.058	.035	-.048	
	Unterstützungsverhalten (LK)	-.072	.036	-.064*	
	Bildungsaspiration (LK)	-.178	.025	-.209**	
	Testleistung	-.662	.037	-.534**	
	Intelligenz (kft)	-.051	.021	-.067*	
<b>Modell 4</b>					.721**
	Bildungsstatus	-.019	.026	-.016	
	Sozialer Status (ISEI)	.032	.022	.030	
	Geschlecht	.080	.031	.048*	
	Schulkontakt (LK)	-.025	.028	-.020	
	Unterstützungsverhalten (LK)	-.014	.028	-.012	
	Bildungsaspiration (LK)	-.083	.020	-.097**	
	Testleistung	-.184	.036	-.148**	
	Intelligenz (kft)	-.022	.017	-.029	
	Begabung Deutsch (LK)	.274	.032	.307**	
	Sprachliche Fähigkeiten (LK)	.314	.031	.354**	
	Vorwissen (LK)	-.055	.028	-.045*	
<b>Modell 5</b>					.733**
	Bildungsstatus	-.029	.025	-.025	
	Sozialer Status (ISEI)	.028	.022	.026	
	Geschlecht	.045	.032	.027	
	Schulkontakt (LK)	-.022	.027	-.018	
	Unterstützungsverhalten (LK)	.035	.029	.031	
	Bildungsaspiration (LK)	-.059	.020	-.069**	
	Testleistung	-.174	.036	-.140**	
	Intelligenz (kft)	-.017	.016	-.022	
	Begabung Deutsch (LK)	.247	.031	.278**	
	Sprachliche Fähigkeiten (LK)	.263	.032	.296**	
	Vorwissen (LK)	-.045	.027	-.037	
	Arbeitsverhalten (LK)	.188	.030	.193**	
	Sozialverhalten (LK)	-.027	.027	-.024	

Anmerkungen: N=823; \* p&lt;.05; \*\* p&lt;.01

Zur Beantwortung der zweiten Fragestellung wurden für das Fach Deutsch schrittweise die Prädiktoren der Deutschnote in das Regressionsmodell aufgenommen. Modell 1 umfasst die strukturellen Merkmale der Familie (Bildungsstatus, sozialer Status), Modell 2 nimmt zusätzlich die von der Lehrkraft eingeschätzten familiären Prozessmerkmale (Schulkontakt, Unterstützungsverhalten, Bildungsaspiration) auf, Modell 3 betrachtet außerdem die kognitiven Testleistungen (Testleistung, kft) der Schüler, Modelle 4 und 5 umfassen zusätzlich die Lehrkräfteeinschätzungen zu kognitiven (Begabung Deutsch, sprachliche Fähigkeiten, Vorwissen) und nicht-kognitiven (Arbeitsverhalten, Sozialverhalten) Schülermerkmalen. Das Geschlecht des Schülers wird in allen Schritten kontrolliert (vgl. [Tabelle 27](#)).

Aus der hierarchischen Regressionsanalyse in [Tabelle 27](#) wird ersichtlich, dass die einbezogenen individuellen und familiären Prädiktoren insgesamt 73,3 Prozent der Varianz in der Deutschnote der vierten Klasse erklären.

Insgesamt erklärt das erste Modell 12,3 Prozent der Varianz in den Deutschnoten der vierten Klasse (ANOVA:  $F(3,819)=39.50, p<.01$ ). In diesem Modell 1, das zunächst ausschließlich die Strukturmerkmale und das Schülergeschlecht berücksichtigt, zeichnet sich der Bildungsstatus der Eltern als einflussstärkster Prädiktor ab ( $\beta = -.282, p<.01$ ), gefolgt von einem signifikanten, schwach negativen Effekt des sozialen Status ( $\beta = -.091, p<.05$ ). Das Geschlecht ergibt für die Deutschnote ebenfalls hochsignifikante, direkte Effekte auf die Notenzusammensetzung ( $\beta = .136, p<.01$ ).

Im zweiten Modell, das zusätzlich die Lehrkräfteeinschätzung der familiären Prozessmerkmale berücksichtigt, ergibt sich eine Varianzaufklärung von 28,6 Prozent (ANOVA:  $F(6,816)=56.00, p<.01$ ). Durch Hinzunahme der eingeschätzten Prozessmerkmale sinkt das Beta-Gewicht der Strukturmerkmale ab, dennoch bleibt ein direkter hochsignifikanter Effekt des Bildungsstatus auf die Notenzusammensetzung bestehen ( $\beta = -.137, p<.01$ ), der soziale Status (ISEI) wird nicht mehr signifikant. Der Geschlechtereffekt bleibt auch im zweiten

Modell hochsignifikant ( $\beta = .113, p < .01$ ). Bezüglich der schulbezogenen Prozessmerkmale ergibt der Schulkontakt nahezu keinen Effekt ( $\beta = -.001, n.s.$ ). Demgegenüber ergeben das eingeschätzte Unterstützungsverhalten ( $\beta = -.133, p < .01$ ) und die Bildungsaspiration ( $\beta = -.362, p < .01$ ) der Eltern hochsignifikant negative Effekte auf die Deutschnote. Durch die Hinzunahme der kognitiven Testleistungen der Schüler im dritten Modell erhöht sich die Varianzaufklärung auf 54,8 Prozent (ANOVA:  $F(8,814)=125.79, p < .01$ ). Unter Kontrolle der Testleistungen sind beide einbezogenen Strukturmerkmale der Familie nicht signifikant, das Geschlecht dagegen nimmt weiterhin schwach Einfluss ( $\beta = .064, p < .01$ ). Bezüglich der schulbezogenen Prozessmerkmale bleibt der Schulkontakt weiterhin irrelevant, das Unterstützungsverhalten ( $\beta = -.064, p < .05$ ) und Bildungsaspiration ( $\beta = -.209, p < .01$ ) bleiben aber weiterhin bedeutsam. Bei gleicher Testleistung nehmen die familiären Strukturmerkmale also keinen direkten Einfluss auf die Notenzusammensetzung im Fach Deutsch, die wahrgenommene Bildungsaspiration und das Unterstützungsverhalten nehmen jedoch über die Testleistung hinaus signifikant Einfluss auf die Notenzusammensetzung. Die kognitive Testleistung des Schülers ist im dritten Modell insgesamt die stärkste Einflusskomponente ( $\beta = -.534, p < .01$ ), aber auch für die Intelligenz ( $g_{ft}$ ) des Schülers ergeben sich schwache Effekte.

Im vierten Modell werden zusätzlich die Lehrkräfteeinschätzungen zu den kognitiven Schülermerkmalen berücksichtigt und erklären über die familiären Hintergrundmerkmale und die Testleistungen hinaus weitere 17,3 Prozent der Notenvarianz im Fach Deutsch (ANOVA:  $F(11,811)=193.65, p < .01$ ). Die beiden familiären Strukturmerkmale bleiben weiterhin nicht signifikant, der Effekt der Kontrollvariable Geschlecht bleibt auch durch die Hinzunahme der kognitiven Fähigkeitseinschätzungen schwach, aber signifikant bestehen ( $\beta = .048, p < .05$ ). Von den Prozessmerkmalen nimmt nur noch die eingeschätzte Bildungsaspiration der Eltern Einfluss auf die Notenzusammensetzung ( $\beta = -.097, p < .01$ ). Durch die Hinzunahme der Lehrkräfteeinschätzungen sinkt das Beta-Gewicht der individuellen Testleistung deutlich ab ( $\beta = -.148, p < .01$ ), was für eine gemeinsame

Varianzaufklärung der Prädiktoren spricht, die Intelligenz ( $kft$ ) des Schülers wird dagegen nicht mehr signifikant. Die drei Prädiktoren der kognitiven Lehrkräfteeinschätzungen sind allesamt signifikant bedeutsam, der stärkste Einfluss geht von der sprachlichen Fähigkeitseinschätzung aus ( $\beta = .354, p < .01$ ), aber auch die Begabungseinschätzung ( $\beta = .307, p < .01$ ) und das Vorwissen ( $\beta = -.045, p < .05$ ) ergeben signifikante Effekte.

Im fünften Gesamtmodell, steigt die Varianzaufklärung unter Einbezug der nicht-kognitiven Fähigkeitseinschätzung um weitere 1,2 Prozent auf 73,3 Prozent an (ANOVA:  $F(13,809)=174.94, p < .01$ ). Von den einbezogenen familiären Strukturmerkmalen ergeben sich weiterhin keine signifikanten direkten Effekte auf die Deutschnote. Der Geschlechtereffekt wird unter zusätzlicher Kontrolle des Arbeits- und Sozialverhaltens im Gesamtmodell nicht mehr signifikant. Für die schulbezogenen Prozessmerkmale bleibt auch unter Kontrolle der kognitiven und nicht-kognitiven Schülermerkmale ein direkter Effekt der Lehrkräfteeinschätzung zur elterlichen Bildungsaspiration auf die Noten in Deutsch am Ende der Grundschulzeit bestehen ( $\beta = -.069, p < .01$ ), die anderen familiären Prozessmerkmale nehmen weiterhin keinen signifikanten Einfluss auf die Notenzusammensetzung. Kinder, deren Eltern – aus Perspektive der Lehrkraft – einen hohen Schulabschluss anstreben, erhalten also bei gleichen Testleistungen und gleichen kognitiven sowie nicht-kognitiven Lehrkräfteeinschätzungen im Durchschnitt eine bessere Note im Fach Deutsch. Durch die Hinzunahme der nicht-kognitiven Schülermerkmale sinkt das Beta-Gewicht der individuellen Testleistung nochmals etwas ab ( $\beta = -.140, p < .01$ ), die Intelligenz ( $kft$ ) des Schülers bleibt weiterhin ohne signifikante Effekte. Während bei gleicher Testleistung die kognitiven Merkmaleinschätzungen der sprachlichen Fähigkeiten ( $\beta = .296, p < .01$ ) und der fachspezifischen Begabung hochsignifikant bleiben ( $\beta = .278, p < .01$ ), ergeben sich für das eingeschätzte Vorwissen keine signifikanten Effekte mehr. Von den nicht-kognitiven Merkmalsurteilen nimmt ausschließlich das Arbeitsverhalten positiv Einfluss auf die Notenzusammensetzung ( $\beta = .193, p < .01$ ), das Sozialverhalten ist dagegen nicht signifikant.

### 7.3.2 Effekte familiärer und individueller Schülermerkmale auf die Mathematiknote

Die Analyse der Effekte familiärer und individueller Schülermerkmale auf die Mathematiknote der vierten Klasse verläuft analog zum Vorgehen in [Kapitel 7.3.1](#) zur Erklärung der Deutschnote. Um zu prüfen, inwiefern die Halbjahresnote in Mathematik in der vierten Jahrgangsstufe von strukturellen und prozessbezogenen familiären Hintergrundmerkmalen beeinflusst wird und inwiefern diese durch den Einbezug der von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale sowie deren Testleistungen in ihren Effekten variieren, wird in [Tabelle 28](#) das Ergebnis der hierarchischen Regressionsanalyse vorgestellt.

Tabelle 28: Regressionsanalyse zur Vorhersage der Mathematiknote in der vierten Klasse durch familiäre und individuelle Schülermerkmale

	Variable	B	SE B	$\beta$	R <sup>2</sup> (korr.)
<b>Modell 1</b>					.095**
	Bildungsstatus	-.327	.045	-.264**	
	Sozialer Status (ISEI)	-.094	.041	-.083*	
	Geschlecht	-.092	.059	-.052	
<b>Modell 2</b>					.253**
	Bildungsstatus	-.152	.043	-.123**	
	Sozialer Status (ISEI)	-.030	.038	-.026	
	Geschlecht	-.133	.053	-.075*	
	Schulkontakt (LK)	.069	.048	.054	
	Unterstützungsverhalten (LK)	-.185	.048	-.156**	
	Bildungsaspiration (LK)	-.324	.033	-.359**	
<b>Modell 3</b>					.510**
	Bildungsstatus	-.034	.036	-.028	
	Sozialer Status (ISEI)	.023	.031	.021	
	Geschlecht	-.216	.043	-.122**	
	Schulkontakt (LK)	.005	.039	.004	
	Unterstützungsverhalten (LK)	-.105	.039	-.088**	
	Bildungsaspiration (LK)	-.197	.028	-.219**	
	Testleistung	-.590	.041	-.449**	
	Intelligenz (kft)	-.145	.023	-.180**	
<b>Modell 4</b>					.678**
	Bildungsstatus	-.014	.029	-.012	
	Sozialer Status (ISEI)	.045	.025	.039	
	Geschlecht	-.036	.037	-.020	
	Schulkontakt (LK)	.011	.031	.008	
	Unterstützungsverhalten (LK)	-.049	.032	-.041	
	Bildungsaspiration (LK)	-.114	.023	-.126**	
	Testleistung	-.177	.042	-.135**	
	Intelligenz (kft)	-.059	.019	-.073**	
	Begabung Mathematik (LK)	.494	.027	.538**	
	Sprachliche Fähigkeiten (LK)	.071	.029	.076*	
	Vorwissen (LK)	-.065	.032	-.050*	
<b>Modell 5</b>					.683**
	Bildungsstatus	-.023	.029	-.019	
	Sozialer Status (ISEI)	.043	.025	.038	
	Geschlecht	-.076	.038	-.043*	
	Schulkontakt (LK)	.014	.031	.011	
	Unterstützungsverhalten (LK)	-.010	.033	-.009	
	Bildungsaspiration (LK)	-.099	.024	-.109**	
	Testleistung	-.173	.041	-.132**	
	Intelligenz (kft)	-.058	.019	-.072**	
	Begabung Mathematik (LK)	.473	.027	.515**	
	Sprachliche Fähigkeiten (LK)	.023	.031	.025	
	Vorwissen (LK)	-.058	.031	-.045	
	Arbeitsverhalten (LK)	.136	.035	.132**	
	Sozialverhalten (LK)	-.003	.031	-.002	

Anmerkungen: N=823; \* p&lt;.05; \*\* p&lt;.01

Zur Beantwortung der zweiten Fragestellung wurden für das Fach Mathematik ebenfalls schrittweise die Prädiktoren der Mathematiknote in das Regressionsmodell aufgenommen. Modell 1 umfasst die strukturellen Merkmale der Familie (Bildungsstatus, sozialer Status), Modell 2 nimmt zusätzlich die von der Lehrkraft eingeschätzten familiären Prozessmerkmale (Schulkontakt, Unterstützungsverhalten, Bildungsaspiration) auf. Modell 3 betrachtet außerdem die kognitiven Testleistungen (Testleistung, kft) der Schüler. Modell 4 und 5 umfassen zusätzlich die Lehrkräfteeinschätzungen zu kognitiven (Begabung Mathematik, sprachliche Fähigkeiten, Vorwissen) und nicht-kognitiven (Arbeitsverhalten, Sozialverhalten) Schülermerkmalen. Das Geschlecht des Schülers wird in allen Schritten kontrolliert (vgl. [Tabelle 28](#)).

Aus der hierarchischen Regressionsanalyse in [Tabelle 28](#) wird ersichtlich, dass die einbezogenen individuellen und familiären Prädiktoren insgesamt 68,3 Prozent der Varianz in der Mathematiknote der vierten Klasse erklären.

Modell 1 erklärt insgesamt lediglich 9,5 Prozent der Varianz in den Mathematiknoten der vierten Klasse (ANOVA:  $F(3,819)=29.62, p<.01$ ). Im ersten Modell, das zunächst ausschließlich die Strukturmerkmale und das Schülergeschlecht berücksichtigt, zeichnet sich der Bildungsstatus der Eltern als einflussstärkster Prädiktor ab ( $\beta = -.264, p<.01$ ), der soziale Status der Familie nimmt zusätzlich schwachen negativen Einfluss ( $\beta = -.083, p<.05$ ). Das Schülergeschlecht ergibt im Gegensatz zur Deutschnote zunächst keine signifikanten Effekte.

In Modell 2, das zusätzlich die familiären Prozessmerkmale berücksichtigt, steigt die Varianzaufklärung auf 25,3 Prozent an (ANOVA:  $F(6,816)=47.47, p<.01$ ). Analog zur Erklärung der Deutschnote sinkt das Beta-Gewicht der Strukturmerkmale durch die Hinzunahme der eingeschätzten Prozessmerkmale ab. Der direkte Effekt des Bildungsstatus bleibt aber weiterhin hochsignifikant (von  $\beta = -.264$  auf  $\beta = -.123, p<.01$ ) sowie der direkte Effekt des sozialen Status nicht signifikant wird

( $\beta = -.026$ , n.s.), das Geschlecht des Schülers nimmt dagegen schwach signifikant Einfluss ( $\beta = -.075$ ,  $p < .05$ ). Ebenfalls ähnlich zur Deutschnote sind die Effekte der familiären Prozessmerkmale auf die Mathematiknote, wovon die von der Lehrkraft eingeschätzte Bildungsaspiration die stärkste Erklärungskraft auszeichnet ( $\beta = -.359$ ,  $p < .01$ ). Auch die Unterstützung der Eltern spiegelt sich in der Notenzusammensetzung wider, der Schulkontakt ist dagegen nicht signifikant.

Durch die Hinzunahme der kognitiven Testleistungen der Schüler in Modell 3 steigt die Varianzaufklärung auf 51 Prozent an (ANOVA:  $F(8,814)=108.06$ ,  $p < .01$ ). Unter Kontrolle der Testleistungen wird auch der zweite strukturelle Faktor der Familie nicht signifikant. Konträr zur Zusammensetzung der Deutschnote (vgl. [Tabelle 27](#)) wird der Geschlechtereffekt auf die Mathematiknote durch die Aufnahme der Testleistungen jedoch hochsignifikant bedeutsamer ( $\beta = -.122$ ,  $p < .01$ ). Bezüglich der schulbezogenen Prozessmerkmale bleiben das Unterstützungsverhalten ( $\beta = -.088$ ,  $p < .01$ ) und die Bildungsaspiration ( $\beta = -.219$ ,  $p < .01$ ) weiterhin bedeutsam. Bei gleicher Testleistung nehmen die familiären Strukturmerkmale also keinen direkten Einfluss auf die Notenzusammensetzung im Fach Mathematik, die wahrgenommene Bildungsaspiration und das Unterstützungsverhalten nehmen jedoch über die Testleistung hinaus hochsignifikant Einfluss auf die Notenzusammensetzung. Zudem ist im Modell 3 die kognitive Testleistung der stärkste Einflussprädiktor ( $\beta = -.449$ ,  $p < .01$ ), die Intelligenz (kft) nimmt ebenfalls hochsignifikant negativen Einfluss ( $\beta = -.180$ ,  $p < .01$ ).

Im vierten Modell werden zusätzlich die Lehrkräfteeinschätzungen zu den kognitiven Schülermerkmalen berücksichtigt und erklären über die familiären Hintergrundmerkmale und die Testleistungen hinaus weitere 16,8 Prozent der Notenvarianz (ANOVA:  $F(11,811)=158.00$ ,  $p < .01$ ). Die beiden familiären Strukturmerkmale bleiben weiterhin nicht signifikant, der signifikante Effekt der Kontrollvariable Geschlecht verschwindet durch die Hinzunahme der kognitiven Fähigkeitseinschätzungen. Bezüglich der familiären Prozessmerkmalen nimmt weiterhin die Bildungsaspiration auch unter Kontrolle der kognitiven Lehr-

krafteinschätzung hochsignifikanten Einfluss auf die Notenzusammensetzung ( $\beta = -.126, p < .01$ ). Erneut reduziert sich durch die Zunahme der kognitiven Merkmaleinschätzungen das Beta-Gewicht der individuellen Testleistungen ( $\beta = -.135, p < .01$ ), was wiederum für eine gemeinsame Varianzaufklärung spricht. Von den kognitiven Lehrkrafteinschätzungen sticht die mathematische Begabungseinschätzung als stärkster Erklärungsprädiktor hervor ( $\beta = .538, p < .01$ ), die sprachlichen Fähigkeiten ( $\beta = .076, p < .05$ ) und das Vorwissen ( $\beta = -.050, p < .05$ ) werden nur knapp signifikant. Im fünften Gesamtmodell, unter Einbezug der nicht-kognitiven Fähigkeitseinschätzung der Lehrkraft, steigt die Varianzaufklärung um weitere 0,5 Prozent auf 68,3 Prozent an (ANOVA:  $F(13,809)=137.55, p < .01$ ).

Von den einbezogenen familiären Strukturmerkmalen ergeben sich weiterhin keine signifikanten direkten Effekte auf die Mathematiknote. Zudem wird der Geschlechtereffekt knapp signifikant ( $\beta = -.043, p < .05$ ). Analog zu den Ergebnissen für die Deutschnote bleibt von den familiären Prozessmerkmalen ein direkter Effekt der Lehrkrafteinschätzung zur elterlichen Bildungsaspiration auf die Notenzusammensetzung in Mathematik am Ende der Grundschulzeit bestehen ( $\beta = -.109, p < .01$ ), die anderen familiären Hintergrundmerkmale bleiben weiterhin nicht signifikant. Während im fünften Modell für die kognitiven Merkmaleinschätzungen nur noch die Begabungseinschätzung hochsignifikant wird ( $\beta = .515, p < .01$ ), ist von den nicht-kognitiven Merkmalsurteilen allein das Arbeitsverhalten bedeutsam ( $\beta = .132, p < .01$ ).

### 7.3.3 Zusammenfassung der bisherigen Ergebnisse

Im Vordergrund der vorangegangenen Analysen stand u.a. die Frage, ob die familiären Hintergrundmerkmale einen eigenständigen Effekt auf die Notenzusammensetzung haben oder ob sie vollständig durch die Einschätzung individueller Schülermerkmale mediiert werden. Ebenso wurde beleuchtet, inwiefern die familiären Strukturmerkmale die familiären Prozessmerkmale beeinflussen und ob sich für die familiären Hintergrundmerkmale unterschiedliche Effekte für Testleistungen, Merkmaleinschätzungen und Schulnoten ergeben.

Fachübergreifend wird ersichtlich, dass die familiären Strukturmerkmale allein circa 10 Prozent der Varianz in der Notenzusammensetzung erklären sowie der Bildungsstatus insgesamt stärker Einfluss nimmt als der soziale Status der Familie (ISEI) (vgl. [Tabelle 27](#) bzw. [Tabelle 28](#)). Durch die Hinzunahme der familiären Prozessmerkmale (Unterstützungsverhalten, Bildungsaspiration, Schulkontakt) wird der direkte Effekt der Strukturmerkmale schwächer (für den sozialen Status zudem nicht signifikant), wodurch von einem vermittelnden Effekt ausgegangen werden kann. Zudem steigt durch Hinzunahme der familiären Prozessmerkmale die Varianzaufklärung für beide Fächer auf knapp unter 30 Prozent an. Fachübergreifend ergeben von den Prozessmerkmalen das von der Lehrkraft eingeschätzte Unterstützungsverhalten und die Bildungsaspiration des Elternhauses signifikante Effekte auf die Notenzusammensetzung und bleiben auch unter Kontrolle der kognitiven Testleistungen bestehen, während die Strukturmerkmale der Familie dann beide nicht signifikant werden (vgl. [Tabelle 27](#) bzw. [Tabelle 28](#), Modell 3). Während die Effektstärke der eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale nahezu identisch zu den regressionsanalytischen Ergebnissen für die Individualmerkmale verläuft (vgl. [Tabelle 23](#) bzw. [Tabelle 25](#)), bleibt für beide Fachnoten im Gesamtmodell unter Kontrolle sämtlicher Individualprädiktoren der hochsignifikante, negative Effekt der Bildungsaspiration auf die Notenzusammensetzung in Deutsch und Mathematik bestehen.

Zusätzlich dazu wurden jedoch auch unter Einbezug der familiären Hintergrundmerkmale fachspezifische Unterschiede in der Notenzusammensetzung der Viertklassnoten in Deutsch und Mathematik gefunden. Während analog zu den Individualmodellen in Mathematik weiterhin die mathematische Begabungseinschätzung und für Deutsch die sprachlichen Fähigkeiten die wichtigsten Prädiktoren darstellen, ergeben sich für die Kontrollvariable Geschlecht unterschiedliche Ergebnisse. Obwohl die Effekte insgesamt schwach ausgeprägt sind, ergeben sich für das Fach Mathematik unter Kontrolle der familiären und individuellen Schülermerkmale schwache, aber signifikante Effekte für das Geschlecht, wohingegen der Effekt für die Deutschnote nach Kontrolle der nicht-kognitiven Fähigkeiten verschwindet.

Nach der Analyse der zweiten Fragestellung, folgt anschließend mithilfe von Strukturgleichungsmodellierungen eine vertiefte Analyse der komplexeren Zusammenhangsstrukturen für die vorliegenden Fragestellungen.

## 7.4 Prüfung des Gesamtmodells

Mithilfe einer Strukturgleichungsmodellierung werden nachfolgend nun die komplexeren Zusammenhangsstrukturen zwischen den bereits einbezogenen Prädiktorvariablen analysiert, um auch die Beziehung zwischen den Einflussvariablen und deren Wechselwirkung für die Notenzusammensetzung zu erfassen. Neben dem Vorteil der gleichzeitigen Berücksichtigung direkter und indirekter Effekte können mithilfe von Strukturgleichungsmodellen komplexe Zusammenhänge auf messfehlerbereinigter Ebene analysiert werden, sowie die hierarchische Struktur in Daten berücksichtigt wird (Geiser, 2011).

Im Fokus der Darstellung steht dabei das theoretisch deklarierte Gesamtmodell aus [Kapitel 4.5](#). Geprüft wird zum einen, ob die Strukturmerkmale der Familie (Bildungsstatus, sozialer Status) direkt und vermittelt über die Prozessmerkmale der Familie (Unterstützungsverhalten, Schulkontakt, Bildungsaspiration) auf die Notenzusammensetzung Einfluss nehmen und inwiefern diese Effekte über die kognitiven und nicht-kognitiven Schülermerkmale vermittelt werden. Diese strukturellen Merkmale der Familie werden als latente Variable mit den Indikatoren Bildungsstatus und sozialer Status (ISEI) im Modell berücksichtigt. Auch die familiären Prozessmerkmale fließen als latentes Konstrukt mit den Indikatoren Unterstützungsverhalten, Schulkontakt und Bildungsaspiration mit in das Gesamtmodell ein. Geprüft wird hier, ob von den familiären Prozessmerkmalen eigenständige Effekte auf die Notenzusammensetzung ausgehen und inwiefern diese über die individuellen Schülermerkmale vermittelt werden.

Gleichzeitig wird auf individueller Schülerebene davon ausgegangen, dass die kognitiven die nicht-kognitiven Schülermerkmale beeinflus-

sen sowie die kognitiven Schülermerkmale nach Erhebungsperspektive (Testleistung des Schülers vs. Lehrkräfteeinschätzung) getrennt analysiert werden. Die kognitiven Merkmale der Schüler werden getrennt nach Testleistung (Testleistung, Intelligenz) und Einschätzung (Begaubung, sprachliche Fähigkeiten, Vorwissen) als zwei latente Variablen in den fachspezifischen Modellen analysiert. Aufgrund der vorangegangenen regressionsanalytischen Auswertung wird in beiden Modellen für die Noten in Deutsch und Mathematik als nicht-kognitives Merkmal allein das Arbeitsverhalten als manifeste Variable aufgenommen sowie das Geschlecht des Schülers im Strukturgleichungsmodell aufgrund der schwachen Effekte nicht mehr berücksichtigt wird. Zudem wird für das Fach Mathematik als Indikator für die kognitiven Merkmaleinschätzungen auf die sprachlichen Fähigkeiten verzichtet. Begründet wird dies damit, dass die sprachliche Fähigkeitseinschätzung in den vorangegangenen Analysen ebenfalls keinerlei Effekte auf die Notenzusammensetzung in Mathematik ergab, sowie theoriegeleitet davon ausgegangen werden kann, dass die sprachlichen Fähigkeiten keinen großen Einfluss auf die Mathematiknote nehmen. Zusätzlich dazu werden die in Kapitel 7.2.1 bzw. 7.2.2 geprüften Interaktionseffekte in den nachfolgenden Strukturgleichungsmodellen aufgrund von Schätzproblemen in Mplus nicht berücksichtigt. Auf weitere Modellvereinfachungen wurde aufgrund der theoretischen Modellüberprüfung verzichtet, zumal davon ausgegangen werden kann, dass auch geringe oder nicht signifikante Prädiktoren in einem latenten Konstrukt durchaus Erklärungswert aufweisen können. Zusätzliche, nicht aufgeführte Strukturgleichungsmodellierungen unter Einbezug der ausschließlich signifikanten Prädiktoren kamen zudem zu ähnlichen Ergebnistendenzen.

Bezüglich der Korrelationen zwischen den betrachteten Prädiktoren wird davon ausgegangen, dass die Lehrkräfteeinschätzungen zu den kognitiven Merkmalen und zum Arbeitsverhalten nicht unabhängig voneinander sind, weswegen die Fehlerkorrelationen zwischen diesen Einschätzungen im Modell zugelassen werden. Diese werden aus Gründen der Übersichtlichkeit für jedes Modell eigens tabellarisch aufgeführt und nicht mit in die Abbildungen aufgenommen.

Der hierarchischen Struktur der Daten wird bei der Strukturgleichungsmodellierung mit der Software Mplus ebenfalls Sorge getragen. Fehlende Werte wurden mithilfe des Full-Information-Maximum-Likelihood-Verfahrens (FIML) in MPLUS geschätzt (Muthén & Muthén, 2012; Reinecke, 2005). Zusätzlich dazu wird gemäß der theoretischen Herleitung die leistungsbezogene und soziale Zusammensetzung der Klasse mithilfe eines Mehrebenenregressionsmodells auf Effekte für die Notenzusammensetzung auf Individualebene geprüft, diese Darstellung folgt im Anschluss der Ergebnisse für die Individualebene in [Kapitel 7.4.3](#).

Im Folgenden werden nun gemäß der Fragestellung die Ergebnisse für die Deutsch- und Mathematiknote nacheinander präsentiert.

#### 7.4.1 Prüfung des Gesamtmodells: Effekte auf die Deutschnote

Im Folgenden werden die Ergebnisse der Überprüfung des theoretischen Gesamtmodells für die Notenzusammensetzung der Viertklassnote im Fach Deutsch präsentiert. [Abbildung 8](#) gibt einen Überblick über die geprüften Zusammenhänge. In [Tabelle 29](#) werden die angesprochenen Fehlerkorrelationen zwischen den Lehrkräfteeinschätzungen präsentiert, [Tabelle 30](#) und [Tabelle 31](#) zeigen die indirekten und totalen Effekte der Strukturgleichungsmodellierung. Anzumerken gilt für das Signifikanzniveau, dass hochsignifikante Pfade ( $p < .01$ ) fett gedruckt, signifikante Pfade dünn durchgezogen ( $p < .05$ ) und nicht signifikante Pfade gestrichelt präsentiert werden.

#### Modellanpassung

Die Güte des Modells aus [Abbildung 8](#) zeigt eine gute Passung auf die Daten (RMSEA = .056; SRMR = .053; CFI = .958). Obwohl der signifikante  $\chi^2$ -Wert mit 207.779,  $df = 38$  ( $p < .01$ ) zunächst gegen die Annahme des Modells zu sprechen scheint, muss für die Interpretation des Kennwerts die Stichprobengröße berücksichtigt werden, da für große Stichproben schon kleine Abweichungen von einer optimalen Modellpassung signifikant werden (Bühner, 2011).

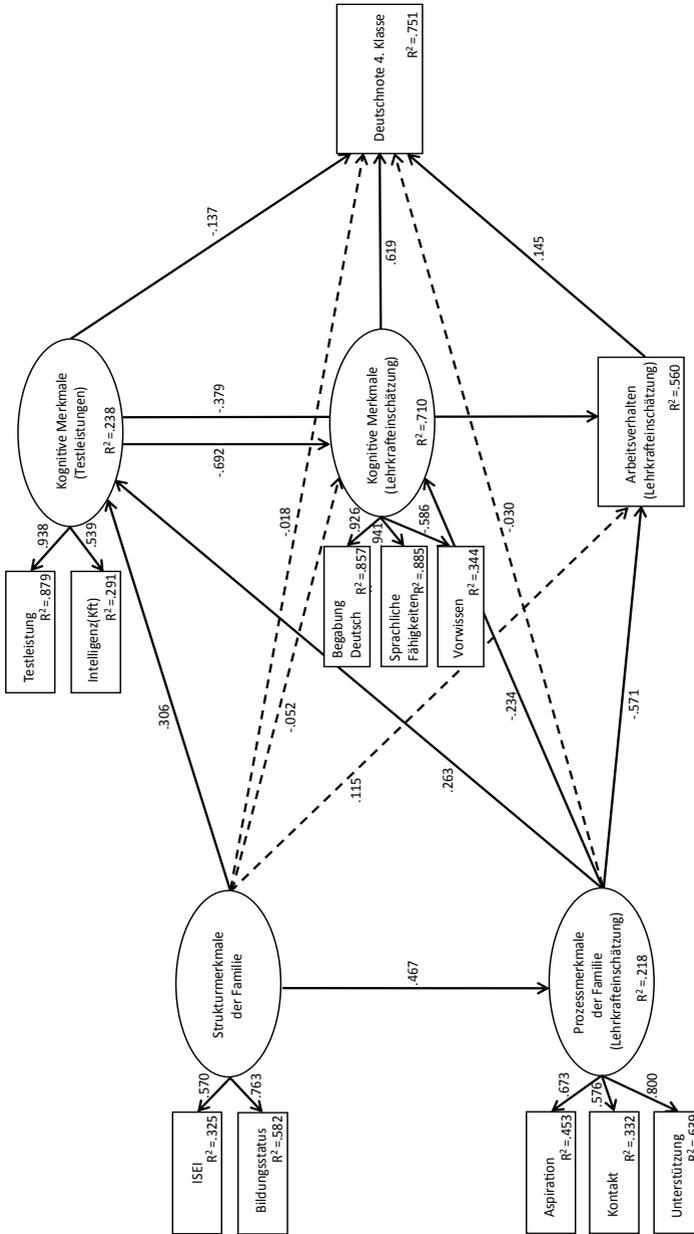


Abbildung 8: Strukturgleichungsmodell zum Einfluss individueller und familiärer Merkmale auf die Deutschnote

Anmerkungen: N=1446; Chi<sup>2</sup>=207.779, df=38, p<.01; RMSEA=.056; SRMR=.053; CFI=.958; Fett gedruckter Pfeil: p<.05; Gestrichelter Pfeil: n.s.

## Direkte Effekte

Wie in [Abbildung 8](#) ersichtlich, ergeben sich für die Strukturmerkmale der Familie keine signifikanten direkten Effekte auf die Deutschnote sowie auf die Lehrkräfteeinschätzungen zu den kognitiven Merkmalen der Schüler und deren Arbeitsverhalten. Hochsignifikante direkte Effekte der familiären Strukturmerkmale ergeben sich dagegen auf die familiären Prozessmerkmale ( $\beta = .467, p < .01$ ) und die kognitiven Testleistungen der Schüler ( $\beta = .306, p < .01$ ). Für die Prozessmerkmale der Familie ergeben sich ebenso keine signifikanten direkten Effekte auf die Deutschnote. Jene eingeschätzten familiären Prozessmerkmale nehmen allerdings direkten, hochsignifikanten Einfluss auf die kognitiven Testleistungen ( $\beta = .263, p < .01$ ), auf die kognitiven Merkmaleinschätzungen ( $\beta = -.234, p < .01$ ) und auf das Arbeitsverhalten ( $\beta = -.571, p < .01$ ) des Schülers. Der positive Koeffizient auf die Testleistung ist im Sinne einer höheren Testleistung zu lesen, für die Merkmaleinschätzung gilt die gleiche Polung wie für eine Notenskala, niedrigere Werte entsprechen besseren Einschätzungen. Je höher (besser) die Lehrkräfte also die familiären Prozessmerkmale einschätzen, desto niedriger (besser) schätzen sie wiederum deren kognitive Merkmale ein.

Bezüglich der individuellen Schülermerkmale wirken sich die kognitiven Testleistungen der Schüler hochsignifikant direkt auf die eingeschätzten kognitiven Merkmale ( $\beta = -.692, p < .01$ ), auf das Arbeitsverhalten ( $\beta = -.379, p < .01$ ) und nicht zuletzt auf die Deutschnote ( $\beta = -.137, p < .01$ ) aus. Der stärkste direkte und hochsignifikante Effekt auf die Deutschnote ergibt sich allerdings für die eingeschätzten kognitiven Schülermerkmale ( $\beta = .619, p < .01$ ), aber auch das wahrgenommene Arbeitsverhalten des Schülers wirkt sich direkt auf dessen Deutschnote aus ( $\beta = .145, p < .01$ ). Je höher (schlechter) die kognitiven Fähigkeiten und das Arbeitsverhalten, desto höher (schlechter) fällt die Note aus.

Tabelle 29: Korrelationen der Fehlerterme im Strukturgleichungsmodell für das Fach Deutsch

Variablenpaar	Fehlerkorrelation
Arbeitsverhalten/kognitive Merkmale (LK)	.357**
Sprachliche Fähigkeiten/Begabung Deutsch	-.362*
Sprachliche Fähigkeiten/Vorwissen	.293**
Vorwissen/Begabung Deutsch	.110

Anmerkungen: \*  $p < .05$ ; \*\*  $p < .01$

### Indirekte Effekte

Ein Vorteil der Strukturgleichungsmodellierung ist die Verfolgung indirekter Effekte zwischen den einbezogenen Untersuchungsvariablen, die über mindestens eine andere Variable vermittelt werden (Reinecke, 2005). [Tabelle 30](#) gibt einen Überblick zu den standardisierten indirekten Effekten, um gemäß den zugrundeliegenden Fragestellungen zu klären, inwiefern sich Vermittlungseffekte zwischen den einbezogenen Variablen auf die Deutschnote ergeben. Geprüft wird dadurch beispielsweise, inwiefern sich die familiären Herkunftsmerkmale über die Lehrkräfteeinschätzungen der individuellen Schülermerkmale auf die Zusammensetzung der Deutschnote auswirken. Indirekte Effekte wurden in Anlehnung an das empirische Modell (vgl. [Abbildung 8](#)) für die Struktur- und Prozessmerkmale der Familie und die kognitiven Testleistungen der Schüler gemessen. Während die nachfolgende Tabelle über jeden einzelnen Vermittlungspfad Auskunft gibt, wird im Text auf auffällige Effekte eingegangen.

Tabelle 30: Standardisierte indirekte Effekte im Strukturgleichungsmodell auf die Deutschnote

<b>Ausgangsvariable</b>	<b>Vermittlungspfad I</b>	<b>Vermittlungspfad II</b>	<b>Vermittlungspfad III</b>	<b>Indirekte Effekte auf die Deutschnote</b>
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	--	--	-.042**
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	-.131**
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	-.017**
Strukturmerkmale der Familie	Kognitive Merkmale (LK)	--	--	-.032
Strukturmerkmale der Familie	Arbeitsverhalten (LK)	--	--	.017
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	--	--	-.014
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	--	-.017*
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	-.053**
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	-.007**
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (LK)	--	-.068**
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Arbeitsverhalten (LK)	--	-.039**
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	--	--	-.036**

Ausgangsvariable	Vermittlungspfad I	Vermittlungspfad II	Vermittlungspfad III	Indirekte Effekte auf die Deutschnote
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	-.113**
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	-.014**
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (LK)	--	--	-.145**
Prozessmerkmale der Familie (LK)	Arbeitsverhalten (LK)	--	--	-.083**
Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	--	-.429**
Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	--	-.055**

Anmerkungen: Der in der linken Spalte aufgeführte Prädiktor (Ausgangsvariable) stellt jeweils den Ausgangspunkt des indirekten Effekts auf die abhängige Variable dar, Vermittlungspfade I-III umfassen die jeweiligen Mediationsschritte. \*  $p < .05$ ; \*\*  $p < .01$

Der totale indirekte Effekt der familiären Strukturmerkmale ist moderat ausgeprägt ( $\beta = -.402$ ,  $p < .01$ ) (vgl. [Tabelle 31](#)). Wie [Tabelle 30](#) zu entnehmen ist, ergeben sich für die Strukturmerkmale die stärksten indirekten Effekte vermittelt über die kognitiven Testleistungen und die kognitiven Merkmaleinschätzungen ( $\beta = -.131$ ,  $p < .01$ ). Die These, dass die familiären Strukturmerkmale direkt über die kognitiven und nicht-kognitiven Lehrkrafeinschätzungen vermittelt werden, findet nur schwache Bestätigung, da die indirekten Effekte der Strukturmerkmale allein über die kognitive Merkmaleinschätzung ( $\beta = -.032$ , n.s.) und das Arbeitsverhalten ( $\beta = .017$ , n.s.) nicht signifikant werden.

Der Vermittlungspfad von Strukturmerkmalen auf die familiären Prozessmerkmale auf die Deutschnote ist ebenfalls nicht signifikant ( $\beta = -.014$ , n.s.), allerdings ergeben sich signifikante Pfade zwischen Struktur- und Prozessmerkmalen und den individuellen Schülermerkmalen auf die Deutschnote (vgl. [Tabelle 30](#)). So zeigt sich beispielsweise ein schwacher, aber hochsignifikanter indirekter Effekt zwischen Strukturmerkmalen, Prozessmerkmalen und der Einschätzung der kognitiven Schülermerkmale ( $\beta = -.068$ ,  $p < .01$ ).

Auf Seiten der familiären Prozessmerkmale ist der totale indirekte Effekt ebenfalls moderat ausgeprägt ( $\beta = -.391$ ,  $p < .01$ ) (vgl. [Tabelle 31](#)). Am stärksten wirken sie über die kognitive Merkmalseinschätzung auf die Deutschnote ( $\beta = -.145$ ,  $p < .01$ ) und deuten damit entgegengesetzt zu den familiären Strukturmerkmalen auf familiär bedingte Merkmalseinschätzungen hin. Zudem wirken sich die familiären Prozessmerkmale ebenfalls über die kognitiven Testleistungen und die kognitive Merkmalseinschätzung auf die Deutschnote aus ( $\beta = -.113$ ,  $p < .01$ ).

Bezüglich der kognitiven Testleistungen ergibt sich ebenfalls ein moderater totaler indirekter Effekt ( $\beta = -.484$ ,  $p < .01$ ) (vgl. [Tabelle 31](#)). Der stärkste indirekte Effekt der kognitiven Testleistungen wird hier über die kognitive Merkmalseinschätzung vermittelt ( $\beta = -.429$ ,  $p < .01$ ).

### **Totale Effekte**

Der totale Effekt einer Strukturgleichungsmodellierung ergibt sich aus der Summe der im Modell aufgeführten direkten und indirekten Effekte (Reinecke, 2005). [Tabelle 31](#) gibt einen Überblick über die standardisierten totalen Effekte aus dem Strukturgleichungsmodell für die Deutschnote am Ende der vierten Jahrgangsstufe.

Tabelle 31: Standardisierte totale Effekte im Strukturgleichungsmodell auf die Deutschnote

Ausgangsvariable	Totaler Effekt	Totaler direkter Effekt	Totaler indirekter Effekt
1. Strukturmerkmale der Familie	-.420**	-.018	-.402**
2. Prozessmerkmale der Familie	-.421**	-.030	-.391**
3. Kognitive Merkmale (Testleistungen)	-.621**	-.137**	-.484**

Anmerkungen: \*  $p < .05$ ; \*\*  $p < .01$

Obwohl die totalen direkten Effekte der familiären Struktur- und Prozessmerkmale schwach ausfallen, ergeben die totalen Effekte sowohl für die familiären Strukturmerkmale ( $\beta = -.420$ ,  $p < .01$ ) als auch für die familiären Prozessmerkmale ( $\beta = -.421$ ,  $p < .01$ ) hochsignifikante, moderat ausgeprägte Effekte auf die Deutschnote. Der mit Abstand stärkste totale Effekt ergibt sich dennoch für die kognitiven Testleistungen auf die Deutschnote ( $\beta = -.621$ ,  $p < .01$ ).

#### 7.4.2 Prüfung des Gesamtmodells: Effekte auf die Mathematiknote

Nach Prüfung der Passung des Gesamtmodells für die Deutschnote werden im Folgenden die Ergebnisse der Überprüfung für die Notenzusammensetzung der Viertklassnote im Fach Mathematik präsentiert. [Abbildung 9](#) gibt einen Überblick über die analysierten Zusammenhänge. In [Tabelle 32](#) werden die angesprochenen Fehlerkorrelationen zwischen den Lehrkräfteeinschätzungen präsentiert, [Tabelle 33](#) und [Tabelle 34](#) zeigen die indirekten und totalen Effekte der Strukturgleichungsmodellierung. Für das Signifikanzniveau ist anzumerken, dass hochsignifikante Pfade ( $p < .01$ ) fett gedruckt, signifikante Pfade dünn durchgezogen ( $p < .05$ ) und nicht signifikante Pfade gestrichelt präsentiert werden.

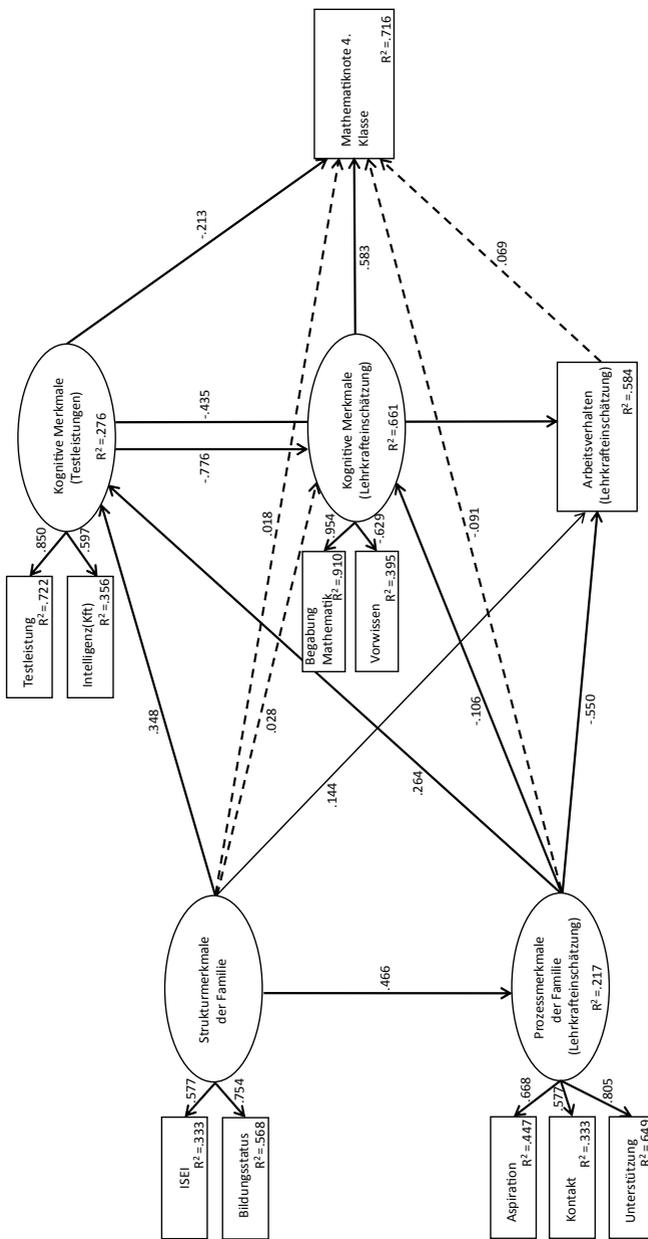


Abbildung 9: Strukturgleichungsmodell zum Einfluss individueller und familiärer Merkmale auf die Mathematiknote

Anmerkungen:  $N=1446$ ;  $\chi^2=178.331$ ,  $df=30$ ,  $p<.01$ ;  $RMSEA=.058$ ;  $SRMR=.052$ ;  $CFI=.945$ ; Fett gedruckter Pfeil:  $p<.05$ ; Gestrichelter Pfeil: n.s.

## Modellanpassung

Auch für die Mathematiknote zeigt die Güte des Modells aus [Abbildung 9](#) eine gute Passung auf die Daten (RMSEA = .058; SRMR = .052; CFI = .945). Obwohl der signifikante  $\chi^2$ -Wert mit 178.331,  $df = 30$  ( $p < .01$ ) zunächst ebenfalls gegen die Annahme des Modells zu sprechen scheint, muss erneut darauf verwiesen werden, dass für die Interpretation des Kennwerts die Stichprobengröße berücksichtigt werden muss, da für große Stichproben schon kleine Abweichungen von einer optimalen Modellpassung signifikant werden (Bühner, 2011).

## Direkte Effekte

Bezüglich der direkten Effekte der einbezogenen Prädiktoren für die Notenzusammensetzung im Fach Mathematik zeigt die [Abbildung 9](#), dass sich auch für die Mathematiknotenzusammensetzung keine signifikanten direkten Effekte der familiären Strukturmerkmale auf die Note und die kognitive Merkmalseinschätzung ergeben. Konträr zur Deutschnote ergibt sich allerdings ein signifikanter direkter Effekt auf die Einschätzung des Arbeitsverhaltens des Schülers ( $\beta = .144$ ,  $p < .05$ ). Hochsignifikante direkte Effekte der familiären Strukturmerkmale ergeben sich analog zur Deutschnote auf die familiären Prozessmerkmale ( $\beta = .466$ ,  $p < .01$ ) und die kognitiven Testleistungen der Schüler ( $\beta = .348$ ,  $p < .01$ ), auch ergeben sich für jene Prozessmerkmale der Familie ebenso keine signifikanten direkten Effekte auf die Mathematiknote. Dennoch wirken die familiären Prozessmerkmale direkt und hochsignifikant auf die kognitiven Testleistungen ( $\beta = .264$ ,  $p < .01$ ), auf die kognitive Merkmalseinschätzung ( $\beta = -.106$ ,  $p < .01$ ) und auf das Arbeitsverhalten ( $\beta = -.550$ ,  $p < .01$ ) des Schülers. Die Effekte fallen hier allerdings etwas schwächer aus als für die Deutschnote. Der positive Koeffizient auf die Testleistung ist im Sinne einer höheren Testleistung zu lesen, für die Merkmalseinschätzung gilt die gleiche Polung wie für eine Notenskala, niedrigere Werte entsprechen besseren Einschätzungen. Je höher (besser) die Lehrkräfte also die familiären Prozessmerkmale einschätzen, desto niedriger (besser) schätzen sie wiederum deren kognitive Merkmale und deren Arbeitsverhalten ein.

Im Gegensatz zur Deutschnote sind die Effekte auf Seiten der individuellen Schülermerkmale für die Mathematiknote etwas stärker. So wirken die kognitiven Testleistungen der Schüler hochsignifikant direkt auf die eingeschätzten kognitiven Merkmale ( $\beta = -.776, p < .01$ ), auf das Arbeitsverhalten ( $\beta = -.435, p < .01$ ) und nicht zuletzt auf die Mathematiknote ( $\beta = -.213, p < .01$ ). Je höher (besser) also die Testleistung des Kindes ist, desto niedriger (besser) fallen deren Merkmalschätzung bzw. deren Note aus. Analog zur Deutschnote ergibt sich der stärkste direkte und hochsignifikante Effekt für die eingeschätzten kognitiven Schülermerkmale auf die Mathematiknote ( $\beta = .583, p < .01$ ), konträr zur Deutschnote wirkt sich das wahrgenommene Arbeitsverhalten des Schülers allerdings nicht signifikant direkt auf die Mathematiknote aus ( $\beta = .069, n.s.$ ).

Tabelle 32: Korrelationen der Fehlerterme im Strukturgleichungsmodell für das Fach Mathematik

Variablenpaar	Fehlerkorrelation
Arbeitsverhalten/kognitive Merkmale (LK)	.015
Vorwissen/Begabung Mathematik	.474*

Anmerkungen: \*  $p < .05$ ; \*\*  $p < .01$

### Indirekte Effekte

Wie in [Kapitel 7.4.1](#) bereits erwähnt, ist ein Vorteil der Strukturgleichungsmodellierung, indirekte Effekte zwischen den einbezogenen Untersuchungsvariablen zu erfassen (Reinecke, 2005).

[Tabelle 33](#) gibt einen Überblick zu den standardisierten indirekten Effekten, um auch für die Mathematiknote der vierten Klasse zu klären, inwiefern sich Vermittlungseffekte zwischen den einbezogenen Variablen für die Notenzusammensetzung ergeben. Geprüft wird beispielsweise, inwiefern sich die familiären Herkunftsmerkmale über die Lehrkräfteeinschätzungen der individuellen Schülermerkmale auf die Zusammensetzung der Mathematiknote auswirken. Indirekte Effekte wurden in Anlehnung an das empirische Modell (vgl. [Abbildung 9](#))

für die Struktur- und Prozessmerkmale der Familie und die kognitiven Testleistungen der Schüler gemessen. Während die nachfolgende Tabelle über jeden einzelnen Vermittlungspfad Auskunft gibt, werden im Text analog zur Deutschnote auffällige Effekte für das Fach Mathematik berichtet.

Tabelle 33: Standardisierte indirekte Effekte im Strukturgleichungsmodell auf die Mathematiknote

Ausgangsvariable	Vermittlungspfad I	Vermittlungspfad II	Vermittlungspfad III	Indirekte Effekte auf die Mathematiknote
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	--	--	-.074**
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	-.157**
Strukturmerkmale der Familie	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	-.010
Strukturmerkmale der Familie	Kognitive Merkmale (LK)	--	--	.016
Strukturmerkmale der Familie	Arbeitsverhalten (LK)	--	--	.010
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	--	--	-.042
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	--	-.026*
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	-.056**
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	-.004

Ausgangsvariable	Vermittlungspfad I	Vermittlungspfad II	Vermittlungspfad III	Indirekte Effekte auf die Mathematiknote
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Kognitive Merkmale (LK)	--	-.029**
Strukturmerkmale der Familie	Prozessmerkmale der Familie (LK)	Arbeitsverhalten (LK)	--	-.018
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	--	--	-.056**
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	-.119**
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	-.008
Prozessmerkmale der Familie (LK)	Kognitive Merkmale (LK)	--	--	-.062**
Prozessmerkmale der Familie (LK)	Arbeitsverhalten (LK)	--	--	-.038
Kognitive Merkmale (Testleistungen)	Kognitive Merkmale (LK)	--	--	-.452**
Kognitive Merkmale (Testleistungen)	Arbeitsverhalten (LK)	--	--	-.030

Anmerkungen: Der in der linken Spalte aufgeführte Prädiktor stellt jeweils den Ausgangspunkt des indirekten Effekts auf die abhängige Variable dar, Vermittlungspfad I-III umfassen die jeweiligen Mediationsschritte. \*  $p < .05$ ; \*\*  $p < .01$ .

Der totale indirekte Effekt der familiären Strukturmerkmale ist moderat ausgeprägt ( $\beta = -.390$ ,  $p < .01$ ), fällt aber etwas schwächer aus als für die Deutschnote (vgl. [Tabelle 34](#)). Wie [Tabelle 33](#) zu entnehmen ist, ergeben sich bei den familiären Strukturmerkmalen auch für die Mathematiknote die stärksten indirekten Effekte vermittelt über die kognitiven Testleistungen und die kognitiven Merkmalseinschätzungen ( $\beta = -.157$ ,  $p < .01$ ). Die These, dass die familiären Strukturmerk-

male über die kognitiven und nicht-kognitiven Lehrkräfteeinschätzungen vermittelt werden, findet auch für die Mathematiknote nur schwach Bestätigung, da die indirekten Effekte der Strukturmerkmale allein über die kognitive Merkmalseinschätzung ( $\beta = .016$ , n.s.) und das Arbeitsverhalten ( $\beta = .010$ , n.s.) nicht signifikant werden.

Der Vermittlungspfad von Strukturmerkmalen auf die familiären Prozessmerkmale auf die Mathematiknote ist ebenfalls nicht signifikant ( $\beta = -.042$ , n.s.), allerdings ergeben sich analog zur Deutschnote signifikante Pfade zwischen Struktur- und Prozessmerkmalen und den individuellen Schülermerkmalen auf die Mathematiknote (vgl. [Tabelle 33](#)). So zeigt sich z.B. auch für die Mathematiknote ein schwacher, aber hochsignifikanter indirekter Effekt zwischen den Struktur- und Prozessmerkmalen und der Einschätzung der kognitiven Schülermerkmale ( $\beta = -.029$ ,  $p < .01$ ), wobei der Effekt noch etwas kleiner ausfällt als für die Deutschnote (vgl. [Tabelle 30](#) bzw. [Tabelle 33](#)).

Auf Seiten der familiären Prozessmerkmale ist der totale indirekte Effekt auf die Mathematiknote ebenfalls moderat ausgeprägt, insgesamt allerdings etwas schwächer als für die Deutschnote ( $\beta = -.283$ ,  $p < .01$ ; vgl. [Tabelle 34](#)). Im Gegensatz zur Deutschnote ergeben sich für die indirekten Effekte der familiären Prozessmerkmale auf die Mathematiknote die stärksten Effekte vermittelt über die kognitiven Testleistungen und die kognitive Merkmalseinschätzung ( $\beta = -.119$ ,  $p < .01$ ). Zudem ergeben sich auch für die anderen Prädiktoren hochsignifikante, aber schwächer ausgeprägte indirekte Effekte auf die Mathematiknote (kognitive Merkmalseinschätzung:  $\beta = -.062$ ,  $p < .01$ ; kognitive Testleistung:  $\beta = -.056$ ,  $p < .01$ ).

Bezüglich der kognitiven Testleistungen ergibt sich ebenfalls ein moderater totaler indirekter Effekt ( $\beta = -.482$ ,  $p < .01$ ; vgl. [Tabelle 34](#)). Analog zur Deutschnote wird auch für die Mathematiknote der stärkste indirekte Effekt für die kognitiven Testleistungen über die kognitive Merkmalseinschätzung vermittelt ( $\beta = -.452$ ,  $p < .01$ ).

### Totale Effekte

Der totale Effekt einer Strukturgleichungsmodellierung ergibt sich aus der Summe der im Modell aufgeführten direkten und indirekten Effekte (Reinecke, 2005). [Tabelle 34](#) gibt einen Überblick über die standardisierten totalen Effekte aus dem Strukturgleichungsmodell für die Mathematiknote am Ende der vierten Jahrgangsstufe.

Tabelle 34: Standardisierte totale Effekte im Strukturgleichungsmodell auf die Mathematiknote

Ausgangvariable	Totaler Effekt	Totaler direkter Effekt	Totaler indirekter Effekt
1. Strukturmerkmale der Familie	-.372**	.018	-.390**
2. Prozessmerkmale der Familie	-.374**	-.091	-.283**
3. Kognitive Merkmale (Testleistungen)	-.696**	-.213**	-.482**

Anmerkungen: \*  $p < .05$ ; \*\*  $p < .01$

Äquivalent zur Deutschnote ergeben die totalen direkten Effekte der familiären Struktur- und Prozessmerkmale schwach ausgeprägte, nicht signifikante Effekte. Dagegen verweisen die Ergebnisse der totalen Effekte sowohl für die familiären Strukturmerkmale ( $\beta = -.372$ ,  $p < .01$ ) als auch für die familiären Prozessmerkmale ( $\beta = -.374$ ,  $p < .01$ ) auf hochsignifikante Effekte auf die Mathematiknote, sie fallen insgesamt aber etwas schwächer aus als für die Deutschnote (vgl. [Tabelle 31](#)). Der mit Abstand stärkste totale Effekt geht auch für die Mathematiknote von den kognitiven Testleistungen der Schüler aus und fällt sogar etwas stärker aus als für die Deutschnote ( $\beta = -.696$ ,  $p < .01$ ).

#### 7.4.3 Effekte der Klassenkomposition auf die Deutsch- und Mathematiknote

Der letzte Punkt des vorliegenden Ergebnisteils beschäftigt sich mit der dritten Fragestellung dieser Arbeit. Obgleich bei den vorangegangenen Analysen der Individualnoten die Klassenzugehörigkeit bereits kontrolliert wurde, sind für die Zusammensetzung einer Note Fra-

gen nach der Bedeutung der Klassenkomposition nahezu unumgänglich. Aus diesem Grund beschäftigt sich dieses Kapitel mit der Frage, inwiefern die leistungsbezogene und soziale Klassenzusammensetzung Unterschiede in der Benotung auf Individualebene erklären können. Nach Schätzung des Nullmodells für die Intraklassenkorrelationen wird eine Mehrebenenanalyse berechnet, die Zentrierung erfolgte um den Gesamtmittelwert der Stichprobe<sup>42</sup> und liefert damit eine direkte Kompositionseffektschätzung (Christ & Schlüter, 2010; Ditton, 1998).

### **Nullmodell: Intraklassenkorrelationen für die Deutsch- und Mathematiknote**

Zu Beginn einer Mehrebenenanalyse bedarf es zunächst einer Prüfung des Grades der Abhängigkeit der Beobachtungen zwischen den interessierenden Variablen, welche durch die sog. Intraklassenkorrelation berechnet wird. Die Intraklassenkorrelation bezieht sich auf den Anteil der Varianz der abhängigen Variable auf der Kontextebene aus der Summe der auf Individual- und Kontextvarianz berechneten Gesamtvarianz (Christ & Schlüter, 2010). Obwohl die vorliegende Untersuchung lediglich die Abhängigkeit der Notenzusammensetzung von der leistungs- und sozialen Klassenzusammensetzung prüft, wird nachfolgend in [Tabelle 35](#) die Intraklassenkorrelation aller berechneten Variablen für das Fach Deutsch und Mathematik vorgestellt. Hier interessiert die Frage, inwiefern sich die Lehrkrifteinschätzungen und die individuellen Testleistungen der Schüler generell zwischen den Klassen unterscheiden, um dann in einem zweiten Schritt, Effekte der leistungsbezogenen und sozialen Klassenzusammensetzung zu prüfen.

---

<sup>42</sup> Die Wahl der Metrik der Prädiktoren ist ein vielfach behandeltes Thema, das in der vorliegenden Arbeit nicht weiter ausgeführt wird. Unterschieden wird meist zwischen der Zentrierung um den Gesamtmittelwert der Stichprobe oder um die Gruppenmittelwerte, für die wiederum unterschiedliche Vor- und Nachteile angeführt werden (Ditton, 1998; Raudenbush & Bryk, 2002). Die Wahl der Zentrierung um den Gesamtmittelwert erfolgte aufgrund der vorliegenden Frage nach kompositionellen Effekten der Klassenzusammensetzung auf die individuellen Noten, für welche mit der gewählten Zentrierung der Kompositionseffekt direkt ermittelt werden kann (Ditton, 1998).

Tabelle 35: Intraklassenkorrelationen für die Deutsch- und Mathematiknote

Schulfächer (4.Klasse)	Variable	Intraklassenkorrelation (ICC)
<b>Deutsch</b>	Deutschnote	.082
	Testleistung	.063
	Intelligenz (kft)	.053
	Sprachliche Fähigkeiten	.071
	Begabung Deutsch	.026
	Vorwissen	.272
	Arbeitsverhalten	.119
	Sozialer Status (ISEI)	.086
	Bildungsstatus	.092
	Kontakt Elternhaus	.477
	Unterstützungsverhalten	.192
Bildungsaspiration	.178	
<b>Mathematik</b>	Mathematiknote	.068
	Testleistung	.054
	Intelligenz (kft)	.050
	Begabung Mathematik	.040
	Vorwissen	.277
	Arbeitsverhalten	.117
	Sozialer Status (ISEI)	.085
	Bildungsstatus	.094
	Kontakt Elternhaus	.477
	Unterstützungsverhalten	.192
	Bildungsaspiration	.180

Anmerkungen: N=1446; Cluster (Klasse)=77; Durchschnittliche Schülerzahl=18.78

Wie aus [Tabelle 35](#) ersichtlich wird, ergibt die Intraklassenkorrelation für die Deutsch- und Mathematiknote der 4.Klasse, dass in Deutsch 8,2 Prozent und in Mathematik 6,8 Prozent der Notenvarianz durch Unterschiede zwischen den Schulklassen erklärt werden können. Bei einer durchschnittlichen Klassengröße von  $N=18,78$  sind auch die individuelle Testleistung, Lehrkräfteeinschätzungen sowie die familiären Struktur- und Prozessmerkmale zwischen den Klassen verschieden. Allerdings ergeben sich für die kognitiven Merkmaleinschätzungen,

mit Ausnahme des Vorwissens, und für die familiären Strukturmerkmale deutlich geringere Intrakorrelationen als für die eingeschätzten familiären Prozessmerkmale und das Arbeitsverhalten des Schülers.

Obwohl die Intraklassenkorrelationen für die beiden Fachnoten auf eine mäßige Varianz zwischen den Klassen hinweisen, können auch kleine Intraklassenkorrelationen relevant werden. Ebenso müssen für die Interpretation der Stärke immer die Besonderheiten der jeweiligen Cluster und deren Größe berücksichtigt werden (Muthén & Satorra, 1995). Im Folgenden werden die Ergebnisse der Mehrebenenanalyse vorgestellt. Zur Beantwortung der Frage nach Effekten der leistungsbezogenen und sozialen Klassenzusammensetzung für die fachspezifische Note wird das Individualmodell aus [Kapitel 7.4](#) verwendet und um die entsprechenden Merkmale auf Klassenebene ergänzt (Ditton, 1998).

### **Effekte der Klassenkomposition auf die Deutsch- und Mathematiknote**

Nachfolgend werden für das Strukturgleichungsmodell zur Erklärung der Deutschnote das durchschnittliche Leistungsniveau (Testleistung) und die durchschnittliche soziale Herkunft (Bildungsstatus, sozialer Status) der Klasse als Aggregatmerkmale in die Analyse miteinbezogen. Angenommen wird, dass die Deutschnote umso höher (schlechter) ausfällt, je höher das Leistungsniveau der Klasse ist. Gleiches gilt für die soziale Klassenzusammensetzung; gemäß dem Forschungsstand müssten Schüler in sozial günstigeren Klassenzusammensetzungen unter Kontrolle der Individualmerkmale weniger gut bewertet werden (vgl. [Kap. 4.4](#)). Die Analysen zeigen, dass sich sowohl für die mittlere soziale Klassenzusammensetzung als auch für das mittlere Leistungsniveau keinerlei signifikante Effekte auf die Deutsch- und Mathematiknote der vierten Klasse ergeben.

Die standardisierten Ergebnisse für die Deutschnote zeigen, dass sowohl der mittlere Leistungsindex der Klasse ( $\beta = .226$ , n.s.) als auch die mittlere bildungsbezogene ( $\beta = -.277$ , n.s.) und die mittlere sozioökonomische Klassenzusammensetzung ( $\beta = .233$ , n.s.) keinen signifikanten Einfluss auf die Noten nehmen. Gleiches gilt für die Mathematiknote, auch

hier ergibt die Prüfung der aggregierten Individualmerkmale keinerlei signifikante Effekte auf die Notenzusammensetzung auf Individual-ebene. Weder für die mittlere Testleistung der Klasse ( $\beta = .074$ , n.s.) noch für die mittlere bildungsbezogene ( $\beta = -.114$ , n.s.) und die mittlere sozioökonomische Klassenzusammensetzung ( $\beta = .018$ , n.s.) ergeben sich signifikante Effekte auf die Noten. Für eine detaillierte Darstellung siehe [Abbildung A-1](#) bzw. [Abbildung A-2](#) im Anhang dieser Arbeit.



## 8 Diskussion

Zum Abschluss der vorliegenden Arbeit werden – nach einem Rückbezug auf die Untersuchungsziele – die ermittelten Ergebnisse zusammengefasst und diskutiert. Hierauf folgt eine abschließende Beantwortung der eingangs formulierten Hypothesen, um dann in einem letzten Schritt die wichtigsten Befunde bilanzierend festzuhalten, kritisch Bilanz zu ziehen und zentrale Implikationen der Arbeit aufzuzeigen.

### 8.1 Untersuchungsziele der Arbeit

Die nach wie vor bestehende Abhängigkeit zwischen familiären Hintergrundmerkmalen und schulischem Bildungserfolg diente der vorliegenden Arbeit als Ausgangspunkt, um einen eigenständigen Beitrag zur Ursachensuche jener Ungleichheit zu leisten.

Im Fokus stand die Notenzusammensetzung in den beiden, für den Sekundarschulerfolg maßgeblichen Fächern Deutsch und Mathematik in der vierten Jahrgangsstufe am Ende der Grundschulzeit. Während die Suche nach Erklärungen von Bildungsdisparitäten am Übergang in die Sekundarstufe I bereits häufig erfolgt, liegen nach wie vor wenige Analysen zur Zusammensetzung jener schulischen Leistung – ausgedrückt in Schulnoten – vor, die letztlich auch den Werdegang des Schülers im Bildungssystem bestimmen. Gefragt wurde einmal, welche individuellen kognitiven und nicht-kognitiven Schülermerkmale es überhaupt sind, die als Leistung bewertet werden und damit den Schulerfolg des Kindes maßgeblich beeinflussen. Geklärt wurde anschließend, in welchem Zusammenhang diese Schülermerkmale wiederum mit den familiären Hintergrundmerkmalen stehen, inwiefern die Schülermerkmale also unabhängig vom familiären Hintergrund in die Notenzusammensetzung einfließen, und ob auch der Klassenzusammensetzung eine Bedeutung zukommt.

Auch wenn es nicht das Ziel dieser Arbeit sein konnte, ein vollständiges Bild der Bedingungsfaktoren einer Schulnote aufzuzeigen, wurde den Fragen sowohl auf theoretischer als auch empirischer Ebene nachgegangen.

Auf theoretischer Basis wurde zunächst das Ansinnen verfolgt, die Begrifflichkeiten um den schulischen Leistungsbegriff zu ordnen und deren Bedeutung für die Notenzusammensetzung in der schulischen Praxis herauszuarbeiten. Ebenfalls stand im Fokus, welche konkreten Anforderungen an die Notengebung gestellt werden: Auf Seiten des Schulrechts wurde nach den institutionalisierten Vorgaben gefragt, um dann deren Anforderungen auf funktionaler und messtheoretischer Ebene zu diskutieren und auf ihre Stimmigkeit mit empirischen Korrelaten zu prüfen.

Neben diesen Anforderungen existieren mittlerweile zahlreiche Modelle, die meist auf Makroebene die multikausale Bedingungsstruktur schulischer Leistung in schematischen Einteilungen systematisieren; Faktoren des Schülers selbst, Faktoren auf Seiten der Familie und der Schule sind dabei die drei Forschungsbereiche, die meist in diesem Zusammenhang analysiert werden. Ausgehend von der theoretischen Annahme, dass sich die Prädiktoren gegenseitig in ihrer Wirksamkeit beeinflussen, war es das Ziel der vorliegenden Arbeit, die Relevanz ausgewählter Prädiktoren für die fachspezifische Notenzusammensetzung zu prüfen, die gemäß der Forschungslage, der Modelle von Helmke und Weinert (1997), Heller (1995) und der dargestellten schulrechtlichen Grundlagen Einfluss auf die Notenzusammensetzung nehmen.

Ausgehend von einem theoretisch hergeleiteten Gesamtmodell wurde mittels Korrelations- und Regressionsanalysen sowie Strukturgleichungsmodellen geprüft, welchen Einfluss individuelle Schülermerkmale auf die Notenzusammensetzung in Deutsch und Mathematik haben. Im Sinne der Ausgangslage wurde ferner geprüft, ob familiäre Strukturmerkmale sowie ausgewählte, schulbezogene Prozessmerkmale der Familie direkt und vermittelt Einfluss auf die fachspezifische Notenzusammensetzung nehmen; inwiefern die von den Lehrkräften

eingeschätzten Schülermerkmale also in Abhängigkeit von familiären Herkunftsmerkmalen variieren. Aufgrund theoretischer und empirischer Erkenntnisse wurde zudem mithilfe einer Mehrebenenanalyse festgestellt, ob die leistungsbezogene und soziale Klassenkomposition, zusätzlich zu den Prädiktoren auf Individualebene, auf die Note Einfluss nimmt.

Eine Besonderheit der vorliegenden Analysen war die gewählte Erhebungsperspektive. Unter Ausnahme der individuellen Testleistungen wurden sowohl für die kognitiven und nicht-kognitiven Schülermerkmale als auch für die familiären Prozessmerkmale Einschätzungen der Lehrkraft in die Analysen miteinbezogen. Diese Auswahl wurde getroffen, um die Sicht aus Lehrkraftperspektive bzw. deren Auswahl der Leistungskomponenten für Schulnoten zu erfassen und im Sinne der schichtspezifischen Funktionalitätsannahme zu prüfen, inwiefern diese institutionell legitimierte Bewertungskultur herkunftsabhängige Bildungsprozesse fördert.

Welche empirischen Antworten die vorliegende Arbeit dazu fand, gilt es im nachfolgenden Kapitel zusammenfassend zu diskutieren.

## 8.2 Zusammenfassung und Diskussion der Ergebnisse

Die Ergebnisse der dargestellten Analysen beziehen sich auf drei große Inhaltsbereiche: Die Bedeutung individueller kognitiver und nicht-kognitiver Schülermerkmale für die fachspezifische Notenzusammensetzung in der vierten Jahrgangsstufe sowie deren Interaktionseffekte (I), die Bedeutung zusätzlich einbezogener Struktur- und Prozessmerkmale der Familie und den vermittelnden Prozessen zwischen Familie, Schüler und Schulnote (II) und das analysierte Gesamtmodell, dass neben den individuellen und familiären Merkmalen auch die Relevanz der leistungsbezogenen und sozialen Klassenkomposition für die Notenzusammensetzung aufgreift (III). Die Hauptergebnisse der drei Bereiche werden an dieser Stelle noch einmal zusammengefasst und

unter Einbezug der eingangs vorgestellten theoretischen Grundlagen diskutiert.

### **Einfluss individueller Schülermerkmale auf die Notenzusammensetzung in Deutsch und Mathematik am Ende der Grundschulzeit**

Bereits auf deskriptiver Basis zeigt sich bezüglich der kognitiven und nicht-kognitiven Schülermerkmale, dass die Merkmalseinschätzungen der fachspezifischen Begabung in Deutsch und Mathematik, sprachlichen Fähigkeiten, Vorwissen, Arbeits- und Sozialverhalten mit Ausnahme des Sozialverhaltens auf mittlerem bis starkem Niveau untereinander und mit der Testleistung des Schülers korrelieren. Bezüglich der Zusammenhänge zwischen fachspezifischen Noten und individuellen Merkmalen ergeben sich insgesamt für das eingeschätzte Sozialverhalten, das eingeschätzte Vorwissen aus der zweiten Klasse und die Testintelligenz (*kft*) die geringsten Zusammenhänge mit den Schulnoten.

Bezüglich der kausalen Effekte der kognitiven und nicht-kognitiven Schülermerkmale für die Notenzusammensetzung in Deutsch und Mathematik ergeben die hierarchischen Regressionsanalysen bezüglich der Hypothese 1 – der Frage nach unterschiedlichen Effekten der Test- und Merkmalseinschätzungen – dass die eingeschätzten kognitiven Schülermerkmale fachübergreifend den stärksten Erklärungswert für die Schulnoten ergeben. Die rasante Abnahme der Effektstärke der kognitiven Testleistungen durch die Hinzunahme der Merkmalseinschätzungen sprechen allerdings für gemeinsame Varianzanteile. Bezüglich der kognitiven Testleistungen ergeben sich unter Kontrolle der anderen individuellen Schülermerkmale für die Intelligenz (*kft*) nur noch sehr schwache Effekte, die Testleistung bleibt in beiden Modellen hochsignifikant relevant. Ebenfalls fachübergreifend sind bezüglich der Vorwissenseinschätzung aus der zweiten Jahrgangsstufe schwache, aber signifikante Effekte auf die Notenzusammensetzung in der vierten Klasse festzustellen. Effekte der nicht-kognitiven Schülermerkmale zeigen für das Arbeitsverhalten die stärkeren, hochsignifikanten Effekte auf die Notenzusammensetzung, während das Sozialverhalten nahezu keinen Einfluss nimmt. Ebenfalls fachübergreifend

nimmt das Geschlecht des Schülers, unter Kontrolle der anderen Prädiktoren, keinen Einfluss auf die Schulnoten.

Zusätzlich ergeben sich fachspezifische Unterschiede in der Notenzusammensetzung, die sich vornehmlich auf die Prädiktorstärke beziehen. Während im Fach Deutsch die sprachliche Fähigkeitseinschätzung gefolgt von der Begabungseinschätzung den stärksten Effekt auf die Notenzusammensetzung aufweist, ist es im Fach Mathematik die mathematische Begabungseinschätzung; die sprachlichen Fähigkeiten sind hier nahezu irrelevant. Zudem ist die Varianzaufklärung der Noten für die einbezogenen Prädiktoren im Fach Deutsch etwas höher (72,9 %) als im Fach Mathematik (67,7%).

Ebenfalls fachspezifisch unterschiedliche Ergebnisse ergeben sich bezüglich der zusätzlich einbezogenen Interaktionseffekte zwischen dem Arbeitsverhalten und den kognitiven Merkmalseinschätzungen. Unter Berücksichtigung der Interaktionen wird für das Fach Deutsch der Haupteffekt des Arbeitsverhaltens hochsignifikant bedeutsamer sowie die Interaktion zwischen sprachlichen Fähigkeiten und Arbeitsverhalten einen hochsignifikant negativen Effekt zeigt. Für das Fach Mathematik wird dagegen unter Berücksichtigung der Interaktionen der Haupteffekt des Arbeitsverhaltens abgeschwächt, die Interaktion zwischen Begabungseinschätzung und Arbeitsverhalten ergibt einen hochsignifikant positiven Koeffizienten.

Inwiefern entsprechen die Befunde nun der theoretischen Basis der Arbeit mit ihren schulrechtlichen Grundlagen und der empirischen Befundlage und welche Erkenntnisse müssen diskutiert werden?

Während zu den begrifflichen Grundlagen bezüglich der einbezogenen Bewertungsmaßstäbe (vgl. [Kap. 2.2](#)) zunächst keine Aussagen gemacht werden können, gilt für die Schulnote festzuhalten (vgl. [Kap. 2.3](#)), dass es sich bei den abhängigen Variablen um Zeugnisnoten handelt, die sich vermutlich aus mündlichen und schriftlichen Einzelnoten zusammensetzen und zum Ende der Grundschulzeit eine Selektions- und Allokationsfunktion zu erfüllen haben. Während über Menge,

Maßstab, Art und Stimmigkeit der einbezogenen Einzelnoten anhand der Daten ebenfalls keine Aussagen getroffen werden können, kann anhand der unterschiedlich einbezogenen Perspektiven geprüft werden, inwiefern die Notenzusammensetzung besser durch individuelle, kognitive Testleistungen oder durch die kognitiven bzw. nicht-kognitiven Schülermerkmale aus Lehrkraftsicht zu erklären sind.

Für den Vergleich zwischen diesen beiden Erhebungsperspektiven ergibt sich im Einklang mit den bisherigen Forschungsbefunden, dass die Noten deutlich besser durch die Hinzunahme der Lehrkräfteeinschätzungen erklärbar sind. Sodass man daraus folgern kann, dass die fachspezifischen Testleistungen allein also eine „unvollständige“ Schulleistung – so wie sie in Schulnoten abgebildet werden – repräsentieren. Daraus ergibt sich dann aber auch die Frage, welche die zusätzlichen Aspekte sind, die in der Notengebung Berücksichtigung finden und ob diese im Einklang mit schulrechtlichen Vorgaben stehen. Unter Bezug auf die theoretischen Marko-Modelle von Helmke und Weinert (1997) und Heller (1995) wurden für die vorliegende Untersuchung einige ausgewählte nicht-kognitive bzw. motivationale und kognitive Schülermerkmale als jene zusätzlichen Aspekte geprüft, die laut Schulrecht aus Sicht der Lehrkraft im Bewertungssystem beachtet werden können.

Es wurde davon ausgegangen, dass sich die nicht-kognitiven bzw. motivationalen Faktoren des Schülers – aus Lehrkraftsicht – in dessen Arbeits- bzw. Lernverhalten zeigen und bewertet werden, was sich dann wiederum in den Schulnoten des Kindes widerspiegelt und auch von einschlägigen Forschungsbefunden in dieser Art interpretiert wird (Heller, 1984; Steinmayr & Meißner, 2013; Trautwein et al., 2014). Denn gemäß der schulrechtlichen Grundlage darf in der Notenzusammensetzung das Arbeitsverhalten des Schülers berücksichtigt werden, das Sozialverhalten gilt es dagegen extra aufzuführen (vgl. [Kap. 2.5](#)). Die regressionsanalytische Beleuchtung unter Einbezug des eingeschätzten Arbeits- und Sozialverhaltens bestätigt diese schulrechtliche Begebenheit. Das eingeschätzte Arbeitsverhalten wird unter Kontrolle der kognitiven Schülermerkmale von der Lehrkraft

in beiden Fächern bei der Notenzusammensetzung hochsignifikant berücksichtigt sowie über diese Einschätzung auch bestehende Geschlechtereffekte verschwinden. Diese Ergebnisse entsprechen den frühen Befunden von Kemmler (1967), aber auch neuere, internationale Abhandlungen bestätigen diese Tendenzen (vgl. z.B. Bennett et al., 1993; Nölle et al., 2009; Stubbe & Bos, 2008). Auch für die vermittelnden Geschlechtereffekte existieren empirische Evidenzen (vgl. z.B. Kuhl & Hannover, 2012). Die Befunde von Schumacher (2002), die in ihrer qualitativen Studie einen Zusammenhang zwischen Lehrkräfteeinschätzungen zum Sozialverhalten und deren Leistungsbewertungen fand, konnten dagegen in den vorliegenden quantitativen Analysen nicht bestätigt werden.

Für die einbezogenen kognitiven Merkmale war die Einschätzung der Lehrkraft zu den sprachlichen Fähigkeiten und der fachspezifischen Begabung das entscheidende Kriterium der Notenzusammensetzung. Entgegen der Forschungsbefunde ergeben sich für die Testintelligenz (kft) unter Kontrolle der anderen Prädiktoren nur sehr schwache Effekte. Der Befund zum Zusammenhang zwischen Intelligenz und Schulnoten reiht sich damit in die widersprüchliche Befundlage der Relevanz der Intelligenzprädiktoren für die schulische Leistung ein (vgl. Kap. 4.2.1) und kann somit nicht als Ursache der Begabungsunterschiede herangezogen werden. Im Einklang mit den bisherigen Forschungsbefunden (vgl. Kap. 4.2.1) zeichnen sich auch für das eingeschätzte Vorwissen aus der zweiten Klassenstufe schwache, aber signifikante Effekte auf die Notenzusammensetzung ab. Dieser Befund kann einmal dahingehend interpretiert werden als dass sich die Lehrkräfte untereinander über die Leistungsentwicklung des Schülers austauschen oder aber, dass das Vorwissen durch die anderen Individualmerkmale vermittelt wird. Gemäß der schulrechtlichen Grundlage (vgl. Kap. 2.5) wäre es zulässig, wenn sich die bewertende Lehrkraft über vorausgegangene Bewertungen des Schülers informiert. Inwiefern diese – sicherlich nie isoliert entstandene – Prüfungsnote dann allerdings von vorgeprägten Erwartungshaltungen gesteuert werden könnte (vgl. Kap. 4.4), bleibt an dieser Stelle offen, gilt es aber dennoch für den Bewertungsvorgang mit zu berücksichtigen.

Für die Deutschnote sind die sprachlichen Fähigkeiten der wichtigste Prädiktor, für die Mathematiknote sind diese nahezu irrelevant, dafür ist hier die mathematische Begabungseinschätzung maßgeblich. Betrachtet man vor dieser Tatsache die schulrechtlichen Grundlagen (vgl. [Kap. 2.5](#)), sind es einmal die fachspezifischen Anforderungen und Bildungsstandards, die in jenen fachspezifischen Noten gemessen werden sollen; sprachliche Fähigkeiten und eine gewisse Begabung benötigt man wohl für jegliche Fachanforderung, auch wenn die sprachlichen Fähigkeiten entgegen der theoretischen Annahmen für die Mathematiknote nahezu irrelevant sind.

Dass die sprachlichen Fähigkeiten – zumindest für die Deutschnote – eine entscheidende Einflussgröße schulischer Leistung zu sein scheinen, deckt sich trotz veränderter Erhebungsperspektive und unterschiedlichen Operationalisierungen mit dem bisherigen Forschungsstand (vgl. [Kap. 4.2.1](#)). Für die Relevanz der Schülerbegabung fanden auch Anders, McElvany und Baumert (2010) für die kindliche Begabungseinschätzung höhere Effekte auf die Mathematiknote als auf die Deutschnote. Zusätzlich dazu wäre es im Sinne einer hohen Konstruktvalidität (vgl. [Kap. 3.3](#)) für die Noten erforderlich, dass die Mathematiknote am besten durch die Mathematikleistung des Schülers vorhergesagt werden kann und die Deutschnote durch die sprachlichen Fähigkeiten. Allerdings bleibt in diesen Analysen zunächst offen, inwiefern die mathematische Begabungseinschätzung – als mit Abstand stärkster Prädiktor im Fach Mathematik – mit der Testleistung gleichzusetzen ist.

Geht man davon aus, dass die Begabungseinschätzung auch unter Kontrolle der individuellen Testleistung für beide Fächer eine entscheidende Einflussdeterminante darstellt, muss im Sinne der Argumentation aus [Kapitel 4.2.1](#) auch nach der Ursache dieser Begabungen gefragt werden. Obgleich durch die Analysen keine endgültige Aussage über die Ursache jener Begabungseinschätzung getroffen werden kann, gilt es doch zu hinterfragen, inwiefern die Begabung unabhängig von sozialen Hintergrundmerkmalen entstanden ist und ob ihre Einschätzung dadurch beeinflusst wird. Denn selbst wenn

die Einschätzung statistisch unabhängig von den familiären Hintergrundmerkmalen sein sollte, gilt es doch zu diskutieren, inwiefern die Vorstellung von Begabung wiederum jenen Aspekten bzw. Fähigkeiten entspricht, die in Familien aus höheren Sozialschichten häufiger vorhanden sind. Im Sinne der Argumentation in [Kapitel 4.2.1](#) zur Umweltdetermination von Begabung müsste man davon ausgehen, dass Begabungsunterschiede im hohen Maße von sozialen Faktoren abhängig sind. Wenn diese maßgeblich die Leistungen von Schülern beeinflussen, wären die als objektiv bezeichneten Leistungskriterien gleichzeitig auch soziale Reproduktionsmechanismen. Dergestalt, dass Kinder aus höheren Schichten mehr von jenen Eigenschaften aufweisen, deren Ursache spezifischen Begabungen zugeschrieben wird und in dessen Folge höhere Leistungen erzielen. Aspekte, die im Gesamtmodell zumindest einführend analysiert werden konnten. Die andere These zur Begabung, nach der die Anlage die maßgebliche Determinante ist (vgl. [Kap. 4.2.1](#)), soll an dieser Stelle jedoch nicht unterschlagen werden. Folgt man dieser Argumentation, ist Begabung maßgeblich durch die inneren Anlagen des Kindes determiniert. Eine Ansicht, die gemäß Roth (1967) allerdings zur Folge haben könnte, die Förderbarkeit und Umweltbedeutung von Begabung zu vernachlässigen, die dann wiederum zu selektiven Förderstrukturen führen könnten. In dem Sinne, dass als begabt angesehene Schüler von vorneherein als *gute* Schüler den Schulweg durchlaufen, während fehlende Begabung als begrenzt förderbar angesehen werden könnte.

Im Sinne der Forschungsbasis wurde zudem angenommen (vgl. [Kap. 4.2.1](#)), dass der Zusammenhang zwischen Testleistungen, Merkmalseinschätzungen und Schulnote im Sinne einer diagnostisch validen Einschätzung relativ hoch sein müsste, was durch die regressionsanalytische Auswertung zunächst nur vorsichtig bestätigt werden kann und im Gesamtmodell ebenfalls geprüft wurde.

Bezüglich des von Heller (1995) konstatierten multikausalen Bedingungsmodells schulischer Leistung wurde in der vorliegenden Untersuchung ebenfalls geprüft, inwiefern die kognitiven Merkmalseinschätzungen in Abhängigkeit vom Arbeitsverhalten des Schülers

unterschiedlich stark auf die fachspezifische Note Einfluss nehmen. Diesbezüglich ergeben sich fachspezifische Unterschiede.

Für die Deutschnote werden unter Berücksichtigung der Interaktionen die direkten Effekte der sprachlichen Fähigkeiten, Begabung und des Arbeitsverhaltens etwas stärker. Zudem zeigt sich ein schwacher, hochsignifikant negativer Effekt des Interaktionsterms zwischen Arbeitsverhalten und sprachlichen Fähigkeiten. Die Ergebnisse weisen ähnliche Tendenzen zu den Befunden von Gut, Reimann und Grob (2012) auf, die zeigen, dass Schüler, die in ihren sprachlichen Fähigkeiten und ihrem Arbeitsverhalten bereits als *gut* eingeschätzt werden, nicht eine noch bessere Note erhalten. Dagegen hat die Berücksichtigung der Interaktion für einen Schüler mit schlechten Einschätzungen eine stärkere kompensatorische Wirkung als für einen Schüler mit ohnehin guten Bewertungen. Der etwas stärkere direkte Effekt des Arbeitsverhaltens kann dahingehend interpretiert werden, dass es für die Viertklassnote im Fach Deutsch wichtig ist, durch welches Arbeitsverhalten sich das Kind auszeichnet. Die sprachliche Fähigkeitseinschätzung nimmt allerdings – unter Berücksichtigung der Interaktionen – weiterhin am stärksten Einfluss auf die Deutschnote.

Dagegen zeigen die Ergebnisse für die Notenzusammensetzung im Fach Mathematik unter Berücksichtigung der Interaktionen einen leicht abgeschwächten direkten Effekt der Begabungseinschätzung auf die Note. Die direkten Effekte der anderen kognitiven Merkmale verändern sich ebenfalls nur schwach. Bezüglich des nicht-kognitiven Merkmals Arbeitsverhalten wird der direkte Effekt unter Berücksichtigung der Interaktionen etwas abgeschwächt. Für die einbezogenen Interaktionen ergibt sich ein schwacher, hochsignifikant positiver Interaktionsterm zwischen Arbeitsverhalten und mathematischer Begabungseinschätzung. Berücksichtigt man im Modell also neben den beiden Haupteffekten (direkte Pfade von Arbeitsverhalten, mathematische Begabung) noch den zusätzlichen Effekt zwischen mathematischer Begabung und Arbeitsverhalten, kann dieser Effekt im Sinne eines verstärkenden Effekts interpretiert werden. Dagegen weisen die etwas abgeschwächten direkten Effekte des Arbeitsverhaltens unter

Berücksichtigung der Interaktionen darauf hin, dass bei mathematisch begabten Schülern das Arbeitsverhalten etwas weniger relevant ist. Im Sinne der Ergebnisse von Steinkamp (1976) ist an dieser Stelle also auch die Begabung des Schülers wichtiger als dessen Arbeitsverhalten. Unter Bezug auf die Argumentation von Solga (2009) in [Kapitel 4.2.1](#), welche die (fehlende) Differenzierung zwischen Begabung und Fleiß diskutiert, wäre an dieser Stelle aufgrund der vorliegenden Analysen vorsichtig zu folgern, dass Begabungen andere Komponenten beinhalten als den Fleiß des Schülers.

Bilanzierend zu den Moderatoreffekten ergeben sich bezüglich der von Heller (1997) deklarierten veränderten Prädiktor-Kriteriums-Zusammenhänge durch den Einbezug von Interaktionseffekten in beiden Fächern leichte Veränderungen in der Stärke der kognitiven Haupteffekte. Zusätzlich dazu ist in diesen Analysen jedoch der leicht veränderte direkte Haupteffekt des nicht-kognitiven Schülermerkmals Arbeitsverhalten hervorzuheben.

Während an dieser Stelle nochmals auf die Feststellung Schröders (1990) verwiesen wird (vgl. [Kap. 4.2.2](#)), der auf mögliche Fehlschlüsse von Prämisse und Konklusion bei der Leistungsbewertung hinweist (hoher Fleiß führt nicht zwangsläufig zu guten Leistungen), diskutiert bereits Heckhausen (1974) in Anlehnung an [Kapitel 2.2](#) die Relevanz der individuellen Anstrengung für die verschiedenen Facetten von Schulleistung unter Bezug auf den Grundsatz der Gleichbehandlung. Demnach widerspreche es jenem Grundsatz, Kinder bei gleicher Schulleistung aufgrund ihrer unterschiedlichen Anstrengung unterschiedlich zu honorieren, müsse man sich doch entscheiden, ob man das objektive Leistungsergebnis oder die individuelle Anstrengung bewerte (Heckhausen, 1974). Diese Feststellung führt zwangsläufig wieder zu den in [Kapitel 2](#) aufgeworfenen Fragen, was nun schlussendlich als schulische Leistung bewertet werden soll und welche (gesellschaftlichen) Funktionen dahinterstehen; Punkte, die am Ende dieses Kapitels sowie im nachfolgenden [Kapitel 8.4](#) abschließend diskutiert werden.

### **Einfluss familiärer und individueller Hintergrundmerkmale auf die Notenzusammensetzung in Deutsch und Mathematik am Ende der Grundschulzeit**

Die ersten deskriptiven Analysen bezüglich der Zusammenhänge zwischen familiären Strukturmerkmalen, schulbezogenen Prozessmerkmalen, individuellen Schülermerkmalen und Schulnoten zeigen bereits, dass von den beiden einbezogenen Strukturmerkmalen der Bildungsstatus der Familie höher mit Schulnoten und Schülermerkmalen korreliert. Bezüglich der schulbezogenen Prozessmerkmale ergeben sich für die eingeschätzte Bildungsaspiration die stärksten korrelativen Zusammenhänge mit Noten und Schülermerkmalen. Ebenfalls auf deskriptiver Basis wurde festgestellt, dass Schüler, deren Eltern über einen hohen Bildungsabschluss verfügen sowie von der Lehrkraft in ihren Bildungsaspirationen hoch eingeschätzt werden, mit Ausnahme des Sozialverhaltens (hier waren die Einschätzungen ausgewogen) bessere Merkmaleinschätzungen aufweisen.

Bezüglich der kausalen Effekte der individuellen Schülermerkmale und familiären Struktur- und Prozessmerkmalen für die Notenzusammensetzung entsprechen die Ergebnisse für die kognitiven und nicht-kognitiven Schülermerkmale in ihren Tendenzen den Ergebnissen der hierarchischen Regressionsanalysen ausschließlich für die individuellen Schülermerkmale: Es zeigen sich für die Begabungseinschätzung in Mathematik und die Einschätzung der sprachlichen Fähigkeiten für Deutsch die stärksten Effekte auf die Notenzusammensetzung.

Geprüft wurde bezüglich der familiären Hintergrundmerkmale erstens, ob diese im Sinne der Ergebnisse von Maaz, Baeriswyl und Trautwein (2011) sowie Stubbe, Bos und Euen (2012) unter Einbezug der individuellen Schülermerkmale weiterhin einen eigenständigen Effekt auf die Notenzusammensetzung ausüben oder ob diese – wie Stahl (2009) und Helmke und Schrader (2010) diskutieren – auch in den vorliegenden Daten durch die Schülermerkmale vermittelt wirken und keinen direkten Einfluss nehmen. Zudem wurde unter Bezug auf das Makro-Modell von Helmke und Weinert (1997) und die Erkenntnissen von Stocké (2010), Sauer und Gattringer (1985) u.a. geprüft,

inwiefern sich für die Struktur- und schulbezogenen Prozessmerkmale unterschiedliche Effekte ergeben und inwiefern zweitens von den familiären Strukturmerkmalen beeinflusst werden.

Die Ergebnisse der hierarchischen Regressionsanalysen für die Notenzusammensetzung am Ende der vierten Jahrgangsstufe in Deutsch und Mathematik ergeben fachübergreifend, dass für die beiden Strukturmerkmale der Bildungsstatus der Eltern stärker Einfluss nimmt als die sozioökonomische Ausstattung der Familie. Zudem wird für beide Fächer der direkte Effekt der Strukturmerkmale unter Kontrolle der schulbezogenen Prozessmerkmale schwächer, was wiederum für vermittelnde Effekte sprechen könnte. Bezüglich der schulbezogenen Prozessmerkmale ergibt sich für die eingeschätzte elterliche Bildungsaspiration der stärkste negative Effekt auf die Notenzusammensetzung in Deutsch und Mathematik, der auch unter Kontrolle der individuellen Testleistung – auch für das eingeschätzte elterliche Unterstützungsverhalten – bestehen bleibt. Bei gleicher Testleistung erhalten Schüler, deren Eltern über hohe Bildungsaspirationen verfügen und ihrem Kind unterstützend zur Seite stehen also niedrigere (bessere) Schulnoten in Deutsch und Mathematik als ihre Mitschüler, bei denen dies nicht zutrifft. Kontrollierte man zusätzlich die kognitiv und nicht-kognitiv eingeschätzten Schülermerkmale, bleibt für beide Schulfächer ein direkter, hochsignifikant schwacher negativer Effekt der eingeschätzten Bildungsaspiration auf die Notenzusammensetzung bestehen. Bezüglich des durch die Lehrkraft eingeschätzten elterlichen Schulkontakts ergeben sich entgegen der Befunde von Stocké (2010) für beide abhängige Variablen keinerlei Effekte. Aber auch das eingeschätzte Unterstützungsverhalten spielt unter Kontrolle aller individuellen Schülermerkmale keine signifikante Rolle mehr. Die Notenzusammensetzung wurde unter Kontrolle der Struktur- und Prozessmerkmale sowie der individuellen Schülermerkmale also nicht durch die Präsenz und Unterstützung der Eltern in der Schule beeinflusst.

Die Ergebnisse bezüglich der schulbezogenen Prozessmerkmale Bildungsaspiration und Unterstützungsverhalten könnten für

sich genommen dennoch bereits für eine Bestätigung der Thesen Bourdieus (1983) bzw. Colemans (1988) im Sinne einer förderlichen Wirkung von kulturellen und sozialen Kapitalien für die Schulleistung sprechen. Auch wenn die einbezogenen Prozessmerkmale sicherlich zeitlichen Veränderungen unterliegen und lediglich eine Einschätzung zu einem einmaligen Zeitpunkt darstellen, könnten sie doch für die aus den theoretischen Grundlagen abgeleitete These Stockés (2010) sprechen. Dergestalt, dass sich das schulbezogene Sozialkapital über die Förder- und Erwartungshaltung der Eltern auf die Schulleistung des Kindes auswirkt; eine Vermittlung über die schulbezogenen Eigenschaften des Kindes wurden im Gesamtmodell ebenfalls geprüft.

### Prüfung des Gesamtmodells

Die abschließend dargestellten Strukturgleichungsmodelle ermöglichten es dann, alle Variablen auf messfehlerbereinigter Ebene in ihren gegenseitigen Abhängigkeiten zu berücksichtigen und deren gemeinsame Wirkmechanismen für die Notenzusammensetzung am Ende der Grundschulzeit zu prüfen. So erlauben die regressionsanalytischen Auswertungen zwar bereits Aussagen zum Zusammenhang zwischen familiären Hintergrundmerkmalen, (eingeschätzten) Schülermerkmalen und Schulnoten, sie lassen jedoch keine gesicherten Aussagen über die indirekten Wirkmechanismen zu sowie die Mehrebenenstruktur der Daten erst durch die Strukturgleichungsmodellierung berücksichtigt werden konnte.

Theoretische Ausgangslage der Prüfung war das in [Kapitel 4.5](#) deklarierte Gesamtmodell zu den Komponenten von Schulleistung in Deutsch und Mathematik in der vierten Jahrgangsstufe. Durch die regressionsanalytischen Auswertungen ergaben sich für die Deutsch- und Mathematiknote leicht abgewandelte, empirische Strukturgleichungsmodelle. Unter Berücksichtigung der hohen Korrelationen zwischen den einzelnen Prädiktoren wurden die kognitiven Merkmale der Schüler getrennt nach Testleistung (Testleistung,  $kft$ ) und Einschätzungen (fachspezifische Begabung, sprachliche Fähigkeiten, Vorwissen) als zwei latente Variablen in den fachspezifischen Modellen analysiert. Für das Fach Mathematik wurden als eingeschätzte kog-

nitiv Merkmale aufgrund der vorangegangenen Analysen nur noch die mathematische Begabung und das Vorwissen des Schülers als Indikatoren berücksichtigt. Als nicht kognitives Schülermerkmal wurde aufgrund der vorangegangenen Analysen lediglich noch das Arbeitsverhalten als manifeste Variable in die Modelle aufgenommen. Auf familiärer Ebene wurden die strukturellen Merkmale der Familie als latente Variable mit den Indikatoren Bildungsstatus und sozialer Status (ISEI) im Modell berücksichtigt. Auch die familiären Prozessmerkmale flossen als latentes Konstrukt mit den Indikatoren Unterstützungsverhalten, Schulkontakt und Bildungsaspiration mit in das Gesamtmodell ein.

Fachübergreifend ergeben sich für die beiden Strukturgleichungsmodelle gute Modellanpassungen.

Auf der Ebene der individuellen Schülermerkmale zeigt sich fachübergreifend bezüglich der direkten Effekte der Testleistungen, eingeschätzter kognitiver Schülermerkmale und dem Arbeitsverhalten auf die Notenzusammensetzung, dass die eingeschätzten kognitiven Schülermerkmale analog zu den Ergebnissen der Regressionsanalysen den stärksten Effekt aufweisen. Die Testleistung nimmt jedoch zusätzlich direkt Einfluss auf die Noten, und im Fach Deutsch zeigt auch das Arbeitsverhalten direkte hochsignifikante Effekte. Bezüglich vermittelnder Mechanismen zeigt sich fachübergreifend, dass die kognitiven Testleistungen in einem hohen Maße durch die kognitive Merkmaleinschätzung vermittelt werden.

Diese Befunde sprechen einmal im Sinne der messtheoretischen Gütekriterien dafür (vgl. [Kap. 3.3](#)), dass die kognitive Merkmaleinschätzung in relativ hohem Maße mit den individuellen Testleistungen übereinstimmt, verschiedene Lehrkräfte also relativ gut die Leistungen analog zu den Testleistungen beurteilten. Sie weisen allerdings im Sinne der Argumentation der theoretischen Basis auch daraufhin, dass die kognitive Merkmaleinschätzung – als stärkster Einflussprädiktor – mehr Komponenten umfasst als durch die individuellen Testleistungen (Testleistung, *kft*) erfasst werden können. Damit ist die These bestä-

tigt, dass Schulnoten immer mehr umfassen als durch eine einmalige Testleistung sichtbar werden kann.

Bezüglich der Effekte der familiären Struktur- und Prozessmerkmale auf die Notenzusammensetzung ergeben sich fachübergreifend keinerlei signifikante direkte Effekte auf die Schulnoten am Ende der vierten Klasse, was ebenfalls den vorangegangenen Ergebnissen der Regressionsanalysen entspricht.

Hochsignifikant direkt positiv nehmen die familiären Strukturmerkmale dagegen auf die schulbezogenen Prozessmerkmale und die Testleistungen der Schüler Einfluss, zudem wird für das Fach Mathematik der direkte Pfad von den Strukturmerkmalen auf das Arbeitsverhalten schwach positiv signifikant. Die Ausprägung der schulbezogenen Prozessmerkmale und der individuellen Testleistungen des Kindes variiert also in Abhängigkeit von der bildungsbezogenen- bzw. sozioökonomischen Ausstattung der Familie. Dieser Befund ist insofern neuartig, als dass die schulbezogenen Prozessmerkmale von der Lehrkraft eingeschätzt werden, diese Einschätzung also nicht unbedeutend von der strukturellen Ausstattung der Familie beeinflusst zu werden scheint. Der Zusammenhang zwischen familiären Strukturmerkmalen und Testleistungen deckt sich dagegen mit der bisherigen Forschungsbasis. Entgegen der Annahmen wurden die kognitiven Merkmaleinschätzungen der Lehrkräfte allerdings nicht direkt von den familiären Strukturmerkmalen beeinflusst. Dies widerspricht zunächst der These von Helmke und Schrader (2010) u.a., dass der familiäre Hintergrund (in diesem Fall die familiären Strukturmerkmale) über die individuellen Schülermerkmale Einfluss nimmt; zumindest in dieser Studie nicht vermittelt über die kognitive Merkmaleinschätzung, wohl aber über die individuellen Testleistungen und die prozessbezogene Ausstattung der Familie.

Dagegen zeigen sich für die familiären, schulbezogenen Prozessmerkmale fachübergreifend direkte hochsignifikante Effekte auf die individuellen Testleistungen (Testleistung, kft), die kognitive Merkmaleinschätzung und das Arbeitsverhalten des Kindes. Hoch (gut)

eingeschätzte schulbezogene Prozessmerkmale beeinflussen also nicht nur die Testleistungen des Kindes, sondern auch die Lehrkräfteeinschätzung der kognitiven und nicht-kognitiven Schülermerkmale. Obgleich anhand der Daten keinerlei Aussagen über die tatsächlich unterschiedlichen Verhaltensweisen der Schüler getätigt werden kann, ist in diesem Zusammenhang festzuhalten, dass die Merkmaleinschätzung durch die Lehrkraft nicht unabhängig von der (prozessbezogenen) familiären Ausstattung erfolgt.

Neben der offenen Frage, inwiefern diese Zusammenhänge zu Bildungsungleichheiten beitragen können, wurde zusätzlich ermittelt, inwieweit die familiären Struktur- und Prozessmerkmale indirekt, vermittelt über die individuellen Schülermerkmale auf die Notenzusammensetzung Einfluss nehmen.

Für die familiären Strukturmerkmale (Bildungsstatus, ISEI) zeichnet sich für den vermittelnden Effekt über die individuellen Testleistungen und die kognitive Merkmaleinschätzung auf die Note in beiden Fächern der stärkste indirekte Effekt ab sowie der totale indirekte Effekt der Strukturmerkmale für beide Fächer moderat ausgeprägt ist.

Bezüglich der schulbezogenen Prozessmerkmale (Bildungsaspiration, Unterstützungsverhalten, Schulkontakt) zeigen sich für die Note in Deutsch die stärksten Vermittlungseffekte über die kognitive Merkmaleinschätzung des Kindes. In Mathematik wirken die schulbezogenen Prozessmerkmale am stärksten vermittelt über die individuellen Testleistungen und die kognitiven Einschätzungen auf die Notenzusammensetzung. Fachübergreifend ist der totale indirekte Effekt der Prozessmerkmale ebenfalls moderat ausgeprägt. Demnach wirken sich im Fach Deutsch die schulbezogenen Prozessmerkmale über die kindliche Fähigkeitseinschätzung hochsignifikant auf die Schulnote aus. Für die Mathematiknote ist dieser Effekt zusätzlich noch über die individuellen Testleistungen vermittelt.

Bezüglich der zusätzlichen Effekte der leistungsbezogenen und sozialen Klassenzusammensetzung auf die Notenzusammensetzung ergaben

sich in den Mehrebenenanalysen keinerlei signifikante Effekte auf die fachspezifischen Schulnoten. Das auf Aggregatebene einbezogene durchschnittliche Leistungsniveau (Testleistung) und die durchschnittliche soziale Herkunft (Bildungsstatus, sozialer Status) der Klasse nimmt also keinerlei Beeinflussung auf die Notenzusammensetzung und widersprechen somit den Befunden in [Kapitel 4.4](#) (vgl. z.B. Maaz et al., 2008; Trautwein & Baeriswyl, 2007).

Diese Ergebnisse auf Individualebene legen den Schluss nahe, dass es für die Notenzusammensetzung im Fach Deutsch wichtig ist, wie die Lehrkraft die Bildungsaspiration, den Schulkontakt und das Unterstützungsverhalten der Eltern einschätzt, da die kognitive Fähigkeitseinschätzung davon beeinflusst wird und diese sich wieder bedeutsam in der Notenzusammensetzung widerspiegelt. Obwohl erneut darauf hingewiesen wird, dass anhand der Daten nicht beurteilt werden kann, inwiefern diese Einschätzung einem tatsächlich beobachteten Sachverhalt entspricht, muss hinterfragt werden, ob diese familiären Eigenschaften für den Erfolg im Schulsystem tatsächlich notwendig sein müssen und ob diese durch das Schulrecht legitimiert werden und so einen Zusammenhang untereinander rechtlich erlauben. Zudem ist zu diskutieren, inwieweit durch diese Einschätzungen sozialschichtspezifische Bildungserfolge vermittelt werden könnten.

Greift man die theoretischen Grundgedanken aus [Kapitel 4.3.1](#) nochmals auf, konstatieren auch einschlägige Befunde, dass Herkunftseffekte im Sinne Boudons (1974) maßgeblich über die primären Herkunftseffekte vermittelt werden (vgl. z.B. Ditton, 2010). Ausschlaggebender Punkt scheint in diesem Zusammenhang allerdings zu sein, welche schulische Leistung man in diesem Fall in den Analysen betrachtet. Während in Boudons (1974) Abhandlung die sozialstratifizierten Schulleistungen durch verbale Tests („*Achievement in a Verbal Test*“) und schulische Leistungen („*school achievement*“) erhoben wurden, wird selten thematisiert, welche Leistungsfaktoren konkret als sozialschichtspezifisch determiniert betrachtet werden; welche ganz konkreten Leistungsvariablen also durch familiäre Vermittlungsmechanismen beeinflusst und als unterschiedliche Leistung bezeichnet

werden. In den vorliegenden Analysen galten hierzu die Zeugnisnoten in Deutsch und Mathematik der vierten Jahrgangsstufe als ausschlagendes Kriterium.

Betrachtet man die gefundenen Zusammenhänge, sind es die kognitiven Merkmalseinschätzungen der Lehrkraft, die maßgeblich in die Note einfließen und damit vermutlich auch den schulrechtlichen Vorgaben entsprechen. Berücksichtigt man zusätzlich deren Relevanz für den anstehenden Übergang in die Sekundarstufe I kann anhand der Befunde festgehalten werden, dass dieser maßgeblich durch die Merkmalseinschätzung der Lehrkraft determiniert wird; diese Einschätzung allerdings nicht unabhängig von den schulbezogenen Prozessmerkmalen der Familie erfolgt. Die individuellen Merkmale selbst sind also sozialschichtspezifisch unterschiedlich ausgeprägt. In Anlehnung an die in [Kapitel 4.3.1](#) vorgestellten theoretischen Erklärungsansätze zu familiär bedingten Ungleichheiten, werden Leistungsunterschiede in Anlehnung an die bourdieusche Reproduktionstheorie laut Maaz, Baumert und Trautwein (2011) vornehmlich durch die unterschiedliche Ausstattung in ökonomischem, kulturellem und sozialem Kapital und deren Wirkmechanismen vermittelt, die u.a. auch das lehrkraftspezifische Entscheidungsverhalten beeinflussen. Diese These der indirekten bzw. direkten Schichteffekte kann anhand der gefundenen Zusammenhänge vorsichtig bestätigt werden, allerdings ist auch festzuhalten, dass sich die Einschätzungen in hohem Maße mit den kognitiven Testleistungen decken, diese jedoch ebenfalls durch die familiären Hintergrundmerkmale beeinflusst werden.

Ebenfalls unter Rückbezug auf die in [Kapitel 2.4](#) besprochenen funktionalen Anforderungen an die schulische Notengebung darf nicht vergessen werden, welche Funktionen die gefundenen Prädiktoren schulischer Leistung haben könnten bzw. ob diese mit den pädagogischen bzw. gesellschaftlichen Funktionszuschreibungen übereinstimmen. Ausgangspunkt war dabei die Frage nach dem *Warum von Leistungsbewertungen*, welche Funktion damit also in der Gesellschaft erfüllt werden soll.

Kritisch beleuchtet wird vor allem die große Spannweite der Anforderungen, welche die Art der Fragen an die Bewertungspraxis bestimmt (Breidenstein et al., 2011; Titze, 2000). Während über die Funktionen der einbezogenen Einzelnoten anhand der Daten keine Aussagen getroffen werden können, kann anhand der gefundenen Prädiktoren in den Zeugnisnoten der vierten Jahrgangsstufe darüber diskutiert werden, inwiefern diese den pädagogischen und gesellschaftlichen Funktionszuweisungen entsprechen könnten.

Im Sinne der gesellschaftlichen Funktionen (vgl. [Kap. 2.4](#)) muss man bei der Selektions- und Allokationsfunktion der Notengebung danach fragen, ob die Prädiktoren der Noten einen diagnostischen und prognostischen Wert aufweisen, wie *valide* die aufgrund dieser Bewertung getroffene Selektion also ist. Damit zusammen hängt selbstverständlich die Frage, welche Leistungskriterien diesem prognostischen Wert entsprechen, gelten diese in Anbetracht der unterschiedlichen Chancenzuweisung doch als gerechter Verteilungsmaßstab. Die Frage nach dem Inhalt des Leistungskriteriums kann in den hier erfolgten Analysen vornehmlich durch die kognitiv eingeschätzten Leistungsaspekte beantwortet werden, die allerdings wiederum nicht unabhängig von den familiären Hintergrundmerkmalen erfolgten. Die Frage nach den Leistungsinhalten führt dann wiederum zu der zweiten gesellschaftlichen Funktion, der Frage nach der Sozialisations- und Qualifikationsfunktion der bewerteten Inhalte. Einmal erfolgt aus Schülerperspektive durch die Bewertung sicherlich eine (unter)bewusste Anpassung dahingehend, welche gezeigten Verhaltensweisen als Leistung wahrgenommen und in der Gesellschaft anerkannt werden. Dabei ist die Frage, ob vornehmlich der funktionale Blick auf die Ausbildung für die Leistungsinhalte bestimmend ist oder ob man einen breiteren, umfassenderen Bildungswert betrachten sollte, also welche Inhalte schlussendlich als Leistung legitimiert werden. Betrachtet man diese Frage erneut vor dem Hintergrund der gefundenen Prädiktoren für die Zeugnisnoten am Ende der Grundschulzeit, sind neben fachspezifischen Inhalten die sprachlichen Fähigkeiten, die fachspezifische Begabung und das Vorwissen des Kindes die bedeutsamsten Leistungskomponenten, aber auch das Arbeitsverhalten des Kindes ist ein relevanter Leistungsas-

pekt, die allesamt nicht unabhängig von familiären Hintergrundmerkmalen scheinen. Während diese individuellen Schülermerkmale wohl vom deutschen Schulrecht als Leistung legitimiert werden, stehen die bewerteten Eigenschaften vermutlich auch mit den gesellschaftlichen und pädagogischen Funktionsanforderungen im Einklang. Inwiefern dadurch allerdings soziale Ungleichheiten reproduziert werden, kann an dieser Stelle durchaus hinterfragt werden.

### 8.3 Abschließende Beantwortung der Hypothesen

Nachdem nun die einzelnen Ergebnisse zusammenfassend aufgeführt und diskutiert wurden, gilt es nun die in [Kapitel 5](#) formulierten Hypothesen anhand der in dieser Arbeit vorgestellten Ergebnisse zu beantworten.

Insgesamt wurden drei Fragestellungen mit entsprechenden Hypothesen formuliert: die erste Fragestellung fokussierte die Gewichtung kognitiver und nicht-kognitiver Schülermerkmale für die Notenzusammensetzung in der vierten Klasse. Zweitens wurde der zusätzlichen Relevanz familiärer Hintergrundmerkmale nachgegangen. Die dritte Fragestellung bezog sich dann auf die Bedeutung der leistungsspezifischen und sozialen Klassenzusammensetzung für die individuelle Notenzusammensetzung.

Zur ersten Fragestellung wurden zwei Hypothesen formuliert.

**Hypothese 1 postulierte, dass die von der Lehrkraft eingeschätzten kognitiven und nicht-kognitiven Schülermerkmale einen stärkeren Einfluss auf die Viertklassnote in Deutsch und Mathematik haben als die individuelle Test- und Intelligenzleistung des Schülers.** Zusätzlich zu dieser Hypothese wurde dieser Zusammenhang dann noch in vier Unterhypothesen konkretisiert. Es wurde davon ausgegangen, dass Schüler mit gleicher Test- und Intelligenzleistung eine bessere fachspezifische Note erhalten, wenn sie in ihren kognitiven Fähigkeiten (Begabung, sprachliche Fähigkeiten, Vorwissen)

positiver eingeschätzt werden (Hypothese 1.1). Zusätzlich dazu wurde davon ausgegangen, dass die kognitiven Schülermerkmale (Testleistung, Intelligenz, Begabung, sprachliche Fähigkeiten, Vorwissen) stärkere Prädiktoren für die fachspezifische Notenzusammensetzung sind als die nicht-kognitiven Schülermerkmale (Arbeits-, Sozialverhalten) (Hypothese 1.2). Bezogen auf die nicht-kognitiven Schülermerkmale wurde zudem postuliert, dass sich das Arbeitsverhalten stärker in der fachspezifischen Notenzusammensetzung widerspiegelt als das Sozialverhalten des Schülers (Hypothese 1.3). Im Sinne einer prognostischen Validitätsprüfung wurde zudem noch gesondert gefragt, inwiefern das eingeschätzte Vorwissen aus der zweiten Klasse zusätzlich zu den Merkmaleinschätzungen aus der vierten Klasse noch einen Einfluss auf die Viertklassnote hat (Hypothese 1.4).

Zur Prüfung dieser Hypothesen wurde zunächst in deskriptiven Analysen geprüft, inwiefern die kognitiven und nicht-kognitiven Schülermerkmale untereinander und mit den Zeugnisnoten der vierten Klasse in Deutsch und Mathematik zusammenhängen. In einem zweiten und dritten Schritt wurde der Einfluss jener individuellen Schülermerkmale auf die Notenzusammensetzung mithilfe von hierarchischen Regressionsanalysen und als Komponenten in Strukturgleichungsmodellierungen analysiert. Mit Ausnahme des eingeschätzten Sozialverhaltens fanden sich bereits auf deskriptiver Basis mittlere bis starke Zusammenhänge zwischen den eingeschätzten (kognitiven und nicht-kognitiven) Schülermerkmalen und den Noten der vierten Jahrgangsstufe; die eingeschätzte fachspezifische Begabung und die sprachlichen Fähigkeiten korrelierten bereits höher mit den Noten als die individuellen Test- und Intelligenzleistungen (vgl. [Kap. 7.1.2](#)).

Die anschließend durchgeführten hierarchischen Regressionsanalysen für die Deutsch- und Mathematiknote ergaben dann ähnliche Tendenzen (vgl. [Kap. 7.1.2](#) bzw. [Kap. 7.2.2](#)). Für die Deutschnote zeigte sich unter Berücksichtigung aller kognitiver und nicht-kognitiver Schülermerkmale, dass die kognitiven Merkmaleinschätzungen zur Begabung und den sprachlichen Fähigkeiten sowie das nicht-kognitive Merkmal Arbeitsverhalten höhere Effekte auf die Notenzusammensetzung erzie-

len als die individuellen Testleistungen des Schülers. Gleiches galt für die Mathematiknote mit Ausnahme der sprachlichen Fähigkeitseinschätzung, die für die Notenzusammensetzung in Mathematik nahezu irrelevant war. Bezüglich des ebenfalls eingeschätzten Vorwissens aus der zweiten Klasse ergaben sich für beide Fachnoten der vierten Klasse ebenfalls signifikante Effekte, allerdings waren diese schwächer ausgeprägt als die Testleistungen der Schüler aus der vierten Klasse. Fachübergreifend ergaben sich für das Sozialverhalten schwache und nicht signifikante Effekte.

Nach diesem zweiten Analyseschritt konnte bereits die erste Hypothese mit ihren vier Unterhypothesen überwiegend bestätigt werden: Mit Ausnahme des eingeschätzten Vorwissens und des Sozialverhaltens nahmen die eingeschätzte Begabung, die sprachlichen Fähigkeiten und das Arbeitsverhalten stärker Einfluss auf die Notenzusammensetzung als die individuelle Test- und Intelligenzleistung des Schülers (Hypothese 1). Zudem wurde deutlich, dass diese auch unter Kontrolle der Testleistungen größeren Einfluss nahmen (Hypothese 1.1) und dass der Einfluss der kognitiven Merkmale den der nicht-kognitiven Merkmale übersteigt (Hypothese 1.2). Ebenfalls Bestätigung fand die These, dass das Arbeitsverhalten – im Einklang mit dem deutschen Schulrecht – stärker in der Notengebung berücksichtigt wird als das Sozialverhalten des Schülers (Hypothese 1.3). Auch der schwache, aber signifikante Effekt der Vorwissenseinschätzung aus der zweiten Klasse auf die Viertklassnote bestätigt in gewissem Sinne die prognostische Validitätsprüfung (Hypothese 1.4).

Unklar war jedoch weiterhin, inwiefern die kognitiven Merkmale – eingeschätzt und durch individuelle Testleistungen erhoben – getrennt voneinander Einfluss nehmen bzw. inwiefern die Testleistungen den Fähigkeitseinschätzungen entsprechen.

In einem letzten Schritt wurde geprüft, inwiefern die kognitiven Schülermerkmale und das Arbeitsverhalten, als latente und manifeste Konstrukte, Effekte auf die Deutsch- und Mathematiknotenzusammensetzung ergeben (vgl. [Kap. 7.4.1](#) bzw. [Kap. 7.4.2](#)). Um die

individuellen Testleistungen und die eingeschätzten kognitiven Schülermerkmale in ihrer unterschiedlichen Wirkweise zu analysieren, wurden zwei latente Variablen gebildet. Eine latente Variable Kognitive Merkmale (Testleistungen) mit den Indikatoren Testleistung und Intelligenz ( $k_{ft}$ ) und eine latente Variable Kognitive Merkmale (Lehrkräfteeinschätzungen) mit den Indikatoren Begabung, Vorwissen, und sprachlichen Fähigkeiten für das Fach Deutsch. Beide wurden als Komponenten in das Strukturgleichungsmodell aufgenommen. Das Arbeitsverhalten floss als manifeste Variable in die Strukturgleichungsmodellierung mit ein. Unter Berücksichtigung der Mehrebenenstruktur der Daten ergab sich für beide Modelle getrennt nach Fach, dass die von der Lehrkraft eingeschätzten kognitiven Merkmale die stärksten direkten Pfade auf die Note ergeben. Die Betrachtung der indirekten Vermittlungspfade zeigte aber auch, dass die individuellen Testleistungen im starken Maße mit den Merkmalseinschätzungen zusammenhängen (vgl. Abb. 8 bzw. Abb. 9). Bezüglich des eingeschätzten Arbeitsverhaltens ergaben sich Unterschiede zwischen den Fächern: Während dieses nicht-kognitive Merkmal im Fach Deutsch sogar stärker direkt Einfluss auf die Notenzusammensetzung nahm als die kognitiven Testleistungen, wurde der direkte Effekt im Modell zur Erklärung der Mathematiknote nicht signifikant. Abschließend kann die erste Hypothese anhand der vorliegenden Ergebnisse für die Deutschnote bestätigt werden. Für die Mathematiknote gilt dies mit Ausnahme des nicht-kognitiven Merkmals Arbeitsverhalten ebenfalls. Die Unterhypothesen konnten für beide Fächer bestätigt werden.

**Hypothese 2 der ersten Fragestellung postulierte, dass der Zusammenhang zwischen den eingeschätzten kognitiven Schülermerkmalen und Schulnoten durch nicht-kognitive Merkmale moderiert wird.**

Angenommen wurde im Sinne der theoretischen Argumentation von Heller (1995), dass sich die kognitiv eingeschätzten Merkmale in Abhängigkeit von dem kindlichen Arbeits- und Sozialverhalten in ihren Effekten auf die Notengebung verändern. Bereits in den deskriptiven Analysen zeigten sich die niedrigsten Zusammenhänge zwischen

dem Sozialverhalten mit den Schulnoten und den anderen Schülermerkmalen (vgl. Kap. 7.1.2). Auch in den Regressionsanalysen zeigten sich dann keinerlei Effekte des Sozialverhaltens auf die Notenzusammensetzung (vgl. Kap. 7.1.2 bzw. Kap. 7.2.2), weswegen die Interaktionen ausschließlich für das eingeschätzte Arbeitsverhalten mit den Einschätzungen zu Begabung, Sprache und Vorwissen zusätzlich in die Analysen aufgenommen wurden.

Unter Berücksichtigung der Interaktionen ergaben sich für beide Fächer leicht veränderten direkte Effekte der kognitiv eingeschätzten Schülermerkmale auf die Notenzusammensetzung: Für die Deutschnote nahmen die direkten Effekte der Begabungseinschätzung und der sprachlichen Fähigkeiten leicht an Stärke zu sowie das eingeschätzte Vorwissen nicht signifikant wurde. Für die Mathematiknote wurde die Begabungseinschätzung etwas schwächer sowie der direkte Effekt der sprachlichen Fähigkeiten nahezu unverändert blieb. Der ohnehin schwache Effekt des Vorwissens wurde etwas stärker. Zusätzlich dazu ist in diesen Analysen jedoch der ebenfalls leicht veränderte direkte Effekt des nicht-kognitiven Schülermerkmals Arbeitsverhalten auffällig, der für die beiden Fächer unterschiedlich verlief. Unter Berücksichtigung der Interaktionen wurde für die Deutschnote der direkte Effekt des Arbeitsverhaltens etwas stärker. Für die Mathematiknote ergab sich dagegen ein leicht abgeschwächter Effekt dieses nicht-kognitiven Merkmals.

Bezüglich der einbezogenen Interaktionsterme zeigte sich für die Deutschnote ein negativer Effekt des Interaktionsterms zwischen Arbeitsverhalten und sprachlichen Fähigkeiten. Der verstärkte direkte Effekt des Arbeitsverhaltens deutet zudem darauf hin, dass es für die Viertklassnote im Fach Deutsch essentiell wichtig war, durch welches Arbeitsverhalten sich das Kind auszeichnete, die sprachliche Fähigkeitseinschätzung war dagegen – unter Berücksichtigung der Interaktionen – weiterhin am bedeutsamsten und nahm sogar an Stärke noch etwas zu.

Zusätzlich dazu ergab sich für die Mathematik eine positive Interaktion zwischen Arbeitsverhalten und mathematischer Begabungseinschätzung. Berücksichtigte man im Modell also neben den beiden Haupteffekten (direkte Pfade von Arbeitsverhalten, mathematische Begabung) noch den zusätzlichen Effekt zwischen mathematischer Begabung und Arbeitsverhalten, konnte dieser Effekt im Sinne eines verstärkenden Effekts interpretiert werden.

Der Einbezug der Interaktionen veränderte also teilweise die direkten Effekte der eingeschätzten kognitiven Merkmale auf die Deutsch- und Mathematiknote, wenn auch nicht besonders stark. Zudem veränderte sich die Relevanz des Arbeitsverhaltens für die Notenzusammensetzung. Von daher ist Hypothese 2 als gültig zu betrachten.

Unter der zweiten Fragestellung wurden ebenfalls zwei Hypothesen formuliert.

**Hypothese 3 postulierte, dass die strukturellen Hintergrundmerkmale der Familie (Bildungs- und sozialer Status) einen signifikanten Einfluss auf die Viertklassnoten in Deutsch und Mathematik haben.** Zusätzlich zu dieser Hypothese wurde der Zusammenhang dann noch in zwei Unterhypothesen konkretisiert. Abgeleitet aus der zweiten Hypothese wurde konkret davon ausgegangen, dass Schüler mit höheren Werten in den Strukturmerkmalen eine positivere kognitive und nicht-kognitive Merkmalseinschätzung erhalten als Schüler mit niedrigeren Werten in den Strukturmerkmalen (Hypothese 3.1). Zusätzlich dazu wurde postuliert, dass der Einfluss der strukturellen Hintergrundmerkmale auf die fachspezifische Schulnote über die individuellen kognitiven und nicht-kognitiven Merkmalseinschätzungen des Kindes vermittelt wird (Hypothese 3.2).

Zur Prüfung dieser Hypothesen wurde zunächst auf deskriptiver Ebene ermittelt, inwiefern die familiären Strukturmerkmale Bildungsstatus und sozialer Status (ISEI) mit den Viertklassnoten in Deutsch und Mathematik und den kognitiven und nicht-kognitiven Schülermerkmalen zusammenhängen. In einem zweiten und dritten Schritt wurden

die Merkmale in hierarchische Regressionsanalysen und Strukturgleichungsmodelle miteinbezogen, um die Hypothesen auf ihre Gültigkeit zu prüfen.

Auf deskriptiver Ebene ergaben sich für die beiden einbezogenen Strukturmerkmale höhere Zusammenhänge zwischen dem höchsten Bildungsstatus der Familie und den Noten in Deutsch und Mathematik, gleiches galt für die individuellen Schülermerkmale (vgl. [Kap. 7.1.3](#) bzw. [Kap. 7.1.4](#)).

Die Ergebnisse der anschließend durchgeführten hierarchischen Regressionsanalysen ergaben fachübergreifend ähnliche Tendenzen (vgl. [Kap. 7.3.1](#) bzw. [Kap. 7.3.2](#)). Für die beiden Strukturmerkmale nahm der Bildungsstatus der Eltern stärker Einfluss als die sozioökonomische Ausstattung der Familie, allerdings zeigten sich für beide Merkmale unter Kontrolle der individuellen Schülermerkmale keinerlei direkte Effekte mehr.

Um den an dieser Stelle vermuteten Mediationseffekt abschließend zu prüfen, wurden die Zusammenhänge mithilfe eines Strukturgleichungsmodells auf jene indirekten Vermittlungsmechanismen geprüft (vgl. [Kap. 7.4.1](#) bzw. [Kap. 7.4.2](#)). Die beiden Indikatoren Bildungsstatus und sozialer Status (ISEI) wurden als Indikatoren für die latente Variable Strukturmerkmale der Familie mit in das Strukturgleichungsmodell aufgenommen.

Im Einklang mit den vorangegangenen Analysen ergaben sich bezüglich der direkten Effekte der familiären Strukturmerkmale auf die Notenzusammensetzung fachübergreifend keinerlei signifikante direkte Effekte auf die Schulnoten am Ende der vierten Klasse, was ebenfalls den vorangegangenen Ergebnissen der Regressionsanalysen entspricht. Ebenfalls entgegen der Annahmen wurden die kognitiven Merkmalseinschätzungen der Lehrkräfte in beiden Fächern nicht direkt von den familiären Strukturmerkmalen beeinflusst. Allerdings ergaben sich Effekte der Strukturmerkmale auf die individuellen Testleistungen des Schülers und im Fach Mathematik – auf schwachem

Niveau – auf das eingeschätzte Arbeitsverhalten. Bezüglich der indirekten Effekte ergaben sich für die familiären Strukturmerkmale über die individuellen Testleistungen und die kognitive Merkmaleinschätzung auf die Note in beiden Fächern der stärkste indirekte Effekt. Der indirekte Pfad ausschließlich über die kognitive Merkmaleinschätzung auf die Noten ergab für beide Fachnoten keine signifikanten Effekte, gleiches galt für das eingeschätzte Arbeitsverhalten.

Hypothese 3 und die beiden Unterhypothesen müssen also mit einigen Einschränkungen abgelehnt werden. Es konnte kein direkter Effekt der familiären Strukturmerkmale und kein indirekter Effekt über die kognitiven und nicht-kognitiven Merkmaleinschätzungen auf die Schulnoten nachgewiesen werden. Allerdings ergaben sich Vermittlungseffekte über die individuellen Testleistungen und die kognitiven Merkmaleinschätzungen sowie im Strukturgleichungsmodell für das Fach Mathematik der direkte Pfad zwischen den Strukturmerkmalen der Familie und dem Arbeitsverhalten schwach signifikant wurde.

Innerhalb der zweiten Fragestellung postulierte **Hypothese 4, dass die schulbezogenen Prozessmerkmale der Familie (Lehrkräfteeinschätzung Unterstützungsverhalten, Schulkontakt, Bildungsaspiration der Eltern) einen signifikanten Einfluss auf die Vierklassnoten in Deutsch und Mathematik haben.** Konkretisiert wurde diese Hypothese erneut durch zwei Unterhypothesen. Es wurde davon ausgegangen, dass die schulbezogenen Prozessmerkmale der Familie den Effekt zwischen familiären Strukturmerkmalen und Schulnoten vermitteln, dass höhere Werte in den Strukturmerkmalen einer Familie also höhere Werte in den schulbezogenen Prozessmerkmalen bedingen und diese wiederum die Schulnote beeinflussen (Hypothese 4.1). Zusätzlich dazu wurde geprüft, inwiefern die schulbezogenen Prozessmerkmale direkt und vermittelt über die Einschätzung der kognitiven und nicht-kognitiven Schülermerkmale auf die fachspezifische Viertklassnote eines Schülers Einfluss nehmen (Hypothese 4.2).

Zur Prüfung dieser Hypothesen wurde auf deskriptiver Ebene festgestellt, inwiefern die familiären Prozessmerkmale Schulkontakt,

Unterstützungsverhalten und Bildungsaspiration der Eltern mit den Viertklassnoten in Deutsch und Mathematik und den kognitiven und nicht-kognitiven Schülermerkmalen zusammenhängen. In einem zweiten und dritten Schritt wurden die Merkmale in hierarchische Regressionsanalysen und Strukturgleichungsmodelle miteinbezogen, um die Hypothesen auf ihre Gültigkeit zu prüfen.

Auf deskriptiver Ebene ergaben sich von den drei einbezogenen Prozessmerkmalen für die eingeschätzte Bildungsaspiration die stärksten korrelativen Zusammenhänge mit Noten und Schülermerkmalen (vgl. [Kap. 7.1.3](#) bzw. [Kap. 7.1.4](#)).

Die Ergebnisse der anschließend durchgeführten hierarchischen Regressionsanalysen ergaben fachübergreifend ähnliche Tendenzen (vgl. [Kap. 7.3.1](#) bzw. [Kap. 7.3.2](#)). Für die schulbezogenen Prozessmerkmale nahm die von Lehrkraft eingeschätzte Bildungsaspiration der Eltern stärker Einfluss auf die Schulnoten als das Unterstützungsverhalten und der Schulkontakt der Eltern. Fachübergreifend blieb der direkte negative Effekt der eingeschätzten Bildungsaspiration auf die Notenzusammensetzung auch unter Kontrolle der individuellen Schülermerkmale bestehen. Zudem wurde bereits in den hierarchischen Regressionsanalysen für beide Fächer der direkte Effekt der Strukturmerkmale unter Kontrolle der schulbezogenen Prozessmerkmale schwächer, was wiederum für vermittelnde Effekte sprechen könnte.

Als Komponente der Strukturgleichungsmodelle wurden dann die indirekten Vermittlungseffekte zwischen Struktur- und Prozessmerkmale und den individuellen Schülermerkmalen auf die Notenzusammensetzung geprüft (vgl. [Kap. 7.4.1](#) bzw. [Kap. 7.4.2](#)).

Bezüglich der direkten Effekte der schulbezogenen Prozessmerkmale auf die Notenzusammensetzung ergaben sich fachübergreifend keinerlei signifikante direkte Effekte auf die Schulnoten am Ende der vierten Klasse, was ebenfalls vornehmlich den vorangegangenen Ergebnissen der Regressionsanalysen entspricht.

Direkt positiv und hochsignifikant nahmen die familiären Strukturmerkmale dagegen auf die von der Lehrkraft eingeschätzten schulbezogenen Prozessmerkmale Einfluss, der indirekte Pfad von den familiären Strukturmerkmalen auf die schulbezogenen Prozessmerkmale auf die Noten wurde allerdings für beide Fächer nicht signifikant.

Dagegen ergaben sich für die familiären, schulbezogenen Prozessmerkmale fachübergreifend direkte, hochsignifikant negative Effekte auf die kognitive Merkmalseinschätzung und das Arbeitsverhalten des Kindes. Ebenfalls zeigten sich bezüglich der indirekten Vermittlungseffekte der schulbezogenen Prozessmerkmale auf die Notenzusammensetzung für beide Fächer hochsignifikant negative Vermittlungseffekte über die kognitive Merkmalseinschätzung des Kindes auf die Noten. Der indirekte Vermittlungspfad über das Arbeitsverhalten auf die Note nahm nur im Fach Deutsch schwach negativ, aber hochsignifikant Einfluss. Demnach wirkten sich im Fach Deutsch die schulbezogenen Prozessmerkmale über die kindliche Fähigkeitseinschätzung hochsignifikant auf dessen Schulnote aus.

Hypothese 4, dass die schulbezogenen Prozessmerkmale der Familie einen signifikanten Einfluss auf die Vierklassnoten in Deutsch und Mathematik haben, kann also indirekt Gültigkeit beanspruchen. Hypothese 4.1 konnte dagegen nicht bestätigt werden. Hypothese 4.2, der Frage nach dem direkten und vermittelnden Einfluss auf bzw. über die Merkmalseinschätzungen konnte dagegen teilweise bestätigt werden: es ergaben sich keine direkten Effekte der schulbezogenen Prozessmerkmale auf die Schulnoten, wohl aber auf und vermittelt über die eingeschätzten Schülermerkmale.

In der dritten und letzten Fragestellung der vorliegenden Arbeit wurden ebenfalls zwei Hypothesen formuliert.

**Hypothese 5 postulierte, dass sich die Merkmale der Schulklasse auf die Viertklassnoten in Deutsch und Mathematik auswirken.** Konkretisiert wird dieser Zusammenhang erneut in zwei Unterhypothesen, die sich auf die Art der Merkmalsbeeinflussung beziehen.

Einmal wurde davon ausgegangen, dass ein Zusammenhang zwischen dem durchschnittlichen Leistungsniveau der Klasse und der Schulnote eines Schülers besteht. Schüler also in einer leistungsstärkeren Klasse schlechtere Noten erhalten als in einer leistungsschwächeren Klasse (Hypothese 5.1). Zudem wurde geprüft, inwiefern ein Zusammenhang zwischen der durchschnittlichen sozialen Klassenzusammensetzung und der Schulnote des Schülers besteht, ob Schüler also in einer sozialschwächeren Klasse bessere Noten erhalten (Hypothese 5.2).

Zur Prüfung dieser Hypothesen wurde zusätzlich zu den fachspezifischen Strukturgleichungsmodellen, die bereits die Mehrebenenstruktur der Daten berücksichtigten, in Mehrebenenanalysen der Effekt der leistungsbezogenen und sozialen Klassenzusammensetzung auf die Noten geprüft. Nach Schätzung der Nullmodelle für die Intraklassenkorrelationen, die schwache Abweichungen zwischen den Klassen ergaben, wurden für die Strukturgleichungsmodelle auf Individualebene das durchschnittliche Leistungsniveau (Testleistung) und die durchschnittliche soziale Herkunft (Bildungsstatus und sozialer Status) der Klasse als Aggregatmerkmale in die Analyse miteinbezogen. Die Ergebnisse wiesen auf keinerlei Beeinflussung der Notenzusammensetzung durch das durchschnittliche Leistungsniveau und durch die durchschnittliche soziale Herkunft der Klasse hin. Somit kann Hypothese 5 nicht bestätigt werden.

**Hypothese 6 postulierte dann sozusagen fragestellungsübergreifend, dass sich für die Zusammensetzung der Viertklassnoten je nach Fach Unterschiede ergeben.**

Auf deskriptiver Ebene zeigten sich für die Notengebung im Fach Deutsch die stärksten positiven Korrelationen mit der Merkmalseinschätzung der Lehrkraft zu den sprachlichen Fähigkeiten. Für die Mathematiknote bestand auf deskriptiver Basis der stärkste Zusammenhang zur mathematischen Begabungseinschätzung, ansonsten ergaben Zusammenhänge zu den einzelnen Noten ähnliche Tendenzen (vgl. [Kap. 7.1.2](#)). Die anschließend durchgeführten hierarchischen Regressionsanalysen für die Deutsch- und Mathematiknote ergaben

dann fachspezifische Unterschiede in der Notenzusammensetzung, die sich vornehmlich auf die Prädiktorstärke beziehen (vgl. [Kap. 7.2.1](#) bzw. [Kap. 7.2.2](#)). Während im Fach Deutsch die sprachliche Fähigkeitseinschätzung gefolgt von der Begabungseinschätzung den stärksten Effekt auf die Notenzusammensetzung aufwies, war im Fach Mathematik die mathematische Begabungseinschätzung am wichtigsten, die sprachlichen Fähigkeiten dagegen nahezu irrelevant. Zudem war der Effekt des Arbeitsverhaltens auf die Deutschnote etwas stärker sowie insgesamt die Varianzaufklärung durch die einbezogenen Prädiktoren im Fach Deutsch etwas höher ausfiel als im Fach Mathematik. Ebenfalls fachspezifisch unterschiedliche Ergebnisse kamen bezüglich der zusätzlich einbezogenen Interaktionseffekte zwischen dem Arbeitsverhalten und den kognitiven Merkmalseinschätzungen zustande. Unter Berücksichtigung der Interaktionen wurde für das Fach Deutsch der Haupteffekt des Arbeitsverhaltens hochsignifikant bedeutsamer sowie die Interaktion zwischen sprachlichen Fähigkeiten und Arbeitsverhalten einen hochsignifikant negativen Effekt zeigte. Für das Fach Mathematik wurde dagegen unter Berücksichtigung der Interaktionen der Haupteffekt des Arbeitsverhaltens abgeschwächt, die Interaktion zwischen Begabungseinschätzung und Arbeitsverhalten ergab einen hochsignifikant positiven Koeffizienten.

Bezüglich der zusätzlich geprüften Effekte der familiären Struktur- und Prozessmerkmale ergaben sich in den hierarchischen Regressionsanalysen keine bedeutsamen Unterschiede zwischen den Fachnoten (vgl. [Kap. 7.3.1](#) bzw. [Kap. 7.3.2](#)). In den anschließend geprüften Strukturgleichungsmodellen zeigten sich dann allerdings fachspezifische Unterschiede (vgl. [Kap. 7.4.1](#) bzw. [Kap. 7.4.2](#)). Für die kognitiven und nicht-kognitiven Schülermerkmale wird der Einfluss des Arbeitsverhaltens auf die Deutschnote hochsignifikant, für das Fach Mathematik ist dies nicht der Fall. Bezüglich der direkten Effekte der familiären Struktur- und Prozessmerkmale zeigten sich nur für die Mathematik schwache signifikante Effekte der Strukturmerkmale auf das eingeschätzte Arbeitsverhalten. Bezüglich der schulbezogenen Prozessmerkmale wurden für die Note in Deutsch die stärksten Vermittlungseffekte über die kognitive Merkmalseinschätzung des Kindes ermittelt. In

Mathematik wirken die schulbezogenen Prozessmerkmale am stärksten vermittelt über die individuellen Testleistungen und die kognitiven Einschätzungen auf die Notenzusammensetzung. Ansonsten unterschieden sich die Effekte in den beiden Gesamtmodellen sowie in den zusätzlich berechneten Mehrebenenanalysen wenig in ihren Ergebnistendenzen. Hypothese 6 kann dennoch mit Einschränkungen Gültigkeit beanspruchen, insgesamt betrachtet ergaben sich durchaus fachspezifische Unterschiede.

## 8.4 Abschließende Bilanz und zentrale Implikationen

In einem letzten Punkt werden nun nochmals die zentralen Ergebnisse der vorliegenden Untersuchung zusammengefasst und anschließend kritisch bilanziert, um dann abschließend zentrale Implikationen für Wissenschaft und Praxis abzuleiten.

### 8.4.1 Bilanz zentraler Befunde

Die vorliegende Arbeit machte es sich zum Ziel, den schulischen Leistungsbegriff wie er in der Schule in Form von Noten verwendet wird anhand seiner beeinflussenden Komponenten greifbar zu machen. Dies umfasste letztlich die Frage, was als Leistung aufgefasst wird, wenn es heißt, dass Lehrkräfte in der Notengebung die Leistung der Schüler berücksichtigen sollen; welche Komponenten die schriftlichen und mündlichen Leistungen also umfassen müssen, welches Verhalten ein Schüler (und seine) Familie im Schulalltag zeigen muss, um eine gute oder schlechte Noten zu erhalten.

In Anlehnung an die Bedingungsmodelle schulischer Leistung von Helmke und Weinert (1997) und Heller (1995) sowie den schulrechtlichen Anforderungen an die Leistungsbewertung wurden zunächst einzelne individuelle Leistungskomponenten in ihrer Wirkung auf die Notenzusammensetzung geprüft. Anschließend wurde im Sinne der theoretischen Argumentation von Bourdieu bzw. Coleman der Frage nachgegangen, inwiefern familiäre Struktur- und Prozessmerkmale die Notengebung im Sinne eines direkten Schichteffekts beeinflussen,

oder inwiefern indirekte Schichteffekte über die kindliche Merkmalseinschätzung durch die Lehrkraft erfolgen.

Auf der Ebene der individuellen Schülermerkmale ergab sich fachübergreifend bezüglich der direkten Effekte der Testleistungen, eingeschätzten kognitiven Schülermerkmalen und dem Arbeitsverhalten auf die Notenzusammensetzung, dass die eingeschätzten kognitiven Schülermerkmale insgesamt die stärksten Effekte aufweisen. Das eingeschätzte Vorwissen und Sozialverhalten sowie die Intelligenz (kff) des Schülers waren dagegen weniger bedeutsam. Die individuelle Testleistung des Schülers nimmt zusätzlich direkt Einfluss auf die Noten. Bezüglich vermittelnder Mechanismen zeigt sich aber auch fachübergreifend, dass die kognitiven Testleistungen in einem hohen Maße durch die kognitive Merkmalseinschätzung vermittelt werden. Von den eingeschätzten Indikatoren der kognitiven Schülermerkmale ergaben sich für die Deutschnote die stärksten Effekte für die sprachliche Fähigkeitseinschätzung, für die Mathematiknote war diese nahezu irrelevant, hier war die mathematische Begabungseinschätzung über das Kind maßgeblich. Das nicht-kognitive Merkmal Arbeitsverhalten wurde ebenfalls für die Notenzusammensetzung relevant, die Effekte waren in Deutsch insgesamt allerdings stärker ausgeprägt. Durch zusätzlich einbezogene Interaktionseffekte zwischen dem Arbeitsverhalten und den kognitiven Merkmalseinschätzungen wurde für das Fach Deutsch unter Berücksichtigung der Interaktionen der direkte Effekt der sprachlichen Fähigkeiten, der Begabungseinschätzung und des Arbeitsverhaltens etwas stärker. Die Interaktion zwischen sprachlichen Fähigkeiten und Arbeitsverhalten zeigte einen hochsignifikant negativen, kompensatorischen Effekt. Für das Fach Mathematik wurde dagegen unter Berücksichtigung der Interaktionen der direkte Effekt der Begabung und des Arbeitsverhaltens etwas abgeschwächt, die Interaktion zwischen Begabungseinschätzung und Arbeitsverhalten ergab einen hochsignifikant positiven Effekt.

Während sich keinerlei direkte Effekte der familiären Struktur- und Prozessmerkmale auf die Schulnoten ergaben, zeigten sich für die familiären Strukturmerkmale die stärksten indirekten Effekte über

die individuelle Testleistung und die kognitive Merkmalseinschätzung des Kindes auf dessen Note. Bezüglich der von der Lehrkraft eingeschätzten schulbezogenen Prozessmerkmale ergaben sich für die Note in Deutsch die stärksten Vermittlungseffekte über die kognitive Merkmalseinschätzung des Kindes. In Mathematik wirkten die schulbezogenen Prozessmerkmale am stärksten vermittelt über die individuellen Testleistungen und die kognitiven Einschätzungen auf die Notenzusammensetzung. Bezüglich der Relevanz der leistungsbezogenen und sozialen Klassenkomposition ergaben sich in den vorliegenden Analysen keinerlei Effekte auf die Notenzusammensetzung.

Welche Bilanz ist nun aus den dargestellten Ergebnissen zu ziehen?

Bezüglich der schulrechtlichen und funktionalen Anforderungen an die Notengebung ist festzuhalten, dass die gefundenen Zusammenhänge zwischen individuellen Schülermerkmalen und Schulnoten wohl allesamt von diesen Vorgaben legitimiert werden. Im Sinne des pädagogischen Bewertungsspielraums ist es der Lehrkraft gestattet, neben fachspezifischen Regelungen, die Auslegung bzw. Zuordnung zu den jeweiligen Notenstufen durch einen verantwortungsvollen – individuellen – pädagogischen Bewertungsvorgang vorzunehmen. Vergleicht man das Ergebnis dieses Bewertungsvorgangs mit unabhängigen Leistungstests und zieht daraus Schlüsse über die Qualität bzw. Validität der Notengebung, wird vermutlich niemals eine vollständige Übereinstimmung zwischen diesen gefunden werden, da Schulnoten nachgewiesenermaßen mehr umfassen als die individuellen Testleistungen des Schülers. Unbedingt sollte darüber diskutiert werden, in welchem Zusammenhang diese geprüften, individuellen (eingeschätzten) Leistungsaspekte mit familiären Hintergrundmerkmalen stehen bzw. stehen sollten. Demnach sind zumindest jene für die Notengebung relevanten, eingeschätzten kognitiven Leistungsaspekte zu einem gewissen Maß sozial stratifiziert. Folglich wurden diese durch die von der Lehrkraft wahrgenommenen schulbezogenen Prozessmerkmale beeinflusst, und auch die indirekten Pfade zwischen Struktur-, Prozess- und eingeschätzten kognitiven Schülermerkmalen zeigten hochsignifikante, schwache Effekte. An dieser Stelle ist zu fragen, inwiefern

diese schulbezogenen Prozessmerkmale für die kindliche Merkmalsseinschätzung eine Rolle spielen sollten sowie hinterfragt werden muss, ob dieser Zusammenhang im Sinne der bourdieuschen Reproduktionstheorie als Vermittlungsmechanismus sozial ungleicher Schulleistung gelten könnte. Wenn man diese Zusammenhänge weiterdenkt und daraus schließt, dass für eine *gute* Leistungseinschätzung diese elterlichen Komponenten vorhanden sein müssen, ist durchaus zu fragen, inwiefern dieser Zusammenhang als leistungsrelevantes Kriterium gelten sollte bzw. ob hier nicht eine sozialschichtspezifische Ungleichheit vorherrscht, die vom Kind selbst nicht beeinflusst werden kann.

#### 8.4.2 Kritische Bilanz und zentrale Implikationen

Neben der bilanzierenden Diskussion zentraler Befunde sollen an dieser Stelle auch kritische Aspekte besprochen werden, die sich in der vorliegenden Arbeit zeigten.

Bezüglich der ausgewählten Komponenten schulischer Leistung muss festgehalten werden, dass diese Auswahl lediglich einen kleinen Teil möglicher Einflusskomponenten darstellt und damit sicherlich noch unzählige (individuelle und kontextuelle) Prädiktoren einen zusätzlichen Erklärungswert für die Notenzusammensetzung aufweisen. Ebenso kann durch die vorliegenden Daten keinerlei Aussage darüber getroffen werden, aus welchen (mündlichen und schriftlichen) Einzelnoten sich die Zeugnisnoten in der vierten Jahrgangsstufe zusammensetzten. Zudem wurden der Bewertungsmaßstab der Lehrkraft und Faktoren der Lehrkraftperson selbst nicht zusätzlich kontrolliert. An die Grenzen der Interpretierbarkeit stoßen die Ergebnisse auch durch die Tatsache, dass es sich vornehmlich um eine Querschnittsansicht handelt. Dass also nur Aussagen zur Zusammensetzung der Leistung in der vierten Grundschulklasse getroffen werden können sowie die Einschätzungen auch zeitlichen Veränderungen unterliegen und nicht uneingeschränkt als konstante Merkmale gelten können. Würde man die Lehrkräfteeinschätzung aus der dritten Klasse für die Aufklärung der Note in der vierten Klasse heranziehen, hätte man einen größeren Zeitraum betrachtet und eher Fragen zur prognostischen Validität in der Notengebung fokussiert. Allerdings war für die

vorliegende Fragestellung relevant, welche Komponenten von Noten sich in der vierten Klasse zeigen, weswegen für die vorliegende Fragestellung die Lehrkräfteeinschätzung aus der vierten Klasse maßgeblich war. Bezüglich der ausgewählten Skalierung der eingeschätzten kognitiven Schülermerkmale und der Schulnoten muss zudem festgehalten werden, dass diese in der vorliegenden Untersuchung so wie in der Forschungsgemeinschaft überwiegend als metrische Variablen behandelt wurden, für eine genauere Passung allerdings ein ordinales Skalenniveau gewählt werden sollte. Bezüglich der gewählten individuellen Schülermerkmale gilt es außerdem festzuhalten, dass diese Einschätzungen sicherlich nicht vollständig trennscharf voneinander sind und so zur verminderten Präzision der Schätzwerte führen können (Backhaus, Erichson & Weiber, 2011). Bezüglich des Umgangs mit Multikollinearität zwischen einbezogenen Prädiktoren halten Weiber und Mühlhaus (2014) jedoch fest, dass dieser im Rahmen von Strukturgleichungsmodellierungen mit reflexiven Messmodellen entgegenwirken kann.

Welche Implikationen lassen sich nun für mögliche Anschlussuntersuchungen sowie für die schulische Praxis ableiten?

Bezüglich der zeitlichen Stabilität der Merkmalseinschätzung wäre eine Folgeuntersuchung denkbar, die sich mit der Veränderung der Zusammenhänge zwischen Einschätzung und Notenzusammensetzung über die Zeit beschäftigt. Allerdings bleibt im Sinne der theoretischen Argumentation auch zu überdenken, wie wünschenswert eine hohe prognostische Validität im Schulsystem ist, würde das doch gleichzeitig bedeuten, dass das Kind über die Zeit keine unvorhersehbaren Leistungssprünge mache sowie im Sinne Fingerhut und Langfeldts (1974) eine hohe prognostische Validität nicht zwangsläufig zu einem stabilen Leistungskonstrukt führe.

Bezüglich des Erhebungszeitpunktes wäre zusätzlich zu überlegen, bei der Analyse von Bewertungsvorgängen zwischen dem Prozess zum Handlungsergebnis und der Bewertung als Endresultat zu unterscheiden. Ebenso wäre von Interesse, welche ganz konkreten

expliziten wie impliziten Informationen (schriftlich, mündlich und deren Beweggründe) die bewertende Lehrkraft für ihre Urteilsbildung heranzieht und zu welchem Zeitpunkt sie dies tut; so beeinflussen diese Informationen doch vermutlich auch wieder das analysierte Schulleistungskonstrukt.

Diese Anregungen führen aber letztlich zu der methodischen Frage, durch welche Erhebungsinstrumente man relevante Leistungsbereiche sinnvoll erfassen kann. Beim Einsatz von Fallvignetten zu Beurteilungsvorgängen ist beispielsweise zu fragen, inwiefern diese Fallbeispiele tatsächlich einer authentischen Bewertungssituation entsprechen. Dagegen würde eine Unterrichtsbeobachtung zumindest das Verhalten des Schülers und die anschließend getroffene Bewertung erfassen, wobei auch hier wieder zu fragen ist, inwiefern dies einer alltagsgetreuen Bewertungssituation entspricht. Durch qualitative Interviews wäre es dagegen möglich, die Urteilsbildung der Lehrkraft konkreter nachzuvollziehen, wobei dieses Erhebungsverfahren lediglich Einzelfälle fokussiert und daraus wiederum keine allgemeingültigen Aussagen getroffen werden können. Dennoch könnte dieses spezifische Erhebungsverfahren beispielsweise für eine Anschlussfrage der Untersuchung genützt werden; der Frage nach dem Inhalt der – für diese Untersuchung bedeutsamen – fachspezifische Begabungseinschätzung, welche unbedingt noch weiterer Analysen bedarf.

Ebenso wäre eine Anschlussuntersuchung denkbar, welche die Lehrkraft selbst mit ihren Hintergrundmerkmalen in den Fokus nimmt und analysiert, inwiefern sich dadurch Unterschiede in den individuellen Bewertungsvorgängen ergeben. Zusätzlich dazu wäre es eine interessante Anschlussfragestellung, Leistungsbereiche herauszukristallisieren, in denen die Lehrkraft eventuell stärker Einfluss nimmt. Beispielsweise wäre denkbar, dass der „normale“ Bewertungsvorgang bei Grenzfällen andere Urteilsgrundlagen heranzieht als bei eindeutigen Leistungssituationen.

Bezüglich der Implikationen für die schulische Praxis wird an dieser Stelle auf die Anmerkung Terharts (2002) verwiesen, der im Zuge der

PISA-Erhebungen darauf hinweist, dass bei Rückmeldungen über Leistungsvergleiche vornehmlich „der Tag danach“ (Terhart, 2002, S.70) entscheidend sei. Entscheidend scheint aber auch in Anbetracht der hier gefundenen Ergebnisse zu sein, welche konstruktiven Schlussfolgerungen man aus den gefundenen Informationen für die (schulische) Bewertungspraxis ableitet.

Mit dem Bewusstsein, dass die Subjektivität in der Notengebung wohl niemals unterbunden werden kann und vermutlich auch nicht sollte, stellt sich eher die Frage, an welcher Stelle man ansetzt und wie man die „Reflexivität“ der bewertenden Personen unterstützen kann, um ungleichheitsfördernde Urteilstendenzen zu unterbinden. Sollte man dabei feststellen, dass kindliche bzw. familiäre Merkmale schul- und grundrechtswidersprechend in die Notengebung einfließen oder je nach familiärem Hintergrund unterschiedlich ausgeprägt sind, ergeben sich für die Pädagogik vermutlich unterschiedliche Handlungsfelder: Während der erste Sachverhalt eher Gegenstand der Lehrerbildung ist, sollte der zweite Punkt zu einer verstärkten Förderung der betroffenen Kinder – in und außerhalb des Schulkontextes – führen.

Erst vor dem Bewusstsein, welche Eigenschaften es überhaupt sind, die schulrechtskonform oder widersetzend in eine Note einfließen, erscheint es sinnvoll, Aussagen über die Qualität von Schulnoten anhand psychometrischer Gütekriterien zu treffen oder gar Bewertungsalternativen zur Notengebung zu entwickeln.

Ogleich die vorliegende Abhandlung lediglich aufzeigen wollte, welche Komponenten schulische Leistungen in der Grundschule umfassen, können die vorgebrachten Argumente durchaus zum Nachdenken anregen. Nicht zuletzt durch die dargestellten Ergebnisse sollte ein Bewusstsein dafür geschaffen werden, welche Leistungskomponenten nun eigentlich für den Erfolg im Schulsystem gewollt und damit legitimiert werden – oder umgekehrt – was schulische Leistung keinesfalls umfassen sollte und wer dies festlegt.



## 9 Literaturverzeichnis

- Anders, Y., McElvany, N. & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule. Wie differenziert urteilen Lehrkräfte? In Bundesministerium für Bildung und Forschung (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (Bildungsforschung, Bd. 34, S. 313–330). Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). Münster: Waxmann.
- Arnold, K.-H. & Jürgens, E. (2001). *Schülerbeurteilung ohne Zensuren*. Neuwied: Luchterhand.
- Arnold, R., Kleß, E. & Prescher, T. (2014). Systematik des Talents und der Begabung aus Sicht der Pädagogischen Psychologie sowie Personal- und Sozialpsychologie. In M. Stamm (Hrsg.), *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik* (S. 45–61). Bern: Huber.
- Aschersleben, K. (1971). Untersuchungen zur Reliabilität von Schulnoten. *Schule und Psychologie*, 18 (5), 147–154.
- Aulinger, J. (2009). *Schulische Leistungen von Kindern mit Migrationshintergrund. Die Rolle des Besuchs vorschulischer Einrichtungen und der Familiensprache*. Berlin: Mensch und Buch Verlag.
- Avenarius, H. & Füssel, H.-P. (Hrsg.). (2010). *Schulrecht. Ein Handbuch für Praxis, Rechtsprechung und Wissenschaft*. Kronach: Link.
- Backhaus, K., Erichson, B. & Weiber, R. (2011). *Fortgeschrittene multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer.

- Baeriswyl, F., Wandeler, C. & Trautwein, U. (2011). Auf einer anderen Schule oder bei einer anderen Lehrkraft hätte es für's Gymnasium gereicht. *Zeitschrift für Pädagogische Psychologie*, 25 (1), 39–47.
- Baeriswyl, F., Trautwein, U., Wandeler, C. & Lüdtke, O. (2010). Wie gut prognostizieren subjektive Lehrerempfehlungen und schulische Testleistungen beim Übertritt die Mathematik- und Deutschleistung in der Sekundarstufe I? In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen. Zeitschrift für Erziehungswissenschaft*, 12, 352–372 [Themenheft]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baerwald, R. (1896). *Theorie der Begabung. Psychologisch-pädagogische Untersuchung über Existenz, Klassifikation, Ursachen, Bildungsamkeit, Wert und Erziehung*. Leipzig: O.R. Reisland.
- Ballauff, T. & Hettwer, H. (Hrsg.). (1967). *Begabungsförderung und Schule*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U. et al. (Hrsg.). (2002). *Pisa 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U. et al. (Hrsg.). (2003). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Baumert, J., Trautwein, U. & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Hrsg.), *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261–331). Opladen: Leske + Budrich.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6 (1), 46–72.
- Baurmann, J. (1976). Der Einfluß von Auswertungsbedingungen, Vorinformationen und Persönlichkeitsmerkmalen auf die Benotung von Deutschaufsätzen. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 117–130). Weinheim: Beltz.

- Bayerisches Staatsministerium für Unterricht und Kultus. (2008). *Schulordnung für die Grundschulen in Bayern (Grundschulordnung-GrSO)*, Bayerische Staatskanzlei. Zugriff am 18.04.2017. Verfügbar unter <http://www.gesetze-bayern.de/Content/Document/BayVSO/true>.
- Becker, D. & Birkelbach, K. W. (2013). Lehrer als Gatekeeper? Eine theoriegeleitete Annäherung an Determinanten und Folgen prognostischer Lehrerurteile. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte* (S. 207–237). Wiesbaden: Springer.
- Becker, R. & Lauterbach, W. (2016). Bildung als Privileg – Ursachen-, Mechanismen, Prozesse und Wirkungen. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (S. 3–53). Wiesbaden: Springer VS.
- Behrens, U. (2000). Lernen statt Begabung: Vorschläge zu einer neuen Herangehensweise an das Problem individuell unterschiedlicher Leistungen. *Forum Psychologie* (42), 90–107.
- Behrens, B. & Solzbacher, C. (2016). *Grundwissen Hochbegabung in der Schule. Theorie und Praxis* (Reihe Hochbegabung und pädagogische Praxis). Weinheim: Beltz.
- Bennett, R. E., Gottesman, R. L., Rock, D. A. & Cerullo, F. (1993). Influence of Behavior Perceptions and Gender on Teachers' Judgments of Students' Academic Skill. *Journal of Educational Psychology*, 85 (2), 247–356.
- Bernstein, B. B. (1972). *Studien zur sprachlichen Sozialisation*. Düsseldorf: Pädagogischer Verlag Schwann.
- Birkel, P. (1978). *Mündliche Prüfungen*. Bochum: Kamp.
- Birkel, P. (1984a). Beurteilung mündlicher Prüfungsleistungen. In K. A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 229–236). Bern: Huber.
- Birkel, P. (1984b). Schriftliche Prüfungen durch Klassenarbeiten. In K. A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 237–244). Bern: Huber.
- Birkel, P. (2005). Beurteilungsübereinstimmungen bei Mathematikarbeiten? *Journal für Mathematik-Didaktik*, 26 (1), 28–51.
- Birkel, P. (2009). Rechtschreibeleistung im Diktat - eine objektive Leistung? *Didaktik Deutsch*, 15 (27), 5–32.

- Birkel, P. & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49 (3), 219–224.
- Birkelbach, K. (2011). Lehrerurteile im Lebensverlauf: Valide Urteile oder Self-Fulfilling Prophecy. *Soziale Welt*, 62 (3), 227–249.
- Blossfeld, H.-P., Bos, W., Lenzen, D., Hannover, B., Müller-Böling, D., Prenzel, M. et al. (2009). *Geschlechterdifferenzen im Bildungssystem – die Bundesländer im Vergleich. Fakten und Daten zum Jahrgutachten 2009*. München: Druck & Medien Schreiber.
- Böhnel, E. (1996). Die Frage der Prognostizierbarkeit von Schulerfolg in der Sekundarstufe I aufgrund der Benotung in der Primarstufe. *Unterrichtswissenschaft*, 24 (4), 343–360.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (Hrsg.). (2004). *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich. Zusammenfassung ausgewählter Ergebnisse*. Münster: Waxmann.
- Bos, W., Voss, A., Lankes, E.-M., Schwippert, K., Thiel, O. & Valtin, R. (2004). Schullaufbahneempfehlung von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 191–228). Münster: Waxmann.
- Boudon, R. (1974). *Education, opportunity, and social inequality: changing prospects in Western society*. New York: Wiley-Interscience.
- Bourdieu, P. (1976). Kulturelle Reproduktion und soziale Reproduktion. In K. H. Hörning (Hrsg.), *Soziale Ungleichheit. Strukturen und Prozesse sozialer Schichtung* (S. 223–231). Darmstadt: H. Luchterhand.
- Bourdieu, P. (1982). *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt am Main: Suhrkamp.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183–198). Göttingen: Schwartz.

- Bourdieu, P. & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit. Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs*. Stuttgart: Klett.
- Breen, R. & Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society*, 9 (3), 275–305.
- Breidenstein, G., Meier, M. & Zaborowski, K. U. (2011). Das Projekt Leistungsbewertung in der Schulklasse. In K. U. Zaborowski, M. Meier & G. Breidenstein (Hrsg.), *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule* (S. 15–38). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Breidenstein, G. (2014). Die Verschiebung von Sinnproblemen des Unterrichts auf die Noten. Studien zur alltäglichen Praxis der Zensurengebung. *Schulverwaltung, Nordrhein-Westfalen*, 25 (4), 104–106.
- Brosius, F. (2013). *SPSS 21. Fundierte Einführung in SPSS und in die Statistik ; alle statistischen Verfahren mit praxisnahen Beispielen ; inklusive CD-ROM*. Heidelberg: Mitp Verl.-Gruppe Hüthig Jehle Rehm.
- Brügelmann, H. (2001). Kontroversen um die Schulleistungsmessung in Deutschland. Eine fiktive Diskussion über Positionen und Perspektiven in verteilten Rollen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 33–44). Weinheim: Beltz.
- Brügelmann, H. (2006). *Sind Noten nützlich und nötig? Ziffernzensuren und ihre Alternativen im empirischen Vergleich*. Frankfurt am Main: Grundschulverband.
- Brügelmann, H. (2015). *Vermessene Schulen - standardisierte Schüler. Zu Risiken und Nebenwirkungen von PISA, Hattie, VerA & Co*. Weinheim: Beltz.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bundesministerium der Justiz. (1949). *Grundgesetz für die Bundesrepublik Deutschland*. Zugriff am 18.06.2013. Verfügbar unter <http://www.gesetze-im-internet.de/gg/BJNR000010949.html#BJNR00010949BJNG000100314>.

- Bundesministerium für Bildung und Forschung. (2015). *Begabte Kinder finden und fördern. Ein Wegweiser für Eltern, Erzieherinnen und Erzieher, Lehrerinnen und Lehrer*. Berlin: Bundesministerium für Bildung und Forschung.
- Bürgermeister, A. (2014). *Leistungsbeurteilung im Mathematikunterricht. Bedingungen und Effekte von Beurteilungspraxis und Beurteilungsgenauigkeit*. Münster: Waxmann.
- Christ, O. & Schlüter, E. (2010). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung*. München: Oldenbourg.
- Chudaske, J. (2012). *Sprache, Migration und schulfachliche Leistung. Einfluss sprachlicher Kompetenz auf Lese-, Rechtschreib- und Mathematikleistungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94, 95–120.
- Coleman, J. S. (1996). Der Verlust sozialen Kapitals und seine Auswirkungen auf die Schule. *Zeitschrift für Pädagogik, Beiheft*, 34, 99–105.
- Dahrendorf, R. (1965). *Bildung ist Bürgerrecht. Plädoyer für eine aktive Bildungspolitik*. Hamburg: Nannen-Verlag GmbH.
- Dann, H.-D. (2008). Lehrerkognitionen und Handlungsentscheidungen. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 177–208). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Datcher-Loury, L. (1989). Family Background and School Achievement among Low Income Blacks. *The Journal of Human Resources*, 24 (3), 528–544.
- Dicker, H. (1976). Die Reliabilität der Beurteilung von Mathematikarbeiten. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 173–176). Weinheim: Beltz.
- Diefenbach, H. (2007). *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dietrich, F. & Fricke, U. (2013). Reproduktion sozialer Ungleichheit im Prozess schulischer Leistungsbewertung. In F. Dietrich, M. Heinrich & N. Thieme (Hrsg.), *Bildungsgerechtigkeit jenseits von Chancen-*

- gleichheit. Theoretische und empirische Ergänzungen und Alternativen zu ‚PISA‘* (S. 259–292). Wiesbaden: Springer.
- Ditton, H. (1992). *Ungleichheit und Mobilität durch Bildung. Theorie und empirische Untersuchung über sozialräumliche Aspekte von Bildungsentscheidungen*. Weinheim: Juventa-Verlag.
- Ditton, H. (1998). *Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells*. Weinheim: Juventa.
- Ditton, H. (2007a). Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (S. 243–271). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, H. (Hrsg.). (2007b). *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen*. Münster: Waxmann.
- Ditton, H. (2010). Schullaufbahnen und soziale Herkunft - eine Frage von Leistung oder Diskriminierung? In S. Aufenanger, F. Hamburger, L. Ludwig & R. Tippelt (Hrsg.), *Bildung in der Demokratie* (S. 79–99). Opladen: Budrich.
- Ditton, H. (2011). Familie und Schule – eine Bestandsaufnahme der bildungssoziologischen Schuleffektforschung von James S. Coleman bis heute. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 245–264). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, H. (2013a). Bildungsgerechtigkeit. Analysen zu einer vergessenen Voraussetzung. In E. Zwick & N. J. Hofmann (Hrsg.), *Dialog der Religionen. Eine interdisziplinäre Annäherung* (S. 155–185). Berlin: Lit.
- Ditton, H. (2013b). Kontexteffekte und Bildungsungleichheit: Mechanismen und Erklärungsmuster. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte* (S. 173–206). Wiesbaden: Springer.
- Ditton, H. (2013c). Wer geht auf die Hauptschule? Primäre und sekundäre Effekte der sozialen Herkunft beim Übergang nach der Grundschule. *Zeitschrift für Erziehungswissenschaft*, 16 (4), 731–749.
- Ditton, H. (2016). Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In R. Becker & W. Lauterbach (Hrsg.),

- Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (S. 281–312). Wiesbaden: Springer VS.
- Ditton, H. & Auling, J. (2011). Schuleffekte und institutionelle Diskriminierung – eine kritische Auseinandersetzung mit Mythen und Legenden in der Schulforschung. In R. Becker (Hrsg.), *Integration durch Bildung* (S. 95–119). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, H. & Krüsken, J. (2006a). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 9 (3), 348–372.
- Ditton, H. & Krüsken, J. (2006b). Sozialer Kontext und schulische Leistungen – zur Bildungsrelevanz segregierter Armut. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 26 (2), 135–157.
- Ditton, H. & Krüsken, J. (2009). Denn wer hat, dem wird gegeben werden? Eine Längsschnittstudie zur Entwicklung schulischer Leistungen und den Effekten der sozialen Herkunft in der Grundschulzeit. *Journal für Bildungsforschung Online*, 1 (1), 33–61.
- Ditton, H., Krüsken, J. & Schauenberg, M. (2005). Bildungsungleichheit - der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8 (2), 285–304.
- Ditton, H. & Maaz, K. (2011). Sozioökonomischer Status und soziale Ungleichheit. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (S. 193–208). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, H. & Maaz, K. (2015). Sozioökonomischer Status und soziale Ungleichheit. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (S. 229–244). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dohse, W. (1967). *Das Schulzeugnis. Sein Wesen und seine Problematik*. Weinheim: Beltz.
- Dohse, W. (1976). Die Funktionen der Zensur. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte* (S. 56–61). Weinheim: Beltz.
- Dummert, F., Endlich, D., Schneider, W. & Schwenck, C. (2014). Entwicklung schriftsprachlicher und mathematischer Leistungen bei Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46 (3), 115–132.

- Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. *Psychologie in Erziehung und Unterricht*, 60 (3), 163–183.
- Eells, W. C. (1976). Die Zuverlässigkeit wiederholter Benotung von aufsatzähnlichen Prüfungsarbeiten. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 167–172). Weinheim: Beltz.
- Engelmayer, O. (1967). Begabung. Zum Begriff der Begabung. In T. Ballauff & H. Hettwer (Hrsg.), *Begabungsförderung und Schule* (S. 37–46). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Erikson, R., Goldthorpe, J. H. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30, 415–441.
- Erikson, R. & Jonsson, J. O. (1996). Explaining Class Inequality in Education: The Swedish Test Case. In R. Erikson (Ed.), *Can education be equalized? The Swedish case in comparative perspective* (pp. 1–63). Boulder: Westview Press.
- Esser, H. (1990). „Habits“, „Frames“ und „Rational Choice“. Die Reichweite von Theorien der rationalen Wahl. *Zeitschrift für Soziologie*, 19 (4), 231–247.
- Esser, H. (1999). *Soziologie. Spezielle Grundlagen*. Frankfurt am Main: Campus-Verlag.
- Esser, H. (2006). *Migration, Sprache und Integration. AKI-Forschungsbilanz 4*. Berlin: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Intergration (AKI).
- Faber, G. & Billmann-Mahecha, E. (2010). *Praxis der Notengebung – Probleme, Erfordernisse und Möglichkeiten aus pädagogisch-psychologischer Sicht*. Zugriff am 18.04.2017. Verfügbar unter <https://www.psychologie.uni-hannover.de/fileadmin/psychologie/Dateien/Institut-Downloads/Billmann/notgeborg.pdf>.
- Fend, H. (1977). *Gesellschaftliche Bedingungen schulischer Sozialisation. Soziologie der Schule I*. Weinheim: Beltz.
- Fend, H. (1981). *Theorie der Schule*. München: Urban and Schwarzenberg.
- Fend, H. (2008). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Fingerhut, W. & Langfeldt, H.-P. (1974). Leistungsbeurteilung durch Notengebung. In K. A. Heller (Hrsg.), *Leistungsbeurteilung in der Schule* (S. 253–270). Heidelberg: Quelle & Meyer.
- Frey, H. (1973). *Intelligenz und mathematische Leistung. Eine empirische Untersuchung über die Zusammenhänge zwischen Intelligenz und Leistung im mathematischen Bereich bei Schülern der Haupt-, Grund- und Vorschule*. Freiburg: Herder.
- Furck, C.-L. (1961). *Das pädagogische Problem der Leistung in der Schule*. Weinheim: Beltz.
- Fürstenberg, F. (1995). *Soziale Handlungsfelder. Strukturen und Orientierungen*. Wiesbaden: Springer.
- Füssel, H.-P. (2010). Leistungsbewertungen. In H. Avenarius & H.-P. Füssel (Hrsg.), *Schulrecht. Ein Handbuch für Praxis, Rechtsprechung und Wissenschaft* (S. 427–463). Kronach: Link.
- Gaedike, A.-K. (1974). Determinanten der Schulleistung. In K. A. Heller (Hrsg.), *Leistungsbeurteilung in der Schule* (S. 46–93). Heidelberg: Quelle & Meyer.
- Ganzeboom, H. B., de Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21 (1), 1–56.
- Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Geißler, R. (2006). Bildungschancen und soziale Herkunft. *Archiv für Wissenschaft und Praxis der sozialen Arbeit*, 4, 34–49.
- Gerleigner, S. (2013). *Familiale Ressourcen als entscheidende Faktoren für Bildungserfolg? Über die Zusammenhänge von sozialer Herkunft und Erfolg in der Grundschule*. München: Herbert Utz Verlag.
- Gerleigner, S. & Aulinger, J. (2017). „Gymnasium? Das ist nichts für uns.“ Zum Einfluss von Rational Choice und Kapitalienausstattung auf die Entscheidung für oder gegen das Gymnasium. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 29–48). Wiesbaden: Springer VS.
- Giesinger, J. (2014). Talent und Begabung in der Philosophie. In M. Stamm (Hrsg.), *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik* (S. 75–82). Bern: Huber.

- Gniewosz, B. (2015). Testverfahren. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (S. 71–81). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gomolla, M. (2012). Leistungsbeurteilung in der Schule: Zwischen Selektion und Förderung, Gerechtigkeitsanspruch und Diskriminierung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Leistungsbeurteilung* (S. 25–50). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gomolla, M. & Radtke, F.-O. (2009). *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gräsel, C., Krolak-Schwerdt, S., Nölle, I. & Hörtermann, T. (2010). Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung. Eine Analyse aus der Perspektive der sozialen Urteilsbildung. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Zeitschrift für Pädagogik*, 56, 286–295 [Themenheft]. Weinheim: Beltz.
- Gresch, C., Baumert, J. & Maaz, K. (2010). Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I. Bildungsentscheidungen und soziale Ungleichheit. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen. Zeitschrift für Erziehungswissenschaft*, 12, 230–256 [Themenheft]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Groffmann, K. J. (1964). Entwicklung der Intelligenzmessung. In K. Gottschaldt (Hrsg.), *Handbuch der Psychologie. Psychologische Diagnostik* (S. 148–199). Göttingen: Hogrefe.
- Grundmann, M. (2011). Sozialisation – Erziehung – Bildung: Eine kritische Begriffsbestimmung. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 63–85). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gut, J., Reimann, G. & Grob, A. (2012). Kognitive, sprachliche, mathematische und sozial-emotionale Kompetenzen als Prädiktoren späterer schulischer Leistungen. *Zeitschrift für Pädagogische Psychologie*, 26 (3), 213–220.

- Hany, E. A. (2012). Zum Verhältnis von Begabung und Leistung. In A. Hackl, C. Pauly, O. Steenbuck & G. Weigand (Hrsg.), *Werte schulischer Begabtenförderung. Begabung und Leistung* (S. 35–40). Frankfurt am Main: Karg-Stiftung.
- Hartfiel, G. (Hrsg.). (1977). *Das Leistungsprinzip. Merkmale, Bedingungen, Probleme*. Opladen: Leske und Budrich.
- Hartog, P. & Rhodes, E. C. (1976). Die Beurteilung mündlicher Prüfungen. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 177–183). Weinheim: Beltz.
- Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.
- Heid, H. (1992). Was „leistet“ das Leistungsprinzip? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 88 (2), 91–108.
- Heller, K. A. (1973). *Intelligenzmessung. Zur Theorie und Praxis der Begabungsdiagnostik in Schule und Sonderpädagogik unter Berücksichtigung neuer Forschungsergebnisse*. Villingen: Neckar-Verlag GmbH.
- Heller, K. A. (Hrsg.). (1974). *Leistungsbeurteilung in der Schule*. Heidelberg: Quelle & Meyer.
- Heller, K. A. (1976). *Intelligenz und Begabung*. München: Ernst Reinhardt Verlag.
- Heller, K. A. (Hrsg.). (1984). *Leistungsdiagnostik in der Schule*. Bern: Huber.
- Heller, K. A. (1995). Schulleistungsprognosen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 983–989). Weinheim: Beltz.
- Heller, K. A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 183–201). Weinheim: Beltz.
- Heller, K. A. & Geisler, H.-J. (1983). *Kognitiver Fähigkeitstest. KFT 1-3*. Weinheim: Beltz.
- Heller, K. A., Reimann, R. & Senfter, A. (2005). *Hochbegabung im Grundschulalter. Erkennen und Fördern*. Münster: Lit.
- Heller, K. A., Rosemann, B. & Steffens, K.-H. (1978). *Prognose des Schulerfolgs. Eine Längsschnittstudie zur Schullaufbahnberatung*. Weinheim: Beltz.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.

- Helmke, A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 203–216). Weinheim: Beltz.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Verlber: Kallmeyer.
- Helmke, A. & Reich, H. H. (2001). Die Bedeutung der sprachlichen Herkunft für die Schulleistung. *Empirische Pädagogik*, 15(4), 567–600.
- Helmke, A. & Schrader, F.-W. (2010). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 90–102). Weinheim: Beltz.
- Helmke, A., Schrader, F.-W. & Hosenfeld, I. (2004). Elterliche Lernunterstützung und Schulleistung ihrer Kinder. *Bildung und Erziehung*, 57(3), 251–278.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistung. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71–176). Göttingen: Hogrefe.
- Herzog, W. (2011). Schule und Schulklasse als soziale Systeme. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 163–202). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hesse, I. & Latzko, B. (2012). Pädagogische-psychologische Diagnostik. In U. Sandfuchs, W. H. Melzer & A. Rausch (Hrsg.), *Handbuch Erziehung* (S. 620–626). Stuttgart: UTB.
- Hochweber, J. (2010). *Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugnissensuren auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe*. Münster: Waxmann.
- Hofer, M. (1986). *Sozialpsychologie erzieherischen Handelns. Wie das Denken und Verhalten von Lehrern organisiert ist*. Göttingen: Hogrefe.
- Höhn, E. (1980). *Der schlechte Schüler. Sozialpsychologische Untersuchungen über das Bild des Schulversagers*. München: Piper.
- Holmeier, M. (2013). *Leistungsbeurteilung im Zentralabitur*. Wiesbaden: Springer.
- Holtappels, H. G. & Heerdegen, M. (2005). Schülerleistungen in unterschiedlichen Lernumwelten im Vergleich zweier Grundschulmodelle in Bremen. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R.

- Valtin & G. Walther (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 361–397). Münster: Waxmann.
- Hoyer, T. (2012). Begabungsbegriff und Leistung. Eine pädagogische Annäherung. In A. Hackl, C. Pauly, O. Steenbuck & G. Weigand (Hrsg.), *Werte schulischer Begabtenförderung. Begabung und Leistung*. (S. 14–22). Frankfurt am Main: Karg-Stiftung.
- Hradil, S. (2005). *Soziale Ungleichheit in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1–55.
- Hurrelmann, K. (1975). *Erziehungssystem und Gesellschaft*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag GmbH.
- Ingenkamp, K. (1967). *Schulleistungen - damals und heute. Meinungen und Untersuchungen zur Veränderung des Leistungsniveaus unserer Schuljugend*. Weinheim: Beltz.
- Ingenkamp, K. (1968). *Untersuchungen zu Übergangsauslese*. Weinheim: Beltz.
- Ingenkamp, K. (Hrsg.). (1976a). *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte*. Weinheim: Beltz.
- Ingenkamp, K. (1976b). Sind Zensuren aus verschiedenen Klassen vergleichbar? In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte* (S. 194–201). Weinheim: Beltz.
- Ingenkamp, K. (1981). *Wert und Wirkungen von Beurteilungsverfahren. Untersuchungen zu den Gütekriterien und der Wirkung diagnostischer Instrumente in der Schule*. Weinheim: Beltz.
- Ingenkamp, K. (1989). *Diagnostik in der Schule. Beiträge zu Schlüsselselfragen der Schülerbeurteilung*. Weinheim: Beltz.
- Ingenkamp, K. (1993). Der Prognosewert von Zensuren, Lehrgutachten, Aufnahmeprüfungen und Tests während der Grundschulzeit für den Sekundarschulerfolg. In R. Olechowski & E. Persy (Hrsg.), *Frühe schulische Auslese* (S. 68–85). Frankfurt am Main: Peter Lang.

- Ingenkamp, K. (1995a). Beurteilungsfehler minimieren! Lernerfolgsmessung durch Schultests. *Pädagogik*, 47 (3), 25–30.
- Ingenkamp, K. (1995b). Erfassung und Rückmeldung des Lernerfolgs. In G. Otto & W. Schulz (Hrsg.), *Methoden und Medien der Erziehung und des Unterrichts* (S. 173–205). Stuttgart: Klett.
- Ingenkamp, K. & Lissmann, U. (2005). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jachmann, M. (2003). *Noten oder Berichte? Die schulische Beurteilungspraxis aus der Sicht von Schülern, Lehrern und Eltern*. Opladen: Leske + Budrich.
- Jerusalem, M. & Schwarzer, J. (1991). Entwicklung des Selbstkonzepts in verschiedenen Lernumwelten. In R. Pekrun & H. Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung. Ein Resümee der Längsschnittforschung* (S. 115–128). Stuttgart: F. Enke Verlag.
- Jung, J. (2013). *Schülerleistungen erkennen, messen, bewerten*. Stuttgart: Kohlhammer.
- Jürgens, E. (2010). *Leistung und Beurteilung in der Schule. Eine Einführung in Leistungs- und Bewertungsfragen aus pädagogischer Sicht*. Sankt Augustin: Academia-Verlag.
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Weinheim: Beltz.
- Jürgens, E. & Sacher, W. (2008). *Leistungserziehung und Pädagogische Diagnostik in der Schule. Grundlagen und Anregungen für die Praxis*. Stuttgart: Kohlhammer.
- Kalthoff, H. (1996). Das Zensurenpanoptikum. Eine ethnographische Studie zur schulischen Bewertungspraxis. *Zeitschrift für Soziologie*, 25 (2), 106–124.
- Kampshoff, M. (2007). *Geschlechterdifferenz und Schulleistung. Deutsche und englische Studien im Vergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kaufmann, A. (2008). *Die Rolle motivationaler Schülermerkmale bei der Entstehung sozialer Disparitäten des Schulerfolgs. Eine Längsschnittuntersuchung an Grundschulen in Bayern und Sachsen*. Berlin: Mensch und Buch Verlag.

- Kemmler, L. (1967). *Erfolg und Versagen in der Grundschule. Empirische Untersuchungen*. Göttingen: Hogrefe.
- Kleinert, H. (Hrsg.). (1951). *Lexikon der Pädagogik*. Bern: A. Francke AG Verlag.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (1968). *Erläuterung der Notenstufen bei Schulzeugnissen und Einzelergebnissen in staatlichen Prüfungszeugnissen*. Beschluss vom 03.10.1968 [Beschluss 675].
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Vereinbarung über Bildungsstandards für den Primarbereich (Jahrgangsstufe 4)*, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Zugriff am 18.04.2017. Verfügbar unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_10\\_15-Bildungsstandards-Primar.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Primar.pdf).
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005a). *Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4)*, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Zugriff am 18.04.2017. Verfügbar unter [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_10\\_15-Bildungsstandards-Deutsch-Primar.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Deutsch-Primar.pdf).
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005b). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4)*, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Zugriff am 18.04.2017. Verfügbar unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_10\\_15-Bildungsstandards-Mathe-Primar.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf).
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2014). *Das Bildungswesen in der Bundesrepublik Deutschland 2012/2013. Darstellung der Kompetenzen, Strukturen und bildungspolitischen Entwicklungen für den Informationsaustausch in Europa*, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Bonn: KMK.

- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2015). *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)*. Informationsschrift des Sekretariats der Kultusministerkonferenz. Stand: 19.02.2015, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Zugriff am 18.04.2017. Verfügbar unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2015/2015\\_02\\_19-Uebergang\\_Grundschule-SI-Orientierungsstufe.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschule-SI-Orientierungsstufe.pdf).
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krapp, A. (1973). *Bedingungen des Schulerfolgs. Empirische Untersuchung in der Grundschule*. München: R. Oldenbourg Verlag GmbH.
- Krapp, A. (1976). Bedingungsfaktoren der Schulleistung. *Psychologie in Erziehung und Unterricht*, 23 (2), 91–109.
- Kristen, C. (1999). *Bildungsentscheidungen und Bildungsungleichheit – ein Überblick über den Forschungsstand*. Arbeitspapier (Nr. 5). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Krüskén, J. (2007a). Entwicklung von Schülerleistungen und Zensuren in der Grundschule. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (S. 41–61). Münster: Waxmann.
- Krüskén, J. (2007b). Design und Stichprobe der Untersuchung. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (S. 25–39). Münster: Waxmann.
- Kuhl, P. & Hannover, B. (2012). Differenzielle Benotung von Mädchen und Jungen. Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44 (3), 153–162.
- Langfeldt, H.-P. (1984). Die klassische Testtheorie als Grundlage normorientierter (standardisierter) Schulleistungstests. In K. A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 65–98). Bern: Huber.

- Langfeldt, H.-P. & Tent, L. (1999). *Pädagogisch-psychologische Diagnostik. Anwendungsbereiche und Praxisfelder*. Göttingen: Hogrefe.
- Lavin, D. E. (1965). *The Prediction of Academic Performance. A Theoretical Analysis and Review of Research*. New York: Russel Sage Foundation.
- Lehmann, R. H. (2001). Messung von Schulleistungen im Primar- und Sekundarbereich. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 131–141). Weinheim: Beltz.
- Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Bericht über die Erhebung im September 1996 (LAU 5)*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lichtenstein-Rother, I. (Hrsg.). (1976). *Schulleistung und Leistungsschule*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lintorf, K. (2012). *Wie vorhersagbar sind Grundschulnoten. Prädiktionskraft individueller und kontextspezifischer Merkmale*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lüders, M. (2001). Dispositionsspielräume im Bereich der Schülerbeurteilung. Auch ein Beitrag zur Professions- und Organisationsforschung. *Zeitschrift für Pädagogik*, 47 (2), 217–234.
- Maaz, K., Baeriswyl, F. & Trautwein, U. (2011). *Herkunft zensiert? Leistungsdiagnostik und soziale Ungleichheiten in der Schule. Eine Studie im Auftrag der Vodafone Stiftung Deutschland*. Düsseldorf: Vodafone Stiftung.
- Maaz, K., Baumert, J. & Trautwein, U. (2010). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit? In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen. Zeitschrift für Erziehungswissenschaft*, 12, 11–46 [Themenheft]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maaz, K., Baumert, J. & Trautwein, U. (2011). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert

- sich soziale Ungleichheit? In H.-H. Krüger, U. Rabe-Kleberg, R.-T. Kramer & J. Budde (Hrsg.), *Bildungsungleichheit revisited. Bildung und soziale Ungleichheit vom Kindergarten bis zur Hochschule* (S. 69–102). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maaz, K., Hausen, C., McElvany, N. & Baumert, J. (2006). Stichwort: Übergänge im Bildungssystem. Theoretische Konzepte und ihre Anwendung in der empirischen Forschung beim Übergang in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft*, 9 (3), 299–327.
- Maaz, K. & Nagy, G. (2010). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems. Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen. Zeitschrift für Erziehungswissenschaft*, 12, 153–182 [Themenheft]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R. & Baumert, J. (2008). Der Übergang von der Grundschule in die weiterführende Schule. Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30 (3), 519–548.
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht. Schülerleistungen messen, bewerten und fördern*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Marsh, H. W. (2005). Big-Fish-Little-Pond Effect on Academic Self-Concept. *Zeitschrift für Pädagogische Psychologie*, 19 (3), 119–127.
- McElvany, N. & Razakowski. (2013). Soziale Ungleichheiten und Schule. Forschungsstand im Überblick und Ansatzpunkte. *Jahrbuch der Schulentwicklung*, 17, 50–79.
- Middendorf, W. (2012). Schulische Leistungsbeurteilung auf dem Prüfstand: einführende Betrachtungen zu aktuellen Aufgaben und Herausforderungen. In C. Fischer (Hrsg.), *Diagnose und Förderung statt Notengebung? Problemfelder schulischer Leistungsbeurteilung* (S. 9–20). Münster: Waxmann.
- Möller, J. & Jerusalem, M. (1997). Attributionsforschung in der Schule. *Zeitschrift für Pädagogische Psychologie*, 11 (3-4), 151–166.

- Mühle, G. (1970). Definitions- und Methodenprobleme der Begabtenforschung. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen* (S. 69–97). Stuttgart: Ernst Klett Verlag.
- Müller, R. (2013). *Bedingungsfaktoren schulischer Leistungen in der Grundschule - (Womit) Kann Schulerfolg prognostiziert werden? Eine Längsschnittuntersuchung an Grundschulen und Schulen der Sekundarstufe I in Bayern und Sachsen*. München: Herbert Utz Verlag.
- Muthén, B. O. & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology*, 25, 267–316.
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus. Statistical Analysis With Latent Variables. User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.
- Naylor, F.D. (1972). *Personality and educational achievement* (Educational achievement, Bd. 2). Sydney: J. Wiley & Sons Australasia Pty.
- Niederbacher, A. & Zimmermann, P. (2011). *Grundwissen Sozialisation. Einführung zur Sozialisation im Kindes- und Jugendalter*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Nölle, I., Hörstermann, T., Krolak-Schwerdt, S. & Gräsel, C. (2009). Relevante diagnostische Informationen bei der Übergangsempfehlung - die Perspektive der Lehrkräfte. *Unterrichtswissenschaft*, 37 (4), 294–310.
- Oevermann, U. (1970). Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen* (S. 297–356). Stuttgart: Ernst Klett Verlag.
- Osnes, J. (1976). Der Einfluß äußerer Faktoren bei der Aufsatzbeurteilung. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 131–147). Weinheim: Beltz.
- Pache, D. (1978). *Einstellungen und Schulleistungen. Vergleichende Untersuchungen zwischen Sportunterricht und Hauptfächern*. München: Ernst Reinhardt Verlag.
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62, 307–332.
- Parsons, T. (1979). *Sozialstruktur und Persönlichkeit*. Frankfurt am Main: Fachbuchhandlung für Psychologie Verlagsabteilung.

- Perleth, C. & Sen, M. A. (2010). Zuverlässigkeit von Schulnoten, kognitiven Fähigkeitstests und Begabungseinschätzung von Eltern für die Wahl der weiteren Schullaufbahn. In S. Lin-Klitzing, D. Di Fuccia & G. Müller-Frerich (Hrsg.), *Übergänge im Schulwesen. Chancen und Probleme aus sozialwissenschaftlicher Sicht* (S. 105–126). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe. Analyse und Dokumentation*. Olten und Freiburg im Breisgau: Walter-Verlag.
- Plake, K. (2010). *Schule als Konstrukt der Öffentlichkeit: Bilder - Strategien - Wirklichkeiten*. Wiesbaden: Springer.
- Pohlmann-Rother, S. (2010). Die Herausbildung der Übergangsempfehlung am Ende der Grundschulzeit. *Zeitschrift für Grundschulforschung*, 3 (2), 136–148.
- Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (Hrsg.). (2013). *PISA 2012. Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.
- Rademacher, S. & Wernet, A. (2015). Struktur, Funktion und Eigenlogik. In J. Böhme, M. Hummrich & R.-T. Kramer (Hrsg.), *Schulkultur. Theoriebildung im Diskurs* (S. 95–115). Wiesbaden: Springer.
- Rademacher, S. (2009). Die Beurteilung schulischer Leistungen aus rechtlicher Sicht. In W. Sacher (Hrsg.), *Leistungen entwickeln, überprüfen und beurteilen. Bewährte und neue Wege für die Primar- und Sekundarstufe* (S. 185–195). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Radtke, F.-O. (2004). Die Illusion der meritokratischen Schule. Lokale Konstellationen der Produktion von Ungleichheit im Erziehungssystem. In K. J. Bade & M. Bommers (Hrsg.), *Migration - Integration - Bildung. Grundfragen und Problembereiche* (S. 143–178). Bad Iburg: Grote.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg.
- Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 175–190). Weinheim: Beltz.

- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 178–186). Göttingen: Hogrefe.
- Richert, P. (2012). *Elternentscheidung versus Lehrerdiagnose. Der Übergang von der Grundschule zur Sekundarschule*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Rieder, K. (1990). Leistung und Funktion der Leistungsbeurteilung. In R. Olechowski & K. Rieder (Hrsg.), *Motivieren ohne Noten* (S. 56–91). Wien: Jugend und Volk.
- Rodax, K. & Spitz, N. (1978). *Sozialstatus und Schulerfolg. Darstellung und Kritik der schichtenspezifischen Sozialisationsforschung*. Heidelberg: Quelle und Meyer.
- Rolff, H.-G. (1997). *Sozialisation und Auslese durch die Schule*. Weinheim: Juventa-Verlag.
- Rosenthal, R. & Jacobson, L. (1971). *Pygmalion im Unterricht. Lehrererwartungen und Intelligenzentwicklung der Schüler*. Weinheim: Beltz.
- Rost, D. H. (2009). *Intelligenz. Fakten und Mythen*. Weinheim: Beltz.
- Roth, E. & Sauer, J. (1981). Über die Entwicklung einiger kognitiver Bedingungen der Schulleistung. In R. Groner & K. Froppa (Hrsg.), *Kognitive Strukturen und ihre Entwicklung* (S. 135–153). Bern: Huber.
- Roth, H. (1967). Begabung und Begaben. Über das Problem der Umwelt in der Begabungsentfaltung. In T. Ballauff & H. Hettwer (Hrsg.), *Begabungsförderung und Schule* (S. 18–36). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Roth, H. (Hrsg.). (1970). *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. Stuttgart: Ernst Klett Verlag.
- Sacher, W. (1985). Falsche Benotung: Unrecht an unseren Schülern. *Ehrenwirth Hauptschulmagazin*, 4, 3–6.
- Sacher, W. (2001). Leistung und Leistungserziehung in der Grundschule. In W. Einsiedler (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (S. 218–229). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Sacher, W. (Hrsg.). (2009). *Leistungen entwickeln, überprüfen und beurteilen. Bewährte und neue Wege für die Primar- und Sekundarstufe*. Bad Heilbrunn: Verlag Julius Klinkhardt.

- Sächsische Staatskanzlei. (2013). *Schulordnung Grundschulen*. Zugriff am 12.04.2016. Verfügbar unter <http://www.revosax.sachsen.de/vorschrift/3886-Schulordnung-Grundschulen>.
- Saldern, M. von. (1997). *Schulleistung in Deutschland - ein Beitrag zur Standortdiskussion*. Münster: Waxmann.
- Sander, E. (1997). Das Stereotyp des schlechten Schülers. Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 261–267). Weinheim: Beltz.
- Sauer, J. & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersehbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg*. Göttingen: Hogrefe.
- Sauer, J. & Gattringer, H. (1985). Soziale, familiäre, kognitive und motivationale Determinanten der Schulleistung. Ein Beitrag zu einem Strukturmodell der Bedingungen des Schulerfolgs bei Grundschulern. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 37(2), 288–309.
- Scharenberg, K. (2014). Macht die Klasse einen Unterschied? Klassenkomposition und Schulleistung am Ende der Grundschulzeit. In K. Drossel, R. Strietholt & W. Bos (Hrsg.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen* (S. 47–64). Münster: Waxmann.
- Schauenberg, M. (2007). *Übertrittsentscheidungen nach der Grundschule. Empirische Analysen zu familialen Lebensbedingungen und Rational-Choice*. München: Utz.
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik*. Berlin: Springer.
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschullempfehlung. *Zeitschrift für Erziehungswissenschaft*, 14(3), 371–396.
- Schrader, F.-W. (1997). Lern- und Leistungsdiagnostik im Unterricht. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 659–699). Göttingen: Hogrefe.
- Schrader, F.-W. (2009). Kognitive Voraussetzungen. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 425–429). Bad Heilbrunn: Verlag Julius Klinkhardt.

- Schrader, F.-W. & Helmke, A. (2008). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 285–302). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schrader, F.-W. & Helmke, A. (2014). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Schrader, F.-W., Helmke, A. & Hosenfeld, I. (2008). Stichwort. Kompetenzentwicklung im Grundschulalter. *Zeitschrift für Erziehungswissenschaft*, 11 (1), 7–29.
- Schröder, H. (1990). *Leistung in der Schule. Begründung, Forderung, Beurteilung*. München: M. Arndt.
- Schumacher, E. (2002). Die soziale Ungleichheit der Lehrer/innen - oder: Gibt es eine Milieuspezifität pädagogischen Handelns? In J. Mägdefrau & E. Schumacher (Hrsg.), *Pädagogik und soziale Ungleichheit. Aktuelle Beiträge-Neue Herausforderungen* (S. 253–270). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Schümer, G. (2004). Zur doppelten Benachteiligung von Schülern aus unterprivilegierten Gesellschaftsschichten im deutschen Schulwesen. In G. Schümer, K.-J. Tillmann & M. Weiß (Hrsg.), *Die Institution Schule und die Lebenswelt der Schüler. Vertiefende Analysen der PISA-2000-Daten zum Kontext von Schülerleistungen* (S. 73–114). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schütte, K., Frenzel, A. C., Aseburg, R. & Pekrun, R. (2007). Schülermerkmale, naturwissenschaftliche Kompetenz und Berufserwartung. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 125–146). Münster: Waxmann.
- Solga, H. (2009). Meritokratie - die moderne Legitimation ungleicher Bildungschancen. In H. Solga, J. Powell & P. A. Berger (Hrsg.), *Soziale Ungleichheit. Klassische Texte zur Sozialstrukturanalyse* (S. 63–72). Frankfurt: Campus Verlag.
- Sommer, W. (1983). *Bewährung des Lehrerurteils. Eine empirische Untersuchung über den Aussagewert des Lehrerurteils für den Bildungs- und Berufserfolg*. Bad Heilbrunn: Verlag Julius Klinkhardt.

- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85–95.
- Staatsinstitut für Schulqualität und Bildungsforschung München. (2005). *Glossar Begriffe im Kontext von Leistungserhebung und Prüfung vor dem Hintergrund des dreigliedrigen Schulsystems in Bayern*, Staatsinstitut für Schulqualität und Bildungsforschung München. Zugriff am 12.04.2016. Verfügbar unter [http://www.isb.bayern.de/download/941/glossar\\_leistungserhebung.pdf](http://www.isb.bayern.de/download/941/glossar_leistungserhebung.pdf).
- Staatsinstitut für Schulqualität und Bildungsforschung München. (2015). *Kompetenzorientierter Unterricht: Leistungen beobachten – erheben – bewerten*, Staatsinstitut für Schulqualität und Bildungsforschung München. München: Staatsinstitut für Schulqualität und Bildungsforschung.
- Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (S. 171–198). Münster: Waxmann.
- Stahl, N. (2009). *Der institutionelle Beitrag zur Entstehung primärer und sekundärer Herkunftseffekte. Eine Untersuchung an bayerischen Grundschulen*. Berlin: Mensch-und-Buch-Verlag.
- Stamm, M. (2007). Begabtenförderung und soziale Herkunft. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27 (3), 227–242.
- Stamm, M. (Hrsg.). (2014). *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik*. Bern: Huber.
- Starch, D. & Elliot, E. C. (1976). Die Verlässlichkeit der Zensuren von Mathematikarbeiten. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte* (S. 81–89). Weinheim: Beltz.
- Steinkamp, G. (1976). Die Rolle des Volksschullehrers im schulischen Selektionsprozeß. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte* (S. 306–326). Weinheim: Beltz.
- Steinmayr, R. & Meißner, A. (2013). Zur Bedeutung der Intelligenz und des Fähigkeitsselbstkonzeptes bei der Vorhersage von Leistungstests

- und Noten in Mathematik. *Zeitschrift für Pädagogische Psychologie*, 27 (4), 273–282.
- Stern, E. (2002). Wie abstrakt lernt das Grundschulkind? Neuere Ergebnisse der entwicklungspsychologischen Forschung. In H. Petillon (Hrsg.), *Individuelles und soziales Lernen in der Grundschule* (S. 27–42). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stern, W. (1967). Psychologische Begabungsforschung und Begabungsdiagnose. In T. Ballauff & H. Hettwer (Hrsg.), *Begabungsförderung und Schule* (S. 3–10). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Stocké, V. (2010). Schulbezogenes Sozialkapital und Schulerfolg der Kinder: Kompetenzvorsprung oder statistische Diskriminierung durch Lehrkräfte? In B. Becker & D. Reimer (Hrsg.), *Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie* (S. 81–115). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Strietholt, R. & Bos, W. (2010). Die Nutzung der Ergebnisse standardisierter Leistungstests und der Zusammenhang zwischen Schülerleistung und Lehrerurteil. In W. Böttcher, J. N. Dicke & N. Hogrebe (Hrsg.), *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit* (S. 165–177). Münster: Waxmann.
- Stubbe, T. C. & Bos, W. (2008). Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik*, 22 (1), 49–63.
- Stubbe, T. C., Bos, W. & Euen, B. (2012). Der Übergang von der Primar- in die Sekundarstufe. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 209–226). Münster: Waxmann.
- Süllwold, F. (1976). *Begabung und Leistung*. Hamburg: Hoffmann und Campe.
- Susteck, H. (1995). Begabung und Schulleistung. *Pädagogische Welt*, 49 (5), 198–202.
- Tacke, V. (2006). Rationalität im Neo-Institutionalismus. Vom exakten Kalkül zum Mythos. In K. Senge & K.-U. Hellmann (Hrsg.), *Einführung in den Neo-Institutionalismus* (S. 89–101). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Tenorth, H.-E. (2007). Begabung - eine Kontroverse zwischen Wissenschaft und Politik. In D. Lemmermöhle & M. Hasselhorn (Hrsg.), *Bildung-Lernen. Humanistische Ideale, gesellschaftliche Notwendigkeiten, wissenschaftliche Erkenntnisse* (S. 117–145). Göttingen: Wallstein.
- Tent, L. & Birkel, P. (2010). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 949–958). Weinheim: Beltz.
- Tent, L., Fingerhut, W. & Langfeldt, H.-P. (1976). *Quellen des Lehrerurteils. Untersuchungen zur Aufklärung der Varianz von Schulnoten*. Weinheim: Beltz.
- Terhart, E. (1999). Zensurenggebung und innerschulisches Selektionsklima – die Rolle der Schulleitung. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 19 (3), 277–292.
- Terhart, E. (2002). *Nach PISA. Bildungsqualität entwickeln*. Hamburg: Sabine Groenewold Verlage.
- Terhart, E. (2011). Lehrerberuf und Professionalität. Gewandeltes Begriffsverständnis – neue Herausforderungen. *Zeitschrift für Pädagogik, Beiheft*, 57, 202–224.
- Thiel, O. & Valtin, R. (2002). Eine Zwei ist eine Drei ist eine Vier. Oder: Sind Zensuren aus verschiedenen Klassen vergleichbar? In R. Valtin (Hrsg.), *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand* (S. 67–76). Weinheim: Juventa.
- Tiedemann, J. & Billmann-Mahecha, E. (2004). Kontextfaktoren der Schulleistung im Grundschulalter. Ergebnisse aus der Hannoverschen Grundschulstudie. *Zeitschrift für Pädagogische Psychologie*, 18 (2), 113–124.
- Tiedemann, J. & Billmann-Mahecha, E. (2007a). Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 10 (1), 108–120.
- Tiedemann, J. & Billmann-Mahecha, E. (2007b). Leseverständnis, Familiensprache und Freizeitsprache. *Zeitschrift für Pädagogische Psychologie*, 21 (1), 41–49.

- Tillmann, K.-J. & Vollstädt, W. (2009). Funktionen der Leistungsbeurteilung. Eine Bestandsaufnahme. In S.-I. Beutel & W. Vollstädt (Hrsg.), *Leistung ermitteln und bewerten* (S. 27–37). Hamburg: Bergmann + Helbig Verlag.
- Titze, H. (2000). Zensuren in der modernen Gesellschaft. Zur Selbstbeurteilung und Fremdbeurteilung schulischer Leistungsdifferenzen. In J. Schlömerkemper (Hrsg.), *Differenzen. Über die politische und pädagogische Bedeutung von Ungleichheiten im Bildungswesen* (S. 49–62). Weinheim: Juventa.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21 (1), 11–27.
- Trautwein, U. & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind. Referenzgruppeneffekte bei Eintrittsentscheidungen. *Zeitschrift für Pädagogische Psychologie*, 21 (2), 119–133.
- Trautwein, U., Maaz, K. & Baeriswyl, F. (2014). Zur Rolle von Begabung beim Übergang in die Sekundarstufe I. In M. Stamm (Hrsg.), *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik* (S. 205–215). Bern: Huber.
- Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O. & Baumert, J. (2006). Tracking, Grading, and Student Motivation. Using Group Composition and Status to Predict Self-Concept and Interest in Ninth-Grade Mathematics. *Journal of Educational Psychology*, 98 (4), 788–806.
- Treutlein, A., Roos, J. & Schöler, H. (2008). Einfluss des Leistungsniiveaus einer Schulklasse auf die Benotung am Ende des 3. Schuljahres. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30 (3), 579–593.
- Valtin, R., Wagner, C. & Schwippert, K. (2005). Schülerinnen und Schüler am Ende der vierten Klasse - schulische Leistungen, lernbezogene Einstellungen und außerschulische Lernbedingungen. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 187–238). Münster: Waxmann.

- Van Ewijk, R. (2011). Same Work, Lower Grade? Student Ethnicity and Teachers' Subjective Assessments. *Economics of Education Review*, 30, 1045–1058.
- Wagner, E. (1980). *Schülerbeurteilung als soziales Handeln. Auswirkungen schulischer Beurteilung auf Alltagswissen und Identität von Schülern*. Weinheim: Beltz.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a Knowledge Base for School Learning. *Review of Educational Research*, 63 (3), 249–294.
- Watermann, R. & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen. Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 61–94). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung*. Berlin, Heidelberg: Springer.
- Weinert, F. E. (2001). Perspektiven der Schulleistungsmessung - mehrperspektivisch betrachtet. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 353–365). Weinheim: Beltz.
- Weinert, F. E. (2012). Begabung und Lernen. Zur Entwicklung geistiger Leistungsunterschiede. In A. Hackl, C. Pauly, O. Steenbuck & G. Weigand (Hrsg.), *Werte schulischer Begabtenförderung. Begabung und Leistung* (S. 23–34). Frankfurt am Main: Karg-Stiftung.
- Weinert, F. E. (2014). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Weinert, F. E. & Helmke, A. (Hrsg.). (1997). *Entwicklung im Grundschulalter*. Weinheim: Beltz.
- Weiß, R. H. (1998). *Grundintelligenz Skala 2. CFT 20*. Göttingen: Hogrefe.
- Weiss, R. (1976a). Aufgaben der Zensuren und Zeugnisse. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 62–65). Weinheim: Beltz.

- Weiss, R. (1976b). Die Zuverlässigkeit der Ziffernbenotung bei Aufsätzen und Rechenarbeiten. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (S. 104–116). Weinheim: Beltz.
- Weiss, R. (1976c). Über die Strenge der Benotung in verschiedenen Unterrichtsgegenständen. In I. Lichtenstein-Rother (Hrsg.), *Schulleistung und Leistungsschule* (S. 63–72). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Weiss, R. (1988). Dimensionen der Leistungsbeurteilung. *Unser Weg*, 43 (2), 47–51.
- Wiese, W. (1982). Elternstatus, Lehrerempfehlung und Schullaufbahn: Eine empirische Analyse des Einflusses des Grundschullehrers auf die Bildungslaufbahn des Schülers. *Zeitschrift für Soziologie*, 11 (1), 49–63.
- Wild, K.-P. & Krapp, A. (2006). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (S. 525–574). Weinheim: Beltz.
- Wohlkinger, F. (2014). *Die Rolle des Schülers bei der Wahl der weiterführenden Schule. Eine vergleichende Untersuchung von Grundschulern aus Bayern und Sachsen*. Wiesbaden: Springer.
- Wollersheim, H. W. (2014). Talent und Begabung in der Pädagogik. In M. Stamm (Hrsg.), *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik* (S. 23–32). Bern: Huber.
- Woolfolk, A. (2008). *Pädagogische Psychologie*. München: Pearson Studium.
- Zaborowski, K. U., Meier, M. & Breidenstein, G. (Hrsg.). (2011). *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ziegenspeck, J. W. & Lehmann, J. (1999). *Handbuch Zensur und Zeugnis in der Schule. Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen. Ein Studien- und Arbeitsbuch*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Ziegler, A. (2009). Hochbegabte und Begabtenförderung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 937–951). Wiesbaden: VS Verlag für Sozialwissenschaften.

---

Zumhasch, C. (2001). Schulleistungsbeurteilung: Leistungen feststellen und bewerten. In W. Einsiedler (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (S. 263–275). Bad Heilbrunn: Verlag Julius Klinkhardt.



# Anhang

## Abbildungsverzeichnis Anhang A

Abbildung A-1: Mehrebenenmodell für das Fach Deutsch

Abbildung A-2: Mehrebenenmodell für das Fach Mathematik

## Tabellenverzeichnis Anhang B

Tabelle B-1: Korrelationen der Fehlerterme im Mehrebenenmodell für  
das Fach Deutsch

Tabelle B-2: Korrelationen der Fehlerterme im Mehrebenenmodell für  
das Fach Mathematik

Anhang A: Mehrebenenmodelle für das Fach Deutsch und Mathematik

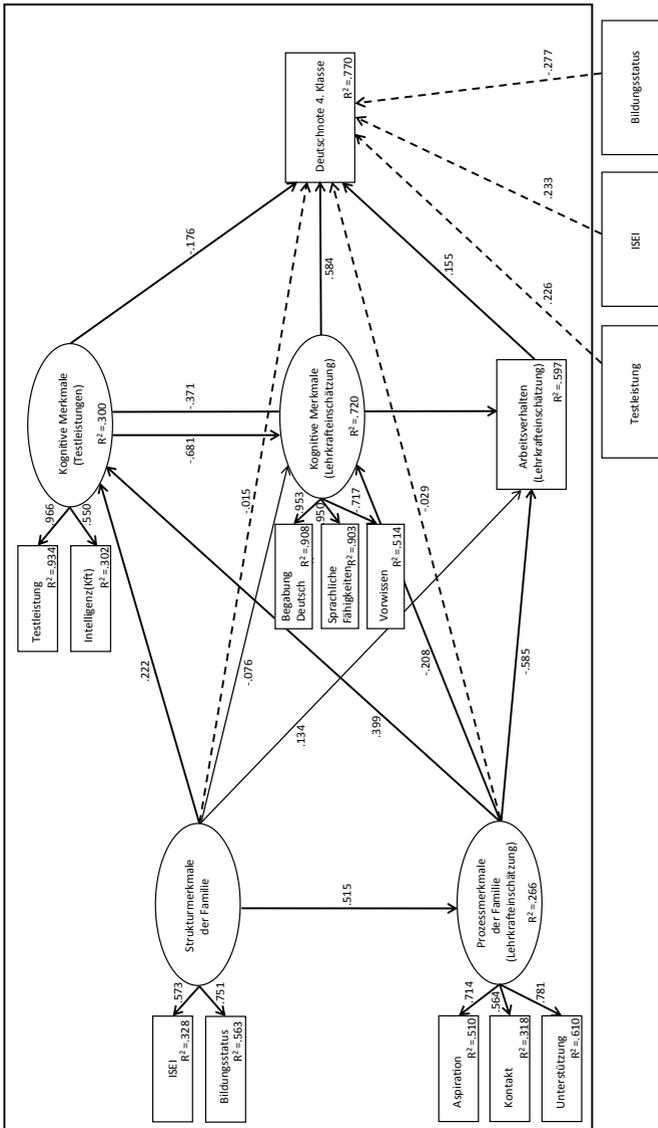


Abbildung A-1: Mehrebenenmodell für das Fach Deutsch

Anmerkungen: N=1446;  $\chi^2=462.882$ ,  $df=101$ ,  $p<.01$ ; RMSEA=.050; SRMR Within=.049; SRMR Between=.316; CF=.948  
 Fett gedruckter Pfeil:  $p<.01$ ; Dünn gedruckter Pfeil:  $p<.05$ ; Gestrichelter Pfeil: n.s.

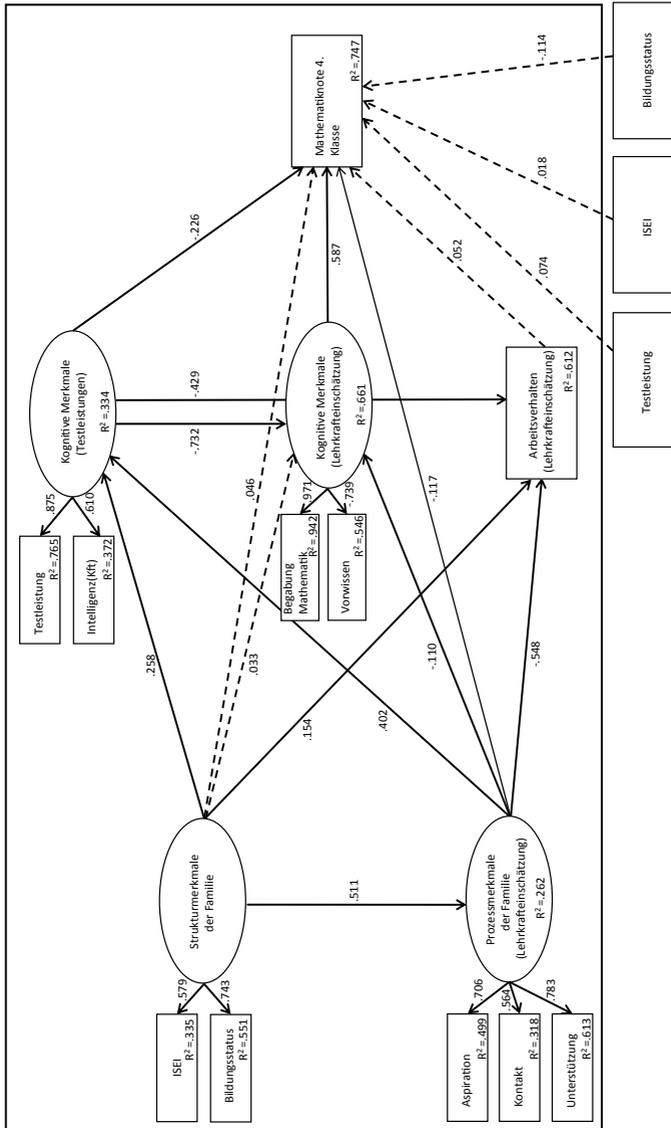


Abbildung A-2: Mehrebenenmodell für das Fach Mathematik

Anmerkungen: N=1446; Chi<sup>2</sup>=412.315, df=82, p<.01; RMSEA=.053; SRMR Within=.050; SRMR Between=.266; CFI=.931  
 Fett gedruckter Pfeil: p<.01; Dünn gedruckter Pfeil: p<.05; Gestrichelter Pfeil: n. s

## Anhang B: Korrelationen der Fehlerterme zu Mehrebenenmodellen für das Fach Deutsch und Mathematik

Tabelle B-1: Korrelationen der Fehlerterme im Mehrebenenmodell für das Fach Deutsch

<b>Variablenpaar</b>	<b>Fehlerkorrelation</b>
Arbeitsverhalten/kognitive Merkmale (LK)	.293**
Sprachliche Fähigkeiten/Begabung Deutsch	-.487**
Sprachliche Fähigkeiten/Vorwissen	.420**
Vorwissen/Begabung Deutsch	.350**

\*\*p <.01 \*p <.05

Tabelle B-2: Korrelationen der Fehlerterme im Mehrebenenmodell für das Fach Mathematik

<b>Variablenpaar</b>	<b>Fehlerkorrelation</b>
Arbeitsverhalten/kognitive Merkmale (LK)	-.010
Vorwissen/Begabung Mathematik	.723*

\*\*p <.01 \*p <.05

Welche Komponenten umfasst die Leistungsbeurteilung am Ende der Grundschulzeit aus Sicht der Lehrkraft? Welche individuellen kognitiven und nicht-kognitiven Schülermerkmale spiegeln sich in der Deutsch- und Mathematiknote der vierten Klasse wider, und inwiefern zeigen sich hier direkte und vermittelnde Effekte der sozialen Herkunft? Welche Rolle spielt bei der Notengebung die leistungsbezogene und soziale Klassenzusammensetzung, und inwiefern werden die einzelnen Einflussfaktoren auf die Notengebung durch schulrechtliche Vorgaben gestützt und können messtheoretischen Gütekriterien entsprechen?

Ausgehend von diesen Fragestellungen zeigen Analysen auf Basis der DFG-geförderten Längsschnitt-Studie KOALA-S, dass die individuellen kognitiven Schülermerkmale den größten direkten Effekt auf die Notengebung vorweisen. Ebenfalls nimmt das Arbeitsverhalten direkten Einfluss auf die Deutschnote. Die familiären Merkmale nehmen unter Kontrolle der Lehrkräfteeinschätzungen zu individuellen Schülermerkmalen keinen direkten Einfluss auf die Deutsch- und Mathematiknote, gleiches gilt für die leistungsbezogene und soziale Klassenzusammensetzung. Sehr wohl aber zeigen sich Vermittlungseffekte der durch die Lehrkraft eingeschätzten familiären Prozessmerkmale (Bildungsaspirationen der Eltern, Schulkontakte, Unterstützung des Kindes) über die individuellen Schülermerkmale auf die Schulnoten.

23,30 €  
ISBN 978-3-95925-092-4

